Welcome to the next step in your career – CPA Program

Today's CPA Program is a globally recognised education program available around the world. All candidates of CPA Australia are required to attain a predetermined level of technical competence before the CPA designation can be awarded. The CPA Program foundation level is designed to provide you with an opportunity to demonstrate knowledge and skills in the core areas of accounting, business and finance.

A pass for each exam is based on a determination of the minimum level of knowledge and skills that candidates must acquire to have a good chance at success in the professional level of the CPA Program.

In 2012 you have more opportunities to sit foundation level exams, allowing you to progress through to the professional level of the CPA Program at your own pace.

The material in this study manual has been prepared based upon standards and legislation in effect as at 1 September 2011. Candidates are advised that they should confirm effective dates of standards or legislation when using additional study resources. Exams for 2012 will be based on the content of this study manual.

Additional Learning Support

A range of quality learning products will be available in the market for you to purchase to further aid your core study program and preparation for exams.

These products will appeal to candidates looking to invest in additional resources other than those provided in this study manual. More information is available on CPA Australia’s website www.cpaaustralia.com.au/learningsupport

You will also be able to source face-to-face and online tuition for CPA Program foundation level exams from registered tuition providers. The tuition provided by these registered parties is based on current CPA Program foundation level learning objectives. A list of current registered providers can be found on CPA Australia’s website. If you are interested you will need to liaise directly with the chosen provider to purchase and enrol in your tuition program.
## Contents

**Introduction**
- Welcome to CPA Australia iii
- Chapter features vi
- Chapter summary viii
- Answering multiple choice questions x
- Learning objectives xi

**Chapter**

**Part 1: Economics**
- 1. Defining economics and the market 3
- 2. Demand, supply and the price mechanism 19
- 3. Elasticity of demand and supply 43
- 4. Costs, revenues and productivity 67
- 5. Market structures 95
- 6. Market failure, externalities and intervention 127
- 7. National income accounting 147
- 8. Determining national income 171
- 9. Macroeconomic concepts – inflation and unemployment 199
- 10. Macroeconomic policy 227
- 11. Government intervention and income distribution 257

**Part 2: Statistics**
- 12. Statistical analysis, data and methods of describing data 277
- 13. Descriptive statistics 317
- 14. Frequency distributions and probability 355
- 15. Hypothesis testing 383
- 16. Linear regression and correlation 417

**Revision questions** 445

**Answers to revision questions** 483

**Before you begin questions: answers and commentary** 505

**Glossary of terms** 523

**Formulae** 537

**Index** 543
## Chapter features

Each chapter contains a number of helpful features to guide you through each topic.

<table>
<thead>
<tr>
<th>Feature</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Learning objectives</td>
<td>Show the referenced CPA Australia learning objectives.</td>
</tr>
<tr>
<td>Topic list</td>
<td>Tells you what you will be studying in this chapter.</td>
</tr>
<tr>
<td>Introduction</td>
<td>Presents a general idea of what is covered in this chapter.</td>
</tr>
<tr>
<td>Chapter summary diagram</td>
<td>Summarises the content of the chapter, helping to set the scene so that you can gain the bigger picture.</td>
</tr>
<tr>
<td>Before you begin</td>
<td>This is a small bank of questions to test any pre-existing knowledge that you may have of the chapter content. If you get them all correct then you may be able to reduce the time you need to spend on the particular chapter. There is a commentary section at the end of the Study Manual called <em>Before you begin: answers and commentary.</em></td>
</tr>
<tr>
<td>Section overview</td>
<td>This summarises the key content of the particular section that you are about to start.</td>
</tr>
<tr>
<td>Learning objective reference</td>
<td>This box indicates the learning objective covered by the section or paragraph to which it relates.</td>
</tr>
<tr>
<td>Definition</td>
<td>Definitions of important concepts. You really need to know and understand these before the exam.</td>
</tr>
<tr>
<td>Exam comments</td>
<td>These highlight points that are likely to be particularly important or relevant to the exam. (Please note that this feature does not apply in every Foundation Level study manual.)</td>
</tr>
<tr>
<td>Worked example</td>
<td>This is an illustration of a particular technique or concept with a solution or explanation provided.</td>
</tr>
<tr>
<td>Question</td>
<td>This is a question that enables you to practise a technique or test your understanding. You will find the solution at the end of the chapter.</td>
</tr>
<tr>
<td>Key chapter points</td>
<td>Review the key areas covered in the chapter.</td>
</tr>
</tbody>
</table>
Quick revision questions

A quick test of your knowledge of the main topics in this chapter.

The quick revision questions are not a representation of the difficulty of the questions which will be in the examination. The quick revision MCQs provide you with an opportunity to revise and assess your knowledge of the key concepts covered in the materials so far. Use these questions as a means to reflect on key concepts and not as the sole revision for the examination.

Revision questions

The revision questions are not a representation of the difficulty of the questions which will be in the examination. The revision MCQs provide you with an opportunity to revise and assess your knowledge of the key concepts covered in the materials so far. Use these questions as a means to reflect on key concepts and not as the sole revision for the examination.

Case study

This is a practical example or illustration, usually involving a real world scenario.

Formula to learn

These are formulae or equations that you need to learn as you may need to apply them in the exam.

Bold text

Throughout the Study Manual you will see that some of the text is in bold type. This is to add emphasis and to help you to grasp the key elements within a sentence and paragraph.
Chapter summary

This summary provides a snapshot of each of the chapters, to help you to put the Study Manual into perspective.

Chapter 1 – Defining economics and the market
Chapter 1 defines key concepts namely economics, production, factors of production, scarcity, resources, opportunity costs and the market. It also defines comparative and absolute advantage.

Chapter 2 – Demand, supply and the price mechanism
Chapter 2 examines the interaction of demand and supply, i.e. the price mechanism and the setting of the equilibrium price. The chapter concludes with an examination of minimum and maximum price setting by both producers and governments.

Chapter 3 – Elasticity of demand and supply
Chapter 3 introduces the key concepts of elasticity of demand and elasticity of supply. It defines demand price elasticity and asks students to perform elasticity calculations. It also requires students to prepare demand curves for necessities and luxuries.

Chapter 4 – Cost, revenues and productivity
Chapter 4 looks firstly at revenues and the calculation of a firm’s revenues. Secondly, it examines the costs of production and the impact of short-run and long-run factors on costs. The chapter concludes with an analysis of individual firm’s productivity based on cost savings and efficiencies.

Chapter 5 – Market structures
Chapter 5 presents common market structures. The two most extreme structures are perfect competition and monopoly. Imperfect market structures include monopolistic competition, oligopoly and duopoly as well as monopsony and oligopsony.

Chapter 6 – Market failure, externalities and intervention
Chapter 6 looks at market imperfections and market failure. It examines the effects of externalities and the impact of government actions and controls used to reduce the misallocation of resources in individual firms.

Chapter 7 – National income accounting
Chapter 7 is the first of the macroeconomic chapters. This chapter and Chapter 8 discuss how to measure the total amount of economic activity of a nation. In Chapter 7 the focus is on calculating Gross Domestic Product (GDP) and Gross National Product (GNP).

Chapter 8 – Determining national income
This chapter follows on from Chapter 7 and examines the determination and calculation of national income. There are two broad theorists: the Keynesians and the monetarists. This chapter examines the basic elements of the Keynesian model for national income determination and equilibrium.
Chapter 9 – Macroeconomic concepts – inflation and unemployment
The first part of Chapter 9 examines two key macroeconomic concepts, namely controlling price inflation and minimising the level of unemployment in a country. The second part of the chapter introduces the concept of money, credit and interest rates, as well as monetary theory.

Chapter 10 – Macroeconomic policy
Chapter 10 gives an overview of the goals of macroeconomic policy by concentrating on two broad types of policy: fiscal policy and monetary policy. It also examines the role of central banks in controlling the supply of money by using the Reserve Bank of Australia (RBA) as an example.

Chapter 11 – Government intervention and income distribution
Chapter 11 outlines the role of government regulation in private markets, privatisation and competitive practices. It concludes with an examination of the government role in measuring income and addressing income inequality.

Chapter 12 – Statistical analysis, data and methods of describing data
This chapter introduces organisational data, which is a collection of raw facts relating to the entity and its environment. It can be classified in a number of ways such as quantitative/qualitative, discrete/continuous, internal/external, formal/informal, primary/secondary.

Data must be processed or analysed in some way to form information that is useful in the decision-making process of the organisation. Much of a manager’s work will involve the use of data and information, collected internally or externally.

You have to analyse and present the data you have collected so that it can be of use and in this chapter we look at how data can be presented in tables and charts.

Chapter 13 – Descriptive statistics
In this chapter we go further than the compilation of a frequency distribution and condense the data into two parameters that characterise the distribution. The first is a measure of central tendency, a typical value round which the various items are grouped i.e. an average. The second is a measure of dispersion i.e., some indication of the way in which these items are spread around the average.

Chapter 14 – Frequency distributions and probability
This chapter introduces probability, which is of fundamental importance in the theory of statistics. Key principles of probability are most easily explained by using examples of coin tossing, dice throwing and games of chance.

Chapter 15 – Hypothesis testing
This chapter explains hypothesis testing, which is a statistical procedure for testing whether chance is a plausible explanation of an experimental finding.

Chapter 16 – Linear regression and correlation
Following our earlier study of correlation and scatter diagrams, in this chapter we look at how the inter-relationship shown between variables in a scatter diagram can be described and calculated. The first two sections deal with correlation, which is concerned with assessing the strength of the relationship between two variables.

We will then see how we can determine the equation of a straight line to represent the relationship between the variables and use that equation to make forecasts or predictions.
Answering multiple choice questions

The questions in your exam will each contain four possible answers. You have to choose the option that best answers the question. The three incorrect options are called distractors. There is a skill in answering MCQs quickly and correctly. By practising MCQs you can develop this skill, giving you a better chance of passing the exam.

You may wish to follow the approach outlined below, or you may prefer to adapt it.

**Step 1** Attempt each question – starting with the easier questions which will be those at the start of the exam. Read the question thoroughly. You may prefer to work out the answer before looking at the options, or you may prefer to look at the options at the beginning. Adopt the method that works best for you.

**Step 2** Read the four options and see if one matches your own answer. Be careful with numerical questions, as the distractors are designed to match answers that incorporate common errors. Check that your calculation is correct. Have you followed the requirement exactly? Have you included every stage of the calculation?

**Step 3** You may find that none of the options matches your answer.
- Re-read the question to ensure that you understand it and are answering the requirement
- Eliminate any obviously wrong answers
- Consider which of the remaining answers is the most likely to be correct and select the option

**Step 4** If you are still unsure make a note and continue to the next question. Some questions will take you longer to answer than others. Try to reduce the average time per question, to allow yourself to revisit problem questions at the end of the exam.

**Step 5** Revisit unanswered questions. When you come back to a question after a break you often find you are able to answer it correctly straight away. If you are still unsure have a guess. You are not penalised for incorrect answers, so never leave a question unanswered!
**Economics and Markets**

**General overview**

This exam covers economics and quantitative methods. In economics, key microeconomics concepts of demand and supply, elasticity, productivity, market structures, and market failure are covered. It also covers macroeconomic concepts of income distribution and the structure of the financial economy including the calculation of key national economic measures. In quantitative methods, key tools of statistical analysis are covered, such as descriptive statistics, frequency distributions and probability, hypothesis testing, simple linear regression and correlation.

**Topics**

<table>
<thead>
<tr>
<th>Learning Objective</th>
<th>Chapter where covered</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Part 1: Economics</strong></td>
<td></td>
</tr>
<tr>
<td><strong>LO1. Defining economics and the market</strong></td>
<td></td>
</tr>
<tr>
<td>LO1.1 Define 'economics' and describe the characteristics of an economic perspective</td>
<td>1</td>
</tr>
<tr>
<td>LO1.2 Distinguish between wants and needs</td>
<td>1</td>
</tr>
<tr>
<td>LO1.3 Explain how consumers allocate resources</td>
<td>1</td>
</tr>
<tr>
<td>LO1.4 Define scarcity</td>
<td>1</td>
</tr>
<tr>
<td>LO1.5 Explain the practical application of the law of marginal utility</td>
<td>1</td>
</tr>
<tr>
<td>LO1.6 Explain the theory of markets</td>
<td>1</td>
</tr>
<tr>
<td>LO1.7 Explain and apply the theory of comparative advantage between products and countries</td>
<td>1</td>
</tr>
<tr>
<td>LO1.8 Analyse in practical terms the advantages and disadvantages of production on the basis of comparative advantage</td>
<td>1</td>
</tr>
<tr>
<td>LO1.9 Identify and describe the factors of production</td>
<td>1</td>
</tr>
<tr>
<td>LO1.10 Explain production and productivity</td>
<td>1</td>
</tr>
<tr>
<td>1.10.1 Prepare and explain the production possibility frontier</td>
<td>1</td>
</tr>
</tbody>
</table>

**LO2. Demand, supply, and the price mechanism**

<table>
<thead>
<tr>
<th>Learning Objective</th>
<th>Chapter where covered</th>
</tr>
</thead>
<tbody>
<tr>
<td>LO2.1 Explain the concepts of demand and supply</td>
<td>2</td>
</tr>
<tr>
<td>LO2.2 Relate consumer indifference to the substitution of goods</td>
<td>2</td>
</tr>
<tr>
<td>LO2.3 Prepare demand curves for normal and inferior goods</td>
<td>2</td>
</tr>
<tr>
<td>LO2.4 Explain the relationship between demand and supply</td>
<td>2</td>
</tr>
<tr>
<td>LO2.5</td>
<td>Distinguish between movement along the demand curve and a shift in the demand curve</td>
</tr>
<tr>
<td>LO2.6</td>
<td>Distinguish between individual and market demand</td>
</tr>
<tr>
<td>LO2.7</td>
<td>Distinguish between firm and industry demand and supply curves</td>
</tr>
<tr>
<td>LO2.8</td>
<td>Distinguish between movement along the supply curve and a shift in the supply curve</td>
</tr>
<tr>
<td>LO2.9</td>
<td>Define market equilibrium price and quantity</td>
</tr>
<tr>
<td>LO2.10</td>
<td>Explain the use of price legislation, including price ceilings and price floors</td>
</tr>
<tr>
<td>LO2.11</td>
<td>Evaluate the process of price stabilisation and price control mechanisms</td>
</tr>
<tr>
<td>LO2.12</td>
<td>Explain and illustrate how an equilibrium price is achieved</td>
</tr>
</tbody>
</table>

**LO3. Elasticity of demand and supply**

| LO3.1 | Explain the concepts of elasticity of demand and elasticity of supply | 3 |
| LO3.2 | Calculate and interpret the elasticity of demand and elasticity of supply | 3 |
| LO3.3 | Prepare demand curves for necessities and luxury goods | 3 |

**LO4. Cost, revenues and productivity**

| LO4.1 | Explain the relationship between marginal cost, total cost, total revenue, marginal revenue, average revenue and price in both the long term and short term | 4 |
| LO4.2 | Apply the concepts of marginal revenue product, marginal product, total product, total cost and marginal cost in an analysis of productivity | 4 |
| LO4.3 | Explain the demand for factors of production | 4 |
| LO4.4 | Explain the concept of diminishing returns for a factor of production | 4 |
| LO4.5 | Explain how a firm can attain an optimal combination of factors of production | 4 |
| LO4.6 | Explain the determinants of elasticity of a factor demand curve | 4 |
| LO4.7 | Explain the causes of a shift of a factor demand curve | 4 |
| LO4.8 | Distinguish between economies of scale and diseconomies of scale | 4 |
## LO5. Market structures

<table>
<thead>
<tr>
<th>LO5.1</th>
<th>Distinguish between perfect competition, monopolistic competition, monopoly, oligopoly, duopoly, and oligopsony</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>5.1.1</td>
<td>Illustrate the relevant demand and supply curves</td>
<td></td>
</tr>
<tr>
<td>LO5.2</td>
<td>Evaluate why monopolistic firms are able to allocate or misallocate scarce resources</td>
<td>5</td>
</tr>
<tr>
<td>LO5.3</td>
<td>Explain the long term pricing approach for a monopolistic firm</td>
<td>5</td>
</tr>
</tbody>
</table>

## LO6. Market failure, externalities and intervention

| LO6.1 | Distinguish between social goods and private goods | 6 |
| LO6.2 | Evaluate the impact of tax, savings and subsidies on the pricing mechanism | 6 |
| LO6.3 | Analyse the implications of spill-overs or externalities using a demand and supply analysis | 6 |

## LO7. National income accounting

| LO7.1 | Distinguish between economic growth and economic development | 7 |
| LO7.2 | Calculate Gross Domestic Product (GDP) and Gross National Product (GNP) | 7 |
| LO7.3 | Perform national accounting calculations | 7 |

## LO8. Determining national income

| LO8.1 | Calculate the National Income equation \( Y = C + G + I + M - X \) | 8 |
| 8.1.1 | Present national income calculations using the IS-LM curve |  |
| 8.1.2 | Calculate marginal efficiency of capital |  |
| 8.1.3 | Apply the multiplier to determine national income |  |
| 8.1.4 | Apply the accelerator principle in the determination of national income |  |
| LO8.2 | Evaluate the implications of the marginal propensity to save (MPS) and the marginal propensity to consume (MPC) on National Income (Y) | 8 |
| LO8.3 | Evaluate the impact of tax, savings and subsidies on National Income | 8 |
| LO8.4 | Explain the relationship between full employment and National Income | 8 |

## LO9. Macroeconomic concepts – inflation and unemployment

| LO9.1 | Describe different types of unemployment | 9 |
| LO9.2 | Describe the causes of inflation and its impact on an economy | 9 |
| LO9.3 | Explain the relationship between rates of employment and the performance of an economy | 9 |
| 9.3.1 | Prepare a Phillips curve |  |
| LO9.4 | Define money | 9 |
| LO9.5 | Explain the structure of interest rates | 9 |
| LO9.6 | Analyse the factors affecting the movement of interest rates | 9 |
| LO9.7 | Explain the Keynesian and Classical theories of money | 9 |
**LO10. Macroeconomic policy**

| LO10.1 | Explain government policy to address the redistribution of income | 10 |
| LO10.2 | Analyse the impact of interest rates on base employment | 10 |
| LO10.3 | Explain the purpose of monetary policy and the implications of holding cash balances | 10 |
| 10.3.1 | Calculate the credit multiplier | |
| LO10.4 | Explain how fiscal policy relates to the stimulation of national income and rates of employment | 10 |
| 10.4.1 | Demonstrate how fiscal policy affects aggregate demand | |
| LO10.5 | Explain the relationship between interest rates, monetary policy, employment and national income | 10 |
| 10.5.1 | Prepare an expectations augmented Phillips curve | |
| LO10.6 | Analyse the role of the monetary authorities (Reserve Banks/Central Banks) in the control of money | 10 |

**LO11. Government intervention and income distribution**

| LO11.1 | Explain how the government may intervene to reduce misallocation of resources | 11 |
| LO11.2 | Analyse ways to redress income inequalities | 11 |
| LO11.3 | Explain the concept of income distribution and describe the Lorenz curve | 11 |
| LO11.4 | Measure income inequality | 11 |

**Part 2: Statistics**

**LO12. Statistical analysis, data, and methods of describing data**

<p>| LO12.1 | Explain the role of statistical analysis in decision making | 12 |
| LO12.2 | Distinguish between quantitative and qualitative data | 12 |
| LO12.3 | Explain and apply the different sampling methods | 12 |
| 12.3.1 | Random sampling | |
| 12.3.2 | Cluster sampling | |
| 12.3.3 | Stratified sampling | |
| LO12.4 | Describe the different methods of collecting data and statistical information | 12 |
| 12.4.1 | Survey | |
| 12.4.2 | Published source | |
| LO12.5 | Explain the different levels of data measurement | 12 |
| 12.5.1 | Nominal level data | |
| 12.5.2 | Ordinal level data | |
| 12.5.3 | Interval level data | |
| 12.5.4 | Ratio level data | |</p>
<table>
<thead>
<tr>
<th>LO12.6</th>
<th>Describe different ways of presenting data</th>
<th>12, 13</th>
</tr>
</thead>
<tbody>
<tr>
<td>12.6.1</td>
<td>Construct a bar graph, a pie chart, a histogram and a scatter diagram from a given set of data</td>
<td></td>
</tr>
<tr>
<td>12.6.2</td>
<td>Interpret data presented in a bar graph, a pie chart, a histogram and a scatter diagram</td>
<td></td>
</tr>
</tbody>
</table>

**LO13. Descriptive statistics**

| LO13.1 | Distinguish between measures of central tendency and measures of variability | 13 |
| LO13.2 | Distinguish between the shapes of a normal distribution, exponential distribution and binomial distribution | 13 |
| LO13.3 | Explain the difference between grouped and ungrouped data | 13 |
| LO13.4 | Calculate and interpret the mean, median, and mode from a given set of data | 13 |
| LO13.5 | Calculate and interpret the range, standard deviation, and variance from a given set of data | 13 |
| LO13.6 | Distinguish between the sample and population standard deviation and the sample and population variance | 13 |
| LO13.7 | Distinguish between kurtosis and skewness | 13 |

**LO14. Frequency distributions and probability**

| LO14.1 | Develop a frequency distribution from a given set of data | 14 |
| LO14.2 | Distinguish between class range, class midpoint, relative frequency, and cumulative frequency | 13, 14 |
| LO14.3 | Define the concept of probability | 14 |
| LO14.4 | Explain the different ways of assigning probability | 14 |
| LO14.5 | Explain and apply marginal, union, joint, and conditional probabilities | 14 |
| LO14.6 | Explain the use of probability matrices to solve probability problems | 14 |

**LO15. Hypothesis testing**

| LO15.1 | Explain the concept of hypothesis testing | 15 |
| LO15.2 | Construct null and alternative hypotheses | 15 |
| LO15.3 | Distinguish between type I and type II errors | 15 |
| LO15.4 | Test population mean using one-tail and two-tail tests | 15 |
| LO15.5 | Test population proportion | 15 |
| LO15.6 | Calculate and interpret the probability value (p-value) in hypothesis testing | 15 |

**LO16. Simple linear regression and correlation**

| LO16.1 | Calculate the equation of a simple regression line from a sample of data | 16 |
| LO16.2 | Explain and interpret the slope and intercept of the equation | 16 |
| LO16.3 | Calculate and interpret estimated values of y using the regression line | 16 |
| LO16.4 | Calculate and interpret the coefficient of correlation | 16 |
| LO16.5 | Calculate and interpret the coefficient of determination | 16 |
## Exam topic exam weightings

### Part 1: Economics

<table>
<thead>
<tr>
<th>Chapter</th>
<th>Topic</th>
<th>Weight</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Defining economics and the market</td>
<td>5%</td>
</tr>
<tr>
<td>2</td>
<td>Demand, supply and the price mechanism</td>
<td>8%</td>
</tr>
<tr>
<td>3</td>
<td>Elasticity of demand and supply</td>
<td>5%</td>
</tr>
<tr>
<td>4</td>
<td>Cost, revenues and productivity</td>
<td>8%</td>
</tr>
<tr>
<td>5</td>
<td>Market structures</td>
<td>4%</td>
</tr>
<tr>
<td>6</td>
<td>Market failures, externalities and intervention</td>
<td>5%</td>
</tr>
<tr>
<td>7</td>
<td>Macroeconomic concepts</td>
<td>5%</td>
</tr>
<tr>
<td>8</td>
<td>Macroeconomic policy</td>
<td>7%</td>
</tr>
<tr>
<td>9</td>
<td>Government intervention and income distribution</td>
<td>5%</td>
</tr>
<tr>
<td>10</td>
<td>National Income accounting</td>
<td>4%</td>
</tr>
<tr>
<td>11</td>
<td>Determining national income</td>
<td>4%</td>
</tr>
</tbody>
</table>

### Part 2: Statistics

<table>
<thead>
<tr>
<th>Chapter</th>
<th>Topic</th>
<th>Weight</th>
</tr>
</thead>
<tbody>
<tr>
<td>12</td>
<td>Statistical analysis, data, and methods of describing data</td>
<td>8%</td>
</tr>
<tr>
<td>13</td>
<td>Descriptive statistics</td>
<td>8%</td>
</tr>
<tr>
<td>14</td>
<td>Frequency distributions and probability</td>
<td>8%</td>
</tr>
<tr>
<td>15</td>
<td>Hypothesis testing</td>
<td>8%</td>
</tr>
<tr>
<td>16</td>
<td>Simple linear regression and correlation</td>
<td>8%</td>
</tr>
</tbody>
</table>

**TOTAL** 100%
Part 1:

Economics
Chapter 1

Defining economics and the market

<table>
<thead>
<tr>
<th>Learning objectives</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Defining economics and the market</td>
<td>LO1</td>
</tr>
<tr>
<td>Define 'economics' and describe the characteristics of an economic perspective</td>
<td>LO1.1</td>
</tr>
<tr>
<td>Distinguish between wants and needs</td>
<td>LO1.2</td>
</tr>
<tr>
<td>Explain how consumers allocate resources</td>
<td>LO1.3</td>
</tr>
<tr>
<td>Define scarcity</td>
<td>LO1.4</td>
</tr>
<tr>
<td>Explain the practical application of the law of marginal utility</td>
<td>LO1.5</td>
</tr>
<tr>
<td>Explain the theory of markets</td>
<td>LO1.6</td>
</tr>
<tr>
<td>Explain and apply the theory of comparative advantage between products and countries</td>
<td>LO1.7</td>
</tr>
<tr>
<td>Analyse in practical terms the advantages and disadvantages of production on the basis of comparative advantage</td>
<td>LO1.8</td>
</tr>
<tr>
<td>Identify and describe the factors of production</td>
<td>LO1.9</td>
</tr>
<tr>
<td>Explain production and productivity</td>
<td>LO1.10</td>
</tr>
<tr>
<td>Prepare and explain the production possibility frontier</td>
<td>LO1.10.1</td>
</tr>
</tbody>
</table>

Topic list

1. Fundamental economic ideas
2. Absolute and comparative advantage
3. The concept of a market
This chapter introduces the basic economic problem, which is: how to use scarce resources to achieve maximum benefits.

It then looks at the choices which result from this problem: what should be produced, how production should be organised, and who should consume the output.

It then goes on to examine the consequence of the economic problem: that using scarce resources for one activity necessarily means they cannot be used for an alternative activity.

It introduces the international market concepts of absolute and comparative advantages and concludes with an exploration of the concepts of market and utility.
If you have studied these topics before, you may wonder whether you need to study this chapter in full. If this is the case, please attempt the questions below, which cover some of the key subjects in the area.

If you answer all these questions successfully, you probably have a reasonably detailed knowledge of the subject matter, but you should still skim through the chapter to ensure that you are familiar with everything covered.

There are references in brackets indicating where in the chapter you can find the information, and you will also find a commentary at the back of the Study Manual.

1. What is covered in the topic economics? (Section 1.1)
2. List three of your own 'needs' and three 'wants'. (Section 1.2)
3. What is the purpose of the production possibility frontier? (Section 1.3)
4. Describe the theory of comparative advantage. (Section 2.2)
5. Define total and marginal utility. (Section 3.3)
1 Fundamental economic ideas

Section overview

- Economics is concerned with how choices are made about the use of resources: what should be produced and who should consume it.
- The need to make such decisions arises because economic resources are scarce. Making decisions involves the sacrifice of benefits that could have been obtained from using resources in an alternative course of action. This is illustrated through a production possibility frontier (or curve). This sacrifice is known as the opportunity cost of an activity.

1.1 Economics as a social science

Economics studies the ways in which society decides what to produce, how to produce it, who to produce it for and how to apportion it. We are all economic agents, and economic activity is what we do to make a living.

Economists assume that people behave rationally at all times and always seek to improve their circumstances. This assumption leads to more specific assumptions:

- Producers will seek to maximise their profits.
- Consumers will seek to maximise the benefits (their 'utility') from their income.
- Governments will seek to maximise the welfare of their populations.

Both the basic assumption of rationality and the more detailed assumptions may be challenged. In particular, we will look again later at the assumption that businesses always seek to maximise their profits. A further complication is that concepts such as utility and welfare are not only open to interpretation, but also that the interpretation will change over time.

The way in which the choices about resource allocation are made, the way value is measured, and the forms of ownership of economic wealth will also vary according to the type of economic system that exists in a society.

(a) In a centrally planned (or command) economy, the decisions and choices about resource allocation are made by the government. Monetary values are attached to resources and to goods and services, but it is the government that decides what resources should be used, how much should be paid for them, what goods should be made and, in turn, what their price should be. This approach is based on the theory that only the government can make fair and proper provision for all members of society.

(b) In a free market economy, the decisions and choices about resource allocation are left to market forces of supply and demand, and the workings of the price mechanism. This approach is based on the observable fact that it generates more wealth in total than the command approach. While there are no instances or unfettered free market economic systems, the United States of America (US) economic system is based on the free market approach.

(c) In a mixed economy the decisions and choices are made partly by free market forces of supply and demand, and partly by government decisions. Economic wealth is divided between the private sector and the public sector. This approach attempts to combine the efficiency of the market system with the centrally planned system’s approach to fair and proper distribution. Australia is an example of a mixed economy.

In practice, the industrialised countries in the developed world have mixed economies, although with differing proportions of free market and centrally planned decision-making from one country to the next. In such economies, the government influences economic activity in a variety of ways and for a variety of purposes.

(a) Direct control over macroeconomic forces can be exercised through policy on tax, spending and interest rates.

(b) Taxes, subsidies and direct controls can affect the relative prices of goods and services.
(c) Government-owned institutions such as Australia’s public health system (Medicare) can provide goods and services directly, free or at low cost at the point of consumption.

(d) Regulation can be used to restrict or prevent the supply of goods and services.

(e) Incomes can be influenced through the tax and social security systems.

**Definitions**

**Microeconomics** is the study of individual economic units; these are called **households** and **firms**.

**Macroeconomics** is the study of the aggregated effects of the decisions of economic units. It looks at a complete national economy, or the international economic system as a whole.

### 1.2 Scarcity of resources

**Definition**

**Scarcity** is the excess of human wants over what can actually be produced. A **scarce resource** is a resource for which the quantity demanded at a nil price would exceed the available supply.

It is a fact of life that the amount of resources available is limited.

(a) For the individual **consumer**, the scarcity of goods and services might seem obvious enough. Most people would like to have more: perhaps a car, or more clothes, or a house of their own. Examples of services which we would like more of include live concerts, public passenger transport and holidays. These are **human wants**.

(b) For the world as a whole, resources available to serve human consumption are limited. For example, the supply of non-renewable energy resources such as coal and oil is, by definition, limited. The amount of many minerals which it is feasible to extract from the earth (for example, metals of various kinds) is also limited. Access to hot water and energy at basic levels is an example of a **human need**.

This idea of scarcity is very important in economics, because it reminds us that producers and consumers have to make **choices** about what to produce or to buy.

In the case of producers, we can identify four types of resource, which are known as **factors of production**. Each of these factors of production has an associated reward which accrues to its owner when it is used.

(a) **Land** is rewarded with **rent**. Although it is easy to think of land as property, the economic definition of land is much broader than this. Land consists not only of property (the land element only: buildings are capital) but also all the natural resources that grow on the land or that are extracted from it, such as timber and coal.

(b) **Labour** is rewarded with **wages** (including salaries). Labour consists of both the mental and the physical resources of human beings. Labour productivity can be improved through training, or by applying capital in the form of machinery.

(c) **Capital** is rewarded with **interest**. It is easy to think of capital as financial resources, and the rate of interest as the price mechanism in balancing the supply and demand for money. However, capital in an economic sense is not ‘money in the bank’. Rather, it refers to man-made items such as plant, machinery and tools which are used to aid the production of other goods and services. As we noted above, buildings – such as factories – are capital items.

(d) **Enterprise**, or entrepreneurship, is the fourth factor of production. An entrepreneur is someone who undertakes the task of organising the other three factors of production in a business enterprise, and in doing so, bears the **risk** of the venture. The entrepreneur creates new business ventures and the reward for the risk associated with this is **profit**.
Since resources for production are scarce and there are not enough goods and services to satisfy the total potential demand, choices must be made. Choice is only necessary because resources are scarce.

(a) Consumers must choose what goods and services they will buy.
(b) Producers must choose how to use their available resources, and what to produce with them.

Economics studies the nature of these choices:

(a) What will be produced?
(b) What will be consumed?
(c) Who will benefit from the consumption?

Making choices about how to use scarce resources is the fundamental problem of economics.

1.3 The production possibility frontier

We can approach this central question of economics (how to allocate scarce resources) by looking first at the possibilities of production.

Definitions

Production is the process and method employed to transform tangible and intangible inputs into goods and services.

Productivity is the measure of efficiency with which output has been produced.

To take a simple example, suppose that an imaginary society can use its available resources to produce two products, wheat and trucks. The society's resources are limited. Therefore, there are restrictions on the amounts of wheat and trucks that can be made. The possible combinations of wheat (A) and trucks (B) is shown on the production possibility frontier below (or curve).

The curve from A, round to B, in Figure 1 shows the maximum of all the various combinations of wheat (A) and trucks (B) that a society can make, given current technology, if it uses its limited resources efficiently.

(a) The society can choose to make up to:
   • A₁ units of wheat (A) and therefore produce no trucks (B).
   • B₁ units of all trucks (B) and no wheat (A).
   • A₃ units of A and B₃ of B (point P on the curve).
   • A₄ units of A and B₁ of B (point Q on the curve).
Defining economics and the market

(b) The combination of $A_x$ units of $A$ and $B_x$ units of $B$ (plotted at point $X$) is inside the production possibility frontier. This illustrates that more than these quantities can be made of either, or both, of $A$ and $B$. Point $X$ is therefore an inefficient production point for the economy, and if the society were to make only $A_x$ of $A$ and $B_x$ of $B$, it would be using its limited resources inefficiently.

(c) Note that the production possibility frontier is just what it says: it defines what is achievable if all productive resources are fully employed. It follows that changes in the level of unemployment have no effect upon it, because the curve represents the position where all labour resources are employed, i.e. there is no unemployment.

Similarly, changes in price levels will affect the monetary value of what can be produced, but not the volume, so they do not affect the curve either.

(d) The curve is normally drawn concave to the origin.

**Question 1: Production possibility curve**

What can you say about the combination of wheat ($A$) and trucks ($B$) indicated by point $Y$ in Figure 1? 

(The answer is at the end of the chapter)

The production possibility frontier is an important idea in economics: it illustrates the need to make choices about what to produce, because it is not possible to have everything.

1.3.1 Opportunity cost: the cost of one use for resources rather than another

Choice involves sacrifice. If there is a choice between having $A$ and having $B$, and a country or individual chooses to have $A$, it will be giving up $B$ to have $A$. The cost of having a certain amount of $A$ can therefore be regarded as the sacrifice of not being able to have the corresponding amount of $B$. There is a sacrifice involved in the choices of consumers and firms (producers), as well as the choices of governments at the level of national economy.

**Definition**

The cost of an item measured in terms of the alternatives forgone is called its opportunity cost.

A production possibility frontier illustrates opportunity costs. For example, if in Figure 1 it is decided to switch from making $A_1$ units of $A$ and $B_3$ units of $B$ (point $Q$) to making $A_2$ units of $A$ and $B_2$ units of $B$ (point $P$), then the opportunity cost of making $(B_2 - B_3)$ more units of $B$ would be the lost production of $(A_3 - A_2)$ units of $A$.

The production possibility line is a concave curve and not a straight line because some resources are more useful for making $A$ than for making $B$, and vice versa. As a result, opportunity costs change as we move away from a situation in which production is wholly devoted to either $A$ or $B$. Thus, as we move away from point $A_1$, and introduce an increasing level of production of $B$, the amount of $B$ that we gain from losing each unit of $A$ progressively diminishes.

At the level of the firm, the production possibility frontier can be seen as showing the maximum output of different goods a firm can produce when all of its resources are fully used. For example, a firm might operate production lines capable of producing washing machines or refrigerators: producing more washing machines bears the opportunity cost of a lower level of production of refrigerators.
2 Absolute and comparative advantage

Section overview
- Countries can have either a comparative or an absolute advantage. World output of goods and services will increase if countries specialise in the production of goods or services in which they have a comparative advantage. Just how this total wealth is shared out between countries depends on circumstances.

Economists distinguish the concepts of comparative advantage and absolute advantage in international trade. Our explanation of this distinction makes the following assumptions:
- There are only two countries, country X and country Y.
- Only two goods are produced (in our example, these are trucks and wheat).
- There are no transport costs and no barriers to trade.
- Resources within each country are easily transferred from one industry to another.

2.1 Absolute advantage
A country is said to have an absolute advantage in the production of a good when it is more efficient than another country in the production of that good; that is, when that country can produce more of a particular good with a given amount of resources than another country. It is a fairly common situation for one country to be more efficient than another in the production of a particular good.

**Worked Example: Absolute advantage**
Assuming that Y produces wheat more efficiently than country X, while country X has an absolute advantage in producing trucks, a simple arithmetical example can illustrate the potential gains from trade. The table below shows the amounts of trucks and wheat that each country can produce per day, assuming that each country has an equal quantity of resources and devotes half of its resources to truck production and half to wheat production.

<table>
<thead>
<tr>
<th>Country</th>
<th>Trucks</th>
<th>Wheat (tonnes)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Country X</td>
<td>20</td>
<td>100</td>
</tr>
<tr>
<td>Country Y</td>
<td>10</td>
<td>150</td>
</tr>
<tr>
<td>World total</td>
<td>30</td>
<td>250</td>
</tr>
</tbody>
</table>

According to the data presented, country X can produce 20 trucks per day while country Y can only produce 10 trucks per day using the same quantity of resources. Similarly, country X can produce 100 tonnes of wheat per day while country Y can produce 150 tonnes of wheat per day using the same quantity of resources. Therefore, country X has an absolute advantage in the production of trucks and country Y has an absolute advantage in the production of wheat.

If country X devotes all of its resources to the production of trucks, it could produce 40 trucks. Similarly, if country Y devotes all of its resources to the production of wheat, it could produce 300 tonnes.

**Question 2: Absolute advantage**
Suppose that each country devotes its entire production resources to the product for which it enjoys an absolute advantage. What will be the total output of trucks and wheat?

(The answer is at the end of the chapter)
By specialising, total world output is now greater. In the simple example we have just looked at, there are ten more trucks and 50 tonnes more wheat now available for consumption. In order to obtain the benefits of specialisation countries X and Y in our example can exchange some part of their individual outputs. It is not possible to specify the exact rate of exchange but the limits of the exchange rate must be somewhere between the domestic opportunity cost ratios of the two countries. These are: for country X, 5 tonnes of wheat per truck and for country Y, 15 tonnes of wheat per truck. One country will not benefit from international trade if the 'exchange rate' is not between these ratios.

2.2 Comparative advantage

Definition

The law of comparative advantage (or comparative costs) states that two countries can gain from trade when each specialises in the industries in which it has lowest opportunity costs.

Introduced by David Ricardo, the theory of comparative advantage is based on the idea of opportunity cost and the production possibility frontier. Within a country, the opportunity cost for any category of product may be established in terms of the next most advantageous use of national resources. If two countries produce different goods most efficiently and can exchange them at an advantageous rate in terms of the comparative opportunity costs of importing and home production, then it will be beneficial for them to specialise and trade. Total production of each good will be higher than if they each produce both goods. This is true even if one country has an absolute advantage in both goods.

The principle of comparative costs can be shown by an arithmetical example.

Worked Example: Comparative costs

It is now assumed that country X is more efficient in the production of both road trucks and wheat. If each country devotes half its resources to each industry the assumed daily production totals are as shown below:

<table>
<thead>
<tr>
<th>Country</th>
<th>Trucks</th>
<th>Wheat (tonnes)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Country X</td>
<td>20</td>
<td>200</td>
</tr>
<tr>
<td>Country Y</td>
<td>10</td>
<td>150</td>
</tr>
<tr>
<td>World total</td>
<td>30</td>
<td>350</td>
</tr>
</tbody>
</table>

In terms of resources used, the costs of production in both industries are lower in country X. If we consider the opportunity costs, however, the picture is rather different. In country X the cost of one truck is ten tonnes of wheat, which means that devoting resources to the production of one truck in country X there is a sacrifice in terms of ten tonnes of wheat forgone. The opportunity cost of one truck in country Y is fifteen tonnes of wheat.

In country X the opportunity cost of a tonne of wheat is now one tenth of a truck, while in country Y the opportunity cost is one fifteenth of a truck. In terms of the output of trucks forgone, wheat is cheaper in country Y than in country X. Country Y has a comparative advantage in wheat. It would now be possible for country Y to buy 10 trucks from country X in exchange for 100 tonnes of wheat. Country X would transfer some of its resources from the production of wheat to the production of trucks, while country Y would put all of its resources into the production of wheat. Total production would now look like this:

<table>
<thead>
<tr>
<th>Country</th>
<th>Trucks</th>
<th>Wheat (tonnes)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Country X</td>
<td>30</td>
<td>100</td>
</tr>
<tr>
<td>Country Y</td>
<td>0</td>
<td>300</td>
</tr>
<tr>
<td>World total</td>
<td>30</td>
<td>400</td>
</tr>
</tbody>
</table>

There is an increase in the world output of wheat.

Alternatively, country X might buy 150 tonnes of wheat from country Y in exchange for 15 trucks. Country X would transfer even more resources to the production of trucks and the total production figures would change again.
There has now been an increase in the world output of trucks.

Clearly, the two countries could adjust their trade between these extremes, achieving overall increases in both types of good. However, the key point is that total production is increased if each country specialises in producing the good for which it has a comparative advantage.

**Exam comments**

Make sure that you are clear about the concept of comparative advantage. Fundamentally, the comparative advantage model explains trade in terms of the benefits of international specialisation. Note that it is trade that leads to specialisation and not the other way round.

### 3 The concept of a market

#### Section overview

- Markets are created when potential buyers and sellers come together to exchange goods or services. A good or service has a price if it is both useful and scarce.
- **Marginal utility** is the extra satisfaction gained by consuming one unit more or the satisfaction forgone by consuming one unit less. Consumers act rationally when they attempt to maximise total utility with their limited income.

#### 3.1 What is a market?

A market involves the buyers and sellers of a good who influence its price. Markets can be worldwide, as in the case of oil, wheat, cotton and copper for example. Others are more localised, such as the housing market or the market for second-hand cars.

**Definition**

A market can be defined as a situation in which potential buyers and potential sellers (suppliers) of a good or service come together for the purpose of exchange.

Suppliers and potential suppliers are referred to in economics as firms. The potential purchasers of consumer goods are known as households.

However, some markets have buyers who are other firms or government authorities. For example, a manufacturing firm buys raw materials and components to go into the products that it makes. Service industries and government departments must similarly buy in supplies in order to do their own work. The demand for goods by firms and government authorities is a derived demand in that it depends on the demand from households for the goods and services that they produce and provide.

Markets for different goods or commodities are often inter-related. All commodities compete for households’ income so that if more is spent in one market, there will be less to spend in other markets. Further, if markets for similar goods are separated geographically, there will be some price differential at which it will be worthwhile for the consumer to buy in the lower price market and pay shipping costs, rather than buy in a geographically nearer market.
3.2 Price theory and the market

Price theory is concerned with how market prices for goods are arrived at, through the interaction of demand and supply.

A good or service has a **price** if it is **useful** as well as **scarce**. Its usefulness is shown by the fact that consumers demand it. In a world populated entirely by vegetarians, meat would not command a price, no matter how few cows or sheep there were because no one would want to eat meat.

3.3 Utility

**Utility** is the word used to describe the pleasure or satisfaction or benefit derived by a person from the consumption of goods. **Total utility** is the total satisfaction that people derive from spending their income and consuming goods.

**Marginal utility** is the *satisfaction gained* from consuming one *additional* unit of a good or the *satisfaction forgone* by consuming one unit *less*. If someone eats six apples and then eats a seventh, total utility refers to the satisfaction he derives from all seven apples together, while marginal utility refers to the additional satisfaction from eating the seventh apple, having already eaten six.

3.4 Assumptions about consumer rationality

Economists assume that consumers act rationally. This means, in turn, that:

(a) Generally, the consumer prefers more goods to less

(b) Substitution is complete. Generally, the consumer is willing to substitute between consumption bundles with differing quantities of goods that provide the same level of satisfaction. This willingness to substitute is a property of the underlying preferences and has little to do with prices. Prices and income will determine the composition of the consumption bundle actually chosen by the individual. The individual compares their willingness to substitute (coming from their preferences) with the market’s rate of substitution (prices). The consumer will seek to maximise their well being subject to their financial constraints.

(c) Choices are transitive. This means that if, at a given time, a commodity A is preferred to B and B is preferred to C then we can conclude that commodity A is preferred to commodity C.

Acting rationally means that the consumer attempts to **maximise the total utility** attainable with a **limited income**. When the consumer decides to buy another unit of a good the customer is deciding that its marginal utility exceeds the marginal utility that would be yielded by any **alternative** use of the price paid.

If a person has maximised total utility, it follows that the expenditure has been allocated in such a way that the utility gained from spending the last penny spent on each good will be equal.
Key chapter points

- Economics is concerned with how choices are made about the use of resources: what should be produced and who should consume it.

- The need to make such decisions arises because economic resources are scarce. Making decisions involves the sacrifice of benefits that could have been obtained from using resources in an alternative course of action. This is illustrated through a production possibility frontier (or curve). This sacrifice is known as the opportunity cost of an activity.

- Countries can have either a comparative or an absolute advantage. World output of goods and services will increase if countries specialise in the production of goods or services in which they have a comparative advantage. Just how this total wealth is shared out between countries depends on circumstances.

- Markets are created when potential buyers and sellers come together to exchange goods or services. A good or service has a price if it is both useful and scarce.

- Marginal utility is the extra satisfaction gained by consuming one unit more or the satisfaction forgone by consuming one unit less. Consumers act rationally when they attempt to maximise total utility with their limited income. Consumers use marginal utility where they decide to/or not to, purchase an additional unit.
Quick revision questions

1. What is the essential feature of a command economy?

2. Macroeconomics is the study of economic units such as households and firms. Is this statement true or false?
   A. true
   B. false

3. Which of the following is not recognised as a factor of production?
   A. capital
   B. management
   C. land
   D. labour

4. The cost of an item measured in terms of the resources used is called its opportunity cost. Is this statement true or false?
   A. true
   B. false

5. Matilda buys four pairs of matching high heels and then buys a fifth pair. Explain the concept of total utility and marginal utility using Matilda’s situation.
Answers to quick revision questions

1. Decisions about resources, production and prices are made by the government.

2. The study of individual economic units is called microeconomics. Macroeconomics is the study of a complete national economy.

3. The fourth factor is enterprise or entrepreneurship.

4. Opportunity cost is defined as the cost of an item in terms of the alternatives forgone. Cost in terms of resources used is a reasonable definition of the accounting concept of ‘full cost’.

5. Total utility is the satisfaction gained from buying all five pairs of shoes. Marginal utility is the satisfaction gained from buying the fifth pair.
1 Point Y lies outside the production possibility frontier. Even with efficient use of resources it is impossible to produce this combination of wheat (A) and trucks (B). To reach point Y, either current resources must be increased or production methods must be improved – perhaps by developments in technology.

2 Total world output will be 40 trucks (produced by country X) and 300 tonnes of wheat (produced by country Y).
Chapter 2
Demand, supply and the price mechanism

<table>
<thead>
<tr>
<th>Learning objectives</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Demand, supply and the price mechanism</td>
<td>LO2</td>
</tr>
<tr>
<td>Explain the concepts of demand and supply</td>
<td>LO2.1</td>
</tr>
<tr>
<td>Relate consumer indifference to the substitution of goods</td>
<td>LO2.2</td>
</tr>
<tr>
<td>Prepare demand curves for normal and inferior goods</td>
<td>LO2.3</td>
</tr>
<tr>
<td>Explain the relationship between demand and supply</td>
<td>LO2.4</td>
</tr>
<tr>
<td>Distinguish between movement along the demand curve and a shift in the demand curve</td>
<td>LO2.5</td>
</tr>
<tr>
<td>Prepare a demand curve showing the impacts of shifts</td>
<td>LO2.5.1</td>
</tr>
<tr>
<td>Distinguish between individual and market demand</td>
<td>LO2.6</td>
</tr>
<tr>
<td>Distinguish between firm and industry demand and supply curves</td>
<td>LO2.7</td>
</tr>
<tr>
<td>Prepare a short run and a long run supply curve</td>
<td>LO2.7.1</td>
</tr>
<tr>
<td>Distinguish between movement along the supply curve and a shift in the supply curve</td>
<td>LO2.8</td>
</tr>
<tr>
<td>Prepare a supply curve showing the impacts of shifts</td>
<td>LO2.8.1</td>
</tr>
<tr>
<td>Define market equilibrium price and quantity</td>
<td>LO2.9</td>
</tr>
<tr>
<td>Explain the use of price legislation, including price ceilings and price floors</td>
<td>LO2.10</td>
</tr>
<tr>
<td>Illustrate the impact of price ceilings and price floors using the demand and supply curves</td>
<td>LO2.10.1</td>
</tr>
<tr>
<td>Evaluate the process of price stabilisation and price control mechanisms</td>
<td>LO2.11</td>
</tr>
<tr>
<td>Explain and illustrate how an equilibrium price is achieved</td>
<td>LO2.12</td>
</tr>
</tbody>
</table>

Topic list

1. Demand
2. Supply
3. The equilibrium price
4. Demand and supply analysis
5. Maximum and minimum prices
The distinction between the microeconomic level and the macroeconomic level was mentioned in Chapter 1. Chapter 1 also examined the concept of a market which, in economics, goes beyond the idea of a single geographical place where people meet to buy and sell goods.

In this chapter, we look in more depth at the microeconomic level of the individual firm, individual markets and consumers (or households). This means looking at what influences the amount of a product which is demanded or supplied and analysing how price and output are determined through the interaction of demand and supply, i.e., the price mechanism and the setting of the equilibrium price. The chapter concludes with an examination of minimum and maximum price setting by both producers and governments.
Before you begin

If you have studied these topics before, you may wonder whether you need to study this chapter in full. If this is the case, please attempt the questions below, which cover some of the key subjects in the area.

If you answer all these questions successfully, you probably have a reasonably detailed knowledge of the subject matter, but you should still skim through the chapter to ensure that you are familiar with everything covered.

There are references in brackets indicating where in the chapter you can find the information, and you will also find a commentary at the back of the Study Manual.

1. What is demand? (Section 1.1)
2. Prepare a normal market demand curve. (Section 1.3)
3. Explain the concept of substitution of goods. (Section 1.4)
4. What is supply? (Section 2.1)
5. What is the equilibrium price? (Section 3.2)
6. What is a price ceiling? (Section 5.2)
7. What is a price floor? (Section 5.3)
8. Does Australia have minimum wage legislation? (Section 5.4)
1 Demand

Section overview

- The position of the demand curve is determined by the demand conditions, which include consumers’ tastes and preferences, and consumers’ incomes. The demand curve can show either individual demand of a specific good or service, or market demand. Market demand is influenced by national distribution of income.
- Substitution of goods will impact the demand of goods, as will complements. The impact of fashion and expectation in affecting demand cannot be underestimated.

1.1 The concept of demand

**Demand** for a good or service is the quantity of that good or service that potential purchasers would be willing and able to buy, or attempt to buy, at any possible price.

Demand might be satisfied, and so actual quantities bought would equal demand. On the other hand, some demand might be unsatisfied, with the number of would-be purchasers trying to buy a good being too great for the supply of that good. In which case, demand is said to exceed supply.

The phrase 'willing and able to buy' in the description above is important. Demand does not mean the quantity that potential purchasers wish they could buy. For example, a million households might wish that they owned a luxury yacht, but there might only be actual attempts to buy one hundred luxury yachts at a given price. Economic demand needs to be effective. That is, it must be supported by available money (i.e. willing and able to buy), rather than just being a general desire for goods or services.

1.2 The demand schedule and the demand curve

The relationship between demand and price can be shown graphically as a demand curve. The demand curve of a single consumer or household is derived by estimating how much of the good the consumer or household would demand at various hypothetical market prices. Suppose that the following demand schedule shows demand for biscuits by one household over a period of one month:

<table>
<thead>
<tr>
<th>Price per kg</th>
<th>Quantity demanded (kg)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$1</td>
<td>9.75</td>
</tr>
<tr>
<td>2</td>
<td>8</td>
</tr>
<tr>
<td>3</td>
<td>6.25</td>
</tr>
<tr>
<td>4</td>
<td>4.5</td>
</tr>
<tr>
<td>5</td>
<td>2.75</td>
</tr>
<tr>
<td>6</td>
<td>1</td>
</tr>
</tbody>
</table>

Notice that we show demand falling as price increases. This is what normally happens with most goods. This is because purchasers have a limited amount of money to spend and must choose between goods that compete for their attention. When the price of one good rises, it is likely that other goods will seem relatively more attractive and so demand will switch away from the more expensive good to the cheaper alternative. So the shape of the demand curve is determined by the consumer acting rationally: with demand tending to be higher at a low price, and lower at a high price for most goods and services.

We can show this schedule graphically, with price on the y axis and quantity demanded on the x axis. If we assume that there is complete divisibility, so that price and quantity can both change in infinitely small steps, we can draw a demand curve by joining the points represented in the schedule by a continuous line (Figure 1). This is the household’s demand curve for biscuits in the particular market we are looking at.
Figure 1 Graph of a demand schedule

The area of each rectangle in Figure 1 represents consumers’ total money outlay at the price in question. For example, at a price of $6, demand would be 1 kilogram and total spending would be $6, represented by rectangle ABCO. Similarly, at a price of $2, demand would be 8 kilograms and the total spending of $16 is represented by rectangle GEFO.

Exam comments

Sketching demand and supply curves is a useful way of preparing for the exam.

In Figure 1, the demand curve happens to be a straight line. Straight line demand curves are often used as an illustration in economics because it is convenient to draw them this way. In reality, a demand curve is more likely to be a curved line convex to the origin. As you will be able to appreciate, such a demand curve means that there are progressively larger increases in quantity demanded as price falls (Figure 2).

Figure 2 Demand curve convex to the origin
Question 1: Demand curve

Refer to Figure 2. The price of the commodity is currently $3 per kilo, and demand is approximately 4 kilograms at that price. What would be the approximate demand for the commodity if the price fell to $2 per kilo? And what would be the demand if the price rose to $4 per kilo?

(The answer is at the end of the chapter)

Note that changes in demand caused by changes in price are represented by movements along the demand curve, from one point to another. These changes in quantity demanded in response to a change in price are called expansions or contractions in demand. The price has changed, and the quantity demanded changes (prompting a movement along the curve), but the demand curve itself remains the same.

1.3 The market demand curve

In the example above, we have been looking at the demand schedule of a single household. A market demand curve is a similar curve, but it expresses the expected total quantity of the good that would be demanded by all consumers together, at any given price.

Market demand is the total quantity of a product that all purchasers would want to buy at each price level. A market demand schedule and a market demand curve are therefore simply the sum of all the individual demand schedules and demand curves put together. Market demand curves would be similar to those in Figures 1 and 2 – sloping downwards from left to right – but with quantities demanded (total market demand) being higher at each price level.

A demand curve normally slopes down from left to right.

(a) As we saw earlier, the curve is downward sloping because progressively larger quantities are demanded as price falls.

(b) A fall in the good’s price means that it becomes cheaper both in relation to the household’s income and also in relation to other (substitute) products. Therefore, the overall size of the market for the good increases. The converse argument applies to an increase in prices; the size of the market will shrink as the good becomes more expensive.

Several factors influence the total market demand for a good. One of these factors is obviously its price, but there are other factors too, and to help you to appreciate some of these other factors, you need to recognise that households buy not just one good with their money but a whole range of goods and services.

Factors determining demand for a good

• The price of the good.
• The size of households’ income (income effect).
• The price of other substitute goods (substitution effect).
• Tastes and fashion.
• Expectations of future price changes.
• The distribution of income among households.

The income effect reflects the impact of a price change on consumers’ income. If the price of a good falls, all other things being equal, consumers become better off as their real income has increased. Therefore, they can afford to buy more of the good after it has fallen in price.

The income effect can also be reinforced by the substitution effect. The substitution effect occurs when consumers buy more of one good and less of another good because of relative price changes between the two goods. For example, if two types of bread are considered substitutes and the price of bread 1 falls relative to the price of bread 2, then consumers will buy more of 1 than 2: they substitute bread 1 for bread 2.

A demand curve shows how the quantity demanded will change in response to a change in price provided that all other conditions affecting demand are unchanged – that is, provided that there is no change...
in the prices of other goods, tastes, expectations or the distribution of household income. (This assumption that 'all other things remain equal' is referred to as ceteris paribus.)

Make sure you remember this point about a movement along the demand curve reflecting a change in price when other factors are unchanged. We will return to it later to examine what happens to the demand curve when the other conditions affecting demand are changed.

1.4 Substitutes and complements

Definitions

Substitute goods are goods that are alternatives to each other, so that an increase in the demand for one is likely to cause a decrease in the demand for another. Switching demand from one good to another 'rival' good is substitution.

Complements are goods that tend to be bought and used together, so that an increase in the demand for one is likely to cause an increase in the demand for the other.

A change in the price of one good will not necessarily change the demand for another good. For example, we would not expect an increase in the price of televisions to affect the demand for bread. However, there are goods for which the market demand is inter-connected. These inter-related goods are referred to as either substitutes or complements.

Examples of substitute goods and services

- Rival brands of the same commodity, like Coca-Cola and Pepsi-Cola.
- Tea and coffee.
- Some different forms of entertainment.

Substitution takes place when the price of one good rises relative to a substitute good.

By contrast, complements are connected in the sense that demand for one is likely to lead to demand for the other.

Examples of complements

- Cups and saucers.
- Bread and butter.
- Motor cars and the components and raw materials that go into their manufacture.

Question 2: Substitutes and complements

What might be the effect of an increase in the ownership of domestic deep freezers on the demand for perishable food products?

(The answer is at the end of the chapter)

1.5 Household income and demand: normal and inferior goods

As you might imagine, more income will give households more to spend, and they will want to buy more products (goods and services) at existing prices. However, a rise in household income will not increase market demand for all goods and services. The effect of a rise in income on demand for an individual good will depend on the nature of the good.

Demand and the level of income may be related in different ways:

(a) We might normally expect a rise in household income to lead to an increase in demand for a good, and goods for which demand rises as household income increases are called normal goods.
(b) Demand may rise with income up to a certain point but then fall as income rises beyond that point. Goods whose demand eventually falls as income rises are called inferior goods: examples might include ‘value’ or ‘basics’ ranges of own-brand supermarket foods (which could be substituted for more expensive ranges), and bus or coach travel (which could be substituted for rail or air travel). The reason for falling demand is that as incomes rise, demand switches to superior products.

1.6 Market demand and the distribution of income

So far we have discussed individual demand. Market demand for a good is influenced by the way in which the national income is shared among households.

In a country with many rich and many poor households and few middle income ones, we might expect a relatively large demand for luxury cars and yachts and also for bread and potatoes. In a country with many middle-income households (such as Australia), we might expect high demand for medium-sized cars and TV sets, and other middle income goods.

Question 3: Income distribution

What do you think might be the demand for swimming pools among a population of five households enjoying total annual income of $1m, if the distribution of income is either as under assumption 1 or as under assumption 2.

<table>
<thead>
<tr>
<th>Household</th>
<th>Assumption 1</th>
<th>Assumption 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>950 000</td>
<td>200 000</td>
</tr>
<tr>
<td>2</td>
<td>12 500</td>
<td>200 000</td>
</tr>
<tr>
<td>3</td>
<td>12 500</td>
<td>200 000</td>
</tr>
<tr>
<td>4</td>
<td>12 500</td>
<td>200 000</td>
</tr>
<tr>
<td>5</td>
<td>12 500</td>
<td>200 000</td>
</tr>
</tbody>
</table>

(The answer is at the end of the chapter)

1.7 Shifts of the demand curve

So far, we have been looking at the way a change in price affects the quantity demanded, depicted as a movement along the demand curve. However, when there is a change in the conditions of demand, the quantity demanded will change even if price remains constant. In this case, there will be a different price/quantity demand schedule and so a different demand curve. We refer to such a change as a shift of the demand curve.

Figure 3 depicts a rise in demand at each price level, with the demand curve shifting to the right, from $D_0$ to $D_1$. For example, at price $P_1$, demand for the good would rise from $X$ to $Y$. This shift could be caused by any of the following conditions of demand:

- A rise in household income (including a reduction in direct taxes).
- A rise in the price of substitutes.
- A fall in the price of complements.
- A change in tastes towards this product.
- An expected increase in the future price of the product.
- An increase in population.

Figure 3 shows an outward shift in the demand curve, but conversely a fall in demand at each price level would be represented by a shift in the opposite direction: to the left of the demand curve. Such a shift may be caused by the opposite of the conditions of demand shown above.
1.8 Demand, fashion and expectations

A change in fashion or tastes will also alter the demand for a product. For example, if it becomes fashionable for middle class households in Australia or the US to drink wine with their meals, expenditure on wine will increase. There may be passing 'crazes', such as roller blades or skateboards. And tastes can be affected by advertisers and suppliers trying to 'create' demand for their products. However, the effect of a product becoming fashionable will be that demand for it rises without its price having to be reduced.

If consumers believe that prices will rise, or that shortages will occur, they may attempt to stock up on the product before these changes occur. Again, this could lead to increases in demand despite the price of the good remaining unchanged.

Exam comments

The difference between a change in demand and a shift of the demand curve is of fundamental importance. Remember:

(a) Movements along a demand curve (contractions or expansions) for a good are caused solely by changes in its price.

(b) Variations in the conditions of demand create shifts in the demand curve.

Question 4: Substitutes and complements

In country X, DVDs are still a popular form of home entertainment. A recent fall in the price of DVDs has seen demand for DVDs increase significantly. However, cinema operators have reported a decline in customer numbers, which they believe is due to people preferring to buy DVDs to watch rather than going to the cinema.

What effect is the fall in the price of DVDs likely to have on the demand curves for:

(a) DVD players?

(b) cinema tickets?

(The answer is at the end of the chapter)
Section overview

- The supply curve shows the quantity of a good which would be supplied by producers at a given price and can be created both for an individual supplier and for all firms producing the good.
- There are short-run and long-run supply curves. Factors affecting supply depend on price and costs. A shift of the supply curve will occur when supply conditions (apart from the price) change.

2.1 The concept of supply

Supply refers to the quantity of a good that existing suppliers or would-be suppliers would want to produce for the market at a given price.

As with demand, supply relates to a period of time – for example, we might refer to an annual rate of supply or to a monthly rate.

The quantity of a good supplied to a market varies up or down for two reasons:

(a) Existing suppliers may increase or reduce their output quantities.
(b) Firms may stop production altogether and leave the market, or new firms may enter the market and start to produce the good.

If the quantity that firms want to produce at a given price exceeds the quantity that households (consumers) would demand, there will be an excess of supply, with firms competing to win what sales demand there is. Over-supply and competition would then be expected to result in price-competitiveness and a fall in prices.

As with demand, a distinction needs to be made:

(a) An individual firm’s supply schedule is the quantity of the good that the individual firm would want to supply to the market at any given price.
(b) Market supply is the total quantity of the good that all firms in the market would want to supply at a given price.

2.2 The supply curve

A supply schedule and supply curve can be created both for an individual supplier and for all firms which produce the good.

A supply curve is constructed in a similar manner to a demand curve (from a schedule of quantities supplied at different prices) but shows the quantity suppliers are willing to produce at different price levels. It is an upward sloping curve from left to right, because greater quantities will be supplied at higher prices.

We usually assume that suppliers aim to maximise their profits, and the upward slope of the supply curve reflects this desire to make profit (i.e. they are prepared to supply more of something the higher the price that is being paid for it).

2.3 Factors influencing the supply quantity

The quantity supplied of a good depends, as you might expect, on prices and costs. More specifically, it depends on the following factors:

(a) The costs of making the good. These include the costs of raw materials, labour and capital, which depend on the prices of the factors of production (wages, interest rates, land rents and profit expectations).
(b) The prices of other goods. When a supplier can switch readily from supplying one good to another, the goods concerned are called substitutes in supply. An increase in the price of one such good would make the supply of another good whose price does not rise less attractive to suppliers. When a production process has two or more distinct and separate outputs, the goods
produced are known as goods in joint supply or complements in production. Goods in joint supply include, for example, meat and hides. If the price of beef rises, more will be supplied and there will be an accompanying increase in the supply of cow hide.

(c) **Expectations of price changes.** If a supplier expects the price of a good to rise, the supplier is likely to try to reduce supply while the price is lower so more of the product or service may be supplied once the price is higher.

(d) **Changes in technology.** Technological developments which reduce costs of production (and increase productivity) will raise the quantity of supply of a good at a given price.

(e) **Other factors,** such as changes in the weather (for example, in the case of agricultural goods), natural disasters or industrial disruption.

The supply curve shows how the quantity supplied will change in response to a change in price. If supply conditions alter, a different supply curve must be drawn. In other words, a change in price will cause a change in supply along the supply curve. A change in other supply conditions will cause a shift in the supply curve itself.

### Exam comments

This distinction between a movement along the supply curve and a shift in the supply curve is just as important as the similar distinction relating to the demand curve.

### 2.4 Shifts of the market supply curve

The market supply curve is the aggregate of all the supply curves of individual firms in the market. A shift of the market supply curve occurs when supply conditions (other than the price of the good itself) change. Figure 4 shows a shift in the supply curve from $S_0$ to $S_1$. A rightward (or downward) shift of the curve shows an expansion of supply and may be caused by the factors below:

(a) **A fall in the cost of factors of production,** for example a reduction in the cost of raw material inputs.

(b) **A fall in the price of other goods.** The production of other goods becomes relatively less attractive as their price falls. Firms are therefore likely to shift resources away from the goods whose price is falling and into the production of higher priced goods that offer increased profits. We therefore expect that *(ceteris paribus)* the supply of one good will rise as the prices of other goods fall (and vice versa).

(c) **Technological progress,** which reduces unit costs and also increases production capabilities.

(d) **Improvements in productivity** or more efficient use of existing factors of production, which again will reduce unit cost.

A shift of the supply curve is the result of changes in costs, either in absolute terms or relative to the costs of other goods. If the price of the good is $P_1$, suppliers would be willing to increase supply from $Q_0$ to $Q_1$ under the new supply conditions (Figure 4).

Conversely, we might see a leftward (or upward) shift in the supply curve if the cost of supply increases. This would mean that at the existing price, a firm’s output will decrease and less will be supplied.

This is also illustrated in Figure 4: at price $P_1$, the quantity supplied now falls from $Q_0$ to $Q_2$, as the supply curve shifts from $S_0$ to $S_2$.

In order for the supplier to restore output levels to the original $Q_0$, price would have to increase to $P_2$.

An upward (leftward) shift ($S_0 \rightarrow S_2$) in supply could be caused by:

(a) **An increase in the cost of factors of production.**

(b) **A rise in the price of other goods** which would make them relatively more attractive to the producer.

(c) **An increase in indirect taxes,** or a reduction in a subsidy, which would make supply at existing prices less profitable.
We need to distinguish between short run and long run responses of both supply and demand. In the short run both supply and demand are relatively unresponsive to changes in price, as compared to the long run.

(a) In the case of supply, changes in the quantity of a good supplied often require the laying off or hiring of new workers, or the installation of new machinery. All of these changes, brought about by management decisions, take some time to implement.

(b) In the case of demand, it takes time for consumers to adjust their buying patterns, although demand will often respond more rapidly than supply to changes in price or other demand conditions.

In some markets, responses to changes in price are relatively rapid. In others, response times are much longer. In stock markets for example, the supply and demand for company shares respond very rapidly to price changes, whereas in the markets for fuel oils or agrichemicals response times are much longer.

Question 5: Quantity supplied

What effect will higher grain prices have on the supply curve of a cereal manufacturer who makes cereals from grain?

(The answer is at the end of the chapter)
(a) Market demand conditions influence the price that a firm will get for its output. Prices act as signals to producers, and changes in prices should stimulate a response from a firm to change its production quantities.

(b) Supply is influenced by production costs and profits. The objective of maximising profits provides the incentive for firms to respond to changes in price or cost by changing their production quantities.

(c) When a firm operates efficiently, responding to changes in market prices and controlling its costs, it is rewarded with profit.

Decisions by firms about what industry to operate in and what markets to produce for will be influenced by the prices obtainable. Although some firms have been established in one industry for many years, others are continually opening up, closing down or switching to new industries and new markets. Over time, firms in an industry might also increase or reduce the volume of goods they sell.

Sometimes, however, price will not represent the economic cost of a good or service. For example, price may be higher than costs due to taxes imposed by government. The sales price of cigarettes and alcohol is much higher than the cost of producing them, due to the imposition of taxes and duty.

By contrast, a subsidy may bring about an artificially low price.

However, in the main, we will be looking at price as an indicator of the exchange value of goods and services, as determined by the market forces of supply and demand.

### 3.2 The price mechanism and the equilibrium price

#### Definitions

The **price mechanism** brings demand and supply into equilibrium, and the **equilibrium price** for a good is the price at which the volume demanded by consumers and the volume that firms would be willing to supply is the same. This is also known as the **market clearing price**, since at this price there will be neither surplus nor shortage in the market.

The way demand and supply interact to come to the equilibrium price can be illustrated by drawing the market demand curve and the market supply curve on the same graph (Figure 5).

![Figure 5 Market equilibrium](image)

Figure 5 shows the planned (or ‘ex ante’) demand and planned supply at a set of prices.

At price $P_1$ in Figure 5, suppliers want to produce a greater quantity than the market demands, meaning that there is excess supply, equal to the distance $AB$. Suppliers would react as the stock of unsold goods accumulates.

(a) They would cut down the current level of production in order to sell unwanted inventories.
(b) They would also reduce prices in order to encourage sales.

The opposite will happen at price $P_2$, where there is an excess of demand over supply shown by the distance CD. Supply and price would increase. Faced with an excess of demand, manufacturers would be able to raise their prices. This would make supplying the good more profitable and supply would increase.

At price $P$ the amount that sellers are willing and able to supply is equal to the amount that customers are willing and able to buy. Consumers will be willing to spend a total of $(P \times Q)$ on buying $Q$ units of the product, and suppliers will be willing to supply $Q$ units to earn revenue of $(P \times Q)$. $P$ is the equilibrium price.

The forces of supply and demand push a market to its equilibrium price and quantity. Note carefully the following key points:

(a) If there is no change in conditions of supply or demand, the equilibrium price will prevail in the market and will remain stable.

(b) If price is not at the equilibrium, the market is in disequilibrium and supply and demand will push prices towards the equilibrium price.

(c) In any market there will only be one equilibrium position where the market is cleared.

(d) Shifts in the supply curve or demand curve will change the equilibrium price (and the quantity traded).

### 3.3 Consumer surplus and producer surplus

The marginal utility derived by different consumers from consumption of a unit quantity of a good will vary and so, therefore, will the price they would offer to buy that good. Because of this, consumers may be able to buy the good at a prevailing market price lower than the price they were prepared to pay.

You will be familiar with this idea from your own experience, where you would have been prepared to pay a certain amount for a holiday or a specific item of clothing, for example, but you have been able to buy them at a lower price than you had anticipated. This situation (which compares the consumer’s maximum price they are willing to pay with the price that the market requires them to pay) is called a consumer surplus, and it is represented as the light shaded triangle in Figure 6 below.

![Figure 6 Consumer surplus and producer surplus](image-url)

In the same way that consumers may be able to buy a good for less than they would be prepared to pay, so producers may be able to sell a good at a higher price than they would have accepted.

In this case, there is a producer surplus. Figure 6 illustrates this. The area of producer surplus on the graph represents the suppliers in the market who would be prepared to sell quantities of the good at less than the market price.
4 Demand and supply analysis

Section overview

- The effects of demand and supply conditions on markets can be analysed by studying the behaviour of both demand and supply curves through the case example of second-hand cars.

In this section we look at a Case study involving the analysis of demand and supply conditions.

We will examine the likely effects on the price and quantity sold of second-hand cars in the event of:

(a) A large increase in petrol prices.
(b) A big increase in the price of new cars.
(c) A massive investment in public transport.

Case study: Second-hand cars

Analysis

Petrol and cars are complementary products; any change in the market for petrol (part (a) of this Case study) would be expected to affect the market for second-hand cars. The demand for petrol, however, is likely to be relatively unresponsive to a change in price, so a change in price will only have a small impact on the quantity demanded. Consequently, a major change in its price will be necessary to affect the demand for any complementary product. Figure 7 assumes that there is a large increase in the price of fuel (petrol) as stated above.

Petrol and cars are complementary products; any change in the market for petrol (part (a) of this Case study) would be expected to affect the market for second-hand cars. The demand for petrol, however, is likely to be relatively unresponsive to a change in price, so a change in price will only have a small impact on the quantity demanded. Consequently, a major change in its price will be necessary to affect the demand for any complementary product. Figure 7 assumes that there is a large increase in the price of fuel (petrol) as stated above.

In this instance, we have assumed that the rise in the price of fuel results from a change in the conditions of supply. This is the basis of the new supply curve $S_1$ shifting to the left of the existing one (Figure 7 (i)). A rise in the price of fuel is a rise in the cost of owning and running a car. There will therefore be a fall in the demand for second-hand cars and a fall in the price and quantity sold (Figure 7 (ii)).

Part (b) of this Case study involves new vehicles and used vehicles as substitute products.

**Figure 7**

In this instance, we have assumed that the rise in the price of fuel results from a change in the conditions of supply. This is the basis of the new supply curve $S_1$ shifting to the left of the existing one (Figure 7 (i)). A rise in the price of fuel is a rise in the cost of owning and running a car. There will therefore be a fall in the demand for second-hand cars and a fall in the price and quantity sold (Figure 7 (ii)).

Part (b) of this Case study involves new vehicles and used vehicles as substitute products.
It is assumed that the increase in the price of new cars is the result of a major increase in supply costs. The rise in price causes a switch of demand into second-hand vehicles, so pushing up their price and leading to an increase in the number sold (Figure 8(ii)).

The increased price of new vehicles could alternatively result from an increase in the demand for them. Case study, part (c) involves another ‘product’ which is in competition with second-hand cars. If there is a reduction in the price of public transport services following the outward shift in supply, this could be the result (Figure 9).

In part (c) the fall in public transport prices leads to an expansion in demand for public transport (Figure 9(i)) while the demand for second-hand cars falls (with a new demand line $D_1$ in Figure 9(ii)) together with a fall in price. However, the relationship between public transport and the market for second-hand cars is likely to be a highly complex and indeterminate one. Therefore, people might make greater use of public transport while the ownership of cars (including second-hand cars) could continue to increase.

### 4.1 Price as a signal

As well as acting as an information system for buyers and sellers in a market, price can also act as a stimulant.

Price information may prompt buyers and sellers to change their behaviour. For example, a price rise may encourage firms to divert resources towards producing a good whose price has risen, in order to obtain a better reward from their resources.
### 5 Maximum and minimum prices

**Section overview**
- Where *maximum prices* (*ceiling prices*) are imposed, there will be excess demand: rationing may be necessary, and black marketeers may seek to operate. Where *minimum prices* (*floor prices*) are imposed, producers will make excess supply.
- The minimum wage is a floor price in the labour market.

#### 5.1 Price regulation

The regulation of prices provides an illustration of how demand and supply analysis can be applied. Governments might try to control prices in two ways:

(a) They might set a **maximum price** *(or price ceiling)* for a good, perhaps as part of an anti-inflationary economic policy. A recent example has been the Venezuelan government setting price ceilings on a variety of essential goods, including basic food. For example the government imposed a ceiling on the price of white rice, a staple diet of many Venezuelans. In 2009, with the annual rate of inflation running at 30 per cent, a price ceiling of 2.15 bolivars per kilo was imposed on white rice, when producers were claiming that the production cost was 4.41 bolivars. To try to ensure supply of white rice, the government was obliged to impose minimum production quotas on the producers.

(b) They might set a **minimum price** *(or price floor)* for a good. A European example is the European Union’s Common Agricultural Policy (CAP), which was first established in 1962. This set minimum prices for certain items of farm produce, and the European Commission (EC) purchased output in the market whenever the price fell to the minimum level. As a result the EC built up large stores of surplus produce, referred to as butter mountains, beef mountains, wine lakes and milk lakes, and so on.

#### 5.2 Maximum prices

Governments may try to prevent prices of goods rising by establishing a price ceiling **below** the equilibrium price. *(Note: the price ceiling has to be below the equilibrium price. If the price ceiling is higher than the equilibrium price, setting a price ceiling will have no effect at all on the operation of market forces. Make sure that you understand why this is so.)*

If the maximum price $M$ is lower than what the equilibrium price would be, there will be an excess of demand over supply (Figure 10). The low price attracts customers, but deters suppliers. Because the price ceiling $M$ is below the equilibrium price $P$, producers will reduce the quantity of goods supplied to the market place from $Q$ to $A$. However, the quantity demanded will increase from $Q$ to $B$ because of the fall in price. The excess quantity demanded is $AB$.

Because the market is now in disequilibrium, the limited supply has to be allocated by a means other than price.

To prevent an unfair allocation of the units of the good that are available, the government might have to introduce **rationing** (as with petrol coupons) or a **waiting list** (as for local authority housing). Rationing and **black marketeers** tend to go together. In Figure 10 consumers demand quantity $B$ but can only get $A$. However, for quantity $A$ they are prepared to pay price $Z$, which is well above the official price $M$. The black marketeers step in to exploit the gap. The commodity may be sold on ration at the official price $M$, but black marketeers may sell illicit production at price $Z$.

Note also that maximum prices can lead to a misallocation of resources. Producers will reduce output of the products subject to price controls because they are now relatively less profitable than those products not subject to price controls.
Figure 10 Maximum price below equilibrium price

Question 6: Equilibrium
Supply of and demand for good Q are initially in equilibrium as shown in the diagram below:

The government introduces a maximum price \( P \). What effect will this have on the quantity of good Q purchased?

A. it will rise from G to E  
B. it will rise from E to H  
C. it will fall from H to G  
D. it will fall from E to G

(The answer is at the end of the chapter)

5.3 Minimum prices
Minimum price legislation aims to ensure that suppliers earn at least the minimum price (or floor price) for each unit of output they sell.

If the minimum price is set below the market equilibrium there is no effect. But if it is set above the market price, it will cause an excess supply (see surplus 'AB' in Figure 11).

This was a serious problem in the Australian wool industry where the government wool buyer, the Australian Wool Corporation (AWC), had been operating a floor price since 1974. Wool which did not meet the floor price was bought by the AWC and stockpiled for sale when market prices rose. This worked effectively for some time until the demand for wool plummeted in the late 1980s. By 1991 the stockpile was approximately 4.7 million tonnes. The floor price scheme was abandoned in 1991, in 1998 the wool market was privatised, and the stockpile was completely sold by 2001.
In Figure 11, the minimum price $Z$ is set above the equilibrium price $P$. The quantity demanded falls from $Q$ to $A$ but the quantity supplied increases to $B$ because the higher price encourages suppliers to supply more. There is excess supply equal to the quantity $AB$.

![Figure 11 Minimum price above equilibrium price](image)

The problem with floor prices is that more of the good will be produced than can be sold at the minimum price, and so surplus quantities will build up, which have to be either stored or destroyed. Either way, the floor price leads to a misallocation of resources.

To try to prevent over-supply and ‘dumping’ of excess supply at low prices, a system of production quotas might be introduced whereby each supplier is only allowed to produce up to a maximum quantity and no more. OPEC, the Organisation of Petroleum Exporting Countries, sets production quotas for its 12 oil-producing member nations. These production quotas are firmly established and tightly controlled by OPEC. In Europe, the European Union (EU) tried to overcome the problem of excess supply on some products by imposing quotas on farmers. From 1992, the EU Common Agricultural Policy started to oblige farmers to take land out of production: this is called ‘set-aside’.

**Question 7: Floor prices**

What are the economic impacts of a floor price scheme such as the Common Agricultural Policy (CAP) of the European Union?

(The answer is at the end of the chapter)

There are also examples of artificially high prices reducing demand for goods where the prices are imposed by companies rather than government or legislation.

The OECD’s (Organisation for Economic Co-operation and Development) food price index details movement of food prices across countries. Analysis in November 2009 found that Australia’s food prices had risen by 41.3 per cent over a ten-year period in comparison to the OECD average of 33 per cent in the same period. Reasons for this included a prolonged drought and increased transport costs, but the fact that the majority of the Australian grocery market is dominated by two large companies was suspected as a major cause of the price rises. However, the reasons for the high price rises were never officially identified.

A European example also relates to the price of food items. A report to the European Union farm ministers in September 2009 found that despite the costs of butter, milk and cheese falling 39 per cent, 31 per cent and 18 per cent respectively in the previous year, the prices that supermarkets charged their customers had only fallen by 2 per cent. Consequently, shoppers’ demand for dairy products was lower than it would have been if more of the price cuts had been passed on to them, and the higher prices had led to the return of a ‘butter mountain’, paid for by EU taxpayers.
5.4 Minimum wages

A minimum wage is an application of floor pricing in the labour market. Over 90 per cent of countries have minimum wage legislation in place including Australia, New Zealand, the US, the UK, China and India. The purpose of a minimum wage is to ensure that low-paid workers earn enough to have an acceptable standard of living. If a minimum wage is enforced by legislation (a statutory minimum wage) or negotiated nationally for an industry by a trade union, the minimum wage will probably be above the current wage level for the jobs concerned. This would have two consequences:

- To raise wage levels for workers employed to a level above the ‘equilibrium’ wage rate. This is shown in Figure 12 by an increase in the wage rate from W to M.
- To create an excess of labour supply over demand for labour. This is shown in Figure 12 as the difference between Qx and Qm. This excess supply is caused partly by an increase in the supply of labour, from Qw to Qx, and partly by a fall in the demand for labour below the ‘equilibrium’ level, from Qw to Qm.

![Figure 12 Minimum wage](image-url)

Without a minimum wage, Qw workers would be employed at wage rate W (Figure 12). The (net) excess supply, i.e. unemployment, coming from the minimum wage, arises from a reduced demand for workers at the minimum wage and an increase in labour force participation.

**Question 8: Minimum wage**

By reference to Figure 12, work out what happens when a minimum wage M, higher than the existing rate W, is imposed.

(The answer is at the end of the chapter)

In practice, there may be a temptation for firms to ignore the floor price, for example by establishing informal arrangements with workers whereby they work for less than the minimum wage. However, while this would provide employment for the workers it raises ethical issues about their treatment.
Key chapter points

- The position of the demand curve is determined by the demand conditions, which include consumers' tastes and preferences, and consumers' incomes. The demand curve can be individual or show market demand. Market demand is influenced by national distribution of income.

- Substitution of goods will impact the demand of goods, as will complements. The impact of fashion and expectation in affecting demand cannot be underestimated.

- The supply curve shows the quantity of a good which would be supplied by producers at a given price and can be created both for an individual supplier and for all firms producing the good.

- There are short-run and long-run supply curves. Factors affecting supply depend on price and costs. A shift of the supply curve will occur when supply conditions (apart from the price) change.

- The competitive market process results in an equilibrium price, which is the price at which market supply and market demand quantities are in balance. In any market, the equilibrium price will change if market demand or supply conditions change.

- The price mechanism may also result in either a consumer or producer surplus.

- The effects of demand and supply conditions on markets can be analysed by studying the behaviour of both demand and supply curves through the case example of second-hand cars.

- Where maximum prices (ceiling prices) are imposed, there will be excess demand: rationing may be necessary, and black marketeers may seek to operate. Where minimum prices (floor prices) are imposed, producers will make excess supply.

- The minimum wage is a floor price in the labour market.
Quick revision questions

1. What factors influence demand for a good?
2. What are (a) substitutes and (b) complements?
3. What factors affect the supply quantity?
4. What is meant by equilibrium price?
5. A demand curve is drawn on all except which of the following assumptions?
   A. incomes do not change
   B. prices of substitutes are fixed
   C. price of the good is constant
   D. there are no changes in tastes and preferences
6. The diagram shown relates to the demand for and supply of Scotch whisky. The market is initially in equilibrium at point X. The government imposes a specific tax on Scotch while at the same time, the price of Irish whiskey (a substitute for Scotch whisky) rises. Which point, A, B, C or D, represents the new market equilibrium?

7. A price ceiling set above the equilibrium market price will result in:
   A. market failure
   B. excess supply over demand
   C. market equilibrium
   D. excess demand over supply
8. Which one of the following would normally cause a rightward shift in the demand curve for a product?
   A. a fall in the price of a substitute product
   B. a reduction in direct taxation on incomes
   C. a reduction in price of the product
   D. an increase in the price of a complementary product
9. What is an inferior good?
   A. a good of such poor quality that demand for it is very weak
   B. a good of lesser quality than a substitute good, so that the price of the substitute is higher
   C. a good for which the cross elasticity of demand with a substitute product is greater than 1
   D. a good for which demand will fall as household income rises
1. The price of the good.
   The price of other goods.
   Household income.
   Taste and fashion.

2. Substitutes are goods that are alternatives to each other (for example, Coca-Cola and Pepsi).
   Complements are goods which are bought and used together (for example, cars and petrol).

3. The price obtainable for the good.
   The prices obtainable for other goods, particularly goods in joint supply.
   The costs of making the good.
   Disruptions such as bad weather and strikes.

4. The price at which the volume of demand and the volume of supply are equal; there is neither surplus nor shortage.

5. Demand curves express the quantity demanded at each given market price. Non-price determinants such as income must be held constant when looking at the effect of price movements in isolation.

6. Supply shifts from $S_0$ to $S_1$, reflecting the per-unit tax. Demand shifts from $D_0$ to $D_1$ as the price of a substitute (Irish whiskey) rises.

7. If the price ceiling is above the equilibrium market price, it will not interfere with the working of the price mechanism. The market will not be forced from its current equilibrium. A price ceiling only affects the workings of the price mechanism if it is set below the equilibrium price.

8. A reduction in income tax will increase 'real' household income, and so demand for normal products will shift to the right, i.e. quantity demanded will be greater at any given price.
   A fall in the price of a substitute good would entice consumers away from the original good.
   This would cause a leftward shift in the demand curve.
   A change in the price of the good itself does not cause a shift in the curve but a movement along it.
   Complementary products tend to be bought and used together, so an increase in the price of one will lead to a reduction in demand for the other, reflected in a leftward shift in the demand curve.

9. Inferior goods are defined in terms of the relationship between quantity demanded and income. The issue of substitutes is not relevant.
1. Demand rises to (approximately) 6 kilos at the reduced price of $2 per kilo. If price rises to $4 per kilo, demand falls to (approximately) 3 kilos.

2. (a) Domestic deep freezers and perishable products are complements because people buy deep freezers to store perishable products.

(b) Perishable products are supplied either as fresh produce (for example, fresh meat and fresh vegetables) or as frozen produce, which can be kept for a short time in a refrigerator but for longer in a freezer. The demand for frozen produce will rise, while the demand for fresh produce will fall.

(c) Wider ownership of deep freezers is likely to increase bulk buying of perishable products. Suppliers can save some packaging costs, and can therefore offer lower prices for bulk purchases.

3. Under assumption 1, the demand for swimming pools will be confined to household 1. Even if this household owns three or four properties, the demand for swimming pools is likely to be less than under assumption 2, where potentially all five households might want one.

4. (a) There is likely to have been an outward shift in the demand curve for DVD players. DVD players are complements to the DVDs themselves because people will need to buy the DVD players in order to watch their DVDs. So, the increased demand for DVDs will lead to an increase in demand for DVD players even though their price may be unchanged. This results in an outward shift in their demand curve (an expansion of demand).

(b) There is likely to be an inward shift in the demand curve for cinema tickets (a contraction of demand). Even though the price of cinema tickets has not changed, people are demanding less of them because they are choosing to watch DVDs instead. Cinema tickets are a substitute product to DVDs, and a fall in the price of a substitute leads to an inward shift of the demand curve for a product.

5. The higher price of grain will cause the supply curve to shift leftwards (or upwards). The increase in grain prices increase the cereal manufacturer’s production costs, making supply at existing prices less profitable.

6. D The initial equilibrium quantity is E (where supply and demand curves intersect). Quantity demanded at the controlled price $P$ will be $H$. However, only quantity $G$ will be supplied and purchases will therefore be limited to this amount. Therefore, the quantity purchased will fall from $E$ to $G$ due to the shortage of supply available.

7. If a floor price scheme guarantees a minimum price above the market equilibrium price, it encourages an excess of supply over demand. This will result in surplus production, and a build-up of surplus inventory, such as a butter mountain or wine lake. To get rid of the surplus inventories, it may be necessary to ‘dump’ them by selling them at very low prices to another country. However, it provides price and income stability for the farmers producing the goods.

8. The supply curve for labour is now the line $MXY$. Demand for labour from employers will fall to $Q_m$, but $Q_m$ workers will at least earn a higher wage. However, ceteris paribus, the imposition of the minimum wage would create unemployment at $Q_w - Q_m$. 

Economics and Markets
Chapter 3

Elasticity of demand and supply

<table>
<thead>
<tr>
<th>Learning objectives</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Elasticity of demand</td>
<td>LO3</td>
</tr>
<tr>
<td>Explain the concepts of elasticity of demand and elasticity of supply</td>
<td>LO3.1</td>
</tr>
<tr>
<td>Calculate and interpret the elasticity of demand and elasticity of supply</td>
<td>LO3.2</td>
</tr>
<tr>
<td>Prepare demand curves for necessities and luxury goods</td>
<td>LO3.3</td>
</tr>
</tbody>
</table>

Topic list

1. Elasticity of demand
2. Income elasticity of demand
3. Elasticity of supply
We discussed in the previous chapter the direction of changes in demand and supply when prices change. When the price of a normal good goes up, the quantity demanded will fall, and the quantity suppliers will be willing to supply will go up.

But if prices are increased or reduced, how much will this affect the amount of revenue from selling a good?

How much will supply of a good change if the price changes?

How far do changes in household income affect the demand for a good?

In this chapter, we consider such questions, and introduce the key concept of elasticity. You need to have a good understanding of this concept, because it could be relevant in a number of different contexts in the exam.
Before you begin

If you have studied these topics before, you may wonder whether you need to study this chapter in full. If this is the case, please attempt the questions below, which cover some of the key subjects in the area.

If you answer all these questions successfully, you probably have a reasonably detailed knowledge of the subject matter, but you should still skim through the chapter to ensure that you are familiar with everything covered.

There are references in brackets indicating where in the chapter you can find the information, and you will also find a commentary at the back of the Study Manual.

1. What is the co-efficient of price elasticity of demand (PED)?
   (Section 1.1)

2. When is demand perfectly inelastic?
   (Section 1.2)

3. Describe the Giffen demand curve.
   (Section 1.5.1)

4. Describe an ostentation demand curve.
   (Section 1.5.2)

5. What is the basic equation for elasticity of supply?
   (Section 3.1)

6. List three factors that affect a firm’s elasticity of supply.
   (Section 3.2)
1 Elasticity of demand

Section overview

- Demand for a good depends largely on price, household income and the relative price of substitutes or complementary goods. Changes in any of these will cause either a movement along the demand curve or a shift in the demand curve.

LOs

3.1
3.2

Price elasticity of demand indicates the responsiveness in the quantity demanded of a good to a change in the price of that good.

1.1 The price elasticity of demand

Exam comments

Elasticity is a common examination topic. It is also of great practical importance in the real world of business and you will need to use it again as you progress in your studies. Make sure you have a good understanding of the concept of elasticity.

We might want to study the extent of the change in demand for a good given a change in its price. It might seem that this can be done simply by looking at the demand curve for the good. For example a steep-sloping demand curve might suggest that demand does not change by much given a change in the price.

However, the slope of the demand curve can be misleading. A demand curve shows units of demand and units of price, and the slope of the demand curve depends on how units of measurement are represented on the graph.

We need a different measure for the responsiveness of demand to any given change in price. This is provided by the price elasticity of demand.

Elasticity, in general, refers to the relationship between two variables. Price elasticity of demand explains the relationship between change in quantity demanded and changes in price. This method of calculation is used as an alternative to the demand curve and measures the elasticity in percentage terms, as opposed to units of measurement.

If prices went up by 10 per cent, would the quantity demanded fall by the same percentage?

Definition

Price elasticity of demand (PED) is a measure of the extent of change in the market demand for a good in response to a change in its price.

The coefficient of PED is measured as:

\[
\text{Percentage change in quantity demanded} \over \text{Percentage change in price}
\]

Since demand usually increases when the price falls, and decreases when the price rises, elasticity has a negative value. However, it is usual to ignore the minus sign and just describe the absolute value of the coefficient.

This can be expressed as:

\[
\frac{\Delta Q}{Q} \times 100 \\
\frac{\Delta P}{P} \times 100
\]

Where: \( \Delta \) is the symbol for ‘change in’

\( Q \) is the quantity demanded of the good

\( P \) is the price of the good
If we are measuring the responsiveness of demand to a large change in price, we can measure elasticity between two points on the demand curve, and the resulting measure is called the arc elasticity of demand. The calculation of elasticity depends upon the reference point chosen. For example, a price decrease that moves down the demand curve from a Point A to Point B would show a different answer if the same price increase moves up the demand curve from Point B to Point A.

The arc elasticity of demand is calculated from the percentage change in quantity relative to average quantity for the relevant range of output and from the percentage price change relative to the average of the corresponding price range.

If we wish to measure the responsiveness of demand at one particular point in the demand curve, we can calculate a point elasticity of demand, without averaging price and quantity over a range. In doing so, it is convenient to assume that the demand curve is a straight line unless told otherwise.

**Worked Example: Arc elasticity of demand**

The price of a good is $1.20 per unit and annual demand is 800,000 units. Market research indicates that an increase in price of 10 cents per unit will result in a fall in annual demand of 70,000.

**Required**

What is the price elasticity of demand measuring the responsiveness of demand over this range of price increase?

**Solution**

Annual demand at $1.20 per unit is 800,000 units.

Annual demand at $1.30 per unit is 730,000 units.

Average quantity over the range is 765,000 units.

Average price is $1.25.

\[
\text{% change in demand} = \left(\frac{-70000}{765000}\right) \times 100\% = 9.15\%
\]

\[
\text{% change in price} = \left(\frac{10c}{125c}\right) \times 100\% = 8\%
\]

Price elasticity of demand = \(-\frac{9.15}{8}\) = –1.14

Ignoring the minus sign, the arc elasticity is 1.14.

The demand for this good, over the range of annual demand 730,000 to 800,000 units, is elastic because the price elasticity of demand is greater than 1. Now try the following question yourself.

**Question 1: Arc price elasticity of demand**

If the price per unit of X rises from $1.40 to $1.60, it is expected that monthly demand will fall from 220,000 units to 200,000 units.

**Required**

What is the arc price elasticity of demand over these ranges of price and output?

(The answer is at the end of the chapter)

**Worked Example: point elasticity of demand**

The price of a good is $1.20 per unit and annual demand is 800,000. Market research indicates that an increase in price of 10 cents per unit will result in a fall in annual demand of 70,000 units.

**Required**

Calculate the elasticity of demand at the current price of $1.20.
Solution

We are asked to calculate the elasticity at a particular price. We assume that the demand curve is a straight line.

At a price of $1.20, annual demand is 800,000 units. For a price rise:

\[
\frac{\text{% change in demand}}{\text{% change in price}} = \frac{\frac{70,000}{800,000} \times 100\%}{\frac{10\text{c}}{120\text{c}} \times 100\%} = \frac{8.75\%}{8.33\%} = -1.05
\]

Ignoring the minus sign, the price elasticity at this point is 1.05. Demand is elastic at this point, because the elasticity is greater than 1.

Question 2: Point price elasticity of demand

If the price per unit of x rises from $1.40 to $1.60, it is expected that monthly demand will fall from 220,000 units to 200,000 units.

What is the point price elasticity of demand when the price is $1.40?

(The answer is at the end of the chapter)

Exam comments

If it is not clear from the details of an examination question whether you need to calculate arc or point elasticity, then calculate the point elasticity.

Question 3: Range of elasticity

A shop sells 100 shirts each month at a price of $20. When the price is increased to $24, the total sales revenue rises by 14%. Within which range does the price elasticity of demand lie?

A  under 0.15
B  greater than 0.15 and less than 0.5
C  greater than 0.5 and less than 1.0
D  greater than 1.0

1.2 Elastic and inelastic demand

The value of demand elasticity may be anything from zero to infinity.

- Demand is inelastic if the absolute value is less than 1.
- Demand is elastic if the absolute value is greater than 1.

Think about what this means if there is an increase in price. Where demand is inelastic, the quantity demanded falls by a smaller percentage than the rise in price. Where demand is elastic, demand falls by a larger percentage than the rise in price.
1.2.1 Price elasticity and the slope of the demand curve

Generally, demand curves slope downwards. Consumers are willing to buy more at lower prices than at higher prices. Except in certain special cases (which we look at below), elasticity will vary in value along the length of a demand curve.

Therefore, it is not possible to state the comparative elasticities of any two curves over different price ranges simply by looking at the slopes of the curves. However, it is possible to say that if a downward sloping demand curve shifts to become steeper over a particular range of quantity, then demand is becoming more inelastic. Conversely, a demand curve becoming shallower over a particular range indicates more elastic demand.

The ranges of price elasticity ($\eta$) at different points on a downward sloping straight line demand curve are illustrated in Figure 1.

![Figure 1 Ranges of price elasticity](image)

At higher prices on a straight line demand curve (the top of the demand curve), small percentage price reductions can bring large percentage increases in quantity demanded. This means that demand is elastic over these ranges.

At lower prices on a straight line demand curve (the bottom of the demand curve), large percentage price reductions can bring small percentage increases in quantity. This means that demand is inelastic over these price ranges.

1.2.2 Special values of price elasticity of demand

There are three special values of price elasticity of demand: 0, 1 and infinity:

(a) **Demand is perfectly inelastic**: $\eta = 0$. There is no change in quantity demanded, regardless of the change in price. When there is a constant elasticity of demand and this is always perfectly inelastic, the demand curve is a **vertical straight line**.

(b) **Perfectly elastic demand**: $\eta = \infty$ (infinitely elastic). Consumers will want to buy an infinite amount, but only up to a particular price level. Any price increase above this level will reduce demand to zero. When there is a constant elasticity of demand and this is perfectly elastic, the demand curve is a **horizontal straight line**.

(c) **Unit elasticity of demand**: $\eta = 1$. Total revenue for suppliers (which is the same as total spending on the product by households) does not change however the price changes. The demand curve of a good whose price elasticity of demand is 1 over its entire range is a **rectangular hyperbola** (Figure 2).

This means that in Figure 2, rectangles OABC, ODEF and OGHJ all have the same area, since the areas of these rectangles represent total spending by customers at each price.
If the selling price were \( D \), total demand would be \( F \) and total spending on the product would be \( D \times F \) (rectangle ODEF).

If the selling price were \( A \), total demand would be \( C \), and total spending on the product would be \( A \times C \) (rectangle OABC).

If the selling price were \( G \), total demand would be \( J \) and total spending on the product would be \( G \times J \) (rectangle OGHJ).

![Figure 2 Unit elasticity of demand](image)

Table 1 summarises the different price elasticities of demand:

<table>
<thead>
<tr>
<th>Level of elasticity</th>
<th>Coefficient value</th>
<th>Actual examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Perfectly inelastic</td>
<td>0 (zero)</td>
<td>-</td>
</tr>
<tr>
<td>Relatively inelastic</td>
<td>Between 0 – 1</td>
<td>Medicines; water</td>
</tr>
<tr>
<td>Unit elasticity</td>
<td>1</td>
<td>-</td>
</tr>
<tr>
<td>Relatively elastic</td>
<td>&gt; 1</td>
<td>Holidays; luxury cars</td>
</tr>
<tr>
<td>Perfectly elastic</td>
<td>( \infty ) (infinity)</td>
<td>-</td>
</tr>
</tbody>
</table>

Table 1: Summary of price elasticity

1.3 Significance of price elasticity of demand

The remainder of Section 1 will enable students to:
- Distinguish between the **factors influencing elasticity**.
- **Measure** price elasticity from given price and demand data, and to draw appropriate conclusions from such information.
- Draw the correct implications for the **total revenue** of the producer of changes in the price of the product.

**The price elasticity of demand is relevant to total spending** on a good or service. Total expenditure is a matter of interest to both suppliers, to whom sales revenue accrues, and to government, who may receive a proportion of total expenditure in the form of taxation.

When demand is **elastic**, an increase in price will result in a greater than proportional fall in the quantity demanded, and **total expenditure will fall**. In Figure 3, total expenditure at price \( P_A \) is represented by the area \( OPA_QA \) and total expenditure at price \( P_B \) is represented by the area \( OP_BQ_B \). Area \( OPA_QA \) is
greater than area $OP_0BQ_0$; this can be seen by observing that area $Y$ (expenditure lost on a rise in price from $A$ to $B$) is greater than area $X$ (expenditure gained).

*Figure 3 Elastic demand*

When demand is inelastic, an increase in price will still result in a fall in quantity demanded, but the fall in quantity demanded will be less than proportional to the rise in price so **total expenditure will rise**. In Figure 4, area $X$ (expenditure gained) is greater than area $Y$ (expenditure lost).

*Figure 4 Inelastic demand*

With **unit elasticity**, expenditure will stay constant regardless of a change in price. In Figure 5, area $X$ and area $Y$ are the same.
Information on price elasticity of demand indicates how consumers can be expected to respond to different prices. Business people can make use of information on how consumers will react to pricing decisions as it is possible to trace the effect of different prices on total revenue and profits. Information on price elasticities of demand will be useful to a business which needs to know the price decrease necessary to clear a surplus (excess supply) or the price increase necessary to eliminate a shortage (excess demand). Elasticity is also useful, more generally, for a firm if it is considering changing the price of a good or service, and considering the impact this change will have on revenues.

Government policy makers can also use information about elasticity, for example when making decisions about indirect taxation. Items with a low price elasticity of demand such as cigarettes and alcohol tend to be useful targets for taxation since by increasing taxes on these, total revenue can be increased. If demand for cigarettes was price elastic, increases in taxation would be counter-productive as they would result in lower government revenue.

Question 4: Elasticity and total revenue

Suppose that there are two products, A and B.

Product A currently sells for $5, and demand at this price is 1,700 units. If the price fell to $4.60, demand would increase to 2,000 units.

Product B currently sells for $8 and demand at this price is 9,500 units. If the price fell to $7.50, demand would increase to 10,000 units.

In each of these cases, calculate:

(a) the price elasticity of demand (PED) for the price changes given.

(b) the effect on total revenue, if demand is met in full at both the old and the new prices, of the change in price.

Use the point price elasticity of demand method.

(The answer is at the end of the chapter)
1.5 Positive price elasticities of demand

When the price of a good rises, there will be a substitution effect: consumers will buy other goods instead because they are now relatively cheaper. But there will also be an income effect in that the rise in price will reduce consumers’ real incomes, and will therefore affect their ability to buy goods and services.

1.5.1 Giffen goods (necessities)

The 19th century economist, Sir Robert Giffen observed that this income effect could be so great for certain goods (called Giffen goods) that the demand curve would be upward sloping. In essence, it is an extreme form of an inferior good (discussed in Chapter 2). The price elasticity of demand for a Giffen good would be positive whereas, mathematically, the price elasticity of demand for normal goods is negative.

Giffen observed that among the labouring classes of his day, consumption of bread rose when its price rose. This could happen because the increase in price of a commodity which made up a high proportion of individuals’ consumption could have a significant effect on real incomes. People would have to increase their consumption of bread because they were no longer able to afford more expensive foods. A more recent example has been witnessed in several provinces in China, with more rice being consumed when its price rose (as when the price was lower, people were able to decrease their consumption of rice in favour of more expensive products such as meat).

1.5.2 Ostentation goods (luxuries)

The demand curve for a good might also slope upwards if it is bought for purposes of ostentation. In this case, having a higher price tag makes the good more desirable to consumers and thus increases demand once price goes above a certain point. At high levels of income, price increases lead to increases in demand through a snob effect operating through the income effect (inferiority). These goods will therefore also have positive elasticities of demand. Examples of such goods are items such as designer handbags, luxury cars and high-end French wines.

1.6 Factors influencing price elasticity of demand for a good

Factors that determine price elasticity of demand (PED) are similar to the factors other than price that affect the volume of demand. The PED is really a measure of the strength of these other influences on demand. Relative size of price elasticity depends on availability of substitutes, percentage of income spent on the good and the income effect. These, and the other main factors affecting PED are discussed below:

- Percentage of income spent on the good.
- Availability of substitutes.
- Necessity versus luxury.
- The time horizon.
- Habit.
- Definition of market.

1.6.1 Percentage of income spent on the good

The percentage of income spent on a good or goods concerns the size of the income effect. If expenditure on a good only constitutes a small proportion of a consumer’s income, then a change in the price of that good will not have much impact on the consumer’s overall real income. Therefore, demand for low price goods (such as safety matches) is likely to be inelastic. By contrast, demand is likely to be elastic for expensive goods. Percentage of overall real income spent will also vary depending on whether it is a small quantity of expensive goods or a large quantity of low cost goods.

1.6.2 Availability of substitutes

The more substitutes there are for a good, especially close substitutes, the more elastic the price elasticity of demand for the good will be. For example, the elasticity of demand for a particular brand of breakfast cereal will be much greater than the elasticity of demand for breakfast cereals as a whole, because the former have both more, and also closer, substitutes. A rise in the price of a particular brand of cereal is likely to result in customers switching their demand to a rival brand. Availability of substitutes is probably the most important influence on price elasticity of demand.
1.6.3 Necessity

Demand for goods which are necessary for everyday life (for example, basic foodstuffs) tends to be relatively inelastic while demand for luxury goods tends to be elastic. If a good is a luxury, and its price rises the rational consumer may well decide he or she no longer needs that good and so demand for it will fall. However, if a good is a necessity, the consumer will have to continue buying it even though its price has increased.

For luxury goods, the income elasticity of demand (see Section 2 below) tends to become more significant than price elasticity of demand.

1.6.4 The time horizon

If the price of a good is increased, there might initially be little change in demand because the consumer may not be fully aware of the increase, or may not have found a suitable substitute for the product. Then, as consumers adjust their buying habits in response to the price increase, demand might fall substantially. The time horizon influences elasticity largely because the longer the period of time which we consider, the greater the knowledge of substitution possibilities by consumers and the provision of substitutes by producers. Therefore, elasticity tends to increase as the time period increases.

1.6.5 Habit

Goods which are habit-forming tend to be inelastic, because the consumer 'needs' the goods despite their increase in price. This pattern can be seen with addictive products such as cigarettes.

1.6.6 Definition of market

If a market is narrowly defined (for example, breakfast cereals) there will be a number of competing brands and substitute products available so these brands will be price elastic. If a market is only broadly defined (for example, food) there will be fewer generic alternatives and so demand will tend to be inelastic.

2 Income elasticity of demand

<table>
<thead>
<tr>
<th>Section overview</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Income elasticity of demand</strong> measures the responsiveness of demand to changes in household income. <strong>Cross elasticity of demand</strong> is determined by the availability of substitute (competitors') products.</td>
</tr>
</tbody>
</table>

It is possible to construct other measures of elasticity than price elasticity, and an important one which you need to know about is *income elasticity of demand*. The income elasticity of demand for a good indicates the responsiveness of demand to changes in *household incomes*.

When household income rises people will not only increase their demand for existing goods but also start to demand new goods which they could not previously afford.

\[
\text{Income elasticity of demand} = \frac{\% \text{ change in quantity demanded}}{\% \text{ change in income}}
\]

(a) Demand for a good is **income elastic** if income elasticity is greater than 1; in other words, if quantity demanded rises by a larger percentage than the rise in income. For example, if the demand for compact discs will rise by 10 per cent if household income rises by 7 per cent, we would say that the demand for compact discs is income elastic.

(b) Demand for a good is **income inelastic** if income elasticity is between 0 and 1, and the quantity demanded rises less than the proportionate increase in income. For example, if the demand for books rises by 6 per cent if household income rises by 10 per cent, we would say that the demand for books is income inelastic.

The change in quantity demanded takes the form of a shift in the position of the demand curve, not a movement along it, since it is not stimulated by a change in price.
Goods whose income elasticity of demand is positive are said to be **normal goods**, meaning that demand for them will rise when household income rises. If income elasticity is **negative**, the commodity is called an **inferior good** since demand for it falls as income rises.

Inferiority in this sense, is an observable fact about the consumer’s demand preferences, rather than a statement about the quality of the good itself.

Inter-city bus travel is an example of an inferior good. Bus travel is cheaper than air or rail travel, but takes longer. When consumers have limited income, they are prepared to forgo the increased time taken to travel in return for the cheaper cost. However, as their income increases, they will choose the more rapid modes of transport over the slower, albeit cheaper, bus travel. Therefore, demand for bus travel will fall as income rises.

Income elasticity of demand can be summarised as follows:

<table>
<thead>
<tr>
<th>Elasticity</th>
<th>Value</th>
<th>Type of good</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>Negative</td>
<td>-ve</td>
<td>Inferior</td>
<td>Inter-city bus travel</td>
</tr>
<tr>
<td>Inelastic</td>
<td>0 – 1</td>
<td>Necessity</td>
<td>Basic food stuffs</td>
</tr>
<tr>
<td>Elastic</td>
<td>&gt; 1</td>
<td>Luxury</td>
<td>Yachts, sports cars</td>
</tr>
</tbody>
</table>

For most commodities, an increase in income will increase demand. The exact effect on demand will depend on the type of product. For example, the demand for some products like bread will not increase much as income rises. Therefore, bread has a low income elasticity of demand. In contrast, the demand for luxuries increases rapidly as income rises and luxury goods therefore have a high income elasticity of demand.

**Question 5: Elasticity of demand and supply**

What will be the effect on price, quantity demanded and quantity supplied of luxury sports cars, given a significant reduction in income tax?  

(The answer is at the end of the chapter)

**2.1 Cross elasticity of demand**

**Definition**

**Cross elasticity of demand** is the responsiveness of quantity demanded for one good following a change in price of another good.

\[
\text{Cross elasticity of demand} = \frac{\% \text{ change in quantity of good A demanded} \times \% \text{ change in price of good B}}{\% \text{ change in quantity of good A demanded} \times \% \text{ change in price of good B}}
\]

*(assuming no change in the price of A)*

The cross elasticity of demand depends upon the degree to which goods are **substitutes or complements**.

(a) If the two goods are **substitutes**, cross elasticity will be **positive** and a fall in the price of one will reduce the amount demanded of the other.

(b) If the goods are **complements**, cross elasticity will be **negative** and a fall in the price of one will raise demand for the other.

For example, assume bread and butter are complements. If the price of bread increases, demand for bread and then, in turn, butter will decrease. So, as a fraction, we have a negative value (quantity of butter demanded) over a positive value (increase in price) meaning the cross elasticity of demand is negative.

Cross elasticity involves a comparison between two products. The concept is a useful one in the context of considering substitutes and complementary products.
The value of the cross-elasticity of demand shows the strength of the relationship between the two goods. Indices tending towards 1 or -1 indicate a strong relationship, statistics tending towards zero indicate a weak relationship.

<table>
<thead>
<tr>
<th>Cross-elasticity</th>
<th>Value</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>Complements</td>
<td>Negative</td>
<td>Bread and butter</td>
</tr>
<tr>
<td>Unrelated products</td>
<td>0</td>
<td>Bread and cars</td>
</tr>
<tr>
<td>Substitutes</td>
<td>Positive</td>
<td></td>
</tr>
</tbody>
</table>

**Exam comments**

Make sure you know how to calculate the different types of elasticity, and how to distinguish between complements and substitutes.

Remember that cross elasticity of demand (like other measures of elasticity) measures the **percentage** change in the quantity demanded for one good, in response to the **percentage** change in the price of another.

### 3 Elasticity of supply

**Section overview**

- Elasticity of supply measures the responsiveness of the quantity of a good supplied following a change in the price of that good. There are various factors affecting a firm’s capacity to supply.
- Elasticity of supply varies according to the timeframe in which it is measured. These timeframes are market, short, long and secular.

#### 3.1 Price elasticity of supply

**Definition**

The **price elasticity of supply** indicates the responsiveness of supply to a change in price.

\[
\text{Elasticity of supply} = \frac{\% \text{ change in quantity supplied}}{\% \text{ change in price}}
\]

Where the supply of goods is **fixed** whatever price is offered, for example in the case of antiques, vintage wines and land, supply is **perfectly inelastic** and the elasticity of supply is **zero**. The **supply curve is a vertical straight line**.

Where the supply of goods **varies proportionately** with the price, **elasticity of supply equals one** and the supply curve is a straight line **passing through the origin**. (Note that a demand curve with unit elasticity along all of its length is not a straight line, but a supply curve with unit elasticity is a straight line.)

Where the producers will **supply any amount at a given price** but none at all at a slightly lower price, elasticity of supply is infinite, or **perfectly elastic**. The **supply curve is a horizontal straight line**.

Perfectly inelastic supply, unit elastic supply and perfectly elastic supply are illustrated in Figure 6.
Elasticity of demand and supply

3. Elasticity of supply (i)

Note that a supply curve with unit elasticity can have many different gradients. The key feature that identifies the unit elasticity is not the gradient of the curve, but the fact that it passes through the origin.

Supply is elastic (greater than 1) when the percentage change in the amount producers want to supply is greater than proportional to the percentage change in price. In Figure 7 below, supply is elastic all the way along the supply curve for a linear supply curve that has a positive value for price when the supply is 0. However, elasticity moves towards 1 as the price rises.

Supply is inelastic (less than 1) when the amount producers want to supply changes by a smaller percentage than the percentage change in price. In Figure 7 below, supply is inelastic all the way along the supply curve for a linear supply curve that has a positive supply quantity when the price is 0. However, elasticity moves towards 1 as the price rises.

3.2 Factors affecting elasticity of supply

Elasticity of supply is a measure of firms’ ability to adjust the quantity of goods they supply. This depends on a number of constraints:

(a) Existence of inventories of finished goods: If a firm has large inventories of finished goods then it can draw on these to increase supply following an increase in the price of the good. Supply will be relatively elastic. Perishability or shelf life are important considerations here though.

(b) Availability of labour: When unemployment is low it may be difficult to find employees with the appropriate skills.

(c) Spare capacity: If a firm has spare capacity (e.g. machinery which is not being fully utilised) it can quickly and easily increase supply following an increase in price. In this way, spare capacity is likely to increase elasticity of supply.
3.2.1 Elasticity of supply and time

As with elasticity of demand, the elasticity of supply of a product varies according to the time period over which it is measured. For analytical purposes, four lengths of time period may be considered:

(a) **The market period** is so short that supplies of the commodity in question are limited to existing inventories. In effect, supply is fixed.

(b) **The short run** is a period long enough for supplies of the commodity to be altered by increases or decreases in current output, but not long enough for the fixed equipment (plant, machinery and so on) used in production to be altered. This means that suppliers can produce larger quantities only if they are not already operating at full capacity; they can reduce output fairly quickly by means of redundancies and laying-off staff.

(c) **The long run** is a period sufficiently long to allow firms’ fixed equipment to be altered. There is time to build new factories and machines, and time for old ones to be closed down. New firms can enter the industry in the long run.

(d) **The secular period** is so long that underlying economic factors such as population growth, supplies of raw materials (such as oil) and the general conditions of capital supply may alter. The secular period is ignored by economists except in the theory of economic growth.

In general, supply tends to be more elastic in longer time periods.

3.3 Response to changes in demand

The price elasticity of supply can be seen, in effect, as a measure of the readiness with which an industry responds following a shift in the demand curve.

Suppose that there is an increase in the demand for restaurant meals in a city, shown by the rightward shift in the demand curve in Figure 8 from $D_1$ to $D_2$. The capacity of the industry is limited in the short run by the number of restaurants in operation. The restaurants can be used more intensively to a certain extent, and so supply ($S_1$) is not perfectly inelastic, but there is a limit to this process. As a result, in the short run there is a large increase in the price from $P_1$ to $P_2$. 
Figure 8 Response to a shift in the demand curve

The rise in price in the short run will encourage entrepreneurs to open new restaurants to take advantage of the profits to be earned. In the long run, supply is consequently more elastic and is shown by supply curve $S_2$. The expanded output in the industry leads to a new equilibrium at a lower price $P_3$ with the new level of output being $Q_3$.

**Question 6: Elasticity of supply**

Which diagram shows perfectly elastic supply?

- A
- B
- C
- D

(The answer is at the end of the chapter)
Key chapter points

- Demand for a good depends largely on price, household income and the relative price of substitutes or complementary goods. Changes in any of these will cause either a movement along the demand curve or a shift in the demand curve.

- Price elasticity of demand indicates the responsiveness in the quantity demanded of a good to a change in the price of that good.

- Candidates will be able to:
  - distinguish between the factors influencing elasticity.
  - measure price elasticity from given price and demand data, and to draw appropriate conclusions from such information.
  - draw the correct implications for the total revenue of the producer of changes in the price of the product.

- Income elasticity of demand measures the responsiveness of demand to changes in household income.

- Cross elasticity of demand is determined by the availability of substitute (competitors') products.

- Elasticity of supply measures the responsiveness of the quantity of a good supplied following a change in the price of that good. There are various factors affecting a firm's capacity to supply.

- Elasticity of supply varies according to the timeframe in which it is measured. These timeframes are market, short, long and secular.
Quick revision questions

1. What is meant by the price elasticity of demand (PED) for a commodity?

2. What is the significance of PED to:
   (a) a manufacturer?
   (b) a national government's treasurer?

3. What determines the cross elasticity of demand between two goods?

4. If the absolute value of the price elasticity of demand for dry white wine is greater than 1, a decrease in the price of all wine would result in:
   A. a more than proportional decrease in the quantity of dry white wine purchased.
   B. a less than proportional decrease in the quantity of dry white wine purchased.
   C. a less than proportional increase in the quantity of dry white wine purchased.
   D. a more than proportional increase in the quantity of dry white wine purchased.

5. Which combination of demand and supply curves would be appropriate for a firm attempting to increase its profits by increasing its market share?
   A. inelastic demand, inelastic supply
   B. elastic demand, elastic supply
   C. inelastic demand, elastic supply
   D. elastic demand, inelastic supply

6. Fish and chips are considered complementary products. If the price of fish rises, what will the impact be in demand for chips?
   A. rises
   B. stays the same
   C. falls
   D. doubles

7. Using the point method, what is the price elasticity of demand of product X as price falls from its current price of $20 to $15?

<table>
<thead>
<tr>
<th>Old</th>
<th>New</th>
</tr>
</thead>
<tbody>
<tr>
<td>Price ($)</td>
<td>20</td>
</tr>
<tr>
<td>Quantity (units)</td>
<td>10</td>
</tr>
</tbody>
</table>

   A. 0.5
   B. 1
   C. 1.5
   D. 2

8. Which of the following statements is true?
   I. If the price elasticity of demand is more than 1, a fall in price will result in a fall in total expenditure on the good.
   II. The income elasticity of demand will only be zero in the case of inferior goods.
   III. The cross-elasticity of demand for complementary goods will always be positive.
   A. none of them is true
   B. statement I only is true
   C. statement II only is true
   D. statement III only is true
9 Elasticity is not constant along a straight line demand curve.
Put the correct values (A, B, C, D or E) for elasticity in the boxes on this diagram.

<table>
<thead>
<tr>
<th>Price</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>2</td>
</tr>
<tr>
<td>0.5 P</td>
<td>3</td>
</tr>
<tr>
<td>0</td>
<td>4</td>
</tr>
<tr>
<td>Q</td>
<td>5</td>
</tr>
</tbody>
</table>

A 1  
B less than 1  
C greater than 1  
D zero  
E infinity

10 Which diagram shows perfectly inelastic supply?

A  
B  
C  
D  

A  
B  
C  
D  

Economics and Markets
Answers to quick revision questions

1 PED is a measure of the extent to which the demand for a commodity changes proportionately in response to a change in its price.

2 If a good has low elasticity of demand, the manufacturer can increase the price without losing much sales revenue, and a government’s treasurer can impose a tax on it and expect to collect revenue. If demand is inelastic, people will buy almost the same amount of the good even if its price goes up.

3 Cross elasticity of demand is the responsiveness of demand for one good to changes in the price of another. It is determined by the extent that the goods are substitutes or complements. Substitutes display positive cross elasticity, and complements have a negative cross elasticity of demand.

4 Assuming a normal good, a decrease in price results in a greater quantity being demanded. Given that demand is price elastic, the increase in quantity will be proportionally greater than the price fall.

5 To increase market share requires greater quantities of the firm’s products to be both demanded and supplied. To sell more, a firm needs to lower price. For this to be profitable, demand must be elastic. To produce more, supply must also be elastic.

6 Because chips are a complementary product to fish, a rise in the price of fish will lead to a fall in the demand for chips.

7 Percentage change in quantity = 50 per cent. Percentage change in price = 25 per cent.

∴ PED = \frac{50\%}{25\%} = 2

8 Statement I is incorrect. When demand is price elastic, a fall in price will increase total spending on the good. Statement II is incorrect, because when household income rises, demand for an inferior good will fall; income elasticity of demand will be negative, not zero. Statement III is incorrect. If goods A and B are complements, a rise in the price of B will cause a fall in the demand for A, and so cross elasticity of demand is negative.

9 1 E per cent change in quantity from or to zero is infinitely large.

2 C per cent change in quantity is larger than per cent change in price.

3 A per cent change in quantity and price are identical.

4 B per cent change in price is larger than per cent change in quantity.

5 D change in price from or to zero is infinitely large.

10 A B is a demand curve and could show unit elastic demand. C is unit elastic supply (two examples). D is perfectly elastic supply (or demand). A shows perfectly inelastic supply (or demand).
Answers to chapter questions

1. Monthly demand at $1.40 per unit = 220 000 units
   Monthly demand at $1.60 per unit = 200 000 units
   Average quantity = 210 000 units
   Average price = $1.50

   % change in demand = \( \frac{20000}{210000} \times 100\% = 9.52\% \)

   % change in price = \( \frac{20}{150} \times 100\% = 13.33\% \)

   Arc price elasticity of demand = \( \frac{-9.52}{13.33} = -0.71\% \)

   Ignoring the minus sign, the arc elasticity is 0.71.

   Demand is inelastic over the demand range considered, because the price elasticity of demand (ignoring the minus sign) is less than 1.

2. We assume that the demand curve is a straight line.
   At a price of $1.40, demand is 220 000 units.
   For a price rise of 20 cents to $1.60:

   % change in demand = \( \frac{20 000}{220 000} \times 100\% = 9.09\% \) (fall)

   % change in price = \( \frac{20c}{140c} \times 100\% = 14.29\% \) (rise)

   Price elasticity of demand = \( \frac{9.09}{-14.29} = -0.64 \)

   or 0.64 ignoring the minus sign.

   Demand is inelastic at this point, because it is less than 1.

3. B
   Total revenue of $20 = 100 \times \$20 = \$2 000
   Total revenue at $24 = \$2 000 \times 1.14 = \$2 280
   Number sold at $24 = \$2 280 \div 24 = 95

   Price elasticity of demand

   **Point method**
   \[
   \frac{5}{100} \times 100 = \frac{5\%}{20\%} = 0.25
   \]

   **Arc method**
   \[
   \frac{5}{97\frac{1}{2}} \times 100 = \frac{5.13\%}{18.2\%} = 0.28
   \]
4  (a)  Product A

At price $5:

Change in quantity \[ \frac{300}{1700} = 17.7\% \]

Change in price \[ \frac{40c}{$5} = 8\% \]

PED = \[ \frac{17.7\%}{8\%} = -2.2 \] (Ignoring the minus sign = 2.2, and so > 1.)

Demand is elastic and a fall in price should result in such a large increase in quantity demanded that total revenue will rise.

Revenue at old price of $5 \( (\times 1700) \)
8 500

Revenue at new price of $4.60 \( (\times 2000) \)
9 200

Increase in total revenue
700

(b)  Product B

At price $8:

Change in quantity \[ \frac{500}{9500} = 5.3\% \]

Change in price \[ \frac{50c}{$8} = 6.25\% \]

PED = \[ \frac{5.3\%}{6.25\%} = -0.85 \] (Ignoring the minus sign = 0.85, and so < 1.)

Demand is inelastic and a fall in price should result in only a relatively small increase in quantity demanded. Total revenue will fall.

Revenue at old price of $8 \( (\times 9500) \)
76 000

Revenue at new price of $7.50 \( (\times 10000) \)
75 000

Fall in total revenue
1 000

5  The demand curve for sports cars will shift to the right. Price, quantity demanded and quantity supplied will all go up.

The effect of a cut in income tax is to leave households with more to spend. Sports cars are a luxury good, so their income elasticity of demand is likely to be quite high. The percentage increase in demand for the cars is therefore likely to be greater than the percentage increase in after-tax household income.

We can illustrate the change diagrammatically:

![Diagram of demand curve shift](image)

The shift in the demand curve (from \( D_0 \) to \( D_1 \)) means that the equilibrium quantity demanded and supplied shifts from \( Q_0 \) to \( Q_1 \) and the equilibrium price increases from \( P_0 \) to \( P_1 \).
A is unit elastic supply (two examples).
B is perfectly inelastic supply (or demand).
D is unit elastic demand.
C shows perfectly elastic supply.
Chapter 4

Cost, revenues and productivity

Learning objectives

<table>
<thead>
<tr>
<th>Costs, revenues and productivity</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Explain the relationship between marginal cost, total cost, total revenue, marginal revenue, average revenue and price in both the long term and short term</td>
<td>LO4.1</td>
</tr>
<tr>
<td>Demonstrate and apply the concept of MC = MR</td>
<td>LO4.1.1</td>
</tr>
<tr>
<td>Apply the concepts of marginal revenue product, marginal product, total product, total cost and marginal cost in an analysis of productivity</td>
<td>LO4.2</td>
</tr>
<tr>
<td>Conduct a break-even analysis</td>
<td>LO4.2.1</td>
</tr>
<tr>
<td>Explain the demand for factors of production</td>
<td>LO4.3</td>
</tr>
<tr>
<td>Explain the concept of diminishing returns for a factor of production</td>
<td>LO4.4</td>
</tr>
<tr>
<td>Calculate the diminishing returns</td>
<td>LO4.4.1</td>
</tr>
<tr>
<td>Explain how a firm can attain an optimal combination of factors of production</td>
<td>LO4.5</td>
</tr>
<tr>
<td>Explain the determinants of elasticity of a factor demand curve</td>
<td>LO4.6</td>
</tr>
<tr>
<td>Explain the causes of a shift of a factor demand curve</td>
<td>LO4.7</td>
</tr>
<tr>
<td>Distinguish between economies of scale and diseconomies of scale</td>
<td>LO4.8</td>
</tr>
</tbody>
</table>

Topic list

1. The firm’s output decision
2. Costs of production
3. Average costs, marginal costs and diminishing returns
4. Economies of scale and long-run costs
In this chapter we will be looking firstly at revenue and the calculation of a firm’s revenues. A key decision for a firm is how much to produce. The outcome of this decision will depend on the firm’s revenues and costs. However, the firm’s revenues themselves will be dependent on the structure and nature of the market in which it operates and the costs of the factors of production.

The chapter then examines costs of production for an individual firm, and how these are affected by both short-run and long-run factors.

We contrast the concept of **opportunity cost** (which was introduced in Chapter 1), with financial cost as seen from the accountant’s point of view.

The aggregate amount of goods supplied by every individual firm adds up to the market supply. By studying an individual firm we are looking at the ‘building blocks’ of market supply.

However, we will also look at how individual firms look to grow in order to benefit from cost savings and efficiencies.

---

**Costs, revenues and productivity**

- **MC = MR**
- **Long-run**
- **Short run**
- **Fixed v variable**
- **Average costs**
- **Marginal costs**
- **Diminishing returns**
- **Economies of scale**
- **Diseconomies of scale**
- **Breakeven**
- **Total revenue**
- **Average revenue**
- **Marginal revenue**
Before you begin

If you have studied these topics before, you may wonder whether you need to study this chapter in full. If this is the case, please attempt the questions below, which cover some of the key subjects in the area.

If you answer all these questions successfully, you probably have a reasonably detailed knowledge of the subject matter, but you should still skim through the chapter to ensure that you are familiar with everything covered.

There are references in brackets indicating where in the chapter you can find the information, and you will also find a commentary at the back of the Study Manual.

1. Define the following: (Section 1.1)
   - Total revenue
   - Average revenue
   - Marginal revenue

2. What is the definition of profit? (Section 1.3)

3. When does break-even occur? (Section 1.5)

4. Define short-run and long-run costs. (Section 2.1)

5. What is an 'economic profit'? (Section 2.4)

6. When examining the relationship between average cost (AC) and marginal cost (MC), when the AC curve is falling, where does MC sit? (Section 3.1)

7. What is the law of diminishing returns? (Section 3.3)

8. Define economies of scale. (Section 4.3)

9. What are the main reasons for diseconomies of scale to occur? (Section 4.9)
1 The firm's output decision

Section overview

• The assumption of profit maximisation provides a basis for beginning to look at the output decisions of individual firms. A firm will maximise its profits where marginal costs equal marginal revenue (MC = MR).

Definition

Profit is equal to total revenue minus total cost of any level of output.

Therefore, in order to understand how firms go about seeking profit, we must examine the ways in which costs and revenues arise. We will start by exploring revenues.

1.1 Total revenue, average revenue and marginal revenue

There are three aspects of revenue to consider:

(a) Total revenue (TR) is the total income obtained from selling a given quantity of output. We can think of this as quantity sold multiplied by the price per unit.

(b) Average revenue (AR) is the total revenue divided by the number of units sold. We can think of this as the price per unit sold.

(c) Marginal revenue (MR) is the addition to total revenue earned from the sale of one extra unit of output. We can think of this as the incremental revenue earned from selling the last unit of output.

1.2 The average revenue curve

When a firm can sell all its extra output at the same price, the AR 'curve' will be a straight horizontal line on a graph. The marginal revenue per unit from selling extra units at a fixed price must be the same as the average revenue (see Figure 1).

If the price per unit must be cut in order to sell more units, then the marginal revenue per unit obtained from selling extra units will be less than the previous price per unit (see Figure 2). In other words, when the AR is falling as more units are sold, the MR must be less than the AR.

Figure 1

In Figure 2, with straight line MR and AR curves, the length OX is exactly half of the length OY.

Exam comments

Figure 2 is a very important diagram and forms the basis of other more detailed illustrations. Make sure you understand the principles behind Figure 2.
Note that in Figure 2, at any given level of sales, **all units are sold at the same price**. The firm has to reduce its selling price to sell more, but the **price must be reduced for all units sold**, not just for the extra units. This is because we are assuming that all output is produced for a single market, where a single price will prevail.

When the price per unit has to be reduced in order to increase the firm’s sales, the marginal revenue can become negative. This happens in Figure 2 at prices below $P_N$ when a reduction in price does not increase output sufficiently to earn the same total revenue as before. In this situation, demand would be price inelastic.

We can illustrate this by using an example. The figures below illustrate the relationship between price (average revenue), total revenue and marginal revenue. Note that although reducing the price from 5 to 4 generates an additional unit of sales (from 6 to 7), the fact that the sale price of all the initial 6 units has also been reduced from 5 to 4 means that total revenue falls as a result.

<table>
<thead>
<tr>
<th>Average revenue (AR) price</th>
<th>Units sold</th>
<th>Total revenue (TR)</th>
<th>Marginal revenue (MR)</th>
</tr>
</thead>
<tbody>
<tr>
<td>11</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>10</td>
<td>1</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td>9</td>
<td>2</td>
<td>18</td>
<td>8</td>
</tr>
<tr>
<td>8</td>
<td>3</td>
<td>24</td>
<td>6</td>
</tr>
<tr>
<td>7</td>
<td>4</td>
<td>28</td>
<td>4</td>
</tr>
<tr>
<td>6</td>
<td>5</td>
<td>30</td>
<td>2</td>
</tr>
<tr>
<td>5</td>
<td>6</td>
<td>30</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>7</td>
<td>28</td>
<td>–2</td>
</tr>
</tbody>
</table>

### 1.3 Profit

We have defined profit as **total revenue** (TR) minus **total cost** (TC).

(a) Figure 3 shows, in simplified form, how TR and TC vary with output. As you might expect, TC increases as output rises. In the diagram, the increase in total cost accelerates as output rises and the TC curve gets steeper. The reason for this rate of increase in TC is diminishing returns as output rises beyond a certain level. Diminishing returns are explained later.

(b) Conversely, the gradient of the TR curve reduces as output and sales increase. This is because most firms operate under the conditions illustrated in Figure 2. That is to say, they must reduce their prices in order to sell more. The rate of growth of TR therefore declines.

(c) Note carefully that the vertical axis of Figure 3 shows total values whereas in Figures 1 and 2, it shows value per unit.

(d) Profits are at a maximum where the vertical distance AB between the TR and TC curves is greatest.
### 1.4 Marginal cost equals marginal revenue

As a firm produces and sells more units, its total costs will increase and its total revenues will also increase (unless demand is price inelastic and MR has become negative).

(a) Provided that the extra cost of making an extra unit is less than the extra revenue obtained from selling it, the firm will increase its profits by making and selling that extra unit.

(b) If the extra cost of making an extra unit of output exceeds the extra revenue obtainable from selling it, the firm’s profits would be reduced by making and selling that extra unit.

(c) If the extra cost of making an extra unit of output is exactly equal to the extra revenue obtainable from selling it, it will be worth the firm’s while to make and sell the extra unit. This is because economic cost includes the cost of entrepreneurship, which is ‘normal’ profit. If the extra cost of an additional unit equals the extra revenue, making and selling the unit is worthwhile, because it will cover all the extra costs of production, including ‘normal’ profit.

Since the extra cost of yet another unit would be higher due to the law of diminishing returns, whereas extra revenue per unit from selling extra units is never higher: this extra unit would generate a loss. Therefore, the profit-maximising output is reached where MC = MR.

The identification of the profit-maximising output (MC = MR) is a crucial concept in economics. Make sure you remember this.

(d) Figures 4 and 5 show the profit maximising output quantity M for the two types of firm shown in Figures 1 and 2. In both cases, the firm increases its profit with each extra item it produces until output M is reached, because MR > MC when output is less than M. At output M, the MC and MR curves cross. The addition to total revenue from subsequent units will be less than the increase in total cost which they cause. This level of output, M, corresponds to the level M shown in Figure 3.

In other words, given the objective of profit maximisation there are three possibilities:

(a) If **MC is less than MR**, profits will be increased by making and selling more.

(b) If **MC is greater than MR**, profits will fall if more units are made and sold, and a profit-maximising firm would not make the extra output. In other words, if MC > MR a firm will look to reduce output.

(c) If **MC = MR**, the profit-maximising output has been reached, and so this is the output quantity that a profit-maximising firm will decide to supply.
Question 1: Revenue and costs

The following data refer to the revenue and costs of a firm:

<table>
<thead>
<tr>
<th>Output</th>
<th>Total revenue</th>
<th>Total costs</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>$  -</td>
<td>$  110</td>
</tr>
<tr>
<td>1</td>
<td>$  50</td>
<td>$  140</td>
</tr>
<tr>
<td>2</td>
<td>$ 100</td>
<td>$  162</td>
</tr>
<tr>
<td>3</td>
<td>$ 150</td>
<td>$  175</td>
</tr>
<tr>
<td>4</td>
<td>$ 200</td>
<td>$  180</td>
</tr>
<tr>
<td>5</td>
<td>$ 250</td>
<td>$  185</td>
</tr>
<tr>
<td>6</td>
<td>$ 300</td>
<td>$  194</td>
</tr>
<tr>
<td>7</td>
<td>$ 350</td>
<td>$  229</td>
</tr>
<tr>
<td>8</td>
<td>$ 400</td>
<td>$  269</td>
</tr>
<tr>
<td>9</td>
<td>$ 450</td>
<td>$  325</td>
</tr>
<tr>
<td>10</td>
<td>$ 500</td>
<td>$  425</td>
</tr>
</tbody>
</table>

Required

(a) Calculate the marginal revenue for the firm.
(b) Calculate the firm’s fixed costs and the marginal cost at each level of output.
(c) What level of output will the firm aim to produce and what amount of profit will it make at this level?

(The answer is at the end of the chapter)

1.5 Break-even analysis

So far we have been looking at the conditions under which a firm maximises its profit.

However, a firm may also want to identify the level of output required for it to break-even.

Break-even occurs where total revenue equals total cost, and therefore by extension, average revenue equals average cost.

We can illustrate the break-even point graphically by adding the average revenue and average cost lines to the Figure 5 we showed earlier.

![Figure 6 Profit maximisation and break-even positions](image)

In economic terms, the **break-even point** also represents the point at which a firm is making a normal profit.

Ultimately, the amount of profit a firm makes, and whether it is normal or supernormal, depends on the **market structure** in which the firm is operating, and the **time period** (short run or long run) under review.

We shall return to both of these points about market structure and time period later.
## Costs of production

### Section overview

- A firm's output decisions can be examined in the **short run** (when some factors of production are fixed) and in the **long run** (when all factors of production can be varied). Total costs can be divided into **fixed** and **variable** elements. These elements have different effects on total cost as output is increased.

- **Economic costs** are different from **accounting costs**, and represent the **opportunity costs** of the factors of production that are used.

### Short-run and long-run costs

Production is carried out by firms using the **factors of production** which must be paid or rewarded for their use. The cost of production is the cost of the factors used.

<table>
<thead>
<tr>
<th>Factor of production</th>
<th>Its cost</th>
</tr>
</thead>
<tbody>
<tr>
<td>Land</td>
<td>Rent</td>
</tr>
<tr>
<td>Labour</td>
<td>Wages</td>
</tr>
<tr>
<td>Capital</td>
<td>Interest</td>
</tr>
<tr>
<td>Enterprise</td>
<td>Normal profit</td>
</tr>
</tbody>
</table>

**Notice that normal profit is viewed as a cost.** This may seem odd to an accountant, who thinks of profit as the excess of revenues over cost, but normal profit is the amount of profit necessary to keep an entrepreneur in his or her present activity. In other words, normal profit is the **opportunity cost** of preventing the entrepreneur investing elsewhere.

Any profit earned in excess of this normal profit is known as **supernormal, abnormal, or excess profit**.

Whether a firm can maintain these profits in the long run will depend on the nature of the industry it is operating in. We will return to this point later in this Study Manual when we look at the differences between perfect competition and monopoly.

### Definition

The **short run** is a time period in which the amount of at least one factor of production (land, labour, capital or enterprise) is fixed. The **long run** is a period sufficiently long to allow full flexibility in all the factors of production used.

### Fixed costs and variable costs

In the **short run**, certain costs are **fixed** because the availability of factors of production is restricted. Decisions must therefore be taken for the short run within the restriction of having some resources in fixed supply. In the **longer run**, however, most costs are **variable**, because the supply of skilled labour, machinery, buildings and so on can all be increased or decreased. Decisions in the long run are therefore subject to fewer restrictions about resource availability.

Inputs are variable at the decision of management. For example, management might decide to buy more raw materials, hire more labour, start overtime working and so on.

(a) Labour is usually considered to be variable in the short run. Inputs which are treated as **fixed** in the short run are likely to include **capital items**, such as buildings and machinery, for which a significant
lead time might be needed before their quantities are changed. There could also be constraints around technology which limits the productivity of other factors of production.

(b) **All inputs are variable in the long run.** A decision to change the quantity of an input variable which is fixed in the short run will involve a change in the **scale of production.**

### 2.3 Short-run costs: total costs, average costs and marginal costs

Let us now turn our attention to short-run costs. Note that in economics the 'short run' is not a time period which can be measured in days or months. The key determinant of a time period being 'short' is that at least one factor of production is fixed.

**Definitions**

- **Total cost** (TC). Total cost is the cost of all the resources needed to produce a given level of output. Total cost comprises total fixed cost (TFC) and total variable cost (TVC).

- **Fixed costs** are costs which do not change when levels of production change, for example, the rent of premises.

- **Variable costs** are costs which change according to the level of output, for example, raw material costs.

- **Average cost** (AC). Average cost for a given level of output is the total cost divided by the total quantity produced.
  
  **Average cost** is made up of an average fixed cost per unit plus an average variable cost per unit.
  
  \[
  AC = \frac{TC}{N} = \frac{TFC}{N} + \frac{TVC}{N}
  \]
  
  \[
  AC = AFC + AVC
  \]

  **Average fixed cost** per unit (AFC) – total fixed costs divided by the number of units – will get smaller as the number of units produced (N) increases. This is because TFC is the same amount regardless of the volume of output, so as N gets bigger, AFC must get smaller.

  **Average variable costs** per unit (AVC) – total variable costs divided by the number of units – will change as output volume increases, but may rise as well as fall.

- **Marginal cost** (MC). This is the extra cost (incremental cost) of producing one more unit of output. For example, the marginal cost for a firm of producing the 50th unit of output is the total cost of making 50 units minus the total cost of making the first 49 units.

**Question 2: Costs**

To test your understanding of these concepts, look at the three definitions given below. Which of them, if any, correctly describes the marginal cost of producing one extra unit of output?

(a) **MC =** increase in total cost of production.

(b) **MC =** increase in variable cost of production.

(c) **MC =** increase in average cost of production.

*(The answer is at the end of the chapter)*
Figure 7 shows how the various elements of cost vary as output changes.

**Worked Example: Short-run costs**

Let us suppose that a firm employs a given amount of capital which is a fixed (invariable) input in the short run. The firm may combine with this capital different amounts of labour, which we assume to be an input which is variable in the short term. Therefore, fixed capital and variable labour can be combined to produce different levels of output.

Here is an illustration of the relationship between the different definitions of the firm’s costs. (The figures used are hypothetical.)

<table>
<thead>
<tr>
<th>Units of output</th>
<th>Fixed costs</th>
<th>Variable costs</th>
<th>Total cost</th>
<th>Average cost</th>
<th>Marginal cost</th>
</tr>
</thead>
<tbody>
<tr>
<td>n</td>
<td>$</td>
<td>$</td>
<td>$</td>
<td>$</td>
<td>$</td>
</tr>
<tr>
<td>1</td>
<td>0.80</td>
<td>0.30</td>
<td>1.10</td>
<td>1.10</td>
<td>1.10</td>
</tr>
<tr>
<td>2</td>
<td>0.80</td>
<td>0.80</td>
<td>1.60</td>
<td>0.80</td>
<td>0.50</td>
</tr>
<tr>
<td>3</td>
<td>0.80</td>
<td>0.95</td>
<td>1.75</td>
<td>0.58</td>
<td>0.15</td>
</tr>
<tr>
<td>4</td>
<td>0.80</td>
<td>1.20</td>
<td>2.00</td>
<td>0.50</td>
<td>0.25</td>
</tr>
<tr>
<td>5</td>
<td>0.80</td>
<td>1.70</td>
<td>2.50</td>
<td>0.50</td>
<td>0.50</td>
</tr>
<tr>
<td>6</td>
<td>0.80</td>
<td>2.32</td>
<td>3.12</td>
<td>0.52</td>
<td>0.62</td>
</tr>
<tr>
<td>7</td>
<td>0.80</td>
<td>3.19</td>
<td>3.99</td>
<td>0.57</td>
<td>0.87</td>
</tr>
<tr>
<td>8</td>
<td>0.80</td>
<td>4.32</td>
<td>5.12</td>
<td>0.64</td>
<td>1.13</td>
</tr>
<tr>
<td>9</td>
<td>0.80</td>
<td>5.50</td>
<td>6.30</td>
<td>0.70</td>
<td>1.18</td>
</tr>
<tr>
<td>10</td>
<td>0.80</td>
<td>7.20</td>
<td>8.00</td>
<td>0.80</td>
<td>1.70</td>
</tr>
</tbody>
</table>

(a) **Total cost** is the sum of capital costs (fixed costs) plus labour costs (variable costs), since these are the only two inputs in this example.

(b) **Average cost** is the cost per unit of output, i.e. \( AC = \frac{TC}{output} = \frac{TC}{n} \)

(c) **Marginal cost** is the total cost of producing \( n \) units minus the total cost of producing one less unit, i.e. \( (n - 1) \) units.
Note the following points on this set of figures.

(a) **Total cost.** Total costs of production carry on rising as more and more units are produced.

(b) **Average cost.** AC changes as output increases. It starts by falling, reaches a lowest level, and then starts rising again.

(c) **Marginal cost.** The MC of each extra unit of output also changes with each unit produced. It too starts by falling, fairly quickly reaches a lowest level, and then starts rising.

(d) **AC and MC compared.** At lowest levels of output, MC is less than AC. At highest levels of output, though, MC is higher than AC. There is a ‘cross-over’ point, where MC is exactly equal to AC. In this example, it is at 5 units of output.

### 2.4 Economists’ and accountants’ concepts of cost

As we have already mentioned, to an economist cost includes an amount for normal profit which is the reward for entrepreneurship. **Normal profit is the opportunity cost of entrepreneurship.** The opportunity cost of entrepreneurship is the amount of profit that an entrepreneur could earn elsewhere and so must be forgone to undertake the current project. In this situation, normal profit is the profit the entrepreneur must earn to persuade him to keep on with his investment in his current enterprise. If he could earn more by undertaking a different enterprise, as a rational decision-maker he will choose to invest in the other enterprise instead.

A further feature of cost accounting is that costs can be divided into fixed costs and variable costs. Total fixed costs per period are a given amount, regardless of the volume of production and sales. Cost accountants usually assume that the variable cost per unit is a **constant amount,** so that the total **variable cost** of sales is directly proportional to the **volume** of sales.

Economists do not take this approach. In the short run, there are **fixed costs** and **variable costs,** but the variable cost of making an extra unit of output need not be the same for each extra unit that is made. As a result, the **marginal cost** of each extra unit is not constant, either.

**Accounting profits** consist of sales revenue minus the **explicit costs** of the business. Explicit costs are those which are clearly stated and recorded, for example:

- Materials costs – prices paid to suppliers.
- Labour costs – wages paid.
- Depreciation costs on non-current assets.
- Other expenses, such as rates and building rental.

**Economic profit** consists of sales revenue minus both the explicit costs and the **implicit costs** of the business. Implicit costs are benefits forgone by not using the factors of production in their next most profitable way (opportunity costs).

It is a well established principle in accounting and economics that relevant costs for decision-making purposes are **future costs incurred as a consequence of the decision.** Past or ‘sunk’ costs are not relevant to our decisions now, because we cannot change them: they have already been incurred. Relevant future costs are the **opportunity costs** of the input resources to be used.

### Worked Example: Economic profits and opportunity costs

Suppose that a sole trader sells goods worth $200 000. He incurs materials costs of $70 000, hired labour costs of $85 000, and other expenses of $20 000. He has no non-current assets other than the building from which he trades, on which depreciation is not charged. In accounting terms, his profit would be as follows:

<table>
<thead>
<tr>
<th>Item</th>
<th>$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sales</td>
<td>200 000</td>
</tr>
<tr>
<td>Materials</td>
<td>70 000</td>
</tr>
<tr>
<td>Labour</td>
<td>85 000</td>
</tr>
<tr>
<td>Other expenses</td>
<td>20 000</td>
</tr>
<tr>
<td><strong>Profit</strong></td>
<td><strong>(175 000)</strong></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Item</th>
<th>$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Profit</td>
<td>25 000</td>
</tr>
</tbody>
</table>
But suppose the buildings he uses in his business could have been put to another use to earn $15,000, and his own labour as business manager could get him a job with a salary of $20,000. The position of the business in economic terms would be as follows:

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Sales less explicit costs</td>
<td>$25,000</td>
</tr>
<tr>
<td>Implicit costs ($15,000 + $20,000)</td>
<td>(35,000)</td>
</tr>
<tr>
<td>Loss</td>
<td>(10,000)</td>
</tr>
</tbody>
</table>

In economic terms, the business has made a loss. It would pay the trader to put his buildings and capital to the alternative use, and employ his own labour another way, working for someone else at a salary of $20,000.

**Question 3: Accounting profit and economic profit**

Wilbur Proffit set up his business one year ago. In that time, his firm has earned total revenue of $160,000, and incurred costs of $125,000, including his own salary of $12,000. Before, he had been a salaried employee of Dead End Ventures, earning an annual salary of $20,000.

To finance the business, Wilbur had to sell his investment of $200,000 in government securities which earned interest of 10 per cent p.a. He used $80,000 of this to buy a warehouse, whose annual commercial rental value would be $11,000 p.a. The remaining $120,000 has been used to finance business operations.

**Required**

Calculate:

- the accounting profit earned by Wilbur in the last year.
- the economic profit or loss earned.

(The answer is at the end of the chapter)

3 **Average costs, marginal costs and diminishing returns**

**Section overview**

- A firm’s short-run average cost (SRAC) curve is U shaped, due to diminishing returns beyond a certain output level.

3.1 **The relationship between AC and MC**

The relationships between average and marginal costs are important.

(a) **When the average cost curve is rising, the marginal cost will always be higher than the average cost.** If the cost of making one extra unit of output exceeds the average cost of making all the previous units, then making the extra unit will clearly cause an increase in the average unit cost.

(b) In our example in Section 2.3 of this chapter, the average cost schedule rises from six units of output onwards and MC is bigger than AC at all these levels of output (6 – 10 units).

(c) **When the average cost curve is falling, marginal cost lies below it.** This follows similar logic. If the cost of making an extra unit is less than the average cost of making all the previous units, the effect of making the extra unit must be a reduction in average unit cost. In our example in Section 2.3, this happens between production of one and four units.

(d) **When the average cost curve is horizontal, marginal cost is equal to it.** In our example in Section 2.3, when there are five units of output, the average cost stays at $0.50 and the marginal cost of the fifth unit is also $0.50.

The relationship between marginal costs and average costs means that the marginal cost curve always cuts through the average cost curve at the **lowest point of the average cost curve** (see Figure 7 earlier).
Question 4: Cost curves

(a) It is possible for the average total cost curve to be falling while the average variable cost curve is rising. True or false?

(b) Marginal fixed costs per unit will fall as output increases. True or false?

(The answers are at the end of the chapter)

3.2 U-shaped short-run average cost curve

The short-run average cost curve (AC in Figure 7) is U shaped. In summary, short-run average total cost will also be eventually upward sloping due to the properties of the short-run average variable cost curve. The minimum point on the short-run average cost curve will occur at an output level that is greater than the short-run AVC due to the presence of fixed costs. The shape of the long-run average cost curve is determined by returns to scale, not diminishing marginal product. Let us now look at this in more detail:

Fixed costs per unit of output, i.e. average fixed costs, will fall as the level of output rises. Therefore, if fixed costs are $20,000 and we make 10,000 units, the average fixed cost (AFC) will be $2 per unit. If output increases to 12,500 units the AFC will fall to $1.60 (20,000 ÷ 12,500) and if output increases to 15,000 units, the AFC will fall again to $1.33 (20,000 ÷ 15,000), and so on. Spreading fixed costs over a larger amount of output is a major reason why short-run average costs per unit fall as output increases.

Variable costs are made up from the cost of the factors of production whose use can be varied in the short run – for example, wages, fuel and raw material purchases. Total variable costs therefore vary with output in the short run as well as in the long run.

(a) The accountant’s assumption about short-run variable costs is that up to a certain level of output, the variable cost per unit is more or less constant (e.g. wages costs and materials costs per unit of output are unchanged). If the average fixed cost per unit is falling as output rises, and the average variable cost per unit is constant, it follows that the average total cost per unit will also be falling as output increases.

(b) However, there are other reasons for the initial fall in average total cost. The first is the effects of the division of labour and specialisation. Imagine a small but fully equipped factory, with a variety of machinery and equipment and a workforce of, say, ten. If each person attempts to perform all the operations on a single item, production is likely to be low.

(i) They will be unable to develop a high level of skill at every one of the jobs.
(ii) Time will be lost as they move from machine to machine.
(iii) Individual variability will produce a high rate of defects, perhaps with each person tending to produce different faults.
(iv) Individuals will work at different rates on different operations: as a result, queues will form at some machines and others will be under-utilised.

If there is a degree of specialisation, expertise and speed will rise, machines will be run at optimum rates and output will rise. Average costs will therefore fall.

(c) The second reason is the utilisation of indivisibilities. If a machine has an output capacity of 100 units per day but is only used to produce 50 units per day, the machinery cost of each of those 50 units will be twice the level it would be if the machine was used to capacity. Operation of a plant below normal output is uneconomical, so there are cost savings as production is increased up to capacity level.

3.3 The law of diminishing returns

Definition

Eventually, as output increases, average costs will tend to rise. The law of diminishing returns says that if one or more factors of production are fixed, but the input of another is increased, the extra output generated by each extra unit of input will eventually begin to fall. In our factory, as we add staff, we start to see queues forming at machines; it becomes more difficult to co-ordinate work; machinery starts to break down through over-use, and there simply is not enough space to work efficiently.
The law of diminishing returns states that, given the present state of technology, as more units of a variable input factor are added to input factors that are in fixed supply in the short run, the resulting increments to total production will eventually and progressively decline. In other words, as more units of a variable factor (e.g. labour) are added to a quantity of a fixed factor (e.g. a factory), there may initially be some increasing returns or constant returns as more units of the variable factor (e.g. labour) are added, but eventually, diminishing marginal returns will set in.

The total product of a variable factor of production identifies what outputs are possible using various levels of the variable factor. The marginal product is the extra output produced by one more unit of a variable factor. Putting more people to work in a factory will increase the yield up to a point, but eventually marginal output will fall, and will ultimately even become negative.

It is important that you appreciate that diminishing returns set in once the rate at which the increase in productivity from adding an extra unit of a factor of production starts to fall. This does not mean, however, that total output has also started to fall. We can illustrate this with a simple example:

<table>
<thead>
<tr>
<th>Workers</th>
<th>Total output</th>
<th>Marginal output</th>
<th>Average output</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>5</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>2</td>
<td>11</td>
<td>6</td>
<td>5.5</td>
</tr>
<tr>
<td>3</td>
<td>18</td>
<td>7</td>
<td>6</td>
</tr>
<tr>
<td>4</td>
<td>24</td>
<td>6</td>
<td>6</td>
</tr>
<tr>
<td>5</td>
<td>29</td>
<td>5</td>
<td>5.8</td>
</tr>
<tr>
<td>6</td>
<td>32</td>
<td>3</td>
<td>5.3</td>
</tr>
<tr>
<td>7</td>
<td>33</td>
<td>1</td>
<td>4.7</td>
</tr>
</tbody>
</table>

In this example, diminishing returns set in once the fourth worker is employed (because the marginal output of adding a fourth worker is lower than the marginal output from adding a third worker). However, total output continues to rise as workers four to seven are added, and only starts to fall when the eighth worker is added.

![Figure 8 U-shaped short-run cost curve and the relationship of AC to MC in the short run](image)

**Exam comments**

Remember that the law of diminishing returns is a short run phenomenon; at least one factor of production is fixed.

The law of diminishing returns is expressed in production quantities, but it obviously has direct implications for short run average and marginal costs. Resources cost money, and the average and marginal costs of output will depend on the quantities of resources needed to produce the given output.

In our simple example, the marginal output from adding an extra worker increased for workers one to three, perhaps as each worker can specialise in specific tasks.

This increasing productivity means it will cost less to produce an extra unit of output, therefore the marginal cost of an extra unit of output falls.
However, once diminishing returns set in, the marginal cost of producing an additional extra unit of output will increase, because trying to add extra workers will now have proportionally less impact on output: for example, because the factory may be getting over-crowded and workers are having to wait for machines to become available for them to use.

A firm is operating at maximum efficiency when average cost is at its minimum. Remember this optimum output point (the bottom point of the AC cost curve) is always found where the marginal cost curve intersects the average cost curve.

**Question 5: Diminishing returns**

![Diagram showing Total output volume versus Number employed]

In the diagram above, from what level of employment do diminishing returns start to occur? (The answer is at the end of the chapter)

### 4 Economies of scale and long-run costs

#### Section overview

- Once all the factors of production become variable, a firm's short-run average cost (SRAC) curve can be shifted, and a firm's minimum achievable average costs at any level of output can be depicted by a long-run average cost (LRAC) curve.

- The shape of the LRAC depends on whether there are increasing, constant or decreasing **returns to scale**. Even if increasing returns to scale are not achievable indefinitely as output rises, up to a certain **minimum efficient scale of production** (MES) there will be increasing returns to scale. Firms will reduce their average costs by producing on a larger scale up to the MES.

#### 4.1 Costs in the long run

We have not yet considered a firm's long-run costs of output. In the long run, all factors of production are, by definition, variable.

There are two direct consequences of this:

(a) Since there are no fixed factors, there can be no fixed costs in the long run.

(b) Because all factors of production are variable, a firm can change its **scale of production** significantly.

These two direct consequences then also generate a third difference between the short run and the long run. Because all inputs are variable in the long run, the problems associated with the diminishing returns for variable factors constrained by fixed factors of production do not arise. In other words, the **law of diminishing returns applies only to short-run costs**.
However, whereas short-run output decisions are concerned with diminishing returns, long-run output decisions are concerned with **economies of scale**.

Output will vary with variations in inputs, such as labour and capital.

(a) If output increases in the **same proportion** as inputs (for example, doubling all inputs doubles output) there are said to be **constant returns to scale**.

(b) If output increases **more than in proportion** to inputs (for example, doubling all inputs trebles output) there are beneficial **economies of scale**. Economies of scale mean that the long-run average costs of production will continue to fall as the volume of output rises.

(c) If output increases **less than proportionally** to inputs (for example, trebling all inputs only doubles output) there are said to be **diseconomies of scale**. Diseconomies of scale mean that the long-run average costs of production will rise as output volume rises.

Returns to scale are, for example, concerned with improvements or declines in productivity by **increasing the scale of production**, for example by mass-producing instead of producing in small batch quantities.

### 4.2 Constant returns to scale

The key feature of constant returns to scale is that long-run average costs and marginal costs per unit remain constant. For example:

<table>
<thead>
<tr>
<th>Output</th>
<th>Total cost (with constant returns)</th>
<th>Average cost per unit</th>
<th>Marginal cost per unit</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>6</td>
<td>6</td>
<td>6</td>
</tr>
<tr>
<td>2</td>
<td>12 (2 × 6)</td>
<td>6</td>
<td>6</td>
</tr>
<tr>
<td>3</td>
<td>18 (3 × 6)</td>
<td>6</td>
<td>6</td>
</tr>
<tr>
<td>4</td>
<td>24 (4 × 6)</td>
<td>6</td>
<td>6</td>
</tr>
</tbody>
</table>

In the real world, the duplication of all inputs might be impossible if one incorporates qualitative as well as quantitative characteristics in inputs. One such input is entrepreneurship. Doubling the size of the firm does not necessarily double the inputs of organisational and managerial skills, even if the firm does hire extra managers and directors. The input of entrepreneurship might be intangible and indivisible.

### 4.3 Economies of scale

**Definition**

**Economies of scale**: factors which cause average cost to decline in the long run as output increases.

The effect of economies of scale is to shift the whole cost structure downwards and to the right on a graph showing costs plus output. A **long-run average cost curve (LRAC)** can be drawn as the ‘envelope’ of all the short-run average cost curves (SRAC) of firms producing on different scales of output. In Figure 9, each SRAC is drawn for a specific size or change in factor of production (possible plant size) and each SRAC curve corresponds to a different plant size level on the LRAC.
4.4 Diseconomies of scale

It may be that the flat part of the LRAC curve is never reached, or it may be that diseconomies of scale are encountered. Diseconomies of scale might arise when a firm gets so large that it cannot operate efficiently or it is too large to manage efficiently, with that average costs begin to rise.

A firm should try to minimise its average costs in the long run, and to do this it ought to try to produce output on a scale where the LRAC curve is at its lowest point. This indicates its minimum efficient scale of production (MES). While there are economies of scale, a firm should always be trying to grow until it reaches its minimum efficient scale of production.

If a firm is suffering diseconomies of scale, in theory it should look to reduce production levels, or seek a more efficient combination of its factors of production.
4.5 Reasons for economies of scale

The economies of scale attainable from large scale production fall into two categories:

(a) **Internal economies**: economies arising within the firm from the organisation of production.

(b) **External economies**: economies attainable by the firm because of the growth of the industry as a whole.

4.6 Internal economies of scale

4.6.1 Technical economies

Technical economies arise in the production process. They are also called **plant economies of scale** because they depend on the size of the factory or piece of equipment.

Large undertakings can make use of **larger and more specialised machinery**. If smaller undertakings tried to use similar machinery, the costs would be excessive because the machines would become obsolete before their physical life ends (that is their economic life would be shorter than their physical life). Obsolescence is caused by falling demand for the product made on the machine, or by the development of newer and better machines.

A large firm also benefits from economies of scale by overcoming **indivisibilities**. For example, a manufacturing company will need a factory unit in which to manufacture its goods. The factory space is a single, indivisible unit, and so the associated costs of the factory (e.g. depreciation or rent) are the same whatever the level of output. Consequently the manufacturing cost per unit will be reduced as production output increases.

We can also see the interaction of economies of scale and indivisibilities in the context of advertising. An advertising campaign would not be worth considering unless there is a certain level of output against which to charge the cost and justify the campaign.

**Dimensional economies of scale** arise from the relationship between the volume of output and the size of equipment (e.g. storage tanks) needed to hold or process the output. The cost of a container for 10 000 gallons of product will be much less than ten times the cost of a container for just 1 000 gallons.

4.6.2 Commercial or marketing economies

**Buying economies** may be available, reducing the cost of material purchases through bulk purchase discounts. Similarly, **stockholding** becomes more efficient. The most economic quantities of inventory to hold increase with the scale of operations, but at a lower proportionate rate of increase. Also, bulk selling will enable a large firm to make relative savings in **distribution** costs, and **advertising** costs.

4.6.3 Organisational economies

When the firm is large, centralisation of functions such as administration, research and development and marketing may reduce the burden of overheads (i.e. the indirect costs of production) on individual operating locations. The need for management and supervisory staff does not increase at the same rate as output.

Equally, large firms can also employ specialist staff in areas such as IT, HR and accountancy and their skills can be fully utilised in their specialist areas.

4.6.4 Financial economies

Large firms may find it easier to borrow money than smaller firms and they may also obtain loan finance at more attractive rates of interest, due to their reputation and asset base. Listed public limited companies can also raise finance by selling shares to the public via a stock exchange.

**Question 6: Economies of scale**

The above list is not exhaustive. Can you add to it?  

(The answer is at the end of the chapter)
4.7 External economies of scale

Whereas internal economies of scale accrue to an individual firm, it is also possible for general advantages to be enjoyed by all of the firms in an industry. These are known as external economies of scale.

External economies of scale occur as an industry grows in size. Here are two examples.

(a) A large skilled labour force is created and educational services can be geared towards training new entrants.

(b) Specialised ancillary industries will develop to provide components, transport finished goods, trade in by-products, provide special services and so on. For instance, law firms may be set up to specialise in the affairs of the industry.

4.8 The effect of size

The extent to which both internal and external economies of scale can be achieved will vary from industry to industry, depending on the conditions in that industry. In other words, large firms are better suited to some industries than others.

(a) Internal economies of scale are potentially more significant than external economies to a supplier of a product or service for which there is a large consumer market. It may be necessary for a firm in such an industry to grow to a certain size in order to benefit fully from potential economies of scale, and thereby be cost-competitive and capable of making profits and surviving.

(b) External economies of scale are potentially significant to smaller firms who specialise in the ancillary services to a larger industry. For example, the development of a large world-wide industry in drilling for oil and natural gas off-shore has led to the creation of many new specialist supplier firms, making drilling rigs, and various types of equipment. Therefore, a specialist firm may benefit more from the market demand created by a large customer industry than from its own internal economies of scale.

4.9 Diseconomies of scale

Economic theory predicts that there will be diseconomies of scale in the long-run costs of a firm, once the firm gets beyond an ideal size. The main reasons for possible diseconomies of scale are human and behavioural problems of managing a large firm. In a large firm employing many people, with many levels in the hierarchy of management, there may be a number of undesirable effects:

(a) Communicating information and instructions may become difficult.

(b) Chains of command may become excessively long, and management will become too remote, and lose control over operations.

(c) Morale and motivation amongst staff may deteriorate, and there may be conflicts between different departments which have different objectives.

(d) Senior management may have difficulty in assimilating all the information they need in sufficient detail to make good quality decisions.

(e) There may be increased levels of bureaucracy.

A firm may also experience technical diseconomies of scale. For example, increasing the size of the plant and equipment may create large administrative overheads thereby increasing total average costs, even though the direct production cost itself is lowered.

These are all internal diseconomies of scale. However, there may also be external diseconomies of scale which affect all firms in an industry as the industry grows. For example, if a natural resource such as oil or gas is over-used then shortages may result. In turn, this would increase the average cost of production.

The implication of diseconomies of scale is that companies should achieve a certain size to benefit fully from scale economies, but should not become too big, because if they do cost controls might slacken and organisational inefficiency may be likely to result.
4.10 Minimum efficient scale

Definition

Given the idea of economies of scale, it is generally accepted that in any industry there is a minimum efficient scale of production which is necessary for a firm to achieve the full potential economies of scale.

The level of the minimum efficient scale (MES) will vary from industry to industry. In the paint manufacturing industry, for example, it might be necessary to have a 15 per cent share of the market in order to achieve maximum scale economies, whereas in frozen food production, a 25 per cent share of the market might be necessary, and so on. If a firm has a production capacity below the minimum economic scale, its unit costs of production will be higher than the unit costs of its bigger competitors, and so it will not compete successfully and it will make lower profits, or even losses. A profit-maximising firm should be attempting to minimise its unit costs, and this means striving to achieve maximum scale economies.

Question 7: Economies of scale and diminishing returns

Explain in detail the difference between diminishing returns and economies of scale.

(The answer is at the end of the chapter)

4.11 Short-run and long-run production decisions

There are some occasions when it is beneficial for a firm to carry on producing even though it is making a loss.

In the short run, a firm will have to pay its fixed costs even though it may not be producing any output. However, it only incurs variable costs once it starts producing output. Therefore, in the short run, a firm will carry on producing provided its total revenue exceeds total variable costs because this means it is making a contribution towards fixed costs.

If total revenue is less than total variable costs, the firm would be better off not producing any output and just paying its fixed costs.

Alternatively, we could look at this short-run cut-off point as being where average revenue (price) is greater than or equal to average variable cost.

However, this cut-off point needs revisiting in the long run.

In the long run, by definition, there are no fixed costs, so all costs are variable. Therefore, in the long run, a firm will only carry on producing if total revenue is greater than or equal to total cost, or if average revenue (price) is greater than or equal to average cost. This means that, in the long run, a firm will only carry on producing if it is making at least a normal profit. If we remember what normal profit is, then this long-run production decision is exactly as it should be. Normal profit is the opportunity cost of enterprise. So if an entrepreneur is making less profit by using his resources in their current way than he could by using them elsewhere, the rational decision will be to shift them to the alternative use which could generate higher profits.
Key chapter points

- The assumption of profit maximisation provides a basis for beginning to look at the output decisions of individual firms. A firm will maximise its profits where marginal costs equal marginal revenue (MC = MR).

- A firm’s output decisions can be examined in the short run, (when some factors of production are fixed), and in the long run, (when all factors of production can be varied). Total costs can be divided into fixed and variable elements. These elements have different effects on total cost as output is increased.

- Economic costs are different from accounting costs, and represent the opportunity costs of the factors of production that are used.

- A firm’s short-run average cost (SRAC) curve is U shaped, due to diminishing returns beyond a certain output level.

- Once all the factors of production become variable, a firm’s short-run average cost (SRAC) curve can be shifted, and a firm’s minimum achievable average costs at any level of output can be depicted by a long-run average cost (LRAC) curve.

- The shape of the LRAC depends on whether there are increasing, constant or decreasing returns to scale. Even if increasing returns to scale are not achievable indefinitely as output rises, up to a certain minimum efficient scale of production (MES) there will be increasing returns to scale. Firms will reduce their average costs by producing on a larger scale up to the MES.
Quick revision questions

1. Explain the distinction between short-run and long-run costs.

2. The diagram shows the cost curves and revenue curves for Hans Tordam Co, a firm of tulip growers. Which of the following statements is true?
   I. Price $P$ and output $Q$ are the profit-maximising price and output levels for the firm.
   II. Price $P$ and output $Q$ are price and output levels at which the firm makes normal profits but no supernormal profits.
   III. Price $P$ and output $Q$ are the revenue-maximising price and output levels.

3. The diagram shows the revenue and cost curves for a profit-maximising monopoly firm, Lord and Masters Co. Which of the following statements are correct?
   I. If the firm has zero marginal costs and 100 per cent fixed costs, its profit-maximising output would be $OZ$.
   II. At profit-maximising output $OY$, supernormal profits for Lord and Masters Co are $STWX$.
   III. If the firm’s fixed costs increased, so that the $AC$ curve rose to a level where it is at a tangent to the $AR$ curve at point $W$, it would cease to make supernormal profit.

4. What is the law of diminishing returns?
5 Why might there be diseconomies of scale?
6 Which of the following is an example of an external economy of scale?
   A Increased wage costs due to falling unemployment in the region.
   B The employment of specialist managers by a firm to cope with higher output levels.
   C The extension of low-cost telecommunication links to an area of the country not previously served by such links.
   D Cheaper finance in recognition of the firm’s increased share of the market and therefore its stability.
7 Which of the following cannot be true? In the short run, as output falls:
   A average variable costs falls
   B average total cost falls
   C average fixed cost falls
   D marginal costs falls
8 The tendency for unit costs to fall as output increases in the short run is due to the operation of:
   A economies of scale
   B the experience of diminishing marginal returns
   C falling marginal revenue
   D increasing marginal returns
9 What is technical efficiency?
10 Harold Ippoli employs 30 people in his factory which manufactures sweets and puddings. He pays them $5 per hour and they all work maximum hours. To employ one more person he would have to raise the wage rate to $5.50 per hour. If all other costs remain constant, the marginal cost of labour per person is:
   A $0.50
   B $5.50
   C $15.00
   D $20.50
11 Which of the statements below best defines the difference between the short run and the long run?
   A Labour costs are fixed in the short run and variable in the long run.
   B Economies of scale are present in the long run but not in the short run.
   C At least one factor of production is fixed in the short run but in the long run it is possible to vary them all.
   D None of the factors of production is fixed in the short run.
12 Which of the following is not true?
   A A firm will carry on producing in the short run provided that price at least equals average variable cost.
   B A firm will carry on producing in the short run provided that price at least equals average fixed cost.
   C A firm will carry on producing in the short run provided that total revenue at least equals total variable cost.
   D A firm will stop producing in the long run if total revenue is less than total cost.
1. The distinction between the short run and the long run is that in the long run all factors of production (land, labour, capital, entrepreneurship) are variable. In the short run, at least one factor of production is fixed. In practice, often only the amount of labour input is variable.

2. B Statement I is incorrect. Profit is maximised where $MC = MR$; this is not at output $Q$ and price $P$ on the graph.

   Statement II is correct. At this price/output level, average cost equals average revenue. Normal profit is included in cost, and so the firm is making normal profits only, but no supernormal profits.

   Statement III is incorrect. Total revenue is maximised where $MR = 0$. This is not at output $Q$ and price $P$ on the graph.

3. D Statement I is correct, because if $MC = 0$, profits would be maximised where $MC = MR$, which would be at output $OZ$, where $MR = 0$. Statement II is correct. Supernormal profits per unit are the difference between $AR$ and $AC$ (price and average cost). This is $(W - X)$ or $(T - S)$. Total supernormal profits for output $OY$ are therefore illustrated by area $STWX$.

   Statement III is probably more difficult to understand but is also correct. If fixed costs increase, but variable costs remain the same, the $MC$ curve will be unchanged, and so the profit-maximising price will still be $OT$ and the profit-maximising output $OY$. But if higher fixed costs have raised average costs ($AC$) to point $W$, at this price and output level $AR = AC$, and so there will be no supernormal profits.

4. If one or more factors of production are fixed, but the input of another is increased, the extra output generated by each extra unit of input will eventually begin to fall. (Note: the law of diminishing returns relates to the marginal output generated by adding an extra unit of input: it does not mean that total output starts to decline.)

5. Diseconomies of scale are problems of size and tend to arise when the firm grows so large that it cannot be managed efficiently. Communications may become difficult, motivation may deteriorate because of alienation and senior management may find it difficult to identify the information they need in the vast volumes available.

6. C This is an external economy of scale.

   A is a diseconomy of scale.
   B is an internal economy of scale.
   D is an internal economy of scale.

7. C If output falls, fixed costs are divided over a smaller number of units, therefore average fixed costs will rise. The key to this question is to draw a diagram of the cost curves.

8. D The benefits of specialisation and the division of labour could allow increasing marginal returns.

   Economies of scale only operate in the long run.
   B results in rising unit costs in the short run.
   C is nothing to do with costs.

9. Technical efficiency means producing the output at which average cost is lowest (i.e. producing at the lowest point on the average cost curve).
10 D

<table>
<thead>
<tr>
<th>Description</th>
<th>$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cost of 31 people (at $5.50 per hour)</td>
<td>170.50</td>
</tr>
<tr>
<td>Cost of 30 people (at $5.00 per hour)</td>
<td>150.00</td>
</tr>
<tr>
<td>Marginal cost</td>
<td>20.50</td>
</tr>
</tbody>
</table>

Note that by increasing the wage for the 31st person to $5.50 per hour, the employer also had to increase the wage for the existing 30 people to $5.50 per hour.

11 C The key difference is that in the short run at least one factor of production is fixed in amount, resulting in diminishing returns. In the long run all factors of production are variable.

12 B A firm will carry on producing in the short run provided total revenue at least equals total variable cost, and is therefore making a contribution towards fixed costs (so C is true).

If total revenue must at least equal total variable cost, price (average revenue) must at least equal average variable cost (so A is true).

In the long run, if total revenue is less than total cost, the firm is not making a normal profit and so will stop producing.
Answers to chapter questions

1

(a) Marginal revenue is the incremental revenue which results from the sale of the last unit of output.

The figures in the table above show that marginal revenue is a constant $50 at all levels of output given. This means that average revenue (price) must also be a constant $50.

(b) The fixed costs of the firm are those costs which do not vary with output. The level of fixed costs are therefore the total costs of $110 at the output level of zero.

Marginal cost is the change in total cost arising from the production of the last unit of output. The marginal cost for each level of output is shown in the table.

(c) The firm will seek to maximise profits by producing at a level of output at which marginal cost equals marginal revenue. It can be seen from the table that this occurs between output level 8 (marginal cost: $40) and output level 9 (marginal cost: $56). Whether the firm can produce output between the values 8 and 9 will depend on the nature of the product. For whole units of output, total profit (total revenues minus total costs) is maximised at an output of 8, where total profit is $131.

2 (a) and (b) are correct; (c) is incorrect. An example might help. Suppose a firm has made 100 units of output, and now goes on to produce one more. The costs might be as follows:

<table>
<thead>
<tr>
<th>Output</th>
<th>Total revenue (TR)</th>
<th>Marginal revenue (TRn - TR-n-1)</th>
<th>Total costs (TC)</th>
<th>Marginal costs (TCn - TCn-1)</th>
<th>Total profit (TR - TC)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>(110)</td>
</tr>
<tr>
<td>1</td>
<td>50</td>
<td>50</td>
<td>140</td>
<td>30</td>
<td>(90)</td>
</tr>
<tr>
<td>2</td>
<td>100</td>
<td>50</td>
<td>162</td>
<td>22</td>
<td>(62)</td>
</tr>
<tr>
<td>3</td>
<td>150</td>
<td>50</td>
<td>175</td>
<td>13</td>
<td>(25)</td>
</tr>
<tr>
<td>4</td>
<td>200</td>
<td>50</td>
<td>180</td>
<td>5</td>
<td>20</td>
</tr>
<tr>
<td>5</td>
<td>250</td>
<td>50</td>
<td>185</td>
<td>5</td>
<td>65</td>
</tr>
<tr>
<td>6</td>
<td>300</td>
<td>50</td>
<td>194</td>
<td>9</td>
<td>106</td>
</tr>
<tr>
<td>7</td>
<td>350</td>
<td>50</td>
<td>229</td>
<td>35</td>
<td>122</td>
</tr>
<tr>
<td>8</td>
<td>400</td>
<td>50</td>
<td>269</td>
<td>40</td>
<td>131(max)</td>
</tr>
<tr>
<td>9</td>
<td>450</td>
<td>50</td>
<td>325</td>
<td>56</td>
<td>125</td>
</tr>
<tr>
<td>10</td>
<td>500</td>
<td>50</td>
<td>425</td>
<td>100</td>
<td>75</td>
</tr>
</tbody>
</table>

Marginal cost = $302 – $300 = $2

Because fixed costs have not changed, the marginal cost is both the increase in the total cost of production and the increase in the variable cost of production.
3 **Accounting profit**

<table>
<thead>
<tr>
<th></th>
<th>$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Revenue</td>
<td>160 000</td>
</tr>
<tr>
<td>Costs</td>
<td>125 000</td>
</tr>
<tr>
<td>Profit</td>
<td>35 000</td>
</tr>
</tbody>
</table>

**Economic profit**

<table>
<thead>
<tr>
<th></th>
<th>$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Revenue</td>
<td></td>
</tr>
<tr>
<td>Accounting costs</td>
<td>125 000</td>
</tr>
<tr>
<td>Opportunity cost of owner’s time – extra salary forgone</td>
<td></td>
</tr>
<tr>
<td>From alternative employment (20 000 − 12 000)</td>
<td>8 000</td>
</tr>
<tr>
<td>Rental of factory (opportunity cost of $80 000)</td>
<td>11 000</td>
</tr>
<tr>
<td>Opportunity cost of other capital tied up in the business (10% of $120 000)</td>
<td>12 000</td>
</tr>
<tr>
<td></td>
<td>156 000</td>
</tr>
<tr>
<td>Economic profit</td>
<td>4 000</td>
</tr>
</tbody>
</table>

4 (a) True. Average total cost (AC) comprises average fixed cost (AFC) and average variable cost (AVC). AFC falls as output rises, and the fall may be sufficient to outweigh a possible increase in AVC. In such a case, AC will fall while AVC rises.

(b) False. It is average fixed costs per unit that fall as output increases. Marginal fixed costs = 0, because fixed costs do not change when one extra unit of output is produced.

5 Diminishing returns occur when the **marginal** physical product of extra units of labour starts to decline. This begins to happen at output W, when the rate of increase in total output starts to decline as numbers employed continue to increase. Note this is a theory of diminishing **marginal** returns, it does not relate to a decline in total output.

6 (a) Large firms attract **better quality employees** if the employees see better career prospects than in a small firm.

(b) Specialisation of labour applies to management, and there are thus **managerial economies**; the cost per unit of management will fall as output rises.

(c) **Marketing economies** are available, because a firm can make more effective use of advertising, specialist salesmen, and specialised channels of distribution.

(d) Large companies are able to devote more resources to **research and development** (R & D). In an industry where R & D is essential for survival, large companies are more likely to prosper.

7 **Diminishing returns**. In the short run, some factors of production are fixed, and some are variable. This means that although a firm can increase the volume of its output in the short run, it can only do so within the constraint of having some fixed factors. As a result, the short-run average cost curve is U shaped, because of increasing and then diminishing marginal returns. Diminishing marginal returns occur within a given production capacity limit.

**Economies of scale**. In the long run, all factors of production are variable and so a firm can increase the scale of its output without any constraints of fixed factors. By increasing output capacity in this way, a firm might be able to reduce its unit costs, for example by mass-producing with bigger and more efficient machines or with more specialised machines. These cost reductions are economies of scale.
Chapter 5

Market structures

<table>
<thead>
<tr>
<th>Learning objectives</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Market structures</strong></td>
<td>LOS</td>
</tr>
<tr>
<td>Distinguish between perfect competition, monopolistic competition, monopoly,</td>
<td>LOS.1</td>
</tr>
<tr>
<td>oligopoly, duopoly, and oligopsony</td>
<td></td>
</tr>
<tr>
<td>Illustrate the relevant demand and supply curves</td>
<td>LOS.1.1</td>
</tr>
<tr>
<td>Evaluate why monopolistic firms are able to allocate or misallocate scarce</td>
<td>LOS.2</td>
</tr>
<tr>
<td>resources</td>
<td></td>
</tr>
<tr>
<td>Explain the long term pricing approach for a monopolistic firm</td>
<td>LOS.3</td>
</tr>
</tbody>
</table>

Topic list

1. Equilibrium under perfect competition
2. Monopoly
3. More about monopoly
4. Monopolistic competition and non-price competition
5. Oligopoly
6. Duopoly
7. Oligopsony and monopsony
Introduction

To understand the immediate environment within which a business operates, you need to be aware of the different market structures in which it might be operating. The purpose of this chapter is to consider output decisions by firms which operate in the forms of market structure characterised as perfect competition and monopoly. These are the two most extreme types of market structure.

Moving away from these extremes leads to the analysis of other forms of imperfect market structure, including monopolistic competition, oligopoly and duopoly as well as monopsony and oligopsony.

These different market structures are defined largely by the number of suppliers in the market. However, in each of the markets, the firm is a price searcher – searching for the price which will maximise profits, given the conditions in the market.
Before you begin

If you have studied these topics before, you may wonder whether you need to study this chapter in full. If this is the case, please attempt the questions below, which cover some of the key subjects in the area.

If you answer all these questions successfully, you probably have a reasonably detailed knowledge of the subject matter, but you should still skim through the chapter to ensure that you are familiar with everything covered.

There are references in brackets indicating where in the chapter you can find the information, and you will also find a commentary at the back of the Study Manual.

1. A pure monopoly is an extreme form of which type of competition? (Section 1)
2. Define monopoly. (Section 2.1)
3. What is price discrimination? (Section 3.1)
4. Give two arguments in favour of, and two arguments against, monopolies. (Sections 3.2.1, 3.2.2)
5. Define monopolistic competition. (Section 4.1)
6. What is an oligopoly? (Section 5.1)
7. Give an example of an oligopoly or duopoly structure that you are familiar with. (Sections 5, 6)
8. What is an oligopsony? (Section 7)
1 Equilibrium under perfect competition

Section overview

• In perfectly competitive markets, firms are price takers, and so their decisions are concerned with what output level will maximise profits. In imperfect competition, firms can influence the market price, and so their decisions are about what price to set as well as what volumes of output to produce. Pure monopoly is an extreme form of imperfect competition.

1.1 Perfect competition

Definition

Perfect competition: a theoretical market structure in which no supplier has an advantage over another.

Perfect competition acts as a useful theoretical benchmark with which to start our review of market structures.

(a) We can use it to judge or predict what firms might do in markets where competition shows some or most of the characteristics of being perfect.

(b) We can also contrast the behaviour of firms in less perfect markets. We shall be looking in this chapter at imperfect types of market structure – namely, monopoly, duopoly, monopolistic competition, oligopoly, monopsony and oligopsony.

Characteristics of perfect competition

• There are a large number of buyers and sellers in the market.

• Firms are ‘price takers’, unable to influence the market price individually. Buyers and sellers can trade as much as they want at the prevailing market price, as determined by the interaction of supply and demand.

• Producers and consumers have the same, perfect, information about the product and the market.

• The product is homogeneous: one unit of the product is the same as any other unit.

• There is free entry of firms into and free exit of firms out of the market: there are no barriers to entry. There are also no restrictions on the mobility of factors of production.

• There are no transport costs or information gathering costs.

• Producers and consumers act rationally. This means that producers will always try to maximise profits.

• Normal profits are earned in the long run.

Exam comments

You should be familiar with the assumptions of the ‘perfect competition’ model for your exam.

Question 1: Perfect market?

Think about the market for a particular product – say motor cars. To what extent is this market ‘perfect’, as defined by the criteria above?

(The answer is at the end of the chapter)
1.2 Equilibrium in the short run

How are price and output determined in the case of the profit-maximising firm operating under conditions of perfect competition in the short run?

We assume that in the short run the number of firms in the market is temporarily fixed. In these circumstances it is possible for firms to make supernormal profits or losses.

Figure 1 shows the cost and demand curves of a firm in the short run making supernormal profits. The demand curve is the horizontal line $D_1$ at price $P_1$. The curve is a horizontal line indicating that the firm has to accept the price that the market as a whole fixes for it. If the firm were to charge a higher price it would lose all its sales because customers, acting rationally and with perfect information, would buy the identical good from another supplier at a lower price. There is no point charging a lower price than the market price because the firm can sell all its output at the given price. Therefore the demand curve for the firm is perfectly elastic at price $P_1$.

Because the demand curve is perfectly elastic, it is also the marginal revenue curve; every new unit sold at price $P_1$ increases total revenue by the same amount, $P_1$.

Figure 1 also shows the average total cost curve (ATC) and the marginal cost curve (MC), with the MC cutting the ATC at the lowest point of the ATC. Given these cost curves and the demand curve $D_1$, the firm will produce the output $Q_1$, at the point where the MC curve cuts the MR horizontal curve which is indicated by point C. As long as $P > ATC$ then a positive supply of output will occur. This is the profit maximising level of output ($MC = MR$).

Note, however, that the profit maximising level of output is not the same as the level of technical efficiency. This would occur at output $E$, where average cost (ATC) is at its minimum. By definition, technical efficiency is achieved if a firm is producing the level of output at which its average costs are minimised.

In Figure 1, average revenue is greater than average cost at the profit-maximising level of output, so the firm is making supernormal profits indicated by the rectangle ABCD. These supernormal profits will attract new firms into the industry and the price will be bid down. This new position is illustrated in Figure 2, where the new price is $P_2$, but here the firm makes a loss shown by the rectangle WXYZ. Once again the firm produces where $MC = MR$ giving an output of $Q_2$. A firm could choose to do this for a short period so long as revenues covered its variable costs; in this case $MC = MR$ is the loss minimising position rather than the profit maximising position.
Figure 2: Losses in the short run

Figure 2 shows losses in the short-run. In the long-run, the entry of new firms will increase supply and cause the price to fall which will allow supernormal profits to be made.

Question 2: Perfect competition

In conditions of perfect competition, the demand curve for a firm’s product is:

(a) identical to the firm’s marginal revenue curve. True or false?
(b) perfectly inelastic. True or false?

(The answer is at the end of the chapter)

Importantly though, it is only the demand curve for the firm which is horizontal because of the firm’s position as a price taker. The demand curve facing the industry may not be horizontal. Its elasticity will vary according to the industry, but it is usually drawn as the normal downward-sloping demand curve.

Similarly, market supply is shown in the traditional way, as an upward-sloping curve.

The demand and supply diagram for the industry is shown as figure 3 below.

Figure 3: Supply and demand in a perfectly competitive industry

The industry supply and demand model also explains why a firm in a perfectly competitive industry can go from making a supernormal profit to making a loss (Figures 1 and 2 above) as new firms enter the industry.

The introduction of new producers leads to an outward shift in the industry supply curve from $S$ to $S_1$ (Figure 4). This shift causes price to fall from $P_1$ to $P_2$, meaning that the individual firm, which takes the market price, moves from making a supernormal profit (Figure 1) to a loss (Figure 2).
1.3 Equilibrium in the long run

In a perfectly competitive market in the long run, the firm cannot influence the market price and its average revenue curve is horizontal. The firm's average cost curve is U-shaped.

In the long run, firms will enter the industry attracted by the opportunity for supernormal profit, and loss-making firms will leave. A long-run equilibrium will be established where there is just enough profit (normal profit) to keep existing firms in the industry. There is no further incentive for firms to enter the industry.

This long-term position is illustrated in Figure 5. Note the following points about Figure 5:

(a) The market price, $P$, is the price which all individual firms in the market must take.

(b) If the firm must accept a given $MR$ (as it must in conditions of perfect competition) and it sets $MR = MC$, then the MC curve is in effect the individual firm's supply curve (Figure 5(b)). The market supply curve in Figure 5(a) is derived by aggregating the individual supply curves of every firm in the industry.

(c) Consumer surplus is represented by the area to the left of the demand curve above $P$.

(d) The firm is operating at its most cost-effective point (the lowest point on the AC curve). The long-run price will equal the minimum value of AC.

Figure 5(b) also shows us that the individual firm's equilibrium position occurs where price equals marginal cost. Since price is a measure of the value of the good to a consumer, and marginal cost
measures the cost to the producer of attracting resources from alternative uses, then the price of the last unit of output is equal to the opportunity cost of its production. This signifies that allocative efficiency is being achieved.

The long-run equilibrium position under perfect competition is unique because it is the only market condition which achieves allocative efficiency.

In the long run, all firms in the industry will have MR = MC = AC = AR = price, as at output Q, in Figure 5(b). Because this position earns the entrepreneur the desired return on their capital (normal profit), ensures allocative efficiency, and means that firms operate their most cost-effective point, long-run equilibrium under perfect competition is held to be a desirable model for an economy.

**Exam comment**

This explanation of long run equilibrium under perfect competition and the accompanying diagrams are fundamental knowledge.

**Question 3: Equilibrium perfect competition**

A perfectly competitive firm will be in equilibrium where price is equal to marginal cost. True or false?  
(The answer is at the end of the chapter)

**Question 4: Small firm**

A small perfectly competitive firm manufactures 200 wooden garden benches each month which it sells for $40 each. The table below shows the firm's costs:

<table>
<thead>
<tr>
<th>Total variable cost</th>
<th>$7 200</th>
</tr>
</thead>
<tbody>
<tr>
<td>Marginal cost</td>
<td>$40</td>
</tr>
<tr>
<td>Total fixed cost</td>
<td>$1 800</td>
</tr>
</tbody>
</table>

What should the firm do in the short term?

A increase output  
B cease production  
C lower its price  
D maintain output at its present level  
(The answer is at the end of the chapter)

**2 Monopoly**

**Section overview**

- Firms will generally try to earn supernormal profits if they can. Competition, though, tends to erode supernormal profits, and firms may have to be satisfied, when equilibrium is reached, with just normal profits. The ability to sustain supernormal profits depends on the nature of competition in the industry.

**2.1 The monopoly market**

**Definition**

In a monopoly, there is only one firm, the sole producer of a good which has no closely competing substitutes.
Whereas perfect competition represents one extreme of the market spectrum (characterised by a large number of suppliers), monopoly represents the other end of that spectrum.

In theory, a monopoly is a market in which there is a single supplier, and many consumers. The single supplier controls market supply, and can control price.

In practice, legislation often deems a firm in the private sector to be a monopoly once its market share exceeds a certain level. For example in the UK a firm is deemed to hold a monopoly if its market share exceeds 25 per cent. (However, this definition can be problematic in economics and is not used in other countries, because it would also include duopolists and some oligopolists.)

A firm's monopolistic position may result from some natural factor which makes it too costly for another firm to enter the industry. For example, in the domestic water supply industry it will normally be too costly for a second firm to lay a second water supply system to compete for part of the business of an existing sole supplier: the sole supplier enjoys a natural monopoly.

In other cases, a monopoly may be formed by mergers of a number of firms in an industry. However it is formed, a monopoly can only exist if potential competitors are kept out of the market by barriers to entry (see below). For a monopoly, the total market supply is identical to the single firm's supply, and the average revenue curve in monopoly is the same as the total market demand curve.

A monopolist can either be a price maker (and so a quantity taker) or set quantity (and take the equilibrium price which results). However, it cannot fix both price and quantity because it cannot control market demand.

If price must be reduced to increase unit sales, average revenue is falling and marginal revenue will always be lower than average revenue. If the monopolist increases output by one unit, the price per unit received will fall, so the extra revenue generated by the sale of the extra unit of the good is less than the price of that unit. The monopolist therefore faces a downward sloping AR curve with an MR curve below the AR curve (Figure 6). For any given price, AR is double the MR on straight-line average revenue curves.

Marginal revenue can be negative. This occurs when demand is price inelastic and, although lowering the price increases sales demand, the volume increase is proportionately less than the price decrease and so total revenue falls.

**Question 5: Price and output level**

Study the diagram in Figure 6 above. At what price and output level would the firm maximise its sales revenue?

(The answer is at the end of the chapter)
It is important that you should understand what the MR and AR (demand) curves are showing us in Figure 6.

(a) At output quantity $X$, the marginal revenue earned from the last unit produced and sold is $MR_X$, but the price at which all the $X$ units would be sold is $P_X$. This is found by looking at the price level on the AR curve associated with output $X$.

(b) Similarly, at output quantity $Y$, the marginal revenue from the last unit produced and sold is $MR_Y$, but the price at which all $Y$ units would be sold on the market is, from the AR curve for $Y$ output, $P_Y$.

(c) At output $Z$, the marginal revenue from the last unit produced is zero, and the price at which all $Z$ units would be sold is $P_z$. Total revenue will be maximised at $Z$. If any more units are sold, MR will be negative, thereby reducing total revenue.

### 2.2 Profit-maximising equilibrium of a monopoly

The condition for profit maximisation is, as we have seen, that marginal revenue should equal marginal cost ($MC = MR$). This is true for any firm. As long as marginal revenue exceeds marginal cost, an increase in output will add more to revenues than to costs, and therefore increase profits.

A monopolist will have the usual U-shaped cost curves.

#### 2.2.1 Monopolist earning normal profits

Figure 7 shows a monopoly equilibrium where the AC curve touches the AR curve at a tangent, at exactly the same output level where $MC = MR$ (output $Q$). Since $AC = AR$ and $AC$ includes normal profits, the monopolist will be earning normal profits but no supernormal profits.

![Figure 7](Equilibrium of a monopoly firm earning normal profits)

In this situation, the monopoly will make a loss by producing at output higher than $Q$, and so it will have to produce at an output level which is well below the capacity at which its average costs are minimised (output $Q_1$).

Monopolies are usually able to earn ‘monopoly’ or supernormal profits in the long run as well as the short run, and the situation illustrated in Figure 7 will be rare for a monopoly, although (as we shall see later in this chapter) this is a long-run equilibrium situation for firms in the type of market structure known as monopolistic competition.

In perfect competition, a firm will not be able to earn supernormal profits in the long run because they would be ‘competed away’ by new entrants to the industry. However, a monopoly firm can earn supernormal profits in the long run as well as in the short run, because there are barriers to entry which prevent rivals entering the market.
2.2.2 Monopolist earning supernormal profits

Figure 8 shows the position of the monopolist earning supernormal profits in the short run. SMC is the short-run marginal cost curve and SAC represents short-run average costs.

![Figure 8 Monopolist’s short-run equilibrium](image)

In Figure 8, the monopolist’s profit is maximised at output $Q_M$, where marginal cost (MC) equals marginal revenue (MR), and the price charged is the average revenue $P_M$. The monopolist is earning supernormal profits represented by the rectangular area $P_M ZYX$.

2.3 Prices and output

LO 5.2

In Figure 9 a perfect competition demand curve has been superimposed on to the monopolist’s demand curve (Figure 8).

The monopolist will choose to produce output $Q_m$ and sell it at price $P_m$ because this is the level of output where its short run marginal cost (SMC) equals marginal revenue MR. If the monopolist had been operating in perfect competition with a horizontal demand and MR curve, it would choose to produce output $Q_{pc}$ and sell it at price $P_{pc}$.

![Figure 9 Monopoly and perfect competition compared](image)

The monopoly therefore has the effect of **restricting output** (the difference between $Q_{pc}$ and $Q_m$ in Figure 9) and **raising the price** (the difference between $P_m$ and $P_{pc}$) compared to levels under perfect competition.
3 More about monopoly

3.1 Price discrimination

Price discrimination occurs when a firm sells the same product at different prices in different markets.

Question 6: Price discrimination

You are likely to have encountered examples of price discrimination in practice. Can you recall any?

(The answer is at the end of the chapter)

Four basic conditions are necessary for price discrimination to be effective and profitable:

(a) The seller must be able to control the supply of the product and keep out any competitors who could undercut the premium price. To this extent, the market must be imperfect.

(b) There must be at least two distinct markets with no cross-over between them. For example, a rail fare will either be for a peak time or off-peak. If a customer buys an off-peak fare he or she cannot use it during a peak period.

(c) The seller must be able to prevent the resale of the good by one buyer to another. The markets must, therefore, be clearly separated so that those paying lower prices cannot resell to those paying higher prices. The ability to prevent resale tends to be associated with the character of the product, or the ability to classify buyers into readily identifiable groups. Services are less easily resold than goods while transportation costs, tariff barriers or import quotas may separate classes of buyers geographically and thus make price discrimination possible.

(d) There must be significant differences in the willingness to pay among the different classes of buyers. In effect this means that the elasticity of demand must be different in at least two of the separate markets so that total profits may be increased by charging different prices.

A monopolist can increase its supernormal profit by selling more of its output in the market which is most inelastic.

We can illustrate this as in Figure 10 below. Figure 10(a) and Figure 10(b) are different markets or different market segments. Demand in market 1 is more inelastic than demand in market 2 (and demand in market 2 is more elastic than demand in market 1). The firm is assumed to have the same marginal costs in each market.

In each market, the profit maximising output is determined by MC = MR, and costs are assumed to be the same in both markets.

The profit-maximising price in market 1 (P₁) is higher than the profit-maximising price in market 2, even though their cost structures are identical. This is because demand is more inelastic in market 1 than in market 2. And the supplier makes more profit in market 1 than market 2.

This is part of the logic behind price discrimination: separating a product into different markets with different elasticities allows the monopolist to increase profits by charging higher prices in the market with inelastic demand.
3.1.1 Perfect price discrimination

We can use a diagram to illustrate how a monopolist seller practising price discrimination can maximise revenue in a situation where it can price discriminate perfectly because it knows the price that each consumer will pay.

Figure 11 Price discrimination

Figure 11 demonstrates firstly the equilibrium position of a monopolist who does not discriminate. He produces at the point C, which is where marginal cost equals marginal revenue, producing output Q3 and selling at price P. His total revenue is given by the rectangle OPBQ3.

If we assume that the monopolist can discriminate perfectly, then he can sell each unit for a different price as indicated on the demand curve. Therefore he can sell the first unit Q1 at the price P1, and the second unit Q2 at the price P2. This follows for all units sold so that the demand curve now becomes the marginal revenue curve; each extra unit sold is sold for the price indicated on the demand curve, each previous unit being sold for the higher price relevant to that unit.

The perfectly discriminating monopolist will still maximise profits by producing at the level of output where MC = MR, but the marginal revenue curve is the demand curve D. The perfectly discriminating monopolist therefore produces at the point E where marginal cost equals the new marginal revenue, producing Q4 units.
Recall that the total revenue for the non-discriminating monopolist by producing at the level of output was equivalent to area OPBQ. The additional revenue of the discriminating monopolist is represented by the areas APB plus QBEQ. The discriminating monopolist has thus maximised his revenue (consistent of course with maximising his profit). If the monopolist did not wish to maximise profit but wished simply to maximise revenue he would expand production to the point Q when his total revenue would be the area OAQ.

Take care not to confuse maximising revenue with maximising profit. Increasing output beyond Q in the example of Figure 11, will not increase profit as marginal costs exceed marginal revenue for each additional unit sold.

### 3.1.2 Examples of price discrimination

Various examples show that the conditions necessary for price discrimination can be met. However, there tend to be four main ways in which price discrimination is applied:

(a) **Time**: Markets may be separated by a time barrier, for example where the cost of telephone calls varies according to the time of day at which they are made. Rail operating companies charge cheaper rates for off peak travel. Holiday companies charging a higher price for a given holiday at certain times of the year is another example. These are examples of services which cannot be transferred from the cheaper to the more expensive market, because they are defined by the time period to which they relate.

(b) **User groups**: Price discrimination also occurs where it is possible to separate customers into clearly defined groups. Industrial users of gas and electricity are able to purchase these fuels more cheaply than are domestic users. Similarly milk is sold more cheaply to industrial users, for example for making into cheese or ice cream, than to private households.

(c) **Income**: A third way price discrimination can be applied is on the basis of income (for example, concessionary travel fares offered to students).

(d) **Place**: Finally, price discrimination could be applied according to place – for example, banks offering different interest rates for online accounts compared to branch-only accounts.

---

**Question 7: Price discrimination**

Explain why it is possible for a railway or airline to charge different fares for passengers using the same service.

(The answer is at the end of the chapter)

### 3.2 Are monopolies beneficial or harmful?

**Section overview**

- **Monopoly** may be beneficial (because of economies of scale) or harmful (because they restrict output and raise prices). Government policies have been directed at the harmful aspects of monopoly.

- There are advantages and disadvantages of monopolies. It may be that monopolies encourage complacency about costs (X-inefficiency) and may produce allocative and technical inefficiency. Goals other than profit maximisation pursued in large companies could also result in inefficiencies.

We have identified two key characteristics of monopolies:

(a) A monopolist is likely to produce less output but charge a higher price for it than a comparable firm operating in conditions of perfect competition, unless the monopolist can achieve economies of scale that a smaller firm could not. This leads to the monopolist earning extra profits and also a social cost or deadweight burden of monopoly.

(b) Monopolists can practise price discrimination.

These two points might suggest that monopolies are a bad thing. But there are economic arguments both for and against monopolies.
3.2.1 Arguments in favour of monopolies

A firm might need a monopoly share of the market if it is to achieve maximum economies of scale. Economies of scale mean lower unit costs, and lower marginal costs of production. Therefore, we could argue monopoly provides a better utilisation of resources and technical efficiency even though it is not operating at the level of allocative efficiency. The consumer is likely to benefit from these cost efficiencies through lower prices from the monopoly supplier. Economies of scale shift the firm’s cost curves to the right, which means that it will maximise profits at a higher output level, and quite possibly at a lower selling price per unit too.

So-called natural monopolies exist because of a very high ratio of fixed costs to variable costs. Think, for example, of the network of pylons needed to supply electricity via a national grid. Such a cost structure makes it very likely that significant economies of scale will exist. Therefore, a monopoly would be the most cost-effective way of organising production.

Monopolies can afford to spend more on research and development, and are able to exploit innovation and technological progress much better than small firms and they can safeguard the rewards of their risks through securing patent rights.

Monopolies may find it easier than small firms to raise new capital on the capital markets, and so they can finance new technology and new products. This may help a country’s economy to grow.

Monopolies will make large profits in the short term, but in many cases their profits will eventually encourage rival firms to break into their market, by developing rival products which might have a better design, better quality or lower price. It can therefore be argued that temporary monopolies can stimulate competition, and are in the longer term interests of consumers.

3.2.2 Arguments against monopolies

Arguments against monopolies include the following:

(a) The profit-maximising output of a monopoly is at a point where total market output is lower and prices are higher than they would be under perfect competition. Consumer surplus is also reduced under monopoly compared to perfect competition, suggesting that the monopolist is benefiting at the expense of society as a whole.

(b) Monopolies do not achieve allocative efficiency since the prices they charge are greater than marginal cost.

(c) Monopolies do not use resources in the most efficient way possible (technical efficiency). Efficient use of resources can be defined as combining factors of production so as to minimise average unit costs. The profit-maximising output of a monopoly is not where average costs (AC) are minimised, and so monopolies are not efficient producers.

(d) Monopolists can carry out restrictive practices, such as price discrimination, to increase their supernormal profits.

(e) The higher prices and supernormal profits encourage firms in competitive markets to want to become monopolies, and they can do this by trying to create product differentiation, by introducing differences between their own products and the products of rival competitors. These differences might be real product design or quality differences, or imaginary differences created by a brand name and a brand image. This can be beneficial for producers, but at the expense of consumers.

(f) Because they are not threatened by competition and can earn supernormal profits, monopolies might become slack about cost control, so that they fail to achieve the lowest unit costs they ought to be capable of. They may also adopt a complacent attitude to innovation. Because a monopolist is able to maintain supernormal profits in the long run (due to barriers to entry) it has less need to innovate than a firm operating in a more competitive market would have.

(g) Monopolies might stifle competition, by taking over smaller competitors who try to enter the market or by exploiting barriers to entry against other firms trying to enter the market.

(h) If a monopoly controls a vital resource, it might make decisions which are damaging to the public interest. This is why the government often chooses to put vital industries under state control (for example, health care, the police service and the postal service).

(i) There might be diseconomies of scale in a large monopoly firm.
3.3 Barriers to entry

**Definition**

**Barriers to entry**: factors which make it difficult for suppliers to enter a market, and therefore allow supernormal profits to be maintained in the long run.

**Barriers to entry** can be classified into several groups:

(a) **Product differentiation barriers.** An existing monopolist would be able to exploit its position as supplier of an established product so that the consumer can be persuaded to believe it is a top quality product. A new entrant to the market would have to design a better product, or convince customers of the product's qualities, and this might involve spending substantial sums of money on research and development, advertising and sales promotion.

(b) **Exclusive control barriers.** These exist where an existing monopolist (or oligopolist) has access to, and exclusive control over, cheaper raw material sources or know-how that the new entrant would not have. This gives the existing monopolist an advantage because his input costs would be cheaper in absolute terms than those of a new entrant.

(c) **Economies of scale.** These exist where the long run average cost curve for firms in the market is downward sloping, and where the minimum level of production needed to achieve the greatest economies of scale is at a high level. New entrants to the market would have to be able to achieve a substantial market share before they could gain full advantage of potential scale economies, and so the existing monopolist would be able to produce its output more cheaply.

(d) The amount of **fixed costs** that a firm would have to sustain, regardless of its market share, could be a significant entry barrier.

(e) **Legal barriers.** These are barriers where a monopoly is fully or partially protected by law. For example, there are some legal monopolies (nationalised industries perhaps) and a company's products might be protected by patent (for example, computer hardware or software).

(f) **Cartel agreements.** If firms work together and agree to co-operate rather than compete they can, in effect, form a monopoly. Such collusion can take the form of **price fixing**.

(g) **Geographical barriers.** In remote areas, the transport costs involved for a supplier to enter a market may prevent it from entering that market. For example in remote villages and towns local corner or town shops have historically had a local monopoly, although the barriers to entry to such a market have been weakened by the growth of the internet and online shopping in less remote areas.

3.4 Allocative inefficiency, technical inefficiency and X-inefficiency

One of the arguments against monopolies is that they are inefficient compared with firms operating under perfect competition because, unlike perfectly competitive firms, they do not produce at an output where price equals marginal costs. Instead, they restrict production and raise price to the level that maximises profit. As a result, less is produced and consumed than would be the case under perfect competition. The resources that would have been used are diverted elsewhere, to produce things that households actually want less than the monopolist's product. This implies that monopolies are inefficient in allocating resources. This is called **allocative inefficiency**. Allocative efficiency is only achieved by producing at the point where price equal marginal cost.

Also, because a monopolist does not produce at the lowest point on its average cost curve, it is producing at a level of **technical inefficiency**.

Another criticism of monopolies is that they are wasteful of costs, and spend more than they need to. The lack of competition, perhaps, makes monopolies complacent, so they do not use resources with **maximum efficiency**. This type of over-spending inefficiency is called **X–inefficiency**.
The difference between technical inefficiency and X-inefficiency is illustrated in Figure 12.

(a) Figure 12(a). If a monopolist maximises profit at output level $Q_2$, there is technical inefficiency because the firm would produce more at lower cost at output $Q_1$, where it achieves the lowest possible cost per unit of output.

(b) Figure 12(b). If a monopolist has an average cost curve $AC_1$, when it ought to use resources more efficiently and have an average cost curve $AC_2$, there is X-inefficiency. The problem is that the whole cost curve is too high, rather than the firm is producing at the wrong level of output.

All monopolies might be accused of some X-inefficiency, but there has been a view that State-owned monopolies have a tendency to be more X-inefficient than monopolies which are private companies. This may be because State-owned monopolies have different objectives from those of private sector organisations.

4 Monopolistic competition and non-price competition

Section overview
- When price competition is restricted, firms usually go in for other forms of competition, such as sales promotion and product differentiation.
- A firm which operates in conditions of monopolistic competition will have a short-run equilibrium in which it can make supernormal profits, and a long-run equilibrium in which it cannot.
- Monopolistic competition is more wasteful than perfect competition.

4.1 Monopolistic competition

Definition
Monopolistic competition is a market structure in which firms’ products are comparable rather than homogeneous. Product differentiation gives the products some market power by acting as a barrier to entry. Monopolistic competition is a market structure which combines features of perfect competition and monopoly.

A firm operating in conditions of monopolistic competition has a downward sloping demand curve like a monopoly: the quantity of output customers demand responds to the price at which the firm is prepared to sell. The downward sloping demand curve is possible because of product differentiation created by the firm. However, unlike a monopoly firm, a firm operating under monopolistic competition is unable to

$\text{LO 5.1}$
utilise barriers to entry against other firms. Indeed, the firm already competes with rivals, which can take retaliatory competitive action if the firm makes big profits. (Remember, the absence of barriers to entry is a feature of perfect competition.)

Firms in monopolistic competition (as well as in oligopoly, which we discuss later in this chapter) will try to avoid competition on price in order to preserve their position as price maker. The market will be characterised by a large number of firms so strategic interactions between the firms are relatively unimportant. They will often resort to non-price competition instead, perhaps through advertising and sales promotion, or through product differentiation. With product differentiation, suppliers try to create differences between their products and other similar products. These differences might be real (for example, design differences) or largely imaginary and created mainly by advertising and brand image (for example, 'designer label' clothing and washing powders).

Question 8: Product differentiation
See if you can think of other examples of product differentiation.

(The answer is at the end of the chapter)

4.2 Profit-maximising equilibrium

A firm which operates in conditions of monopolistic competition will have a short-run equilibrium in which it can make supernormal profits, and a long-run equilibrium in which it cannot. In the long run, the monopolistic competitor cannot earn supernormal profits since there are no entry barriers. Its short-run supernormal profits will be competed away by new entrants. As a result of competition, the demand curve will move to the left and the firm will eventually be able to achieve normal profits only.

However, in the short run, a firm in monopolistic competition is very similar to a monopoly.

The short-run equilibrium for a firm in monopolistic competition is illustrated in Figure 13 below. This is the same as the equilibrium of a monopoly firm earning supernormal profits. The firm makes supernormal profits of \((P - A) \times Q\) units, shown by the area of the rectangle PQBA.

Figure 13 The short-run equilibrium of a firm in monopolistic competition

Note that although the short-run equilibrium of a firm in monopolistic competition resembles that of a monopolist's, the demand curve (average revenue curve) for the firm in monopolistic competition is likely
to be more elastic. This is because the customer in the market has a choice between products, and can be tempted to move between products by advertising campaigns (which are a feature of this type of market structure).

The **long-run equilibrium** for a firm in monopolistic competition is illustrated by Figure 14. This is the same as the equilibrium of a monopoly firm which earns no supernormal profits, but **normal profits only**. The supernormal profits earned in the short run have attracted new entrants and so have been competed away.

---

**Figure 14 The long-run equilibrium of a firm in monopolistic competition**

The competitive rivalry resulting from the new entrants to the market causes the firm to lose some of its customers, but not all. This is an important point: because the firm has established a **brand loyalty** it will be able to retain some of its customers despite the entry of the new competitors into the market.

However, the loss of customers is reflected in a leftward shift of the demand curve, and the reduction of profit levels to normal profit only.

However, although monopolistic competition creates **normal profit** in the long run like perfect competition, unlike perfect competition it **does not achieve allocative or technical efficiency**.

Allocative efficiency generally occurs where price equals marginal cost. However, in Figure 14 price ($P_1$) actually sits higher than marginal cost. This means that output could be increased so some people could be made better off without others suffering. Allocative efficiency in this instance is achieved at quantity $Q_A$ and price $A$.

Technical efficiency is not achieved because the average cost of the equilibrium output is not at the lowest point on the average cost curve ($T$). This would be achieved where output quantity equals $Q_T$. Figure 14 also illustrates that the firm has excess capacity, because output at the level of technical efficiency ($Q_T$), which is the same as the level of $Q_A$, is greater than current output of $Q_1$.

As a result of the allocative and technical inefficiency, monopolistic competition still gives rise to higher prices and lower outputs than perfect competition.

---

**4.3 Implications of monopolistic competition**

Because profit-maximising output under monopolistic competition is lower than it would be under perfect competition and is at a point where average costs are not minimised, monopolistic competition, like monopoly, is arguably more **wasteful of resources** than perfect competition.

Since firms in monopolistic competition cannot expand their output to the level of minimum average cost output without making a loss, industries marked by monopolistic competition will always tend to have
excess capacity. (Check this in Figure 14: profit is maximised at output $Q_1$ but output $Q_1$ is lower than the output level where AC would be minimised ($Q_T$). However, at the lowest point on the AC curve, $AC > AR$. The firm will be loss-making at this point ($Q_T$) so will not produce here.)

It can also be argued that it is wasteful to produce a wide variety of differentiated versions of the same product. If a single version of the same product were made, firms might be able to achieve economies of scale with large-volume production (and so shift their cost curves to the right).

Some methods that are used to create product differentiation are a waste of resources. Advertising costs are arguably an example of this, although some would argue that promotional activity actually adds utility to a product. Similarly, it is debatable how much value packaging adds to a product, although packaging helps distinguish one brand from another.

Nonetheless, while there are some indications that monopolistic competition is wasteful, there are other reasons to argue that monopolistic competition is not so wasteful of resources.

(a) Some product differentiation is 'real', where there are technical differences between similar goods from rival firms. In such situations, consumers therefore have more options to choose from when there is product differentiation and this means their requirements are likely to be satisfied better than if there were just a single, basic, low-price good available, without any choice.

(b) If product differentiation is entirely imaginary, created by brand image and advertising when the goods of rival firms are exactly the same, rational consumers should opt for the least-cost good anyway.

5 Oligopoly

Section overview

- Oligopolies might collude and make a formal or informal cartel agreement on the price for the industry and output levels for each firm. The kinked oligopoly demand curve may explain why there is price stability (and non-price competition) in many oligopoly markets.

5.1 The nature of oligopoly

Definition

Oligopoly: a market structure where a few large suppliers dominate the market.

Oligopoly differs from monopoly in that there is more than one firm in the market and from monopolistic competition because in oligopoly the number of rival firms is small. The UK supermarket industry can be seen as an oligopoly, because it is dominated by a small number of key players. For example, at the end of 2008, the 'Big Four’ – Tesco, Asda, Sainsbury and Morrison – accounted for approximately 75 per cent of the UK grocery market. In Australia, the two major players, Coles and Woolworths, accounted for approximately 80 per cent of the market at the end of 2009. The Australian situation is often also labelled a duopoly (discussed in Section 6).

The size of the existing firms in an oligopoly is likely to act as a barrier to entry to potential new entrants, and can allow oligopolists to sustain abnormal profits in the long run.

Oligopolists may produce a homogeneous product (oil, for example) or there may be product differentiation (cigarettes and cars, for example). An oligopoly will exhibit either price or non-price competition between firms.

Another key feature of oligopoly is that firms' production decisions are interdependent. One firm cannot set price and output without considering how its rivals' response will affect its own profits. An oligopolist's pricing and output decisions will usually depend on what assumptions they make about their competitors' behaviour.
Exam comment

The examiner regards this interdependence of decision making as fundamentally important to any discussion of oligopoly.

One strategy which an oligopolist might adopt is to co-operate with other firms, and such a strategy will give rise to a cartel.

5.2 Price cartels by oligopolist producers

A price cartel or price ring is created when a group of oligopoly firms combine to agree on a price at which they will sell their product to the market. The market might be willing to demand more of the product at a lower price, while the cartel attempts to impose a higher price (for higher unit profits) by restricting supply to the market.

Each oligopoly firm could increase its profits if all the firms together control prices and output as if the market were a monopoly, and split the output between them. This is known as collusion, which can either be tacit or openly admitted. Collusion usually leads to higher prices and lower outputs than the free market equilibrium, and so reduces consumer surplus and consumer sovereignty.

Cartels are illegal but difficult to prevent. There might still be price leadership. This occurs when all firms realise that one of them is initiating a price change that will be of benefit to them all, and so follow the leader and change their own price in the same way.

Figure 15 shows that in a competitive market, with a market supply curve $S_1$ and demand curve $D$, the price would be $P_1$ and output $Q_1$. A cartel of producers might agree to fix the market price at $P_2$, higher than $P_1$. But to do so, the cartel must also agree to cut market supply from $Q_1$ to $Q_2$, and so fix the market supply curve at $S_2$.

5.2.1 Establishing a cartel

Establishing a cartel depends on two things:

- The firms in the cartel must be able to control supply to the market.
- The firms must agree on a price and on the output each should produce.

In Figure 15, if the market price is fixed at $P_2$, firms would want to supply output $Z$ in a free market. This cannot be allowed to happen; otherwise market price $P_2$ could not be sustained.

The main weakness with cartels is that each firm is still seeking the best results for itself, and so there is an incentive for an individual firm to break the cartel agreement by secretly increasing its output and selling it at the fixed cartel price. However, if all firms increased their output in this way, the cartel would collapse.
because the high price could not be sustained without a restricted output, and excess supply on the market would force down the price.

This has been the common experience of the oil-producing countries of the Organisation of Petroleum Exporting Countries (OPEC). Attempts to agree on a restricted output quota for each country in order to push up oil prices have often broken down because some member countries exceeded their quota, or sold below the cartel’s agreed price.

The success of a price cartel will depend on several factors:
(a) Whether it consists of most or all of the producers of the product.
(b) Whether or not there are close substitutes for the product. For example, a price cartel by taxi drivers might lead to a shift in demand for buses and trains, because these are possible substitutes.
(c) The ease with which supply can be regulated. In the case of primary commodities, such as wheat, rice, tea and coffee, total supply is dependent on weather conditions and even political events in the producing country.
(d) The price elasticity of demand for the product. Cartels are likely to be most effective for goods which are inelastic. An attempt to raise prices by cutting output of an elastic good might result in such a large a fall in demand and such a small rise in price that the total income of producers also falls.
(e) Whether producers can agree on their individual shares of the total restricted supply to the market. This is often the greatest difficulty of all.

5.3 The kinked oligopoly demand curve

Price cartels do not always exist in an oligopoly market. So how does an oligopoly firm which is competing with rival oligopoly firms decide on its price and output level? A feature of oligopoly markets, remember, is that each firm’s pricing and output decisions are influenced by what its rivals might do.

When demand conditions are stable, the major problem confronting an oligopolist in fixing his price or output is judging the response of his competitor(s) to the prices he has set. An oligopolist is faced with a downward sloping demand curve, but the nature of the demand curve is dependent on the reactions of his rivals. Any change in price will invite a competitive response. This situation is described by the kinked oligopoly demand curve, shown in Figure 16.

![Figure 16 Kinked oligopoly demand curve](image)

The kinked demand curve is used to explain how an oligopolist might have to accept price stability in the market.
Working with the knowledge that oligopolists set prices, let us assume that the oligopolist is currently charging price $P$, and producing output $Q$, meaning he is producing at the kink on the demand curve.

(a) If the oligopolist were to raise his prices above $P$, his competitors are likely to keep their price lower so that many consumers would buy from them instead. An example is the difficulty which individual petrol companies have in raising the price of petrol at garages. If competitors do not raise their prices too, the firm usually soon has to restore its prices to their previous level. In other words, the demand curve would be quite elastic at these increased prices, because consumers will switch to an alternative supplier.

(b) If, on the other hand, the oligopolist were to reduce his prices below $P$, competitors would probably do the same. Total market demand might rise, but the increase in demand for the individual oligopolist’s products would probably be quite low. Demand is therefore likely to be inelastic at prices below $P$, hence the ‘kink’ in the demand curve, with the curve being elastic at prices above $P$ and inelastic at prices below it.

The marginal revenue (MR) curve is discontinuous at the output level where there is the kink in the demand curve. The kink in the demand curve explains the nature of the marginal revenue curve MR. At price $P$, output $OQ$, the MR curve falls vertically because at higher prices the MR curve is based on the more elastic demand curve, and at prices below $P$ the MR curve is based on the less elastic demand.

### 5.4 Profit maximisation

A firm maximises its profit at the point where $MR = MC$. However, the discontinuity in the MR curves means that there will be a number of points where $MR = MC$ (represented by the range $XY$ on Figure 16). There is thus a wide range of possible positions for the MC curve that produce the same profit maximising level of output.

The oligopolist’s cost structure can change, with worsening or improved efficiencies, but as long as the MC curve cuts the MR curve through its vertical portion $XY$, the oligopolist’s price and output decision should not alter. If the MC curve (in Figure 16) shifts from $MC$ to $MC_2$, price ($P$) and quantity ($Q$) remain unchanged. The important implication of this is that there will be price and output stability, with cost changes for the oligopoly firm, which change its MC curve, not affecting output and price. The discontinuity in the MR curve resulting from the kinked demand curve causes the price stability, because there are a range of points where $MC = MR$. This situation is unique to an oligopoly.

Only if marginal costs rise far enough for the MC curve to pass through the MR curve above point $X$ in Figure 16 is there a case for raising price, and only if MC falls far enough to pass through the MR curve below point $Y$ is there a case for lowering price.

In general, oligopoly prices will only rise if all the firms follow the lead of a rival in raising its price, so that the AR curve shifts outwards. The kink rises to the new common price level, which is again stable. The converse holds for price falls, with all prices being reduced to a new level, perhaps due to technological advances reducing costs in the industry.

The kinked oligopoly demand curve, like other theories of oligopoly behaviour, does not state exactly how oligopolists set their price. There is no single theory of oligopoly behaviour that perfectly captures all the features of an oligopoly market structure. For this reason there are differing oligopoly models involve different assumptions and conclusions about how oligopoly firms set their prices.

### 5.5 Price leadership and price wars

In oligopoly markets there is a tendency for one firm to set the general industry price, with the other firms following suit. This is called price leadership.

If rivals follow suit when a price leader increases price, then the consumer will suffer because the market effectively becomes a monopoly with prices raised above the initial equilibrium level.

However, if the price leader cuts prices (perhaps to try to increase market share) but rivals follow suit, then a price war could be created. Each firm is prepared to cut its own prices in order to preserve its share of the market. In this case, customers will benefit from the action, as prices will fall across the market as a whole.
However, although the effect of price wars is usually beneficial to consumers, such periods of price cutting are usually of limited duration because it is not in the interests of oligopolists to sustain them for long.

Economists sometimes model the strategies of oligopolists and market participants in other types of market structure using game theory, which involves examining participants' strategies according to what they stand to gain or lose from each strategy.

There are four strategies about pricing which an oligopolist might adopt:

(a) Co-operate with the other large firms to agree a common policy on pricing and market sharing. This is a collusive oligopoly.

(b) Make their own decisions and ignore rivals (to become a price leader). The effect of this will depend on how rivals react, as illustrated by the kinked demand curve.

(c) Become a price follower (price taker) and respond to the actions of a price leader.

(d) Do nothing. The firm may feel it would be disadvantageous to change its price, again based on the kinked demand curve it faces.

Unfortunately, however, the economic theories about oligopolies do not explain how an oligopolist sets its price in the first place.

**Question 9: Cartels and oligopoly**

(a) Draw a diagram to show the effect of a cartel on price and output to the market.

(b) Draw a diagram to show a kinked oligopoly demand curve.

Compare your diagrams with Figures 15 and 16 respectively in this chapter.

---

**6 Duopoly**

**Section overview**

- A duopoly is an oligopoly market structure containing only two firms. The two firms can either collude or compete; and if they compete one firm can only gain at the expense of the other firm. This decision is game theory.

Duopoly is a useful model for analysing oligopoly behaviour, especially the interdependence between firms. In a duopoly, if the firms compete, then, assuming the market size remains constant (ceteris paribus), the gains of one firm must be matched exactly by the losses of the other firm. So a firm must decide whether to try to increase its market share or to allow market share to remain constant.

This decision can be considered as game theory.

**6.1 Game theory example**

For example, suppose that firms A and B share the market for soft drinks. As they have a duopoly it would be better for both A and B to not advertise in the market. They may have an agreement in place confirming this. However, both A and B have an incentive to cheat on this agreement. Accordingly, one firm might decide unilaterally to advertise, when the other firm does not. Another possibility is that both firms will decide to advertise. Neither firm knows what the other will do. So what decision should each firm take?

At the moment, both A and B make profits of $250m per year. If one firm advertises but the other does not, it will increase its net profit and the profits of the other firm will fall (because its sales will fall.)

If both firms advertise, their sales will be unaffected but profits will fall because of the advertising costs.
Some illustrative figures are shown in the table below.

<table>
<thead>
<tr>
<th>Option</th>
<th>Aʼs profit</th>
<th>Bʼs profit</th>
<th>Industry profit</th>
</tr>
</thead>
<tbody>
<tr>
<td>Currently (no advertising)</td>
<td>$250m</td>
<td>$250m</td>
<td>$500m</td>
</tr>
<tr>
<td>A advertises, B does not</td>
<td>$350m</td>
<td>$100m</td>
<td>$450m</td>
</tr>
<tr>
<td>Advertises, A does not</td>
<td>$100m</td>
<td>$350m</td>
<td>$450m</td>
</tr>
<tr>
<td>Both advertise</td>
<td>$200m</td>
<td>$200m</td>
<td>$400m</td>
</tr>
</tbody>
</table>

Interestingly, after the advertising campaigns both firms are worse off than they were before.

The figures show that although one firm can gain in the short run from a competitive strategy, in the long run both firms are likely to be better off by working together and not advertising (colluding) rather than competing with each other.

However, game theory presupposes that the firms do not have a collusive agreement and do not know what the other is going to do. So A and B must select their strategies based solely on the outcome which is best for them regardless of the decision made by their rival.

Under these circumstances, both firms will choose to advertise. Individually, they increase profitability $100m by advertising. But collectively this course of action causes them both to lose $50m.

Game theory illustrates the key problem of interdependent decision-making in duopoly and oligopoly. Competition among the firms in a market can lead to inefficiency and competitive actions that use resources without generating corresponding benefits.

7 Oligopsony and monopsony

Section overview

- An oligopsony operates where there are a small number of buyers and a large number of sellers in the market. A monopsony exists where there is one buyer and many sellers. Under both structures the seller can be severely disadvantaged. World cocoa production and the Australian Wool Corporation are examples of oligopsonies.

Definitions

Oligopsony: a market structure where there are few buyers and many sellers.

Monopsony: a market structure where there is one buyer and many sellers.

An oligopsony is a market structure where there are a small number of buyers and a large number of sellers. This is in contrast to an oligopoly or duopoly structure which has many buyers and two or more suppliers. A monopsony is where there is only one buyer and a large number of sellers, which is the opposite of a monopoly. An oligopsony and a monopsony are both forms of imperfect competition.

National publicly-funded health care systems, such as Medicare in Australia and the UK’s National Health Service (NHS) are examples of a monopsony as the government is effectively the only ‘buyer’ of health services.

The world market for cocoa beans is an example of an oligopsony. There are only a small number of intermediary firms who purchase the large majority of world cocoa beans from many sellers, mostly cocoa farmers in developing countries. Another example is the Australian Wool Corporation (AWC). The AWC was the government’s wool buyer until it was privatised in 1998 and operated as the main buyer of wool from the many Australian farmers. This operated effectively using a floor price strategy until world demand for wool fell in the 1980s with disastrous results. This is discussed in Chapter 2.
Sellers can be at a major disadvantage under an oligopsony structure. These disadvantages include:

- Buyers can set one seller off against another seller for the same product thereby lowering the purchase price paid to all sellers.
- Buyers can dictate costs of sellers to a certain extent through imposing exact specifications including quantity, quality, suppliers, wages, labour demand and innovation.
- Buyers are able to pass on risk inherent in the product. This is particularly so for agricultural products where sellers carry the full risks of natural loss, cyclical demand and overproduction.

**Exam comments**

The examiner will test your knowledge of both oligopoly and oligopsony so read the question carefully to ensure you do not confuse the two structures.
Key chapter points

- In **perfectly competitive markets**, firms are **price takers**, and so their decisions are concerned with what **output level** will maximise profits. In **imperfect competition**, firms can influence the market price, and so their decisions are about what **price to set** as well as what volumes of output to produce which are determined by residual demand. **Pure monopoly** is an extreme form of **imperfect competition**.

- Firms will generally try to earn **supernormal profits** if they can. Competition, though, tends to erode supernormal profits, and firms may have to be satisfied, when equilibrium is reached, with just **normal** profits. The ability to sustain supernormal profits depends on the nature of competition in the industry.

- **Monopoly** may be **beneficial** (because of economies of scale) or **harmful** because they restrict output and raise prices. Government policies have been directed at the harmful aspects of monopoly.

- It may be that monopolies encourage **complacency** about costs (X-inefficiency) and may produce allocative and technical inefficiency. Goals other than profit maximisation pursued in large companies could also result in **inefficiencies**.

- When **price competition** is restricted, firms usually go in for other forms of competition, such as **sales promotion** and **product differentiation**.

- A firm which operates in conditions of monopolistic competition will have a **short-run** equilibrium in which it can make **supernormal profits**, and a **long-run** equilibrium in which it cannot.

- Monopolistic competition is more wasteful than perfect competition.

- **Oligopolies** might collude and make a formal or informal **cartel** agreement on the price for the industry and output levels for each firm. The **kinked oligopoly demand curve** may explain why there is price stability (and non-price competition) in many oligopoly markets.

- A **duopoly** is an oligopoly market structure containing only two firms. The two firms can either collude or compete; and if they compete one firm can only gain at the expense of the other firm. This decision is game theory.

- An **oligopsony** operates where there are a small number of buyers and a large number of sellers in the market. A monopsony exists where there is one buyer and many sellers. Under both structures the seller can be severely disadvantaged. World cocoa production and the Australian Wool Corporation are examples of oligopsonies.
Quick revision questions

1. How can a firm in perfect competition make supernormal profits?

2. The necessary conditions for a firm to be able to discriminate on price are:
   I. The firm is a price setter
   II. The markets are kept separate
   III. Price elasticity of demand is different in each market
   IV. Customers in each market are not aware of the prices charged in other markets
   A. I, II and III only
   B. I, II and IV only
   C. II, III and IV only
   D. all of them

3. Which of the following defines the long-run equilibrium position of a firm operating under conditions of perfect competition?
   A. $MC = MR$, $AC < AR$, $MR < AR$
   B. $MC = MR$, $AC = AR$, $MR < AR$
   C. $MC > MR$, $AC = AR$, $MR = AR$
   D. $MC = MR$, $AC = AR$, $MR = AR$

4. These diagrams show long term equilibrium under perfect competition for both the firm and the industry.
   Label these diagrams.
   A. firm's average cost
   B. output of the firm
   C. market supply
   D. firm's marginal cost
   E. market demand
   F. total industry output

5. Why do perfectly competitive markets achieve allocative efficiency?
6 What are the implications of the kinked oligopoly demand curve for price and output by an oligopoly firm?

7 Which one of the following statements about price discrimination is incorrect?
   A Dumping is a form of price discrimination.
   B For price discrimination to be possible, the seller must be able to control the supply of the product.
   C Price discrimination is only profitable where the elasticity of demand is different in at least two of the markets.
   D An example of price discrimination is the sale of first class and second class tickets on an aeroplane journey.

8 The oligopolist is least likely to compete through
   A advertising
   B improving product quality
   C cutting price
   D providing incidental services as an 'add-on' to the basic good

9 This question consists of two statements. Which, if either, is correct?
   I In conditions of monopolistic competition, firms will eventually reach an equilibrium output which is less than the output level at which average total cost is at a minimum.
   II In perfect competition, at the output level where marginal revenue equals marginal cost, a firm's average variable costs are minimised.
   A both statements are correct
   B the first statement is correct but the second statement is false
   C the first statement is false but the second statement is correct
   D both statements are false

10 Which of the following factors would weaken the long-term survival of a cartel?
    A greater price elasticity of demand for the product in the long run
    B a high concentration of production in the hands of a few firms
    C substantial costs associated with entry into the industry
    D broadly similar cost structures between industry members
1. In the short run, the number of firms in the market is fixed. If the prevailing market price is above the lowest point on a firm’s average total cost curve, it will make supernormal profits. This will continue until new entrants are attracted into the market and drive the market price down by increasing supply.

2. A. To be able to discriminate on price a firm must be a price setter, must have at least two separate markets, and the price elasticity of demand must be different in each market. Although customers may be aware of different prices in other markets, because the markets are separate they have to take the price prevailing in their market. (IV) is incorrect.


4. 1. E
   2. C
   3. D. The profit maximising level of output is where marginal costs equals marginal revenue.
   4. A. Long-term equilibrium is at an output where average cost equals marginal cost and marginal revenue.

5. B.
6. F

5. Allocative efficiency is only achieved if a firm produces at the point where price equals marginal cost. Profit maximising firms will seek to produce where MC = MR. Under perfect competition, because price (AR) = MR, price also equals MC.

6. The kinked demand curve illustrates the tendency to stability of prices in oligopoly markets. Oligopolists avoid price competition since a price cut will be matched by competitors and produce little lasting benefit.

7. D. First and second class tickets are not an example of price discrimination, because even though they are tickets for the same aeroplane journey, they are different products – e.g. in terms of service and travel comfort – rather than the same product being sold at two or more different prices. All the other statements are true.

8. C. Oligopoly is usually characterised by price stability, as illustrated by the so-called kinked oligopoly demand curve. Oligopolists are unlikely to cut prices, and are more likely to resort to non-price competition such as advertising and sales promotion, innovation and technical differences and incidental services.

9. B. In monopolistic competition, a firm’s equilibrium is where MR = MC, and this is at an output level below minimum AC. Statement II is false because although average total cost is minimised, average variable costs are not at a minimum and diminishing returns already apply.

10. A. Few suppliers, barriers to entry and similar cost structures are prerequisites for an effective collusive oligopoly. Greater elasticity of demand for a product in the long run would affect the cartel’s ability to charge a higher price while maintaining the same volume of sales. This would lead to firms leaving the cartel and cutting prices in order to sell more of their own production.
Answers to chapter questions

1. (a) There is a huge number of buyers, and many sellers too. For any given model of car, a particular dealer is likely to be a price taker.

(b) Communication is generally good. Product features are well known and list prices are freely available. And discount levels too are widely commented on, in the press and by word of mouth.

(c) The product is very far from homogeneous: different makes and models of car differ significantly from one another.

(d) Entry to the market is not easy, whether we are talking about manufacturers of motor cars (very high start-up costs), or dealers.

(e) Transport costs are not absent. On the contrary, significant geographical price differentiation is possible because of the high transport costs involved.

(f) Consumers don’t always act rationally. Some might be attracted to a car with a higher price, even if it is no better than other cheaper cars. Some may not shop around at different dealers even though it could save them money.

(g) Manufacturers can sustain large accounting profits in the long run, so it is likely these are supernormal economic profits.

2. (a) True. The firm can sell whatever output it produces at the market price (D = AR = MR).

(b) False. (a) above implies that the demand curve is perfectly elastic.

3. True. Under perfect competition, price, average revenue and marginal revenue are all the same. Therefore, given that the firm will be in equilibrium at MC = MR, MC is also equal to price.

4. D The firm is producing and selling at a level of output where marginal cost ($40) is equal to marginal revenue ($40). It is therefore already maximising its profit or minimising its loss. Its monthly total revenue is $40 \times 200$ units = $8000. This covers the variable costs and makes a contribution of $800 towards its fixed costs. Ceasing production would cause this contribution to be lost. There would be no point to reducing price since under perfect competition it can sell as much as it can produce at the prevailing market price. Equally, there would be no point increasing output, because if the firm increased production, it would find that its marginal cost rose.

5. At the point where MR = 0. Price $P_z$ and output $Z$. Further sales will lead to negative MR, and hence a reduction in total revenue.

6. You might have thought of:
   - Telephone calls (different prices for peak and off-peak calls).
   - Rail travel (there are many different tickets you can buy for an identical journey).
   - Package holidays (more expensive during school holidays).

7. (a) The nature of the good. Prices can be varied according to the time of day or day of the week because there are distinct business and leisure markets. Many customers will be forced to travel at peak times and pay top prices and some will switch to travelling at a cheaper time when the railway or airline has spare capacity to be filled up. Demand for journeys at peak times will be relatively inelastic, since demand will be from commuters who must travel in these periods in order to reach their workplace on time. Demand for leisure journeys is much more elastic.

   (b) The seller can prevent the resale of tickets. A cheaper rate might be offered to children. Since a child cannot transfer his ticket to an adult there is no danger that adults can buy cheaper tickets by using children to obtain tickets on their behalf.

   (c) Geographical separation of market segments: an airline can sell cheap travel to customers travelling from say, Los Angeles to Sydney, but still charge full rates to customers travelling from Sydney to Los Angeles.

8. Possible examples could be restaurants, cereals, clothing and shoes – they are all goods or services which are advertised on non-price grounds.

9. See Figures 15 and 16.
Chapter 6
Market failure, externalities and intervention

<table>
<thead>
<tr>
<th>Learning objectives list</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Market failure, externalities and intervention</td>
<td>LO6</td>
</tr>
<tr>
<td>Distinguish between social goods and private goods</td>
<td>LO6.1</td>
</tr>
<tr>
<td>Evaluate the impact of tax, savings and subsidies on the pricing mechanism</td>
<td>LO6.2</td>
</tr>
<tr>
<td>Analyse the implications of spill-overs or externalities using a demand and supply analysis</td>
<td>LO6.3</td>
</tr>
</tbody>
</table>

Topic list
1. Market failure
2. Externalities and government intervention
3. Indirect taxes and subsidies
In this chapter, we are concerned with why a free market would result in an allocation of resources that is not optimal – that is, not the best possible use of scarce resources.

A market imperfection is any situation where actual behaviour in the market differs from what it would be if there were perfect competition in the market.

Trading goods and services, and economic activity in general, may have effects beyond the participants in the market. We are all affected by pollution from cars for example, even if we do not have a car ourselves. Such effects are called ‘externalities’.

Equally, we are all affected by government actions and governmental controls used to reduce misallocation of resources. We examine some aspects of these government actions and controls in this chapter.
Before you begin

If you have studied these topics before, you may wonder whether you need to study this chapter in full. If this is the case, please attempt the questions below, which cover some of the key subjects in the area.

If you answer all these questions successfully, you probably have a reasonably detailed knowledge of the subject matter, but you should still skim through the chapter to ensure that you are familiar with everything covered.

There are references in brackets indicating where in the chapter you can find the information, and you will also find a commentary at the back of the Study Manual.

1. When does market failure occur?  (Section 1.2)
2. Define the following:  (Sections 2.1, 2.2)
   - Private good
   - Social cost
   - Private cost
   - Social benefit
   - Private benefit
3. What are externalities?  (Section 2.3)
4. Define the following:  (Sections 2.4, 2.5)
   - Social good
   - Merit good
5. Explain the supply curve impact of indirect taxation placed on a product.  (Section 3.1)
6. List two reasons why governments may choose to pay a subsidy.  (Section 3.3)
1 Market failure

Section overview

- Free markets may not lead to an ideal allocation of resources, and there are various methods of regulating markets.
- Free markets assume perfect competition in the market which is subject to certain conditions.
- Market failure occurs when the free market is unable to produce the most efficient allocation of scarce resources.

1.1 The case for a free market

What is the general case for allowing market forces to set prices? Advocates of the free market system put forward the following arguments:

(a) Free markets are **efficient**. Suppliers and consumers react fairly quickly to changes in market conditions in making their output and purchasing decisions; **resource allocation** within the economy is quick to adapt to the new conditions.

(b) The market is **impersonal**. Prices and levels of output are arrived at as a result of numerous decisions of consumers and suppliers, and not as the result of bureaucratic or political regulation.

Advocates of a free market economy argue that the market forces of supply and demand will result in an **efficient allocation of economic resources**.

(a) Consumers will want lower prices and producers will want higher prices; a balance of supply and demand is struck in the market through the price mechanism.

(b) Producers will decide what goods to produce, and in what quantities, by relating their prices to the costs of production (and the costs of the scarce resources needed to produce them).

(c) If the price of a product is too high, consumers will want to buy less of it. If the price is too low, producers will make less of it and switch their production resources into making something different which generates a higher price.

However, the arguments in favour of a free market are based on the assumption that there is **perfect competition** in the market. Perfect competition has a number of prerequisites:

(a) Markets each have a large number of competing firms, each producing a homogeneous product, and each having only a small share of the market.

(b) Producers are price-takers. They have to sell at the ruling market price: they cannot impose their own price at which to sell goods or services.

(c) Consumers and producers have perfect information about markets and prices.

(d) There is perfect mobility of factors of production, so factors of production can be switched easily from making one type of good into making another. Perfect competition also assumes there are no barriers to entry or exit for firms wanting to enter (or leave) a market.

In reality, these assumptions are not often completely valid. However, the markets for many goods approximate to conditions of perfect competition.

1.2 The concept of market failure

**Definition**

Market failure occurs when a free market mechanism fails to produce the most efficient allocation of scarce resources.
Remember, the idea of how resources are used lies at the heart of economics. The central economic problem is how scarce resources should be allocated in order to satisfy human wants and needs. Therefore, market failure – and the inefficient allocation of resources – is also a very important issue, because it means these human wants and needs are not being met as efficiently as they could be.

Market failure is caused by a number of factors:

- Imperfections in a market.
- Divergence between private costs and social costs (externalities).
- The need to provide social goods.
- The need to consider non-market goals, such as the consumption of merit goods.

The following are examples of market imperfections:

(a) If a monopoly firm controls a market, it might prevent other firms from entering the market (for example, by claiming patent rights, or launching a strong marketing campaign with the intention of keeping customers away from the new firms). By restricting supply in this way, the monopolist may keep prices higher than they would be in a competitive market.

(b) Just as monopolies are firms which dominate supply to a market, monopsony buyers are large individual buyers who dominate demand in a market. As discussed in Chapter 5, monopsonists may exert control over the market, exacting low prices or other favourable conditions from suppliers.

(c) Consumers may make bad purchasing decisions because they do not have complete and accurate information about all goods and services that are available.

(d) It takes time for the price mechanism to work. Firms cannot suddenly enter a new market or shut down operations. The slow response of the price mechanism to changes in demand creates some short term inefficiency in resource allocation.

### 2 Externalities and government intervention

#### Section overview

This section covers the following concepts:

- **Private good**: must be both exclusive and rivalrous and generally produced for profit by private firms.
- **Social costs**: the total costs to society as a whole of using economic resources.
- **Private costs**: the cost to the firm of the resources used to produce a private good.
- **Social benefits**: the total gains to society as a whole flowing from an economic decision.
- **Private benefit**: the direct gain made by a supplier or a consumer for a private good.
- **Externalities**: the differences between private and social costs.
- **Social goods**: goods which cannot be provided privately because if they are, everyone will benefit from them; regardless of whether they have paid for them or not. As a result, individuals would have no incentive to pay for these goods.
- **Merit goods**: social goods which need to be provided in the long-term public interest.

#### 2.1 Social costs and private costs

In a free market, suppliers and households make their output and buying decisions for their own private benefit, and these decisions determine how the economy’s scarce resources will be allocated to production and consumption. **Private costs** and **private benefits** therefore determine what **private goods** are made and bought in a free market.
Definitions

A private good must be both exclusive and rivalrous.

Exclusivity means that it is reasonable to prevent a class of consumers (or firms) from consuming the good. Rivalrous means that consumption by one party prevents simultaneous consumption by another. A private good satisfies an individual want and is almost always produced for profit.

A piece of clothing is an example of a private good – the purchase of it by one consumer excludes another consumer from purchasing it (exclusivity), and when it is worn it cannot be worn simultaneously by another consumer (rivalrous).

However, these private costs and benefits are not necessarily the same as the social costs and benefits from using the resources (i.e. the costs and benefits to society as a whole).

- **Private cost** measures the cost to the firm of the resources it uses to produce a good.
- **Social cost** measures the cost to society as a whole of the resources that a firm uses.
- **Private benefit** measures the benefit obtained directly by a supplier or by a consumer.
- **Social benefit** measures the total benefit to society from a transaction.

It can be argued that a free market system would result in a satisfactory allocation of resources, provided that private costs are the same as social costs and private benefits are the same as social benefits. In this situation, suppliers will maximise profits by supplying goods and services that benefit customers, and that customers want to buy. By producing their goods and services, suppliers are giving benefit to both themselves and the community.

However, there are other possibilities:

(a) Members of the economy (suppliers or households) may do things which give benefit to others, but yield no reward to themselves.

(b) Members of the economy may do things which are harmful to others, but which have no cost to themselves.

When private benefit is not the same as social benefit, or when private cost is not the same as social cost, an allocation of resources which only reflects private costs and benefits may not be socially acceptable.

Here are some examples of situations where private cost and social cost differ:

(a) A firm produces a good and, during the production process, pollution is discharged into the air. The private cost to the firm is the cost of the resources needed to make the good. The social cost consists of the private cost plus the additional ‘costs’ incurred by other members of society, who suffer from the pollution.

(b) The private cost of transporting goods by road is the cost to the haulage firm of the resources to provide the transport. The social cost of road haulage would consist of the private cost plus the cost of repairs and maintenance of the road system (which sustains serious damage from heavy goods vehicles) plus any environmental costs, such as harm to wildlife habitats from road building or pollution emitted by the transport lorry.

2.2 Private benefit and social benefit

Here are some examples of situations where private benefit and social benefit differ:

(a) Customers at an open air café benefit from the entertainment provided by professional musicians, who are hired by the café. The customers of the café are paying for the service in the prices they pay, and they obtain a private benefit from it. At the same time, other people passing by, who are not customers of the café, might stop and listen to the music. They will obtain a benefit, but at no cost to themselves. They are free riders, taking advantage of the service without contributing to its cost. The social benefit from the musicians’ service is greater than the private benefit to the café’s customers.

(b) Suppose that a large firm pays for the training of employees as accountants, expecting a certain proportion of these employees to leave the firm in search of a better job once they have qualified. The private benefits to the firm are the benefits of the training of those employees who continue to work for it. The total social benefit includes the enhanced economic output resulting from the training of those employees who go to work for other firms.
Question 1: Private and public costs and benefits

Think of some situations other than those mentioned above in which private costs differ from social costs and private benefits differ from social benefits. How might these differences be prevented or compensated for in each situation?

(The answer is at the end of the chapter)

2.3 Externalities

Definition

Externalities are the spill-over effects of a transaction which extend beyond the parties to the transaction and affect society as a whole. In other words, externalities are the differences between the private and the social costs, or benefits, arising from an activity.

An 'externality' is a cost or benefit which the market mechanism fails to take into account, because the market responds purely to private signals.

So far we have looked at the price mechanism as being driven purely by the costs and benefits accruing to the parties to the transaction. However, if pricing policy is to maximise net social benefit then it also needs to include externalities when calculating costs.

We can use demand and supply analysis to illustrate the consequences of externalities. If a negative externality exists, (the social cost of supplying a good is greater than the private cost to the supplier firm), then a supply curve which reflects total social costs will be above the (private cost) market supply curve.

Figure 1 shows two possibilities:

(a) If a free market exists, the amount of the good produced will be determined by the interaction of demand (curve D) and supply curve S. Here, output would be Y, at price P_y.

(b) If social costs are taken into account, and the market operates successfully, the supply curve should shift leftwards, and the amount of the good produced should be X, at price P_x.

The optimum level of production still occurs when MC = MR, but now marginal cost also includes the cost to society of producing an extra unit of output. This is the concept of social marginal costs.

Figure 1 Externalities

Given a free market, output of the good will exceed what it ideally should be (by Y – X in Figure 1 above), and so resources will have been over-allocated to production of this particular good.
2.4 Social goods

Definition

Some goods, by their very nature, involve so much ‘spill-over’ of externalities that they are difficult to provide except as social goods whose production is organised by the government.

Unlike private goods, the consumption of a social good (or service) by one individual or group does not significantly reduce the amount available for others. Furthermore, it is often difficult or impossible to exclude anyone from its benefits, once the good has been provided. In other words, the good is non-exclusive. As a result, in a free market, individuals benefiting from the good would have no economic incentive to pay for them, since they might as well be free riders if they can, enjoying the benefits of the good while others pay for it. Social goods are also commonly referred to as public goods.

National defence is perhaps the most obvious examples of a social good. It is clearly not practicable for individuals to buy their own defence systems. Equally, if some people paid for the service, they couldn’t stop others who hadn’t paid for it from benefiting from it. Therefore, the government provides defence at zero price at the point of consumption, although taxpayers indirectly fund its provision through their tax payments. Street lighting and policing are sometimes cited as other examples of social goods, although the growth of private security firms in the private sector illustrates how some areas of policing are now becoming privatised.

2.5 Merit goods

The existence of market failure and of externalities suggests the need for intervention in markets by the government, in order to improve the allocation of resources. Another possible reason for intervention is to increase the consumption of merit goods.

Definition

Merit goods are considered to be worth providing to everyone irrespective of whether everyone can afford to pay for them, because their consumption is in the long-term public interest. Education is one of the chief examples of a merit good.

Merit goods are different from social goods in that they are divisible. For example, some consumers possess the means to buy merit goods, such as education and healthcare, and they are willing to do so. However, while the people who choose to pay for the good will benefit from it, in doing so they will use up the supply of the good so that free riders cannot also benefit from it.

Although national governments provide a service of merit goods in the interests of the nation’s well-being, the private sector provides an alternative supply of them – for example, in Australia, the UK and the US there are private schools, and private health care schemes. However, what distinguishes a merit good from a social good is that where people have elected for a private supply the benefit is restricted to them. Thus only people who pay for private schools or private health care will benefit from them.

Because merit goods provided by the State are offered on a large scale, governments can achieve economies of scale in their provision.

Governments also provide merit goods (such as education and health) because it is in the nation’s interests as a whole that such things are provided.

On the other hand, many governments want to see less consumption of certain demerit goods, such as tobacco.
Apart from providing social goods and merit goods, a government might choose to intervene in the workings of markets by other methods:

(a) **Controlling the means of production** (for example, through State ownership of industries).

(b) **Influencing markets** through legislation and regulation (regulation of monopolies, bans on dangerous drugs, enforcement of the use of some goods such as car seat belts, laws on pollution control and so on) or by persuasion (for example, using anti-tobacco advertising).

(c) **Redistributing wealth**, perhaps by taxing relatively wealthy members of society and redistributing this tax income so as to benefit the poorer members.

(d) **Influencing market supply and demand** through:
   - (i) Price legislation.
   - (ii) Indirect taxation.
   - (iii) Subsidies.

(e) **Creating a demand for output that creates employment**. A free market system would match supply with demand. Demand would thus lead to employment because of the needs of suppliers, but the level of demand might not be high enough to ensure full employment. Government might therefore wish to intervene to create a demand for output in order to create more jobs.

Some externalities, particularly the problems of pollution and the environment, appear to call for co-operation between governments. The Kyoto Protocol (treaty) adopted in 1997 aimed to stabilise the levels of greenhouse gases in the atmosphere to control the threat of global warming. (By 2011, 191 countries had ratified the treaty. However, the USA, having signed the treaty, has refused to ratify it; and in 2011 France, Russia, Canada and Japan all announced that they would not be joining a second round of carbon cuts under the Protocol.)

**Question 2: Pollution**

An industrial company alters its production methods to reduce the amount of waste discharged from its factory into the local river. What will be the effect (increase or decrease) on:

(a) private costs?
(b) external benefits?

(The answer is at the end of the chapter)

**2.6 Pollution policy**

One area often discussed in relation to externalities is that of pollution. If polluters take little or no account of their actions on others, this generally results in the output of polluting industries being greater than is optimal. If polluters were forced to pay for any externalities they imposed on society, producers would almost certainly reduce production or change their production techniques so as to minimise pollution.

Equally consumers would be likely choose to consume less of those goods which cause pollution because the producers are likely to pass the cost of any changes on to the consumers.

One solution to reflect the social cost of polluting activity is to levy a tax on polluters equal to the cost of removing the effect of the externality they generate: this is called the ‘polluter pays’ principle. This approach is generally held to be preferable to regulation, as this can be difficult to enforce and provides less incentive to reduce pollution levels permanently.

Apart from the imposition of a tax, there are a number of other measures open to the government in attempting to reduce pollution. One of the main measures available is the application of subsidies which may be used either to persuade polluters to reduce output and hence pollution, or to assist with expenditure on production processes, such as new machinery and air cleaning equipment, which reduce levels of pollution.

A problem with using subsidies is that, unlike taxes, they do not provide an incentive to reduce pollution any further: indeed, profits are increased under subsidies which may have the perverse effect of encouraging more pollution to be generated in order to qualify for a subsidy. In addition, this is likely to be an expensive option for the government whereas imposing a tax actually provides the government with additional revenue.
Two other options which could be considered to help deal with externalities are:

**Government regulation:** Governments could set limits for the maximum permitted levels of emissions or minimum levels of environmental quality. Firms which breach these limits could then be fined.

** Tradable permits:** Maximum permitted levels are set for specific pollutants and a firm (or country) is given a permit to emit up to that level of pollution. If the firm (or country) emits below its limit, it is given a credit for the difference which it can then sell to a firm (or country) which needs to go above its permitted levels.

 Tradable permits are used in relation to carbon dioxide emissions, with the result that firms can now engage in carbon trading.

## 3 Indirect taxes and subsidies

### Section overview

- **Demand and supply analysis** can be used to examine the effects on a market of government intervention through imposing an **indirect tax** or a **subsidy**.

### 3.1 Indirect taxes

#### Definition

**Indirect taxes** are levied on expenditure on goods or services as opposed to direct taxation which is applied to incomes. A **selective** indirect tax is imposed on some goods but not on others (or is imposed at a higher rate on some goods than on others).

We looked at the effects of one form of government intervention in markets – price regulation – earlier in the Study Manual. An alternative form of price and output regulation is **indirect taxation**.

Indirect taxation may be used to improve the **allocation of resources** when there are damaging externalities.

If an indirect tax is imposed on a good, the tax will shift the supply curve upwards (leftwards) by the amount the tax adds to the price of each item. This is because although the price to **consumers** includes the tax, the revenue the suppliers receive is only **net-of-tax price**. For example, in Figure 2:

- the supply curve without any tax is \( S_0 \).
- the supply curve including the cost of the tax is \( S_1 \).
- the tax is equal to \( P_1 - P_2 \) or the distance \( A - B \).

Before the tax was imposed, quantity supplied and demanded was \( X_0 \), but once the tax has been imposed, the equilibrium quantity is only \( X_1 \).

At this point (demand = \( X_1 \)), the price the **consumer pays** is \( P_1 \), but the amount that the **supplier receives** is only \( P_2 \).

\( P_1 - P_2 \) is the amount of tax payable.
Figure 2 The effect of an indirect tax

Without the tax, output would be $X_0$ and price $P_0$. Total expenditure (before the tax was imposed) can be shown by the rectangle $OP_0TX_0$.

(a) After the tax has been imposed, output falls to $X_1$ and price with tax rises to $P_1$. Total expenditure is $OP_1AX_1$, of which $P_1P_1AB$ is tax revenue and $OP_2BX_1$ is producers' total revenue.

(b) A new price equilibrium arises at point A.

(i) Price to the customer has risen from $P_0$ to $P_1$.

(ii) Average revenue received by producers has fallen from $P_0$ to $P_2$. Producers suffer lower profits because of the lower price they receive and also the lower output.

(iii) The tax burden is shared between the producers and consumers, with $CB$ borne by the supplier and $AC$ borne by consumers.

Consumers pay $P_0P_1AC$ of total tax revenue and producers pay $P_1P_2CB$.

3.2 Elasticity effects

The proportion of the tax which is passed on to the consumer rather than being borne by the supplier depends upon the elasticities of demand and supply in the market.

Figures 3(a) and 3(b) illustrate the extreme cases of perfectly elastic demand and perfectly inelastic demand respectively.

Figure 3 Elasticity of demand
Question 3: Burden of taxation

Try to work out yourself (from general principles, or from study of Figure 3) who bears the burden of taxation in each of the extreme cases illustrated in Figure 3.

(The answer is at the end of the chapter)

The elasticity of supply is also relevant. Figure 4 shows that for a given demand curve, the more inelastic the supply curve, the greater the proportion of the tax that is borne by the supplier.

(a) Figure 4(a) shows a relatively inelastic supply curve (S). Imposition of the tax shifts the supply curve vertically upwards to \( S_1 \) and the equilibrium price rises from \( P_1 \) to \( P_2 \). The price to the consumer rises by \( AB \) per unit, while the supply price to the producer falls by \( BC \) per unit. Therefore, the greater burden is borne by the supplier (because the distance \( BC \) is greater than the distance \( AB \)).

(b) By contrast, Figure 4(b) shows a relatively elastic supply curve \( S \). With the imposition of the tax, the supply curve shifts to \( S_1 \) and the equilibrium price rises to \( P_2 \). The price to the consumer rises by \( AB \) per unit, and the supply price to the producer falls by \( BC \) per unit. The greater burden is borne by the consumer (because \( AB \) is greater than \( BC \)).

So, Figure 4 shows that the consumer bears a greater proportion of the tax burden the more elastic is the supply curve. Figure 4 also shows that, for any given demand curve, the price rise and the fall in the equilibrium quantity will also both be greater when the supply curve is elastic than when it is inelastic.

Figure 4 The effect of elasticity illustrated

It is also important to consider the relationship between the elasticity of demand and the elasticity of supply when assessing how the burden of tax will be split between producer and consumer.

5(a) Inelastic demand, elastic supply

5(b) Elastic demand, inelastic supply

Figure 5: The effect of the relative elasticities of supply and demand

If demand is less elastic than supply – Figure 5(a) – the consumer will bear the greater proportion of the tax burden.

If demand is more elastic than supply – Figure 5(b) – the producer will bear the greater proportion of the tax burden.
Exam comments

Make sure you know how the relative elasticities of supply and demand affect the way the burden of tax is distributed between producers and consumers.

As a final point on this topic, though, be aware that while indirect tax can help adjust for externalities, it could affect the competitiveness of individual firms or industries. **Since such a tax reduces output, it may be harmful to an industry.** For some companies, the reduction in quantities produced may lead to significant rises in the unit costs of production. This could have adverse consequences on the competitive position of the firm if it competes (either domestically or internationally) with foreign firms which are not subject to the same tax.

### 3.3 Subsidies

A subsidy is a payment to the supplier of a good by the government. The payment may be made for a variety of reasons:

- **To encourage more production of the good**, by offering a further incentive to suppliers.
- **To keep prices lower for socially desirable goods** whose production the government wishes to encourage.
- **To protect a vital industry** such as agriculture, when demand in the short term is low and threatening to cause an excessive contraction of the industry.

![Figure 6 Subsidy](image)

**A subsidy is rather like indirect taxation in reverse.**

- In Figure 6, supply curve $S_0$ shows what the supply would be if no subsidy existed.
- Payment of the subsidy moves the supply curve downwards (outwards) to $S_1$.

If there were no subsidy, the free market equilibrium price would be $P_0$, and output $Q_0$. A subsidy per unit equivalent to $AB$ is introduced, such that suppliers would now be willing to produce $Q_2$ at a lower price ($P_1$ rather than at $P_0$). In other words, the supply curve shifts outwards from $S_0$ to $S_1$. As a result, there will be a shift in the equilibrium quantity produced to $Q_2$, which can be sold on the market for $P_2$. Therefore, the subsidy will have two effects:

- The amount supplied in equilibrium will increase (from $Q_0$ to $Q_2$).
- The price will fall (from $P_0$ to $P_2$), but the decrease in price will be less than the value of the subsidy itself ($P_0 - P_1$).
Question 4: Subsidy

By reference to Figure 6, analyse the extent to which the benefit of the subsidy falls to:

(a) the consumer.
(b) the supplier.

Who bears the cost of the subsidy?  
(The answer is at the end of the chapter)
Key chapter points

- Free markets may not lead to an ideal allocation of resources, and there are various methods of regulating markets.
- Free markets assume perfect competition in the market which is subject to certain conditions.
- Market failure occurs when the free market is unable to produce the most efficient allocation of scarce resources.
- Private goods: must be both exclusive and rivalrous, and generally produced for profit by private firms.
- Social costs are the total costs to society as a whole of using economic resources.
- Private costs are the cost to the firm of the resources used to produce a private good.
- Social benefits are the total gains to society as a whole flowing from an economic decision.
- Private benefit are the direct gain made by a supplier or a consumer for a private good.
- Externalities are the differences between private and social costs.
- Social goods: goods which cannot be provided privately because if they are, everyone will benefit from them; regardless of whether they have paid for them or not. As a result, individuals would have no incentive to pay for these goods.
- Merit goods: social goods which need to be provided in the long-term public interest.
- Demand and supply analysis can be used to examine the effects on a market of government intervention through imposing an indirect tax or a subsidy.
Quick revision questions

1. What is the general case in favour of allowing a free market to operate?
2. What is market failure, and what are its main causes?
3. What is an externality?
4. List the various forms of government intervention in markets.
5. Which of the following are imperfections in a market?
   I. Consumer brand loyalty to a firm's branded goods, regardless of price.
   II. The lack of completely accurate information for consumers about all goods and services available.
   III. The slow response of firms to price changes and the relatively inelastic supply of a good in the short run.
   A. items I and II only
   B. items II and III only
   C. items I and III only
   D. items I, II and III
6. Which of the following are weaknesses of a completely free-enterprise economic system?
   I. It only reflects private costs and private benefits.
   II. It may lead to serious inequalities in the distribution of income and wealth.
   III. It may lead to production inefficiencies and a wastage of resources.
   A. I and II only
   B. II and III only
   C. I and III only
   D. I, II and III
7. Muddy Waters Co is an industrial company which has altered its production methods so that it has reduced the amount of waste discharged from its factory into the local river. Which of the following is most likely to be reduced?
   A. total private costs
   B. social costs
   C. external benefit
   D. variable costs
8. Much Wapping is a small town where a municipal swimming pool and sports centre have just been built by a private firm, Builder Co. Which of the following is an external benefit of the project?
   A. the increased trade of local shops
   B. the increased traffic in the neighbourhood
   C. the increased profits for the sports firm
   D. the increased building on previous open land
9 The government has just increased the tax on tobacco. Assuming that the demand for cigarettes is completely inelastic, who pays the tax?
   A it is shared between supplier and consumer in proportions equal to the relative prices before and after the increase.
   B the supplier
   C the consumer
   D it is shared between supplier and consumer in proportions equal to the relative quantities sold before and after the increase.

10 Which of the following statements is always true if an indirect tax is imposed on a good or service?
   A The price will rise by an amount equal to the tax.
   B The producer will bear more of the tax than the consumer.
   C The price rise will be smaller the greater the price elasticity of demand is.
   D The price rise will be greater the greater the price elasticity of demand is.
1. Free markets are efficient in that they adjust quickly to changing demand and supply and they operate automatically, without need for direction or control.

2. Market failure occurs when a free market mechanism produces an allocation of resources which can be criticised on efficiency, social or political grounds.

3. An externality is an effect caused by an economic transaction which extends beyond the parties to the transaction.


5. Brand loyalty can make consumers pay more for a good, without getting any greater total satisfaction from consuming it. Lack of information to consumers will result in ‘bad’ purchasing decisions. The slowness to price changes is a further market imperfection.

6. The need to limit or avoid these weaknesses is the chief argument in favour of some government involvement in the allocation of economic resources – i.e. in favour of a mixed economy or even a command economy.

7. Social cost is the sum of the private cost to a firm plus the external cost to society as a whole. Here, social cost is the sum of production costs (private costs) plus the cost of pollution (external cost). The firm’s private costs might have been increased by the measures to reduce pollution, but the external costs will have fallen, so that total social costs should have fallen too.

8. This is correct because the benefits to local shops are additional to the private benefits of the sports firm and as such are external benefits.

   B is an external cost of the project, since increased volumes of traffic are harmful to the environment.

   C is a private benefit for the firm.

   D would only be an external benefit if a building is better for society than the use of open land, which is unlikely.

9. As the consumer’s consumption is not altered by the price rise, the supplier can pass the price rise on in full.

10. The price rise will be lower for products with a higher price elasticity of demand. In the extreme case, if demand is perfectly elastic, there will be no increase in the price at all.

    Option A would be true if the good or service had a perfectly inelastic demand, but that is the only condition under which it would be true. Equally, Option B would be true if demand was relatively more elastic than supply, but it will not always be true.
1 Your answer will depend on the situations you thought of.

2 (a) Private costs of the company will presumably increase: the anti-pollution measures will have involved a financial outlay.

(b) External benefits will presumably increase: the public will benefit from a cleaner river.

3 In Figure 3(a), with perfectly elastic demand, demand falls to zero if the price is raised. Consequently, the supplier must bear the full burden of the tax. In spite of the imposition of the tax, the market price remains the same but there is a fall in the quantity supplied from $Q_1$ to $Q_2$. The supplier only receives $P_2$ after paying $P_1 - P_2$ in tax.

In the case of perfectly inelastic demand (Figure 3(b)), the supplier can pass on the full amount of the tax to the consumer by increasing the price from $P_1$ to $P_2$. Demand is not reduced at all by the increase in price. Because the supplier can charge a higher amount to compensate for the tax payable, the quantity supplied remains unchanged.

4 The benefit of the subsidy will be shared between the consumer and the supplier.

(a) Consumers benefit by the lowering of prices from $P_0$ to $P_2$.

(b) Suppliers benefit because although they receive a lower price, $P_2$, they receive the subsidy $AB$ per unit.

The cost of the subsidy is borne by the government (in effect, the taxpayer).
Chapter 7
National income accounting

Learning objectives

<table>
<thead>
<tr>
<th>National income accounting</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Distinguish between economic growth and economic development</td>
<td>LO7.1</td>
</tr>
<tr>
<td>Calculate Gross Domestic Product (GDP) and Gross National Product (GNP)</td>
<td>LO7.2</td>
</tr>
<tr>
<td>Perform national accounting calculations</td>
<td>LO7.3</td>
</tr>
</tbody>
</table>

Topic list

1. Measuring national economic activity
2. National income accounting
3. Key measures of national economic output
4. The circular flow of income in the economy
5. The three approaches to national income accounting
6. Interpretation of national income accounting information
Businesses operate in the economy and as such, changes in the macroeconomic environment can have major implications for them.

There is a clear distinction between economic growth and economic development. Figures for the level of economic activity and economic growth are monitored closely by economists and by financial institutions because they are indicators of the economic health of a country.

In this chapter and in Chapter 8 we look at how we can measure the total amount of economic activity of a nation.

There are different approaches to measuring economic activity. There are also different measures, with specific definitions. The process of measurement involves some estimates and there are a number of sources of possible inaccuracy in the figures.
Before you begin

If you have studied these topics before, you may wonder whether you need to study this chapter in full. If this is the case, please attempt the questions below, which cover some of the key subjects in the area.

If you answer all these questions successfully, you probably have a reasonably detailed knowledge of the subject matter, but you should still skim through the chapter to ensure that you are familiar with everything covered.

There are references in brackets indicating where in the chapter you can find the information, and you will also find a commentary at the back of the Study Manual.

1. Define economic development. (Section 1)
2. What are the three measures of national economic output? (Section 3)
3. What is Gross Domestic Product (GDP) and what does it measure? (Section 3.2)
4. What is Gross National Product (GNP) and what does it measure? (Section 3.3)
5. Give a brief explanation of the following approaches to national income accounting: (Section 5)
   (a) Expenditure approach.
   (b) Income approach.
   (c) Value-added approach.
6. List three items which are not included in a nation's GDP. (Section 6.3)
1 Measuring national economic activity

Section overview
• National income indicates the performance of an economy. It shows total value of goods and services of the economy’s resources and total income of residents.
• It measures economic growth not the wider-ranging measure of economic development.

Measuring national income can indicate how an economy is performing, because national income shows the total value of the goods and services produced by an economy’s resources, and the total incomes of residents in an economy.

National income is important because of the economic activity in a country:
(a) It is an aggregate of personal incomes. The bigger the aggregate of personal incomes in a country, the more income its individual inhabitants will be earning on average.
(b) More income means more spending on the output of firms, and more spending (ignoring inflation) means that a higher output of goods and services is required to be produced.
(c) Economic growth is an economic policy objective of most, if not all, governments. The growth potential of an economy will depend on the amount of economic resources (factors of production) available, and the productivity of them. Because economic growth is an important factor in raising standards of living in a country – overall economic development – governments will be keen to promote it. They can do this either by increasing aggregate demand in an economy or encouraging aggregate supply, which is discussed in later detail in Chapters 9 and 10 of the Study Manual.

Economic growth is a quantitative measure and does not consider qualitative factors in a economy such as human development, social progress and technological advancement. Economic growth and productivity is measured using gross domestic product (GDP) and GDP excludes non-financial indicators such as human and leisure costs, nonmarket production and economic welfare. In contrast, economic development assesses the non-financial indicators of a economy and the overall impact of economic growth on human development, social progress and technological advancement. Improvement in key indicators such as poverty, literacy, health, life expectancy, environment, political freedom and social justice are all examples of economic development.

2 National income accounting

Section overview
• National income accounting is the system through which economic activity on a national scale is measured.
• Economic activity can be measured by output, income and expenditure.

The concept of national income accounting is based on the underlying principle that economic activity in a country can be measured in terms of three things:
(a) The amount of output produced in the country.
(b) The income received by those producing the output.
(c) The amount of expenditure incurred by those purchasing the output.

National income accounting can be viewed from these three different aspects:
(a) Output. The firms (or government departments and corporations) which produce the goods or services in the national economy.
(b) Income. The factors of production, which earn factor incomes.
(c) Expenditure. The people or organisations that spend money to buy the goods and services such as consumers (or households), the government and foreign buyers (the overseas sector).
Question 1: Factors of production
Before proceeding, recall from Chapter 1 what the factors of production are, and what reward is earned by each.

(The answer is at the end of the chapter)

As we shall see later on, the three approaches to the creation of economic wealth give rise to three ways of measuring national economic output:

• The expenditure approach.
• The income approach.
• The value added approach (also called the output approach).

2.1 National income accounting identity
The national income accounting identity follows from the three approaches to the creation of economic wealth and states that:

\[
\text{Income} = \text{Output} = \text{Expenditure}
\]

- Total of all factor income earned during the year
- The total value of final output created during the year
- The total spending on final goods and services during the year

The identity is illustrated diagrammatically below:

Figure 1 The circular flow of money
Figure 1 illustrates three approaches to measuring the same thing.

The diagram illustrates the circular flow of money round the economy. In its simplest form, the national income accounting model shows that, if there are no leakages from the system, the income people receive, equals the amount they spend on goods and services (expenditure) which is in turn equal to the prices of all the goods and services produced (output).

Official national income accounts show each approach to the measurement of economic activity (i.e. the output account, the income account and the expenditure account) as a separate account. In practice, the figures from the three methods differ, so an averaging procedure is often used to determine the overall figure.

2.2 Uses of national income accounting measures
(a) To assess the state of the economy for purposes of constructing economic policy.
(b) To monitor the impact of government economic policies.
(c) To enable government and business to compare the progress and prospects of different economies.
(d) To establish the material standard of living in an economy.
3 Key measures of national economic output

Section overview
There are three key measures of national economic output.

- National income.
- Gross national product (GNP).
- Gross domestic product (GDP).

Question 2: National output
The three terms above are often encountered in news reports, and yet are often only vaguely understood. Jot down what you think is the meaning of each, and review what you have written once you come to the end of the chapter.

(The answer is at the end of the chapter)

3.1 Definition of national income

Definition
National income: the sum of all incomes which arise as a result of economic activity, that is from the production of domestically owned goods and services.

The incomes which make up national income, which include rent, employment income, interest and profit, are known as factor incomes because they are earned by the factors of production:

- Land earns rent.
- Labour earns wages.
- Capital earns interest.
- Entrepreneurship earns profit.

National income is also called net national product.

(a) The terms income and product are just two different aspects of the same circular flow of income.

(b) The term net means ‘after deducting an amount for capital consumption or depreciation of capital assets’. (We shall return to this point later.)

3.2 Gross domestic product (GDP)

Most of a country’s national income is derived from economic activity within that country. Economic activity within that country is referred to as total domestic income or domestic product. It is measured gross, i.e. before deducting an amount for capital consumption or depreciation of capital assets. So, in Australia for example, gross domestic product therefore refers to the total value of income/production from economic activity within Australia. GDP is the most appropriate measure for assessing the productivity of an economy.

3.3 Gross national product (GNP)

Some of a country’s national income arises from overseas investments, while some of the income generated within that country is earned by non-residents. The difference between these items is net property income from abroad.

Gross national product (GNP) is therefore the gross domestic product (GDP) plus the net property income from abroad – or after subtracting the net property income from abroad, if it is a negative value.

For example, income earned in Australia by overseas companies and remitted back to their overseas country of origin would be a negative value for Australian GNP.
We can illustrate the difference between GDP and GNP with a simple example. Nissan is a Japanese car manufacturer with production plants in Thailand. The profits from its Thai activities contribute to Japanese GNP, but, because the activities occur in Thailand, they count towards Thailand’s GDP.

**Definitions**

- **Gross domestic product**: the value of the goods and services produced by an economy in a given period.
- **Gross national product**: GDP plus income accruing to domestic residents from investments abroad less income accruing to foreign residents from investments in the domestic economy.

**3.4 The relationship between GDP, GNP and national income**

In the same way that individual companies show depreciation as a charge in their income statements so national economies show the cost of depreciation on their capital assets.

This deduction for depreciation is known as **capital consumption**.

The value of GNP less capital consumption is known as **national income** (or net national product).

As well as recognising capital consumption, we also recognise that there are new investments. **Gross capital formation** illustrates new investment. The difference between gross capital formation and capital consumption is **net investment**.

The level of net investment in an economy is an indication of the productive potential of an economy. We would expect an economy with high net investment to have higher potential productivity going forward than one with low net investment.

Therefore, the relationship between GDP, GNP and national income is this:

\[
\text{GDP} + \text{Net property income from abroad} = \text{GNP} - \text{Capital consumption} = \text{Net national income or net national product}
\]

**National income** is GNP minus an allowance for depreciation of the nation’s capital.

**3.4.1 Adjustments to GDP at market prices**

It is important to note that GDP, GNP and national income as defined above are expressed at market prices.

Since the prices of many goods and services are distorted by sales taxes (for example, on the purchase of alcohol and cigarettes) and some are distorted by subsidies (for example, on many agricultural products), we often wish to view the situation without these distortions and convert GDP at market prices to **gross value added at basic prices** (which used to be known as GDP at factor cost).

Therefore, Gross value added (GVA) at basic prices = GDP at Market prices – Indirect taxes + Subsidies.

**Question 3: Effects on national income**

Which of the following will cause a rise in national income?

A. an increase in capital consumption (depreciation)
B. a rise in imports
C. a rise in subsidies
D. a rise in indirect taxes

(The answer is at the end of the chapter)
3.5 Summary of national income accounting

<table>
<thead>
<tr>
<th>Indirect taxes – subsidies</th>
<th>+ Net income from abroad</th>
<th>– Capital consumption</th>
</tr>
</thead>
<tbody>
<tr>
<td>GVA @ basic prices (GDP @ factor cost)</td>
<td>GDP @ Market prices</td>
<td>GNI @ Market prices</td>
</tr>
<tr>
<td></td>
<td>National income @ Market prices</td>
<td></td>
</tr>
</tbody>
</table>

The first column shows Gross Value Added (GVA) which is the output measure of GDP. Adding indirect taxes and subtracting subsidies produces GDP at market prices. Adding the net income from abroad to GDP (column 2) results in the Gross National Income (column 3). Finally, subtracting capital consumption from the Gross National Income results in the National Income at market prices (column 4).

3.6 Personal income

Personal income is defined as the total income received by individuals available for consumption, saving and payment of personal taxes. It is calculated as national income less corporate profits and social security taxes (since these are not paid out as income to investors), plus transfer payments, net interest and dividends (since these are paid to investors).

3.7 Disposable income

Disposable income is defined as income available to individuals after payment of personal taxes. It may be consumed or saved.

4 The circular flow of income in the economy

Section overview

- There is a circular flow of income in an economy, which means that expenditure, output and income will all have the same total value.
- The government has an important role in the circular flow of income.
- There are withdrawals from the circular flow of income (savings, taxation, import expenditure) and injections into the circular flow (investment, government spending, export income).

4.1 Income and expenditure flows

Firms must pay households for the factors of production, and households must pay firms for goods. The income of firms is the sales revenue from the sales of goods and services.

This creates a circular flow of income and expenditure, as illustrated in Figure 2. This is a basic closed economy, without foreign trade. It assumes the economy has only two sectors (firms and households), with no government intervention and no imports or exports. In this model, we assume all household income is spent on consumption, and all the firms’ goods and services are sold to the households.
Figure 2 The circular flow of income and expenditure

Households earn income because they have provided the factors of production which enable firms to provide goods and services. The income earned is used as expenditure on these goods and services that are made.

(a) The total sales value of goods produced should equal the total expenditure on goods, assuming that all goods that are produced are also sold.

(b) The amount of expenditure should also equal the total income of households, because it is households that consume the goods and they must have income to afford to pay for them.

At this stage we are assuming there are no withdrawals from, or no injections into, the circular flow of income.

4.2 The government and the circular flow of income

The government has several functions within the national economy, and so plays several different roles in the circular flow of income.

(a) It acts as the producer of certain goods and services instead of privately-owned firms, and the production of public administration services, education and health services, the police force, armed forces, fire services and public transport are all aspects of output. The government in this respect acts, like firms, as a producer and must also pay wages to its employees.

(b) It acts as the purchaser of final goods and services and adds to total consumption expenditure. National and local government obtain funds from the firms or households of the economy in the form of taxation and then use these funds to buy goods and services from other firms.

(c) It invests by purchasing capital goods, for example building roads, schools and hospitals.

(d) It makes transfer payments from one section of economy to another, for example by taxing working households and paying pensions, and by paying unemployment benefits and social security benefits.

4.3 Withdrawals and injections into the circular flow of income

Our simplified diagram of the circular flow of income in Figure 1 needs to be amended to allow for two things:

- **Withdrawals** (W) from the circular flow of income.
- **Injections** (J) into the circular flow of income.
Definitions

- **Withdrawals**: movements of funds out of the cycle of income and expenditure between firms and households.
- **Injections**: movements of funds in the other direction.

### 4.3.1 Withdrawals (W) from the circular flow of income

(a) **Savings (S)**. Households do not spend all of their income. They save some, and these savings out of income are withdrawals from the circular flow of income.

(b) **Taxation (T)**. Households must pay some of their income to the government, as taxation. Taxes cannot be spent by households.

(c) **Imports (M)**. When we consider national income, we are interested in the economic wealth that a particular country is earning.
   (i) Spending on imports is expenditure, but on goods made by firms in other countries.
   (ii) The payments for imports go to firms in other countries, for output created in other countries.
   (iii) Therefore, spending on imports therefore withdraws funds out of a country’s circular flow of income.

Be aware that **saving** is different from **investment**; saving simply means withdrawing money from circulation. Think of it as cash kept in a money box rather than being put into a bank to earn interest.

### 4.3.2 Injections (J) into the circular flow of income

(a) **Investment (I)**. Investment in capital goods is a form of spending on output, which is additional to expenditure by households. Just as savings are a withdrawal of funds, investment is an injection of funds into the circular flow of income, adding to the total economic wealth that is being created by the country.

(b) **Government spending (G)**. Government spending is also an injection into the circular flow of income. In most mixed economies, total spending by the government on goods and services represents a large proportion of total national expenditure. The funds to spend come from either taxation income or government borrowing.

(c) **Exports (X)**. Firms produce goods and services for export. Exports earn income from abroad, and therefore provide an injection into a country’s circular flow of income.

The economy is said to be in equilibrium when injections (J) equal withdrawals (W). So written out in full, equilibrium is reached in an economy when:

\[ I + G + X = S + T + M \]

Investment + Government spending + Exports = Savings + Taxation + Imports

However, if injections are greater than withdrawals (J > W) the level of national income will rise. Equally, if withdrawals are greater than injections (W > J) then the level of national income will fall.

### 4.3.3 The open economy

Figure 3 shows the circular flow of income, taking account of withdrawals and injections. This is an open economy, since it participates in foreign trade.
5 The three approaches to national income accounting

Our basic model of circular flow shows that Output = Expenditure = Income. Therefore, GDP can therefore be measured by focusing on three approaches – output, income or expenditure.

We have already identified that there are three approaches to national income accounting. We are now going to look at them in more detail.

(a) The expenditure approach. The economic wealth created in a period can be measured by the amount of expenditure on the goods and services that are produced by the nation’s economy.

(i) The expenditures will be incurred by consumers, the government and foreign buyers of exports. Expenditures on imports represent wealth created by other countries, and so the value of expenditure on imports must be deducted from the total expenditure figure.

(ii) Expenditures by firms are excluded, to avoid double-counting. Firms buy goods and services which become costs of the goods or services that they produce and sell themselves. If we included expenditure by firms, we would be double-counting the value of the wealth created by the suppliers of raw materials and components and the providers of services to other firms.

(b) The income approach. This approach measures the income of individuals from employment and from self-employment, the profits of firms and public corporations and rent on property. (Interest earnings will be included within the profits of companies or the income of individuals.)

(c) The value added or output approach. This approach is to measure the value added by all activities which produce goods and services, that is their net output in the period. However, we must take the incremental value added at each stage of production to avoid double counting. (One firm’s output might become another firm’s input, so if gross totals were used the gross value would be over-stated.)
Question 4: Measuring national income

We stated above that, in the expenditure approach to measuring national income, expenditures by firms are excluded to avoid double-counting.

Think carefully about this and ensure that you understand exactly what is meant by it. Jot down an explanation before comparing it to the one given at the end of the chapter.

(The answer is at the end of the chapter)

It will be useful to look at some simplified examples which illustrate the three approaches to measuring national income.

Worked Example: National income

Suppose that a small national economy consists of one firm. During a certain period of time, the firm undertakes certain transactions.

- It imports raw materials from abroad, costing $4 000.
- It hires labour, who are paid wages of $9 000.
- It sells all its output for $20 000 and so makes a profit of $7 000.
- It pays its post-tax profits of $4 000 to shareholders as dividends.

The country’s government taxes the labour force $2,000 and the company $3,000.

The firm’s sales of $20 000 are to three types of customer:

(a) Domestic consumers spend $11 000. This $11 000 is the post-tax wages earned by the labour force ($7 000) plus the $4 000 in dividends earned by the company’s shareholders.

(b) The government spends the $5 000 it has raised in taxes.

(c) Foreign buyers spend $4 000.

Required

Calculate the gross domestic product.

Solution

As we have seen, there are three ways of calculating national income:

(a) **The expenditure approach**

<table>
<thead>
<tr>
<th>Description</th>
<th>Amount</th>
</tr>
</thead>
<tbody>
<tr>
<td>Consumers’ expenditure</td>
<td>$11 000</td>
</tr>
<tr>
<td>Government expenditure</td>
<td>$5 000</td>
</tr>
<tr>
<td>Add exports</td>
<td>$16 000</td>
</tr>
<tr>
<td>Subtract imports</td>
<td>$4 000</td>
</tr>
<tr>
<td>GDP</td>
<td>$20 000</td>
</tr>
</tbody>
</table>

(b) **The income approach**

<table>
<thead>
<tr>
<th>Description</th>
<th>Amount</th>
</tr>
</thead>
<tbody>
<tr>
<td>Income from employment (here pre-tax wages)</td>
<td>$9 000</td>
</tr>
<tr>
<td>Gross (pre-tax) profit of the firm</td>
<td>$7 000</td>
</tr>
<tr>
<td>GDP</td>
<td>$16 000</td>
</tr>
</tbody>
</table>

The income is measured before deducting tax.

(c) **The value added or output approach**

<table>
<thead>
<tr>
<th>Description</th>
<th>Amount</th>
</tr>
</thead>
<tbody>
<tr>
<td>Output of firm at sales value</td>
<td>$20 000</td>
</tr>
<tr>
<td>Less cost (sales value) of goods or services purchased from outside firms</td>
<td>$(4 000)</td>
</tr>
<tr>
<td>GDP</td>
<td>$16 000</td>
</tr>
</tbody>
</table>
The cost of goods and services purchased from outside firms – here just the imported materials of $4,000 – has to be subtracted so as either to avoid the double-counting of output, or to remove the value of output produced by firms in other countries.

5.1 The expenditure approach to measuring national income

Probably the most widely used measure of national income is the measurement of total spending or expenditure.

The table below shows figures for the UK as an example. This example can be extrapolated for Australian, New Zealand, US and all other economies.

**UK national income 2007: expenditure approach**

<table>
<thead>
<tr>
<th>At current market prices</th>
<th>£bn</th>
</tr>
</thead>
<tbody>
<tr>
<td>Households’ expenditure</td>
<td>859</td>
</tr>
<tr>
<td>Non-profit institutions’ consumption</td>
<td>35</td>
</tr>
<tr>
<td>General government consumption</td>
<td>296</td>
</tr>
<tr>
<td>Gross domestic fixed capital formation</td>
<td>257</td>
</tr>
<tr>
<td>Value of increase/(decrease) in inventories and work in progress</td>
<td>1</td>
</tr>
<tr>
<td>Total domestic expenditure</td>
<td>1,448</td>
</tr>
<tr>
<td>Exports of goods and services</td>
<td>368</td>
</tr>
<tr>
<td>Imports of goods and services</td>
<td>(416)</td>
</tr>
<tr>
<td>Statistical discrepancy</td>
<td>1</td>
</tr>
<tr>
<td><strong>Gross domestic product (GDP) at current market prices</strong></td>
<td>1,401</td>
</tr>
<tr>
<td>Net taxes on expenditure (indirect taxes less subsidies)</td>
<td>(154)</td>
</tr>
<tr>
<td>Gross value added at basic prices (GDP at factor cost)</td>
<td>1,247</td>
</tr>
<tr>
<td>Net property income from abroad</td>
<td>9</td>
</tr>
<tr>
<td><strong>Gross national product (GNP) at current factor cost</strong></td>
<td>1,256</td>
</tr>
</tbody>
</table>

Source: United Kingdom National Accounts: The Blue Book

(Note how market prices are converted to factor cost by adjusting for taxes and subsidies.)

**The expenditure approach**

![Diagram showing the expenditure approach to measuring national income]

**Figure 4 The expenditure approach to measuring national income**

**Note**: Transfer payments (such as income support or State pensions) need to be excluded from government expenditure, because the government is effectively only transferring these payments from taxpayers to the recipients.

The Australian Bureau of Statistics publishes quarterly figures for GDP. The table below summarises the GDP figures for the first quarter of 2011:
Australia national income, 1st quarter of 2011: expenditure approach

At current market prices  $m
Households’ expenditure  177 419
National government consumption  23 383
State and local government consumption  35 363
Value of increase in inventories  1 436
Gross domestic fixed capital formation  94 290
Total domestic expenditure  331 891
Exports of goods and services  73 533
Imports of goods and services  (80 309)
Gross domestic product (GDP) at current market prices*  325 115

*Ignoring statistical discrepancies

5.1.1 Features of expenditure approach

(a) Accounts broken down by purchaser.
(b) Inserts estimated expenditure on non-traded outputs (e.g. imputed rent) to balance with output approach.
(c) Adjusted for balance of trade and for market to factor cost.
(d) Useful for detecting trends in spending patterns (for example, public versus private sector, or consumption versus investment).
(e) May inflate the real value of national income. Certain public expenditure (e.g. defence) is not really adding value and some economists argue that such expenditure leads to an overestimate of national income.

5.2 The income approach to measuring national income

The second method of calculating national income is the income method. Since money spent by an individual or firm must become income to another, the results of the two methods should be the same.

The income based approach covers several separate items of income:

(a) Income from employment (i.e. wages and salaries before deducting tax and including employers' national insurance contributions).
(b) Income from self-employment.
(c) Pre-tax profits of companies.
(d) Pre-tax profits of public corporations.
(e) The pre-tax 'surplus' of other government enterprises.

Interest earned by individuals and companies on any investments they hold is included in the first two figures.

These income components do not include two elements:

(a) Income from government pensions or social security payments which are transfer payments.
(b) Any value for work done by individuals for no monetary reward, such as housework done by private individuals or do-it-yourself home improvements are activities for which no money value can be given, and so are not economic activities.

Transfer payments are payments such as State pensions and benefits that are made by government, where the recipient does not make any contribution to output in return. They are payments which involve the transfer of wealth, rather than a reward for creating new economic wealth.

Transfer payments do not lead directly to any increase in marketable output of goods and are therefore excluded from the income figures.
5.2.1 Features of income approach

(a) Accounts broken down by source of income (e.g. employed vs self-employed).
(b) Counts all incomes gross of tax (i.e. pre-tax).
(c) Ignores transfer payments (payments for no productive efforts, eg welfare benefits grants) to avoid double-counting.
(d) Adjusted for imputed values (e.g. rent from owner-occupied houses).
(e) Useful for detecting trends in employment patterns (for example, shift from paid employment to self-employment).
(f) May understate national income if income from self-employed work is not declared, or if income is earned through the black economy.

Question 5: Transfer payments

Which of the following is or are transfer payments?

(a) Salaries paid to Members of Parliament.
(b) Incapacity benefit.

(The answer is at the end of the chapter)

5.3 The output or value added method of measuring national income

The third method of calculating national income is the value added or output method.

Since the goods and services we spend our money on must have been produced by some industry or another, it is not surprising to find the amount we have all spent is the same as the total value of the output goods and services produced. This total value of output can be calculated by adding up the ‘values added’ at the various stages of the production and distribution process.

This seeks to measure value of final products: i.e. products or services that are not subsequently used as inputs to other production processes.

However, note this method avoids double-counting by using a value added approach, and only adding the incremental value added at each stage of production.
Figure 6 Example showing that total value added equals total output

Alternatively, rather than adding the incremental value at each stage of the production process, we could measure the total value of just the final output. The two figures will be the same.

Features of output approach
(a) Accounts broken down by industrial sector.
(b) Useful in showing trends in contribution to national income from different sectors of the economy (agriculture, manufacturing, services and so on).
(c) No statistical discrepancy adjustment. The remaining two accounts made to balance to the output account by having balancing figures inserted.
(d) May understate the real output of the economy because it will exclude output which is not sold for a market price (for example, DIY or housework).
(e) The value of some goods and services (for example education, health care) have to be eliminated because they do not have a market price.

Question 6: GDP and GNP
The following data relates to the economy of a country over a one-year period:

<table>
<thead>
<tr>
<th>Description</th>
<th>$million</th>
</tr>
</thead>
<tbody>
<tr>
<td>Consumers' expenditure</td>
<td>18 400</td>
</tr>
<tr>
<td>General government final consumption</td>
<td>4 540</td>
</tr>
<tr>
<td>Gross domestic fixed capital formation</td>
<td>4 920</td>
</tr>
<tr>
<td>Value of physical decrease in inventories</td>
<td>20</td>
</tr>
<tr>
<td>Exports of goods and services</td>
<td>6 450</td>
</tr>
<tr>
<td>Imports of goods and services</td>
<td>6 240</td>
</tr>
<tr>
<td>Taxes on expenditure</td>
<td>2 760</td>
</tr>
<tr>
<td>Subsidies</td>
<td>300</td>
</tr>
<tr>
<td>Net income from abroad</td>
<td>210</td>
</tr>
<tr>
<td>Capital consumption</td>
<td>1 750</td>
</tr>
</tbody>
</table>

Required
Calculate the following from the above data:
(a) gross value added at basic prices (GDP at factor cost).
(b) gross national product (GNP) at factor cost.
(c) national income at factor cost.

(The answer is at the end of the chapter)
6 Interpretation of national income accounting information

Section overview
- To see if there has been any real change in the level of activity we must deduct any influence due to inflation. The standard method for turning money GDP or GNP into real measures is to use what is called the GDP deflator in order to take inflation out of the figures.
- National income has serious limitations as a measure of economic wealth and economic development. Nonetheless, it remains an important indicator.

6.1 The purposes of calculating national income

Calculating the national income serves several purposes:
(a) It measures the standard of living in a country (national income per head).
(b) It allows comparisons between the wealth of different countries.
(c) It makes it possible to measure the improvement (or deterioration) in national wealth and the standard of living.
(d) It assists central government in its economic planning.

National income or GNP per head of the population gives an indication of the trend over time in a country’s standard of living, because GNP is measured consistently from year to year, whatever the weaknesses of the measurement system that is used.

For example, in publishing its statistics for GDP in the first quarter of 2011, the Australian Bureau of Statistics commented that GDP had fallen by 1.2% in the first quarter of the year, following flooding and cyclones in Queensland and Western Australia that began in December 2010. The main reason for the fall in GDP was the consequent fall in exports in the period.

National income statistics are also useful for comparing the income and standards of living in different countries using the average GDP per head of the population (‘per capita’) as the comparative measure. For example in 2010, the USA had the fifth largest GDP per capita, at about US$33 000. The comparative figure for Australia was about US$19 000.

6.2 National income and inflation

6.2.1 Real and nominal GDP

Inflation is a particular problem in using national income as a measure of national wealth. Price inflation increases the money value of national income. We should be careful not to interpret this as meaning that there is more economic activity going on in an economy. All that has happened is that the prices of the things we are measuring have increased.

Calculating real GDP

In terms of measuring domestic income and comparing it year on year, real GDP is a more relevant measure. The real GDP is calculated by using a price index referred to as the GDP deflator. This measures the level of base prices as a proportion of a selected base year, which is measured as 100 on the scale. The real GDP for a year, on the same basis as the base year, can be calculated by adjusting the nominal GDP for inflation over the period, as follows:

\[ \text{Real GDP}_{\text{now}} = \text{Nominal GDP}_{\text{now}} \times \frac{\text{GDP}_{\text{base}}}{\text{GDP}_{\text{now}}} \]

where GDP_{base} = GDP deflator in the base year (= 100)
GDP_{now} = GDP deflator for current year
Note that the GDP deflator is not the same as the consumer price index (CPI). The CPI focuses only on a relatively narrow basket of consumer goods, whereas the GDP deflator aims to take account of all goods in the economy. Generally, the CPI is more commonly used. It measures the impact of rising prices on the income of consumers. However, the GDP deflator is more used as an economy-wide measure.

**Worked Example: Real and nominal GDP**

Nominal GDP is $6,000bn. The GDP deflator for the year is 130 (base = 100). What is real GDP?

**Solution**

Real GDP = \( \frac{6000 \times 100}{130} = 4615 \) billion dollars

### 6.3 Problems with the use of GDP

**Nonmarket production**

Work done where no charge is made will not be included in GDP. This will include, for example, housework, gardening, DIY and similar activities which are not sold for a market price. This makes comparisons of GDP over time less meaningful, particularly since women now do more market production, which is included in GDP, along with the previous household chores. The household chores may take less time with time-saving appliances, the cost of which is again included as part of GDP. In the past, there were more hours of unpaid labour in the house that were not included in GDP.

As a result, GDP tends to overstate the output of developed countries compared to less developed countries.

**The underground economy**

Transactions in the underground economy are not reported, typically for tax evasion or other illegal purposes, understating GDP.

**Leisure and human costs**

Increased leisure time and improved safety conditions at work, for example, are not reflected in GDP.

**Quality variation and new goods**

Improvement in the quality of goods (e.g. dental treatment) and new goods not previously available (such as i-Pods replacing CD players) are not taken into account. The failure to allow for increased quality tends to result in inflation being overstated.

**Economic ‘bads’**

Harmful side effects, such as pollution, are ignored in GDP. In addition, costs of rebuilding houses after flooding or earthquakes, is treated as national income for the year, even though this quite clearly is a cost rather than being a benefit.

### 6.3.1 Economic development

GDP does not measure **economic welfare** or **economic development**, but rather measures the value of goods and services produced in the year. If consumers spent $1bn on alcohol or on schooling for their children, it would have the same value for GDP.

It is possible to adjust GDP to attempt to reach some measure of economic welfare or development by deducting the cost of economic bads, deducting expenditure spent on regrettable necessities (e.g. defence) and adding in the value of non-market production.

A second argument against traditional measures of national income is that they do not provide an accurate indication of **sustainable economic development**. For example, mineral extraction will be included in national income. However, mineral stocks (coal, oil and so on) are a finite resource and so by extracting
them now we are depleting the stocks of resources for the future. So the reduction (and eventually, exhaustion) of finite resources will have a negative impact on future sustainable growth.

However, this future negative impact is not allowed for in the current measure of national income. And as we have already noted, nor are pollution costs.

Critics also argue that defensive expenditure should also be adjusted for in calculating economic development. For example, the number of asthma sufferers has increased due to increased levels of air pollution; consequently spending on asthma treatments have also increased.

To take account of these adjustments, there is an alternative measure of economic growth called an Index of Sustainable Economic Welfare. This deducts the costs of depreciation of natural capital, environmental degradation, and defensive expenditure from traditional GDP or GNP statistics to give a measure of sustainable economic development.

6.4 Using national income figures to compare economies

The main conventions used to make figures more comparable are:

(a) Uses of real national income per capita (i.e. national income divided by population) to get a measure of ‘average’ of how well off people are.

(b) All values equated using exchange rate or expressed in a single currency (eg dollar).

(c) Values converted from nominal values (i.e. measured in the prices at the time) to real values or constant values by eliminating inflation using a GDP deflator.

6.4.1 Comparisons of national incomes

Analysis of trends in national income figures will be important for both firms and governments. There are two ways of looking at these trends: over time, and between countries. However, we need to be aware of the difficulties involved in these comparisons.

Time

National income statistics are an indicator of changing living standards in a country over time. Increases in real national income per head over time will indicate rising living standards.

Note that real (rather than nominal) figures have to be used to allow for the impact of inflation on the headline figures. Similarly, per capita figures need to be used to adjust for the impact of changes in population size.

However, although the overall average real national income per head may have increased over time, this increase may not be evenly distributed across everyone in a country. So policy makers need to be aware of the distribution of income within the economy.

Similarly, they need to be aware of the ways in which national income figures may be distorted. For example, if government expenditure increased to fund a war effort, this would increase national income, but it would not indicate an increased standard of living or an increase in the level of economic welfare.

Between countries

Comparing national incomes between countries provides an indicator of differences in economic development and economic growth.

However, again the overall picture may be distorted by the distribution of income within an economy. For example, the oil-rich countries of the Middle East have relatively high average incomes per head, but the wealth is very unevenly distributed across their populations.

In order to compare national income figures across countries, the figures also need to be converted into a common currency. However, the figures may still be distorted by movements in exchange rates rather than underlying national income.
As discussed in Section 1, there may be non-economic considerations which may influence people's living standards, and which make it difficult to compare living standards through GDP figures alone:

- Quality of life indicators such as crime, or education.
- Political system and stability.
- Climate and geographical infrastructure such as transport networks.

However, despite these limitations, national income statistics can still be a useful indicator for firms when they are considering international expansion – either in respect of sales and marketing growth, in terms of relocating production facilities overseas.
Key chapter points

- National income indicates the performance of an economy. It shows total value of goods and services of the economy's resources and total income of residents.
- It measures economic growth not the wider-ranging measure of economic development.
- National income accounting is the system through which economic activity on a national scale is measured.
- Economic activity can be measured by output, income and expenditure.
- There are three key measures of national economic output.
  - National income.
  - Gross national product (GNP).
  - Gross domestic product (GDP).
- There is a circular flow of income in an economy, which means that expenditure, output and income will all have the same total value.
- The government has an important role in the circular flow of income.
- There are withdrawals from the circular flow of income (savings, taxation, import expenditure) and injections into the circular flow (investment, government spending, export income).
- Our basic model of circular flow shows that Output = Expenditure = Income. GDP can therefore be measured by focusing on three approaches – output, income or expenditure.
- To see if there has been any real change in the level of activity we must deduct any influence due to inflation. The standard method for turning money GDP or GNP into real measures is to use what is called the GDP deflator in order to take inflation out of the figures.
- National income has serious limitations as a measure of economic wealth and economic development. It remains an important indicator nonetheless.
Quick revision questions

1. Which of the following are withdrawals from the circular flow of income?
   A. exports
   B. savings
   C. investment
   D. government spending

2. Define the income approach to measuring national income.

3. Which of the following define GNP at factor cost?
   I. Consumers' total expenditure on domestically produced goods and services
   II. The value of total output produced domestically plus net property income from abroad, minus capital consumption
   III. The total income received by residents in a country in return for factor services provided domestically and abroad
   A. I and III
   B. II and III
   C. III only
   D. I and II

4. Fill in the gaps in the following proforma for calculating national income:
   Consumption
   + Government expenditure
   + Gross domestic capital formation
   + Exports
   – Imports
   = (i)
   – Taxes
   + Subsidies
   = (ii)
   + Net property income from abroad
   = (iii)
   – Capital consumption
   = (iv)

5. What is capital consumption?

6. What is the equilibrium point of national income in an open economy?
1 B only. The rest are injections into the circular flow of income.

2 National income is the aggregate of individuals’ income from employment and self employment, the profits of entities and rent on property.

3 C GNP is a measure of all incomes received by residents, including income for services provided abroad as well as domestically, so option III is correct. Option I is incorrect because it is looking at domestic production only. Option II is incorrect because it deducts capital consumption and so is a measure of net, not gross, national product.

4 (i) GDP at market prices.
(ii) GDP at factor cost.
(iii) GNP at factor cost.
(iv) National income at factor cost.

5 The allowance for depreciation on an economy’s capital assets which is deducted from GNP to calculate national income (or net national product).

6 \[ I + G + X = S + T + M \]
Investment + Government spending + Exports = Savings + Taxation + Imports
Answers to chapter questions

1. The factors of production are as follows:
   (a) Land is rewarded with rent.
   (b) Labour is rewarded with wages.
   (c) Capital is rewarded with interest.
   (d) Enterprise is rewarded with profit.

2. This is a reinforcement activity. Did your review find that your understanding of the three terms had deepened?

3. D A rise in indirect taxes. National income is measured at market prices which are increased by the indirect taxes.

4. Firms buy goods and services which become costs of the goods or services that they produce and sell themselves.
   If we included expenditure by firms, we would be double-counting the value of the wealth created by the suppliers of raw materials and components and the providers of services to other firms.

5. (a) MPs’ salaries are not transfer payments. MPs are like any other employees – they just happen to be employed by the government.
   (b) This falls within the category of social security payments, which are transfer payments. The recipient receives the benefit, but is not making any contribution to output in return.

6. The calculation below also shows GDP and GNP at market prices, although these are not required in the exercise.

<table>
<thead>
<tr>
<th></th>
<th>$million</th>
</tr>
</thead>
<tbody>
<tr>
<td>Consumers’ expenditure</td>
<td>18 400</td>
</tr>
<tr>
<td>General government final consumption</td>
<td>4 540</td>
</tr>
<tr>
<td>Gross domestic fixed capital formation</td>
<td>4 920</td>
</tr>
<tr>
<td>Value of physical decrease in inventories</td>
<td>(20)</td>
</tr>
<tr>
<td>Total domestic expenditure</td>
<td>27 840</td>
</tr>
<tr>
<td>Exports</td>
<td>6 450</td>
</tr>
<tr>
<td>Imports</td>
<td>(6 240)</td>
</tr>
<tr>
<td>GDP at market prices</td>
<td>28 050</td>
</tr>
<tr>
<td>Net property income from abroad</td>
<td>210</td>
</tr>
<tr>
<td>GNP at market prices</td>
<td>28 260</td>
</tr>
</tbody>
</table>

   | GDP at market prices (see above)      | 28 050   |
   | Factor cost adjustment                |          |
   | Taxes on expenditure                  | (2 760)  |
   | Subsidies                             | 300      |
   | Gross value added at basic prices (GDP at factor cost) | (a) 25 590 |
   | Net property income from abroad       | 210      |
   | GNP at factor cost                    | (b) 25 800 |
   | Capital consumption                   | (1 750)  |
   | National income at factor cost        | (c) 24 050 |
Chapter 8
Determining national income

<table>
<thead>
<tr>
<th>Learning objectives</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Determining national income</td>
<td>LO8</td>
</tr>
<tr>
<td>Calculate the national income equation $Y = C + G + I + M - X$</td>
<td>LO8.1</td>
</tr>
<tr>
<td>• Present national income calculations using the IS – LM curve</td>
<td>LO8.1.1</td>
</tr>
<tr>
<td>• Calculate marginal efficiency of capital</td>
<td>LO8.1.2</td>
</tr>
<tr>
<td>• Apply the multiplier to determine national income</td>
<td>LO8.1.3</td>
</tr>
<tr>
<td>• Apply the accelerator principle in the determination of national income</td>
<td>LO8.1.4</td>
</tr>
<tr>
<td>Evaluate the implications of the marginal propensity to save (MPS) and the marginal</td>
<td>LO8.2</td>
</tr>
<tr>
<td>propensity to consume (MPC) on national income ($Y$)</td>
<td></td>
</tr>
<tr>
<td>Evaluate the impact of tax, savings and subsidies on national income</td>
<td>LO8.3</td>
</tr>
<tr>
<td>Explain the relationship between full employment and national income</td>
<td>LO8.4</td>
</tr>
</tbody>
</table>

Topic list

1. The Keynesian approach
2. Aggregate demand analysis: consumption, savings and investment
3. The multiplier and the accelerator
4. Aggregate supply analysis
5. The determination of national income
This chapter follows on from Chapter 7 and examines the determination and calculation of national income. Broadly speaking, macroeconomists divide into two camps when examining these and other decisions: the Keynesians and the monetarists. These two camps have had differing ideas about how national income can be made to grow, how full employment can be achieved, and how booms and slumps of trade cycles can be smoothed out.

Keynes argued that the economy should be managed by controlling aggregate demand, and in this chapter we will study the basic elements of the Keynesian model for national income determination and equilibrium.
Before you begin

If you have studied these topics before, you may wonder whether you need to study this chapter in full. If this is the case, please attempt the questions below, which cover some of the key subjects in the area.

If you answer all these questions successfully, you probably have a reasonably detailed knowledge of the subject matter, but you should still skim through the chapter to ensure that you are familiar with everything covered.

There are references in brackets indicating where in the chapter you can find the information, and you will also find a commentary at the back of the Study Manual.

1. In the national income equation: (Sections 1.2, 2.1)
   \[ Y = \ldots \]
   \[ C = \ldots \]
   \[ G = \ldots \]
   \[ I = \ldots \]
   \[ M = \ldots \]
   \[ X = \ldots \]

2. What are the marginal propensity to consume (MPC) and the marginal propensity to save (MPS)? (Section 2.1.2)

3. What is the multiplier effect? (Section 3.1)

4. What is the accelerator principle? (Section 3.6)

5. Explain the relationship between national income and full employment. (Section 5.2)
1 The Keynesian approach

Section overview

The Keynesian model provides a way of explaining how national income is determined, and how national income equilibrium is reached.

1.1 Aggregate demand and aggregate supply

In our review of microeconomics earlier in the Study Manual, we identified how the price mechanism allows an equilibrium to be established between supply and demand. However, this idea of equilibrium can also be applied at the macroeconomic level to the economy as a whole.

The idea of the circular flow of income in the economy (which we examined in Chapter 7) illustrates that an economy will be in equilibrium when injections equal withdrawals. If injections are greater than withdrawals, the economy is out of equilibrium and national income will rise. Similarly, if injections are less than withdrawals, national income will fall.

You may notice the similarities between the relationship between injections and withdrawals in determining the level of national income, and the interaction of supply and demand in determining equilibrium prices. This similarity was highlighted by John Maynard Keynes, and he incorporated it into his analysis of macroeconomic activity.

Keynes argued that demand and supply analysis could be applied to macroeconomic activity as well as microeconomic activity.

Definitions

Aggregate demand (AD) means the total demand for goods and services in the economy.

Aggregate supply (AS) means the total supply of goods and services in the economy.

Aggregate supply depends on physical production conditions – the availability and cost of factors of production and technical know-how. However, Keynes was concerned with short-run measures to affect the economy, and he also wrote in a period of high unemployment when there was obviously no constraint on the availability of factors of production. His analysis therefore concentrated on the demand side and how aggregate demand could be managed.

1.1.1 Aggregate demand and national income

For Keynesian analysis to have practical value for the management of a national economy, it is necessary to establish how aggregate demand can be shifted.

To understand shifts in AD, we need to turn our attention to expenditure in the economy. A formula for total expenditure (gross national product) is:

\[ E = C + I + G + (X - M) \]

where:

- **E** is the total national expenditure (GNP)
- **C** is the total domestic consumption (money spent on consumer goods)
- **I** is the total industrial investment (money spent by private sector firms and the public sector on capital items) *
- **G** is the total government spending (government ‘current’ or ‘consumption’ spending)
- **X** is the total exports (including income from property abroad)
- **M** is the total imports (including money paid as income to residents in other countries for property they hold in the country)

* Alternatively, government investment spending on capital items can be included in G leaving I to represent investment by firms only.
Definition

Demand management policies involve the manipulation of total national expenditure (E) by influencing C, I, G or net exports (X – M).

If we ignore capital consumption, we can equate GNP with national income. This is what we shall do in our analysis of the Keynesian model.

1.2 Withdrawals and injections

In the previous chapter on national income, the different approaches to calculating national income – the expenditure, income and value added approaches – were explained in terms of the circular flow of income around the economy.

For a national economy, there are certain withdrawals from and injections into this circular flow of income. Withdrawals divert funds out of the circular flow and injections add funds into it.

(a) **Withdrawals** from the circular flow of income (W) consist of imports (M), taxation (T) and savings (S).

(b) **Injections** into the circular flow of income (J) consist of exports (X), government spending (G), and investment spending by firms (I).

Keynes argued that for an equilibrium to be reached in the national income, not only must AD = AS, but also total **planned withdrawals** from the circular flow of funds must be equal to total **planned injections**. Thus, for equilibrium:

\[ W = J, \text{ and so } M + T + S = X + G + I \]

In the long term W will always equal J.

(a) The difference between the value of imports M and the value of exports X is the **balance of payments deficit** (or surplus). Even in the short term, this difference must be balanced by borrowing (or lending abroad).

(b) The difference between government spending and taxation can only be made up by government borrowing. Loans are eventually repaid.

(c) In the long run, savings will also equal investments, even though the people who save and the firms who invest are not the same.

However, although W and J will be equal retrospectively and in the long run, it does not follow that planned J and planned W will equal each other **in the short run**, since injections and withdrawals are made by different people.

This frustration of plans in the short run causes national income to change over time. The imbalance between J and W creates factors which can make the level of national income change. Keynes argued that the imbalance between planned withdrawals and planned injections explained **trade cycles** – the fluctuations in national income which give rise to booms and slumps – which prevent an economy from settling down at an equilibrium level.
2 Aggregate demand analysis: consumption, savings and investment

Section overview

- **Consumption expenditure** depends on income. It might be possible for a government to take measures to boost aggregate demand in the economy, although some price inflation will probably result. When there is inflation in the economy, measures could be taken to suppress aggregate demand.

- **Autonomous consumption** is the basic level of consumption by a household irrespective of income. **Non-autonomous consumption** is where the amount spent depends on the level of income. The proportion of income spent on non-autonomous consumption is the marginal propensity to consume (MPC) and the proportion saved is the marginal propensity to save (MPS).

2.1 Consumption and savings (C and S)

Let us now go into a bit more detail on Keynesian analysis, and concentrate particularly on consumption, savings and investment. To simplify our analysis, we shall ignore government spending, taxation, imports and exports for the time being. By ignoring imports and exports, we are concentrating on a **closed economy** which is not in any way dependent on foreign trade.

If we ignore G, T, X, and M, we can look at a circular flow of income in which households divide all their income between two uses: consumption and saving. (Spending by households in the circular flow of income is known as consumption.)

**Provided that national income is in equilibrium, we will have:**

\[
Y = C + S
\]

where:

- **Y** = national income
- **C** = consumption
- **S** = saving

This should seem logical to you. Income (Y) can only be either spent (C) or saved (S). Since we have a closed economy, consumption must be of goods produced by the economy itself.

2.1.1 Savings

There are two ways of saving. One is to hold the income as money (banknotes and coin, or in a current bank account). The other way is to put money into some form of interest-bearing investment. In the long run, there is no reason for people to hold banknotes or keep money in a current bank account, unless they intend to spend it fairly soon. If this is so, income that is not spent will be saved and income that is saved will, eventually, be invested. (The people who put their money into interest-bearing savings are not making any investment themselves in capital goods, but the institutions with whom they save will use the deposits to lend to investors and so indirectly there will be a real increase in investment when people save money in this way.)

**Question 1: Factors influencing savings**

What do you think are the main factors influencing the amount that people will save?

(The answer is at the end of the chapter)
We can therefore conclude that in **conditions of equilibrium** for national income:

\[
Y = C + S
\]

and,

\[
Y = C + I
\]

and so,

\[
I = S
\]

In the short run, however, savings and investment might not be equal and so there might not be equilibrium.

### 2.1.2 The propensities to consume and save

Even when a household is receiving “zero” income, it will still spend. This spending will be financed by earlier individual **savings** or from government **welfare receipts** such as unemployment benefit. There is thus a constant, basic level of consumption which does not vary whatever the level of income. This is called **autonomous** consumption.

However, alongside this autonomous consumption there is also **non-autonomous consumption**, in which the amount spent will vary according to the income earned. When the household receives an income, some will be spent and some will be saved. The **proportion** of the income which is spent on non-autonomous consumption depends on the marginal propensity to consume (MPC) while the **proportion** which is saved is determined by the marginal propensity to save (MPS).

MPC is calculated as: $\text{MPC} = \frac{\text{Change in consumption}}{\text{Change in income}}$

**Question 2: Marginal propensity to consume**

A household’s disposable income has increased from $200 to $250 per week, and consumption has increased from $180 to $220.

What is the household’s marginal propensity to consume?  

(The answer is at the end of the chapter)

The marginal propensity to save is the amount that is saved out of additional income.

\[\text{MPS} = \frac{\text{Change in amount saved}}{\text{Change in income}}\]

In our analysis (ignoring G, T, X and M) saving and consumption are the only two uses for income, so (MPC + MPS) must = 1.

Taking autonomous and non-autonomous consumption together, we can say that a household’s expenditure in a given period is made up of two elements:

(a) A fixed amount (a) which is the autonomous consumption.

(b) A further constant percentage of its income (b% of Y) representing the MPC.

Figure 1 shows how autonomous and non-autonomous consumption combine to give total consumption. The gradient of the consumption function (C) reflects the MPC.

The 45° line $Y = C$ maps the points where income = consumption (expenditure), and therefore where average propensity to consume (APC) is 1.

If $C > Y$ then we are witnessing **dis-saving**, and APC is greater than 1: i.e. $\frac{\text{Consumption}}{\text{Income}} > 1$.

Where $Y > C$ then a household is making savings over and above their consumption. APC is less than 1.
Similarly, a national economy as a whole will spend a fixed amount, \( 'a' \), plus a constant percentage (\( b\% \)) of national income \( Y \).
We can then state the overall consumption function as \( C = a + bY \).
The marginal propensity to consume (MPC) will be represented by the gradient 'b', because it is the proportion of an extra $1 earned that is spent as consumption.

The average propensity to consume (APC) will be the ratio of consumption to income:

\[
\frac{C}{Y} = \frac{a + bY}{Y}
\]

For example, suppose an individual household has fixed spending of $100 per month, plus extra spending equal to 80% of its monthly income.
When its monthly income is $800, its consumption will be: $100 + 80% of $800 = $740
When its monthly income is $1,000 its consumption will be: $100 + 80% of $1,000 = $900
Because the household's marginal propensity to consume is 80%, consumption has increased by $160 following a rise in income of $200.

**Question 3: Average propensity to consume**

Using the above figures, calculate the household's average propensity to consume:

(a) When its income is $800
(b) When its income is $1,000

(The answer is at the end of the chapter)

Changes in the marginal propensity to consume and the marginal propensity to save will involve a change of preference by households between current consumption and saving for future benefits. A cause of such a change might be a change in interest rates, which makes the investment of savings more or less attractive than before.

### 2.2 What factors influence the amount of consumption?

There will always be a minimum fixed amount of total consumption. Total consumption by households, however, is affected by six influences:

(a) **Changes in disposable income, and the marginal propensity to consume.** Changes in disposable income are affected by matters such as pay rises and changes in tax rates. An increase in household wealth, or expectations of a future increase in household wealth, may increase consumption levels.
(b) **Changes in the distribution of national income.** Some sections of the population will have a higher marginal propensity to consume than others, and so a redistribution of wealth might affect consumption. (A redistribution of wealth might be accomplished by taxing the rich and giving to the poor in the form of more government allowances.)

(c) **Government policy.** Government can influence consumption levels through taxation, subsidies and/or public spending. For example, an increase in direct taxation will reduce disposable income and therefore will also reduce consumption.

(d) **The development of major new products.** When such developments happen, they can create a significant increase in spending by consumers who want to buy the goods or services.

(e) **Interest rates.** Changes in interest rates will influence the amount of income that households decide to save, and also the amount that they might elect to borrow for spending. High interest rates will make saving more attractive. Conversely, low interest rates will reduce the cost of credit and will therefore increase levels of consumption.

(f) **Price expectations.** Expectations of price increases may increase current consumption while expectations of price reductions may have the opposite effect.

Among the determinants of the MPC are taste and attitude. If a household believes that saving is a virtue it will save as much as possible and spend as little as possible. In the economy as a whole, a general belief in the value of thrift may mean that the MPC is low. Nowadays, however, the prestige attached to the possession of consumer goods may have overcome the admiration for thrift, making the MPC higher than it once was.

### 2.3 What factors influence the amount of saving?

Saving is the amount of income which is not consumed. Therefore, not surprisingly, the influences which affect savings are very similar to those that affect consumption – but in mirror image.

(a) **Income.** The level of income will be a key determinant in the level of savings. We have already noted that MPC + MPS = 1, but it is also true that APC + APS (average propensity to consume + average propensity to save) = 1.

(b) **Interest rates.** If interest rates rise, saving becomes more attractive relative to consumption.

(c) **Cost and availability of credit.** Similarly, as the cost of credit rises, consumption becomes less attractive, meaning that people will save more. When credit is easily available, people will be encouraged to spend on credit. In which case, they might acquire as much credit as they are saving, meaning that there is no net saving.

(d) **Inflation.** If inflation rates are high, then nominal interest rates are also likely to be high, even though the real rates (adjusted for inflation) may be low. Consumers may be attracted by the money illusion of high nominal interest rates into saving more.

(e) **Long-term savings.** A large amount of household savings goes into long-term, contractual savings, such as pension schemes. The savings may be less likely to vary with income than with the demographics – for example, the level of savings into pension schemes have risen alongside increases in life expectancy in developed countries.

### 2.4 Marginal propensity to withdraw

Remember that savings are not the only withdrawals from the circular flow of income. In a model that includes savings (S), taxes (T) and imports (M), the marginal propensity to withdraw (MPW) is given by:

\[
MPW = MPS + MPT + MPM
\]

The MPW is the proportion of additional (marginal) national income that is withdrawn from the circular flow of income.

- MPS, the marginal propensity to save, has been explained already
- MPT is the marginal propensity to tax: this is the proportion of any additional national income that will be paid in taxes and so be withdrawn from the circular flow of income
- MPM is the marginal propensity to import: this is the proportion of any additional national income that will be spent on imports and so be withdrawn from the circular flow of income
Chapter 2.5 Investment

Investment (I)
The total volume of desired investment in the economy depends on factors similar to those influencing ‘micro-level’ investment decisions by firms:

- The rate of interest on capital.
- The marginal efficiency of capital (MEC) invested. (We will look at MEC in more detail later, but it is the rate of return at which the NPV of a project is zero.)
- Expectations about the future and business confidence, including expectations about future cash flows and profit flows arising from the investment.
- The strength of consumer demand for goods.
- Opportunity cost of investment.

The demand for funds to invest by firms and the willingness of investors to lend their savings for investment (the supply of funds) should adjust to one another through the price mechanism of the interest rate.

(a) Higher interest rates should make firms less willing to invest, because the marginal efficiency of capital will have to be higher to justify the higher interest cost. However, firms cannot always cut their investment plans quickly and at short notice.

Higher interest rates should have two other effects:

(i) They will tempt individuals to consume less of their income and save more, with a view to investing it.

(ii) They will also tempt individuals to invest more of their savings – that is, to hold less cash and more interest-bearing investments.

(b) Lower interest rates should have the opposite effect.

An investment involves the acquisition of more fixed capital (buildings, machinery, plant and equipment) or inventories of goods. Therefore, the importance of the interest rate for investment should therefore be apparent in the marginal efficiency of capital. Firms should go on adding to their capital provided that the marginal efficiency of capital exceeds the interest rate, which is its marginal cost.

2.5.1 New technology and investment

When new technology emerges which changes methods of production (such as robotics) or provides opportunities to produce new types of good, there will be a boost to investment.

(a) New technology which reduces the unit costs of production will increase profitability. The supply curve for the goods that are affected by the new production methods will shift to the right. Firms will invest in the new technology in order to achieve lower costs and remain competitive.

(b) New technology which leads to new types of good will give a stimulus to consumption demand. Firms will invest to make the product and meet the consumer demand.

2.5.2 Investment and economic recovery

Investment represents one of the major injections in the circular flow of income. Variations in the level of investment can, as we will see later, through the multiplier, affect the level of national income and the level of aggregate demand in the economy. A major conclusion from Keynesian analysis is that in order to achieve economic recovery from a recession, there should be major investment.

Investment can be in either the public or the private sector of the economy, although the money to finance the investment might need to come from different sources:

(a) Private sector investment will come from retained profits, new issues of shares, or borrowing. However, in an economic recession profits might be low, and investors might lack confidence in a recovery, so that new share issues may be difficult on a large scale.

(b) Public sector investment might be financed by higher taxation, or by an increased deficit between government income and expenditure. In the UK the term for this deficit is the public sector net cash requirement (PSNCR). There is no equivalent measure of public finances in Australia which uses the generic terms public deficit or surplus as a comparative measure.
(c) Public sector spending should have socially valuable spin-off effects, such as improved roads, sewers and public buildings.

(d) However, a high government deficit, meaning large government borrowings, might force up interest rates in the capital markets and crowd out private sector investment, by making it too expensive for firms to borrow and invest profitably. (Keynesian economists deny that such crowding out takes place, however.)

There are two additional reasons why investment is very important:

(a) The act of investment represents consumption forgone now in order to increase the capacity to produce, and therefore to consume, in the future. It is through investment (or lack of it) that the future shape and pattern of economic activity is pre-determined.

(b) The growth rate of the economy is determined not only by the technological progress or the increases in the size and quality of the labour force but also by the rate at which the capital stock is increased or replaced. Investment represents an addition to the existing capital stock. If that addition is greater than the amount by which the capital stock depreciates, then the capital stock of the economy is growing and so is the capacity of the economy to produce more goods and services. Therefore, investment is an important determinant of the long-term growth rate of an economy.

We can illustrate the relationship between investment and national income as in Figure 2.

For the moment, we assume that we are dealing with a closed economy (no imports or exports) and with no government intervention (no taxation or subsidies).

We also assume that investment is constant (autonomous) and not related to national income levels.

Aggregate demand (AD) in the economy is represented by C + I.

The economy will be in equilibrium where aggregate demand equals the output of the economy, or aggregate supply (AS).

When AD > AS, national income will need to rise to restore the equilibrium position. This can be achieved through new injections into the economy – through investment.

We can see this at Y, where AD > AS, but I > S. However, the excess of investment over savings (I > S) will provide the injection necessary for the economy to grow and restore equilibrium, at Y_e. At this equilibrium point, savings and investment are equal. Because we are looking at a closed economy with no government intervention, saving is the only withdrawal and investment is the only injection. Therefore, at the equilibrium point injections equal withdrawals. This is shown in Figure 2.

Figure 2 Equilibrium national income in a closed economy
2.6 Determinants of investment

In Figure 2 we have showed a constant level of investment. However, this is unrealistic because, in practice, investment will be determined by the levels of return firms expect to make from them.

2.6.1 Discounted cash flow and net present value

When considering an investment, a firm will be concerned with the net rate of return it will receive from its investment.

In this way, the firm will consider not only the values of future sales and costs, but also the timings of them and interest rate levels.

Interest rates are crucial to the investment decision, because if the funds used for the investment had not been invested they could have earned interest elsewhere.

Consequently, the future income stream which an investment is expected to generate has to be discounted to allow for the interest income forgone, and to convert it to its net present value.

Net present value (NPV) is the value obtained by discounting all cash outflows and inflows of a capital investment project by a chosen target rate of return or cost of capital.

The present value (PV) of a project can be shown as:

\[
PV = \frac{Q}{1 + r} + \frac{Q}{(1 + r)^2} + \ldots + \frac{Q}{(1 + r)^n}
\]

Where \(Q\) is the anticipated annual inflow, \(r\) is the interest rate, and \(n\) is the time period when the inflow will occur.

From this we can see that the PV will vary inversely to interest rates, meaning that if interest rates increase, the PV of the investment will fall, ceteris paribus.

2.6.2 Marginal efficiency of capital

Using the NPV method, present values are calculated by discounting at a target rate of return, and the difference between the PV of costs and the PV of benefits is the net present value.

In contrast, the marginal efficiency of capital approach (as introduced by Keynes) calculates the exact rate of return which a project is expected to achieve, in other words, the rate at which the NPV is zero.

This rate of return is termed the marginal efficiency of capital (MEC). If the MEC (or internal rate of return (IRR)) exceeds the current rate of interest, then an investment will be profitable.

Keynes used this concept to argue that there is a downward sloping MEC curve, that is levels of investment will be higher at lower rates of interest.

A change in interest rates is likely to induce a movement along the MEC curve and prompt a change in levels of investment.

However, we could also witness shifts in the MEC curve, notably from changes in business confidence. If businesses are optimistic about the future they are more likely to invest. This would lead to an outward shift in the MEC curve (MEC to MEC1 in Figure 3). The MEC curve could also be shifted by technological innovation (making capital more productive) or government policy (for example, a reduction in taxes could encourage investment). A final factor which could lead to a shift in the MEC curve is the substitution of other factors of production. For example, if wage costs rise, a business may look to substitute capital for labour, in effect shifting the MEC curve to the right.
3 The multiplier and the accelerator

Section overview

- Changes in national income begin with a small change in expenditure, leading to an even larger eventual change in national income, due to the multiplier effect.
- Changes in national income can be explained partially by the accelerator principle, which is the mechanism by which changes in investment spending are proportionally greater than changes in consumption spending, and therefore investment spending is more susceptible to bigger upturns and downturns.

3.1 The multiplier

The multiplier describes the process of circulation of income in the national economy, whereby an injection of a certain size leads to a much larger increase in national income. The firms or households receiving the injection use at least part of the money to increase their own consumption. This provides money for other firms and households to repeat the process and so on.

The level of national income might increase or decrease for a number of reasons; for example, there might be an increase in productivity or an increase in the country's exports. Keynes showed that if there is an initial change in expenditure due to an increase in exports, government spending, investment or consumer spending, a new equilibrium national income level will eventually be reached.

The eventual total increase in national income will be greater in size than the initial increase in expenditure. This is because of the continuing circulation of the funds concerned.

Definition

The ratio of the total increase in national income to an initial injection is called the multiplier.

\[
\text{Multiplier} = \frac{\text{Total increase in national income}}{\text{Initial increase in national income}}
\]

The multiplier can be defined as a measure of the effect on total national income of a unit change in a component of aggregate demand: I, G or X.

Multiplier values can therefore be measured specifically for each of these separately:

\[
\text{Investment multiplier} = \frac{\text{Eventual change in national income}}{\text{Initial change in investment spending}}
\]
Government spending multiplier = \[\frac{\text{Eventual change in national income}}{\text{Initial change in government spending}}\]

Export multiplier = \[\frac{\text{Eventual change in national income}}{\text{Initial change in exports}}\]

**Worked Example: The multiplier**

A numerical illustration of the multiplier might help to explain it more clearly. In this example, we shall again ignore taxes, government spending, exports and imports, and assume a simple closed economy in which all income is either spent on consumption (C) or saved (S). Let us suppose that in this closed economy, marginal propensity to consume (MPC) is 0.9. This means that out of any addition to household income, 90% is consumed and 10% saved.

(a) If income goes up by $200, $180 would be spent on consumption, and $20 saved.
(b) Because of the circular flow, the $180 spent on consumption in turn increases the income of other people, who spend 90% of it ($162) and save $18.
(c) This $162 in turn becomes additional income to others. We can see that a snowball effect on consumption (and income and output) occurs, as follows:

<table>
<thead>
<tr>
<th>Stage</th>
<th>Income rises</th>
<th>90 per cent is consumed</th>
<th>A further 90 per cent is consumed</th>
<th>90 per cent of $162 is consumed</th>
<th>90 per cent is consumed</th>
<th>etc</th>
<th>Total increase in income</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>$200.00</td>
<td>$180.00</td>
<td>$162.00</td>
<td>$145.80</td>
<td>$131.22</td>
<td></td>
<td>$2 000.00</td>
</tr>
<tr>
<td>2</td>
<td>$180.00</td>
<td>$162.00</td>
<td>$145.80</td>
<td>$131.22</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>$162.00</td>
<td>$145.80</td>
<td>$131.22</td>
<td>$117.10</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>$145.80</td>
<td>$131.22</td>
<td>$117.10</td>
<td>$103.38</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>$131.22</td>
<td>$117.10</td>
<td>$103.38</td>
<td>$90.00</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

However, we have identified that a ‘snowball effect’ has been created, it will not continue indefinitely, because at each stage a proportion of the extra income is lost through withdrawals (savings). In this example, the initial increase in income of $200 will result in a final increase in national income of $2 000. The multiplier is 10.

3.1.1 The marginal propensity to save

In this very simple example, the multiplier is the reciprocal of the marginal propensity to save. Since MPC = 0.9, MPS = 0.1.

\[
\text{Multiplier} = \frac{1}{1 - \text{MPC}} = \frac{1}{0.1} = 10
\]

\[
\text{Increase in national income} = \frac{\text{Initial increase in expenditure}}{\text{MPS}} = \frac{\$200}{0.1} = 2 000
\]

Note that at the new equilibrium, savings of $200 equal the initial increase in expenditure of $200 but national income has risen $2 000.

If the marginal propensity to consume were 80%, the marginal propensity to save would be 20% and the multiplier would only be 5. Because people have saved more of their extra income than they would have done if MPS had been 10% the total increase in national income through extra consumption will be less.

3.1.2 The multiplier in the national economy

The multiplier in a national economy works in the same way. An initial increase in expenditure will have a snowball effect, leading to further and further expenditures in the economy. Since total expenditure in the economy is one way of measuring national income, it follows that an initial increase in
expenditure will cause an even larger increase in national income. The increase in national income will be a multiplier of the initial increase in spending, with the size of the multiple depending on factors which include the marginal propensity to save.

If you find this hard to visualise, think of an increase in government spending on the construction of roads. The government would spend money paying firms of road contractors, who in turn will purchase raw materials from suppliers, and sub-contract other work. All these firms employ workers who will receive wages that they can spend on goods and services of other firms. The new roads in turn might stimulate new economic activity, for example among road hauliers, housebuilders and estate agents.

Depending on the size of the multiplier, an increase in investment would therefore have repercussions throughout the economy, increasing the size of the national income by a multiple of the size of the original increase in investment.

If, for example, the national income were $10 000 million and the average and the marginal propensity to consume were both 75%, in equilibrium, ignoring G, T, X and M:

\[
\begin{align*}
Y &= \$10\ 000\ million \\
C &= \$7\ 500\ million \\
I &= S = \$2\ 500\ million
\end{align*}
\]

Since MPC = 75 per cent, MPS = 25 per cent, and the multiplier is 4.

An increase in investment of $1 000 million would upset the equilibrium, which would not be restored until the multiplier had taken effect, and national income increased by 4 × $1 000 million = $4 000 million, with:

\[
\begin{align*}
Y &= \$14\ 000\ million \\
C &= \$10\ 500\ million\ (75\%) \\
I &= S = \$3\ 500\ million\ (25\%)
\end{align*}
\]

A downward multiplier or *demultiplier* effect also exists. A reduction in investment will have repercussions throughout the economy, so that a small disinvestment (reduction in expenditure/output) will result in a multiplied reduction in national income.

### 3.1.3 Factors influencing MPC and MPS

You should also be aware of factors which might influence the MPC and MPS in any economy, for example: the age distribution of the population, the income distribution of the population, expectations of the future and other socio-economic factors.

### 3.2 The importance of the multiplier

The importance of the multiplier is that an increase in one of the components of aggregate demand will increase national income by more than the initial increase itself. Therefore, if the government takes any action to increase expenditure (for example, by raising government current expenditure, or lowering interest rates to raise investment) it will set off a general expansionary process, and the eventual rise in national income will exceed the initial increase in aggregate demand.

This can have important implications for a government when it is planning for growth in national income through a demand management policy, using levels of taxation and government expenditure to influence levels of aggregate demand in the economy. By an initial increase in expenditure, a government can ‘engineer’ an even greater increase in national income, (provided that the country's industries can increase their output capacity), depending on the size of the multiplier.

### 3.3 The multiplier in an open economy

So far we have been considering a simplified economy in which income is either saved or spent on domestic production, and in which there is no government intervention. The real world is more complex, though, and we must now consider the effect of taxation and imports. Like savings, these are *withdrawals from the circular flow* and they therefore affect the multiplier. Thus, in an open economy, the value of the multiplier depends on three things:

(a) The marginal propensity to save.
(b) The marginal propensity to import, because imports reduce national income, and if households spend much of their extra income on imports, the snowball increase in total national income will be restricted because imports are a withdrawal out of the circular flow of income. One of the reasons for a low multiplier in the UK is the high marginal propensity to import.

(c) The level of taxes, because taxes reduce people’s ability to consume and so are likely to affect the marginal propensity to consume and the marginal propensity to save.

Whereas the multiplier in a closed economy is the reciprocal of the marginal propensity to save, the multiplier in an open economy, taking into account government spending and taxation, and imports and exports, will be less. This is because government taxation and spending on imports reduces the multiplier effect on a country’s economy.

For an open economy:

\[ \text{Multiplier} = \frac{1}{s + m + t} \]

Where: 
- \( s \) is the marginal propensity to save
- \( m \) is the marginal propensity to import
- \( t \) is the marginal propensity to tax – i.e. the amount of any increase in income that will be paid in taxes

The multiplier as defined in this way may still be represented as below:

\[ \text{Multiplier} = \frac{1}{1 - \text{MPC}} \]

but this is now the same as \( \frac{1}{\text{MPW}} \) (reflecting the impact of withdrawals as a whole).

For example, if in a country the marginal propensity to save is 10 per cent, the marginal propensity to import is 45 per cent and the marginal propensity to tax is 25%, the size of the multiplier would be:

\[ \frac{1}{0.1 + 0.45 + 0.25} = \frac{1}{0.80} = 1.25 \]

### 3.4 Changes in equilibrium national income and the multiplier: a graphical representation

It is possible to show the multiplier effect in the form of a diagram.

In Figure 4, the horizontal axis represents national income (\( Y \)). The vertical axis represents planned or desired expenditure. The national economy is in equilibrium when actual output (which is the same as national income) is equal to desired expenditure. This occurs at any point along the 45 degree line \( Y = E \).

Onto this basic picture we have superimposed two other lines:

(a) The lower is the consumption function we described in Section 2. This consists of autonomous expenditure (\( a \)), which occurs even when income is zero, plus the proportion of income which is spent in accordance with the marginal propensity to consume (\( bY \)).

(b) Desired expenditure within the economy is not limited to consumption; we must also consider the injections, government spending (\( G \)), investment (\( I \)), and net exports (\( X - M \)). If we assume these are constant, total actual expenditure for any level of national income is shown by the upper line \( E = C + G + I + (X - M) \).
Figure 4 Equilibrium national income

Equilibrium national income in Figure 4 is at $Y_e$, where $Y = E = C + G + I + (X - M)$.

But what will happen if the economy has unemployed resources at national income $Y_e$, and total expenditure is increased – that is, if $C$, $G$, $I$ or $(X - M)$ increases? By how much will national income increase? This is shown in Figure 5.

Figure 5 Multiplier effect

If injections increased (for example, through an increase in government spending or extra exports) then there would be a shift upwards in the $E$ curve from $E_1$ to $E_2$. In Figure 5, the equilibrium level of income has now increased from $Y_1$ to $Y_2$. Note however that the increase in the level of income from $Y_1$ to $Y_2$ is **bigger than the increase in injections** ($E_1$ to $E_2$). In other words, total national income has increased by more than the amount of the initial increase in expenditure, and this is a portrayal of the multiplier effect.

### 3.5 Limitations of the multiplier

Keynes developed the concept of the multiplier in order to argue that extra government spending on public works, financed by a budget deficit, would have a stimulating effect on a demand-deficient economy.

(a) Demand would be increased, and national income would increase by more than the amount of the initial injection into the economy of the extra government spending.

(b) Because demand would be increased, unemployment would be reduced.
However, there are several important factors that limit the significance of the multiplier:

(a) It is relevant to a demand-deficient economy with high unemployment of resources. **If there is full employment, any increase in demand will lead to inflation** rather than a growth in the economy.

(b) The **leakages** from the circular flow of income might make the value of the multiplier very low. This is relevant to the UK, for example, where there is a high marginal propensity to import.

(c) There may be a **long period of adjustment** before the benefits of the multiplier are felt. If a government wants immediate action to improve the economy, relying on demand management and the multiplier could be too slow.

(d) The consumption function in advanced economies is probably more volatile than Keynes believed. If consumption is unpredictable, measures to influence national income through the multiplier will be impossible to predict too.

### 3.6 The accelerator principle

**Definition**

**Accelerator principle:** the theory that investment changes disproportionately in response to change in output.

The accelerator principle assumes that if there is a **small change** in the output of **consumer** goods, there will be a **much greater change** in the output of **capital** equipment required to make those consumer goods. This change in production of capital equipment (investment spending) speeds up the rate of economic growth, or slump.

A numerical example might help to illustrate this principle. Suppose that a firm makes biscuits and has 100 ovens in operation. The life of each oven is five years.

(a) If the demand for biscuits is constant, on average, 20 ovens must be replaced each year.

(b) If the demand for biscuits now rises by, say, 10 per cent the firm will need 110 ovens in operation. During the first year of the increase, the demand for ovens will be 30 units (instead of the 20 which it would have bought under its usual replacement cycle). This is made up of replacement of 20 ovens and an extra requirement of 10 ovens to bring the total to 110.

So a 10 per cent rise in demand for consumer goods results in a 50 per cent rise in demand for capital goods in the short term. This is an example of the accelerator at work! The accelerator principle indicates how, when the demand for consumer goods rises, there will be an even greater proportional increase in the demand for capital goods. This speeds up growth in national income.

(a) If demand for biscuits now remains constant at the new level, annual replacement of capital equipment will average 22. There is consequently the danger that there will be over-capacity in the oven-making industry because the short-term peak demand of 30 ovens per annum is not maintained.

(b) This means that unless the rate of increase in consumer demand is maintained, over-capacity in capital goods industries is likely to occur.

### 3.7 The accelerator in reverse

The accelerator also works in reverse. A decline in demand for consumer goods will result in a much sharper decline in demand for the capital goods which make them.

The accelerator implies that investment, and therefore national income, remain high only as long as consumption is rising.

So as income approaches the peak level dictated by available capacity, new investment will fall towards zero, reducing aggregate demand and hence national income. (The sharp fall in investment caused by the fall in
consumption, due to the accelerator effect, will be compounded by the ‘demultiplier’, so that the accelerator and the demultiplier will combine to reduce national income more severely than the initial fall in consumption. The recovery in investment when demand stops falling will stimulate the economy again and cause income and thus demand to rise again.)

Note carefully that the accelerator comes into effect as a consequence of changes in the rate of consumer demand.

The extent of the change in investment depends on two things:

(a) The size of the change in consumer demand.

(b) The capital-output ratio. This is the ratio of capital investment to the volume of output, in other words how much capital investment is needed to produce a quantity of output. For example, if the capital output ratio is 1:3, it would need capital investment of $1 to produce an extra $3 of output per year and so if demand went up by say, $3 billion, it would need an extra $1 billion of investment to produce the extra output to meet the demand. Accelerator theory assumes firms maintain a constant capital/output ratio.

Note: Accelerator theory also assumes that firms replace worn out capital each year (replacement investment).

Total investment therefore equals additional investment required to meet additional demand plus replacement investment.

4 Aggregate supply analysis

Section overview

• Aggregate supply determines the way aggregate demand influences inflation and unemployment.

So far in this chapter we have concentrated on issues related to aggregate demand, but it is important to remember that the equilibrium level of national income is determined by both aggregate supply and aggregate demand.

The aggregate supply curve shows the amount of goods and services that producers in the economy would supply at a price level. It expresses the ability of the economy to produce goods and services. The aggregate supply curve can take a number of shapes.

The aggregate supply curve will be flat if supply can increase without raising the price level. This could be the case when there are idle resources in the economy. However, the aggregate supply curve will usually be upward sloping, for the reasons applying to the microeconomic supply curves mentioned in earlier chapters. A higher price means that it is worthwhile for firms to hire more labour and produce more because of the higher revenue-earning capability. So at the macroeconomic level, an increasing price level implies that many firms will be receiving higher prices for their products and will increase their output.

In the economy as a whole, supply will at some point reach a labour constraint, when the entire labour force is employed. When there is full employment, and firms cannot find extra labour to hire, they cannot produce any more, even when prices rise, unless there is some technical progress in production methods. The aggregate supply curve will therefore rise vertically when the full employment level of output is reached.

5 The determination of national income

Section overview

• Equilibrium national income is determined using aggregate supply and aggregate demand analysis which is used to show the ideal situation of full employment.
5.1 Aggregate demand and supply equilibrium

Aggregate demand (AD) is the total planned or desired consumption demand in the economy for consumer goods and services and also for capital goods, no matter whether the buyers are households, firms or government. AD is a concept of fundamental importance in Keynesian economic analysis and shows the relationship between national income and prices. Keynes believed that national economy could be managed by taking measures to influence AD up or down.

The AD curve will be downward sloping because at higher prices, total quantities demanded will be less. Keynes argued that a national economy will reach equilibrium where the aggregate demand curve and aggregate supply curve intersect (AD = AS).

The value of the aggregate demand and supply model is that it allows us to show the impact of changes in AD or AS on output and price level; therefore (un)employment and inflation. We can assume that the level of employment in a country will be a function of the level of national income, so ‘output’ in the model will also be an indicator of employment.

Figure 6 Equilibrium national income, using aggregate supply and aggregate demand analysis

Figure 6 illustrates that the equilibrium level of national income will be at the intersection of the AD curve and AS curves – i.e. at Y. The difference between the equilibrium national income Y and the full employment national income Y_F shows how much national income could be increased with the resources at the economy’s disposal. Y therefore represents the level of satisfied demand in the economy. (Note that the aggregate demand function assumes constant prices.)

Two points follow on immediately from this initial analysis:

(a) Equilibrium national income Y might be at a level of national income below full employment national income Y_F.

(b) On the other hand, the AD curve might cut the AS curve above the point at which it becomes vertical, in which case the economy will be fully employed, but price levels will be higher than they need to be. There will be inflationary pressures in the economy.

5.2 Full-employment national income

If one aim of a country’s economic policy is to achieve full employment, then the ideal equilibrium level of national income will be where AD and AS are in balance at the full employment level of national income, without any inflationary gap – in other words, where aggregate demand at current price levels is exactly sufficient to encourage firms to produce at an output capacity where the country’s resources are fully employed. This is shown in Figure 7, where equilibrium output will be Y_F (full employment level) with price level P_F.
A country will usually continue to seek economic growth, but to achieve a real increase in living standards rather than simply to cause inflation, both AD and AS curves in Figure 7 will now have to shift to the right.

### 5.3 Inflationary gaps

In a situation where resources are already fully employed, there may be an inflationary gap, since increases in aggregate demand will cause price changes and not variations in real output. An inflationary gap can be described as the extent to which the aggregate demand function would have to shift downward to produce the full employment level of national income without inflation.

You should also note that a shift in the AD curve or the AS curve will not only change the national income, it will also change price levels (P). In Figure 8, an inflationary gap can be removed by shifting the aggregate demand curve to the left, from $AD_1$ to $AD_2$.

If you are not sure about this point, a simple numerical example might help to explain it better. Suppose that in Ruritania there is full employment and all other economic resources are fully employed. The country produces 1 000 units of output with these resources. Total expenditure (that is, aggregate demand) in the economy is 100 000 Ruritanian dollars, or 100 dollars per unit. The country does not have any external trade, and so it cannot obtain extra goods by importing them. Because of pay rises and easier credit terms for consumers, total expenditure now rises to 120 000 Ruritanian dollars. The economy is fully employed,
and cannot produce more than 1,000 units. If expenditure rises by 20 per cent, to buy the same number of units, it follows that prices must rise by 20 per cent too. In other words, when an economy is at full employment, any increase in aggregate demand will result in price inflation.

The Keynesian argument is that if a country’s economy is going to move from one equilibrium to a different equilibrium, there needs to be a shift in the aggregate demand curve. To achieve equilibrium at the full employment level of national income, it may therefore be necessary to shift the AD curve to the right (upward) or the left (downwards).

### 5.4 Deflationary gap

In a situation where there is unemployment of resources, there is said to be a deflationary gap (Figure 9). Prices are fairly constant and real output changes as aggregate demand varies. A deflationary gap can be described as the extent to which the aggregate demand function will have to shift upward to produce the full employment level of national income. The economy is currently in equilibrium at $Y_e$ (with aggregate demand $AD_1$) but it would need to be at $Y_f$ (with aggregate demand $AD_2$) to achieve full employment. This is shown in Figure 9 below.

![Figure 9 Deflationary gap](image)

In Figures 8 and 9 we showed how changes in aggregate demand could affect the economy. However, there may also be changes in aggregate supply. For example, a major rise in raw material prices or labour costs could lead to a reduction in supplies, as could a fall in productivity due to technological problems.

### 5.5 Stagflation

In the 1970s there was a problem with stagflation: a combination of unacceptably high unemployment and unacceptably high inflation. One of the causes was diagnosed as the major rises in the price of crude oil that took place. The cost of energy rose and this had the effect of rendering some production unprofitable. The supply curve shifted to the left as a result, leading to a rise in prices and a contraction in output. The way changes in aggregate supply lead to stagnation is illustrated in Figure 10.
Determining national income

Figure 10 Stagflation

National income falls and both prices and unemployment rise. Any long term major increase in costs (a price shock) is likely to have this effect.

Conversely, there could be situations where the supply curve shifts outwards. For example, deliberate supply-side government policies such as tax reductions or subsidies would lead to an expansion of output and lower prices.

Equally, the supply curve could be shifted to the right following improvements as a result of technological change, or due to falls in the cost of inputs such as energy and raw materials.

5.6 The IS-LM curve

An alternative method of presenting national income is to show IS and LM curves (standing for 'Investment Savings' and 'Liquidity preference money supply') on a graph.

Figure 11 The IS-LM model

The IS curve shows all points of equilibrium in the real (i.e., non-financial) economy. Low interest rates will encourage investment and lead to high national income. High interest rates will depress investment and lead to low national income.
The LM curve shows all points of equilibrium in the money market, i.e., combinations of interest rates and levels of income for which money demand equals money supply, so that liquidity preference & money supply (hence the name of the curve).

The intersection point of the IS curve and the LM curve shows the unique point in the economy at which both the non-financial sector and the financial sector will be in equilibrium; it represents the 'general equilibrium' point at which the interest rate in the economy and the level of national income are in equilibrium.

Shifts of either the IS curve on the LM curve will lead to a new general equilibrium point in the economy. For example, a government might decide to reduce taxes; this will increase consumer spending at each interest rate and cause the IS curve to shift to the right.

![Figure 12 Shift in IS curve](image-url)

The effect is to raise national income (from $Y_1$ to $Y_2$) but also to raise the interest rate in the economy (from $i_1$ to $i_2$). This illustrates a major problem associated with cutting taxes to stimulate the economy: rising interest rates will **crowd out**, i.e., discourage, private sector investment, as described earlier.

In conclusion, the IS-LM model of the economy shows the relationships between interest rates and national income. It is a simple model that can be useful in explaining short run effects in an economy. However, it assumes a fixed level of prices in the basic model, so it is not valid in the long run as price levels change over time.
Key chapter points

- The Keynesian model provides a way of explaining how national income is determined, and how national income equilibrium is reached.
- Consumption expenditure depends on income. It might be possible for a government to take measures to boost aggregate demand in the economy, although some price inflation will probably result. When there is inflation in the economy, measures could be taken to suppress aggregate demand.
- Autonomous consumption is the basic level of consumption by a household irrespective of income. Non-autonomous consumption is where the amount spent depends on the level of income. The proportion of income spent on non-autonomous consumption is the marginal propensity to consume (MPC) and the proportion saved is the marginal propensity to save (MPS).
- Changes in national income begin with a small change in expenditure, leading to a larger eventual change in national income, due to the multiplier effect.
- Changes in national income can be explained partially by the accelerator principle, which is the mechanism by which changes in investment spending are proportionally greater than changes in consumption spending, and therefore investment spending is more susceptible to bigger upturns and downturns.
- Aggregate supply determines the way aggregate demand influences inflation and unemployment.
- Equilibrium national income is determined using aggregate supply and aggregate demand analysis which is used to show the ideal situation of full employment.
Quick revision questions

1. If total national expenditure is 'E', what is the equation used to show how 'E' is calculated?

2. How might a government try to influence the volume of investment by firms?

3. Injections into the economy are:
   A. consumption and investment.
   B. investment and government expenditure.
   C. investment, government expenditure and export demand.
   D. consumption, investment, government expenditure and export demand.

4. If a consumption function has the formula C = 750 + 0.4Y where Y is the change in national income, and injections are 500, then equilibrium national income will be at:
   A. 833
   B. 1 250
   C. 2 083
   D. 3 125

5. If the MPC is greater for the poor than the rich then a redistribution of national income in favour of the rich will:
   A. raise savings out of a given income.
   B. increase the multiplier.
   C. decrease the MPS.
   D. stimulate import demand.

6. A deflationary gap occurs when:
   A. a government is cutting its level of expenditure.
   B. a government attempts to spend its way out of recession.
   C. aggregate demand is more than sufficient to buy up all the goods and services produced by an economy.
   D. aggregate demand is insufficient to buy up all the goods and services the company is capable of producing.

7. In an aggregate demand and supply diagram, what would be the consequences if the aggregate supply curve shifted inwards:
   A. prices would rise and national income would rise
   B. prices would fall and national income would rise
   C. prices would fall and national income would fall
   D. prices would rise and national income would fall

8. In an economy, the marginal propensity to consume is 0.85. What is the multiplier in that economy?
1. \[ E = C + I + G + (X - M) \]

2. Lower interest rates, investment grants and tax incentives may encourage investment. Governments can also stimulate demand by tax cuts or lower interest rates and improve business confidence by business friendly and growth enhancing policies like deregulation and controlling inflation. Policies to encourage technological development may also lead to increased investment.

3. **C** Injections are any additional expenditures which do not arise from the circular flow of income itself. Consumption is part of the circular flow so it is not an injection.

4. **C** Equilibrium occurs when \( E = Y \). \[ E = C + J = 750 + 0.4Y + 500 \]. Therefore \( 0.6Y = 1250 \), therefore \( Y = 2083 \).

5. **A** The rich will save more, not spend. The MPC is greater for the poor than the rich.

6. **D** A deflationary gap occurs when the aggregate demand in the economy is insufficient to keep all the resources in the economy fully employed, therefore D is correct.

   Options A and B could lead to deflation or inflation depending on the levels of the other variables in the economy.

   Option C describes ‘too much money chasing too few goods’ which produces an inflationary gap leading to inflation.

7. **D** If the aggregate supply curve shifts to the left, national income will fall. Because the aggregate demand curve is downward sloping, shifting the supply curve to the left will mean the intersection between supply and demand is at a higher price: prices will rise.

   This combination of rising prices and falling national income is characteristic of stagflation.

8. **6.67** The multiplier is \[ \frac{1}{1 - \text{MPC}} \]

\[ \frac{1}{1 - 0.85} = \frac{1}{0.15} = 6.67 \]
Answers to chapter questions

1. The amount that people save will depend on:
   (a) how much income they are getting, and how much of this they want to spend on consumption.
   (b) how much income they want to save for precautionary reasons (for example, unexpected bills) for the future.
   (c) interest rates. If the interest rate goes up we would expect people to consume less of their income, and to be willing to save and invest more.

2. \[
\frac{\text{Change in consumption}}{\text{Change in income}} = \frac{40}{50} = 0.8
\]

3. (a) \[
\text{APC} = \frac{740}{800} = 92.5\%
\]
   (b) \[
\text{APC} = \frac{900}{1000} = 90\%
\]

Note: If MPC < APC an increase in income will lead to a reduction in APC. But if MPC > APC an increase in income will lead to an increase in APC.
Chapter 9

Macroeconomic concepts – inflation and unemployment

<table>
<thead>
<tr>
<th>Learning objectives</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Macroeconomic concepts – inflation and unemployment</td>
<td>LO9</td>
</tr>
<tr>
<td>Describe different types of unemployment</td>
<td>LO9.1</td>
</tr>
<tr>
<td>Describe the causes of inflation and its impact on an economy</td>
<td>LO9.2</td>
</tr>
<tr>
<td>Explain the relationship between rates of employment and the performance of an economy</td>
<td>LO9.3</td>
</tr>
<tr>
<td>Prepare a Phillips curve</td>
<td>LO9.3.1</td>
</tr>
<tr>
<td>Define money</td>
<td>LO9.4</td>
</tr>
<tr>
<td>Explain the structure of interest rates</td>
<td>LO9.5</td>
</tr>
<tr>
<td>Analyse the factors affecting the movement of interest rates</td>
<td>LO9.6</td>
</tr>
<tr>
<td>Explain the Keynesian and Classical theories of money</td>
<td>LO9.7</td>
</tr>
</tbody>
</table>

Topic list

1. Unemployment
2. Inflation
3. Causes of inflation
4. Unemployment and inflation
5. Money
6. Measuring the money supply
7. Credit and interest rates
8. Monetary theory
Two of the key aims of macroeconomic policy are controlling price inflation and minimising the level of unemployment in a country.

The first section of this chapter introduces the characteristics and consequences of inflation and unemployment, and then looks at how they can be managed. Not surprisingly, there are differing views about the causes of inflation, the extent to which inflation creates unemployment and prevents economic growth, and the effectiveness of government measures to stimulate the economy.

The second section of the chapter introduces the concept of money, credit and interest rates. The final section presents monetary theory which is essential knowledge for all macroeconomic studies.
Before you begin

If you have studied these topics before, you may wonder whether you need to study this chapter in full. If this is the case, please attempt the questions below, which cover some of the key subjects in the area.

If you answer all these questions successfully, you probably have a reasonably detailed knowledge of the subject matter, but you should still skim through the chapter to ensure that you are familiar with everything covered.

There are references in brackets indicating where in the chapter you can find the information, and you will also find a commentary at the back of the Study Manual.

1. What is the calculation of the rate of unemployment in an economy? (Section 1.1)
2. List three types of unemployment. (Section 1.3)
3. Define full employment. (Sections 1.6, 4.2)
4. What is the underlying rate of inflation? (Section 2.3)
5. Distinguish between 'narrow' money and 'broad' money. (Section 6.1)
6. List three factors that affect the movement of interest rates. (Section 7.4)
7. What is the Classical theory of money? (Section 8.1)
8. What is the Keynesian theory of money? (Section 8.2)
1 Unemployment

Section overview

- The monetarist concept of a stable equilibrium implies that with zero price inflation, there is a natural optimal level of unemployment, and a rate of economic growth and balance of trade position from which the economy will not deviate. Monetarism focuses on economic stability in the medium to long term, which can only be achieved by abandoning short-term demand management goals.

- The consequences of unemployment for the national economy are loss of output, loss of human capital, an increase in income inequalities in the population, social issues such as theft and suffering and increased social security payments.

- Types of unemployment are frictional, seasonal, technological, cyclical, real wage unemployment and voluntary unemployment.

- The causes of unemployment are manifold and can be summarised into three reasons – demand-deficient, structural change, and supply side unemployment.

- According to Keynes, full employment is when the country’s economic resources are fully employed and a percentage of the population may be unemployed at any time.

1.1 The rate of unemployment

The rate of unemployment in an economy can be calculated as:

\[ \text{Rate of unemployment} = \frac{\text{Number of unemployed}}{\text{Total workforce}} \times 100\% \]

The number of unemployed at any time is measured by government statistics. If the flow of workers through unemployment is constant then the size of the unemployed labour force will also be constant.

Flows into unemployment are:

(a) Members of the working labour force becoming unemployed:
   - Redundancies.
   - Lay-offs.
   - Voluntary quitting from a job.

(b) People out of the labour force joining the unemployed:
   - School leavers or university graduates without a job.
   - Others (for example, carers) rejoining the workforce but having no job yet.

Flows out of unemployment are:

- Unemployed people finding jobs.
- Laid-off workers being re-employed.
- Unemployed people stopping the search for work.

In Australia, the seasonally adjusted unemployment statistics are published by the Australian Bureau of Statistics (ABS). Australia uses the international definition of unemployment for its statistics – a person is unemployed if they did not work for at least one paid hour of work in the previous week, if they are actively seeking work, and if they are able to accept a job in the next week were one to become available. The statistics are conducted via telephone survey.

In the UK, the monthly unemployment statistics are published by the UK’s Office for National Statistics (ONS) and count only the jobless who receive benefits. The ONS also produces figures based on a quarterly survey of the labour force known as the International Labour Organisation measure (ILO measure) that provides seasonally adjusted monthly data. This figure is considered to be more useful because it is also an internationally comparable measure.
1.2 Consequences of unemployment

Unemployment results in the following problems in the performance of the national economy:

(a) **Loss of output.** If labour is unemployed, the economy is not producing as much output as it could. Thus, total national income is less than it could be, because economic resources are not being fully used.

(b) **Loss of human capital.** If there is unemployment, the unemployed labour will gradually lose its skills, because skills can only be maintained by working.

(c) **Increasing inequalities in the distribution of income.** Unemployed people earn less than employed people, and so when unemployment is increasing, the poor get poorer.

(d) **Social costs.** Unemployment brings social problems of personal suffering and distress, and possibly also increases in crime such as theft and vandalism.

(e) **Increased burden of social security payments.** This can have a major impact on government fiscal policy, because governments will have to pay out more in State benefits while collecting less through tax revenue.

1.3 Types of unemployment

Unemployment may be classified into several categories depending on the underlying causes:

<table>
<thead>
<tr>
<th>Category</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Frictional unemployment</td>
<td>It is inevitable that some unemployment is caused not because there are not enough jobs to go round, but because of the friction in the labour market (difficulty in quickly matching workers with jobs), caused perhaps by a lack of knowledge about job opportunities. These are imperfections in the labour market. In general, it takes time to match prospective employees with employers, and individuals will be unemployed during the search period for a new job. Frictional unemployment is temporary, lasting for the period of transition from one job to the next.</td>
</tr>
<tr>
<td>Seasonal</td>
<td>This occurs in certain industries, for example building, tourism and farming, where the demand for labour fluctuates in seasonal patterns throughout the year and where staff are often employed on temporary contracts.</td>
</tr>
<tr>
<td>Structural</td>
<td>This occurs where long-term changes occur in the conditions of an industry. It is likely to result from either a long-term fall in demand for the good or service, or from changes in production methods which mean that labour-intensive production is replaced by capital-intensive production (technology). A feature of structural unemployment is high regional unemployment in the location of the industry affected.</td>
</tr>
<tr>
<td>Technological</td>
<td>This is a form of structural unemployment, which occurs when new technologies are introduced.</td>
</tr>
<tr>
<td></td>
<td>(a) Old skills are no longer required.</td>
</tr>
<tr>
<td></td>
<td>(b) There is likely to be a labour saving aspect, with machines doing the job that people used to do.</td>
</tr>
<tr>
<td></td>
<td>With automation, employment levels in an industry can fall sharply, even when the industry’s total output is increasing.</td>
</tr>
<tr>
<td>Cyclical or demand-deficient</td>
<td>Past experience has shown that domestic and foreign trade go through cycles of boom, decline, recession, recovery, then boom again, and the cycle continues.</td>
</tr>
<tr>
<td></td>
<td>(a) During recovery and boom years, the demand for output and jobs is high, and unemployment is low.</td>
</tr>
<tr>
<td></td>
<td>(b) During decline and recession years, the demand for output and jobs falls, and unemployment rises to a high level. Keynes introduced the concept of ‘demand-deficient’ unemployment to illustrate that unemployment was a result of insufficient aggregate demand in the economy.</td>
</tr>
<tr>
<td></td>
<td>Cyclical unemployment can be long-term, and a government might try to reduce it by doing what it can to minimise a recession or to encourage faster economic growth.</td>
</tr>
</tbody>
</table>
### Causes of unemployment

The **causes of unemployment** can be summarised as follows:

<table>
<thead>
<tr>
<th>Category</th>
<th>Comments</th>
</tr>
</thead>
</table>
| **Demand-deficient**         | A fall in demand for goods and services means that the equilibrium level of output below the full employment position.  
Th e fall in aggregate demand could come from a fall in consumer expenditure; a fall in business investment (which will reduce aggregate demand); a decline in exports or a fall in government expenditure.  
In the 2008/09 recession, many governments have been faced with demand-deficient unemployment. As a result, they have adopted expansionary fiscal and monetary policies (e.g. lowering interest rates) to try to boost aggregate demand and reduce the extent of demand-deficient unemployment. |
| **Structural change**        | The demand for labour falls as a result of a significant change in the structure of an industry.  
This may result from:  
• Technological change (with machines and computers replacing people).  
• Shifts in the industry (for example, the decline in manufacturing industry in developed countries as it was relocated to cheaper, developing countries).  
Note that service industries have now replaced manufacturing industries as the major source of employment in developed countries. This is indicative of structural change. New industries replace the declining ones, and provide new employment opportunities – provided the labour force have the skills required. |
| **Supply side problems**     | In these situations, the aggregate supply curve shifts to the left, creating an equilibrium level of national income below the level of full employment.  
This shift could come from trade unions demanding higher wages for their members, labour market regulations such as minimum wage rates, or social security systems which allow workers to be better off by not working than working.  
Workers may also effectively be excluded from the labour market due to poor education and a lack of skills.  
It is possible that a rise in the cost of imported goods (particularly oil) could also lead to supply side unemployment. A steep rise in oil prices would raise production costs across an economy, and so could cause the aggregate supply curve to shift to the left. This leftward shift will lead to an increase in unemployment. |

**Exam comments**

You need to be familiar with all the different types of unemployment.
1.5 Government employment policies

Job creation and reducing unemployment should often mean the same thing, but it is possible to create more jobs without reducing unemployment.

(a) This can happen when there are a greater number of people entering the jobs market than there are new jobs being created. For example, if 500,000 new jobs are created during the course of one year, but 750,000 extra school leavers are looking for jobs, there will be an increase in unemployment of 250,000.

(b) It is also possible to reduce the official unemployment figures without creating jobs. For example, individuals who enrol for a government financed training scheme are taken off the unemployment register, even though they do not have full-time jobs.

A government can try several options to create jobs or reduce unemployment:

(a) **Spending more money directly on jobs** (for example, hiring more public servants).

(b) **Encouraging growth** in the private sector of the economy. When aggregate demand is growing, firms will probably want to increase output to meet demand, and so will hire more labour.

(c) **Encouraging training in job skills**. There might be a high level of unemployment among unskilled workers, and at the same time a shortage of skilled workers. A government can help to finance training schemes, in order to provide a 'pool' of workers who have the skills that firms need and will pay for.

(d) **Offering grant assistance to employers** in key regional areas.

(e) **Encouraging labour mobility** by offering individuals financial assistance with relocation expenses, and improving the flow of information on vacancies.

Other policies may be directed at **reducing real wages to market clearing levels**:

(a) Abolishing 'closed shop' agreements, which restrict certain jobs to trade union members.

(b) Reviewing **minimum wage regulations**, to assess whether the level set for the minimum wage is preventing employers taking on new staff.

1.6 The Keynesian approach


Pre-Keynesian economists had tried to explain unemployment as a temporary phenomenon. They believed that if there is a surplus of labour available (unemployment) then the forces of demand and supply, through the wages (price) mechanism, would restore equilibrium by bringing down wage levels, thus stimulating demand for labour. Any unemployment would only last as long as the labour market was adjusting to new equilibrium conditions. The pre-Keynesian theory was challenged during the 1930s. If pre-Keynesian theory was right, wages should have fallen and full employment should have been restored. However, this did not happen, and the depression continued for a long time.

It is instructive to note that it was during this economic situation that Keynes put forward his new theory. Its fundamental advance on earlier theory was to explain how **equilibrium could exist in the macroeconomy, but there could still be persistent unemployment and slow growth**.

The term **full employment national income** is used to describe the total national income that a country must earn in order to achieve full employment. By **full employment** we mean that the country’s economic resources are fully employed. However, as far as labour is concerned, full employment does not mean that everyone has a job all the time. There will always be some **normal or transitional** unemployment as people lose their job or give up one job for another, and so full employment might mean, say, that three to five per cent of the total working population is unemployed at any time.
Question 1: Types of unemployment

Match the terms (1), (2) and (3) below with definitions A, B and C.

(1) Structural unemployment
(2) Cyclical unemployment
(3) Frictional unemployment

A unemployment arising from a temporary difficulty in matching unemployed workers with available jobs
B unemployment occurring in the downswing of an economy in between two booms
C unemployment arising from a long-term decline in a particular industry

(The answer is at the end of the chapter)

2 Inflation

Section overview

- High rates of inflation are harmful to an economy. Inflation redistributes income and wealth. Uncertainty about the value of money makes business planning more difficult. Constantly changing prices impose extra costs.
- Consumer price indices measure inflation and the underlying rate of inflation is the CPI adjusted to exclude mortgage costs.

Definition

Inflation is the name given to an increase in price levels generally. It is also manifest in the decline in the purchasing power of money.

Historically, there have been very few periods when inflation has not been present in national economics. We discuss below why high rates of inflation are considered to be harmful. However, it is important to remember that deflation (falling prices) is normally associated with low rates of growth and even recession. It would seem that a healthy economy may require some inflation. This is recognised in the Australian current target rate of between 2 to 3 per cent, the UK inflation target of 2 per cent, and the European Central Bank’s target of 2 per cent. Certainly, if an economy is to grow, the money supply must expand, and the presence of a low level of inflation will ensure that growth is not hampered by a shortage of liquid funds.

2.1 Why is inflation a problem?

An economic policy objective which now has a central place in the policy approaches of the governments of many developed countries is that of stable prices. But why is a high rate of price inflation harmful and undesirable?

(a) Redistribution of income and wealth

Inflation leads to a redistribution of income and wealth in ways which may be undesirable. Redistribution of wealth might take place from suppliers to customers. This is because amounts payable or receivable lose ‘real’ value with inflation. For example, if you owed $1 000, and prices then doubled, you would still owe $1 000, but the real value of your debt would have been halved. In general, in times of inflation those with economic power tend to gain at the expense of the weak, particularly those on fixed incomes. Their nominal income will stay the same but the amount of goods and services they can buy with that income (its purchasing power) will fall.
(b) **Balance of payments effects**

If a country has a higher rate of inflation than its major trading partners, its exports will become relatively expensive and imports into it will be relatively cheap. As a result, the balance of trade will suffer, affecting employment in exporting industries and in industries producing import-substitutes. Eventually, the exchange rate will be affected.

(c) **Uncertainty of the value of money and prices**

Under perfect competition, prices convey correct signals regarding the efficient allocation of resources. However when there is uncertainty over inflation, prices do not necessarily convey the correct signals.

As a result, no one has certain knowledge of the value of money or of the real meaning of prices. If the rate of inflation becomes excessive, and there is 'hyperinflation', this problem becomes so exaggerated that money becomes worthless, so that people are unwilling to use it and are forced to resort to barter. In less extreme circumstances, the results are less dramatic, but the same problem exists. As prices convey less information, the process of resource allocation is less efficient and rational decision-making become much harder. Uncertainty about prices may also undermine business confidence because planning and forecasting may become less accurate.

(d) **Wage bargaining**

Wage demands (particularly from trades' unions) will be increased in times of high inflation. If they are successful then a wage/price spiral will take hold, which will reinforce the problem.

(e) **Consumer behaviour**

People may stockpile goods fearing price increases later. This could create shortages for other people who haven’t already stockpiled themselves.

### 2.2 Consumer price indices

We have already referred to the way in which inflation erodes the real value of money. In order to measure changes in the real value of money as a single figure, we need to group all goods and services into a single price index.

A consumer price index is based on a chosen 'basket' of items which consumers purchase. A weighting is decided for each item according to the average spending on the item by consumers.

Consumer price indices may be used for several purposes, for example as an indicator of inflationary pressures in the economy, as a benchmark for wage negotiations, and to determine annual increases in government benefits payments. Countries commonly have more than one consumer price index because one composite index may be considered too wide a grouping for different purposes.

Australia has a single consumer price index (the **Australian CPI**) which is calculated and published quarterly by the Australian Bureau of Statistics (ABS). It is designed to give a general measure of inflation for the private household sector as a whole. There are eleven 'baskets' of commodities used in the CPI namely food, alcohol and tobacco, clothing and footwear, housing, household contents and services, health, transportation, communication, recreation, education and financial and insurance services.

The UK is an example of a country where more than one consumer price index is used. One important measure of the general rate of inflation in the UK used over many years has been the **Retail Prices Index (RPI)**. The RPI measures the percentage changes month by month in the average level of prices of the commodities and services, including housing costs, purchased by the great majority of households in the UK. The items of expenditure within the RPI are intended to be a representative list of items, current prices for which are collected at regular intervals. The UK also uses the standardised European measure of inflation, sometimes called the Harmonised Index of Consumer Prices (HICP) as the basis for the UK’s inflation target. The UK HICP is called the **Consumer Prices Index (CPI)**. The CPI excludes most housing costs.

There are some core biases to consumer price indices. The calculation of the CPI does not allow for consumer substitution which occurs when consumers switch from goods which have become relatively more expensive to those which are relatively less expensive. The index also ignores the introduction of new goods which act as a substitute and quality changes in existing goods. It also discounts 'outlet bias' where consumers show a change in shopping patterns from expensive retailers to low-cost, bulk retailers (for example, a shift from Myer to K-mart for the same consumer good).
2.3 The underlying rate of inflation

The term underlying rate of inflation is usually used to refer to the CPI (in Australia) and the RPI (in the UK) adjusted to exclude mortgage costs and often elements as well. For example, the Australian underlying rate of inflation excludes 49% of the CPI due to seasonality, volatility or policy. Exclusions include clothing, education and local government rates. The effects of interest rate changes on mortgage costs (as well as other changes due to season, volatility or policy) help to make the CPI fluctuate more widely than the underlying rate of inflation.

3 Causes of inflation

<table>
<thead>
<tr>
<th>Section overview</th>
</tr>
</thead>
<tbody>
<tr>
<td>There are three causes of inflation – cost push inflation, demand pull inflation and the quantity theory of money.</td>
</tr>
<tr>
<td>Cost push inflation arises from increases in the costs of production. Demand pull inflation arises from an excess of aggregate demand over the productive capacity of the economy. The quantity theory argues that inflation is caused by an increase in the supply of money (the 'money supply').</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>LO 9.2</th>
</tr>
</thead>
<tbody>
<tr>
<td>There are three main causes of inflation:</td>
</tr>
<tr>
<td>- Demand pull factors.</td>
</tr>
<tr>
<td>- Cost push factors.</td>
</tr>
<tr>
<td>- Excessive growth in the money supply (quantity theory of money).</td>
</tr>
</tbody>
</table>

3.1 Demand pull inflation

**Definition**

Demand pull inflation: inflation resulting from a persistent excess of aggregate demand over aggregate supply, due to constraints on supply in the economy, for example full employment levels have been reached.

Demand pull inflation occurs when the economy is buoyant and there is a high aggregate demand, in excess of the economy’s ability to supply.

(a) Because aggregate demand exceeds supply, prices rise.

(b) Since supply needs to be raised to meet the higher demand, there will be an increase in demand for factors of production, and so factor rewards (wages, interest rates, and so on) will also rise.

(c) Since aggregate demand exceeds the output capability of the economy, it should follow that demand pull inflation can only exist when unemployment is low.

There are two main causes of demand pull inflation:

(a) **Fiscal.** Government policies affect aggregate demand in an economy. For example, an increase in government spending or a reduction in taxes and interest rates will raise demand in the economy. However, if supply is relatively inelastic and expands to meet the increased demand, the result will be higher prices rather than the economic growth the government had intended.

(b) **Credit.** If levels of credit extended to customers increase, expenditure is likely to follow suit. In this case, inflation is likely to be accompanied by customers increasing their debt burdens.

Traditionally Keynesian economists saw inflation as being caused by Demand pull factors. However, they now accept that Cost push factors are involved as well.
3.2 Cost push inflation

Cost push inflation occurs where the costs of factors of production rise regardless of whether or not they are in short supply, and where the rise in costs is not matched by an increase in productivity. This appears to be particularly the case with wages.

Definition

Cost push inflation: inflation resulting from an increase in the costs of production of goods and services, for example, through escalating prices of imported raw materials or from wage increases.

In this context it is important to distinguish between exogenous and endogenous factors. Exogenous factors are things that cause inflation without being related to the level of aggregate demand in the economy. Endogenous factors are directly related to the level of aggregate demand in the economy. Therefore, if wages rise even though the demand for labour has not increased, the wage rises on an exogenous factor which causes cost plus inflation.

3.2.1 Import cost factors

Import cost push inflation occurs when the cost of essential imports rise regardless of whether or not they are in short supply. This has occurred in the past with the oil price rises of the 1970s. Additionally, a fall in the value of a country’s currency will have import cost push effects since a weakening currency increases the price of imports. This is another example of exogenous cost-push inflation, because the cost rise comes from outside the AD and AS system.

3.3 Expectations and inflation

A further problem is that once the rate of inflation has begun to increase, a serious danger of expectational inflation will occur. This means, regardless of whether the factors that have caused inflation are still persistent or not, there will be a perception of what inflation is likely to be, and so, to protect future income, wages and prices will be raised by the expected amount of future inflation. This can lead to the vicious circle known as the wage-price spiral, in which inflation becomes a relatively permanent feature, because of people’s expectations that it will occur.

3.4 Money supply growth

In contrast to the cost push theory, monetarists argue that inflation is not driven by cost factors, but rather it is caused by increases in the supply of money which lead in turn to excess demand for goods and services. There is a considerable debate as to whether increases in the money supply are a cause of inflation or whether increases in the money supply are a symptom of inflation. Monetarists have argued that since inflation is caused by an increase in the money supply, inflation can be brought under control by reducing the rate of growth of the money supply.

Definition

Monetarism as a theory advocates that the level of national income is determined by the quantity of money in circulation (the quantity theory of money). This is discussed in greater detail in Section 8 of this chapter.
4 Unemployment and inflation

Section overview

- There appears to be a connection between the rate of inflation and unemployment.
- The Phillips curve has been used to show the relationship between wage inflation and unemployment.
- The expectations augmented Phillips curve shows that they may be a short-term trade-off between inflation and unemployment but in the long-run the economy will have a vertical Phillips curve.

Managing unemployment and inflation are two of the key aspects that government try to manage, and the two are often thought to be linked. It has been found that boosting demand to increase the level of employment can cause a higher rate of inflation. However, growth in unemployment can also be associated with a rising rate of inflation.

As discussed in Section 1.6 above the term full employment does not mean a situation in which everyone has a job. There will always be at least a certain natural rate of unemployment, which is the minimum level of unemployment that an economy can expect to achieve.

An aim of government policy might be to reduce unemployment to this minimum natural rate, and so get as close as possible to the goal of full employment. On the basis that unemployment cannot be kept below its natural rate without causing inflation, the natural rate of unemployment is sometimes called the non-accelerating inflation rate of unemployment (NAIRU). But in order to understand the idea of a natural rate of unemployment more fully, we need to examine the trade-off between unemployment and inflation more closely.

4.1 Inflationary gaps and deflationary gaps

As we saw when looking at national income in Chapters 7 and 8, the equilibrium national income can be shown using an aggregate demand curve and an aggregate supply curve.

According to Keynes, demand management by the government could be based on government spending and taxation policies (fiscal policy). These could be used for two purposes:

(a) To eliminate a deflationary gap and create full employment. A small initial increase in government spending will start off a multiplier-accelerator effect, and so the actual government spending required to eliminate a deflationary gap should be less than the size of the gap itself.

(b) To eliminate an inflationary gap and take inflation out of the economy. This can be done by reducing government spending, or by increasing total taxation and not spending the taxes raised.

Keynesians accept that reductions in unemployment can only be achieved if prices are allowed to rise: reducing unemployment goes hand in hand with allowing some inflation.

4.2 The Phillips curve

In 1958, A W Phillips found a statistical relationship between unemployment and the rate of money wage inflation which implied that, in general, the rate of inflation falls unemploymment rose and vice versa. A curve, known as a Phillips curve, can be drawn linking inflation and unemployment (Figure 1).

**Definition**

Phillips curve: a graphical illustration of the inverse relationship which historically existed between the rate of wage inflation and the rate of unemployment.
Note the following three points about the Phillips curve:

(a) The curve crosses the horizontal axis at a positive value for the unemployment rate. This means that zero inflation will be associated with some unemployment; it is not possible to achieve zero inflation and zero unemployment at the same time.

(b) The shape of the curve means that the lower the level of unemployment, the higher the rate of increase in inflation.

(c) The curve essentially identifies a trade-off between inflation and unemployment. This trade-off is more binding at higher inflation rates (lower unemployment) due to capacity constraints. At a point of higher unemployment there is available capacity, showing that expanding employment puts less pressure on prices.

The existence of a relationship between inflation and unemployment of the type indicated by the Phillips curve suggests that the government should be able to use demand management policies to take the economy to acceptable levels of inflation and unemployment.

This re-emphasises the argument of Keynesian economists that in order to achieve full employment, some inflation is unavoidable. If achieving full employment is an economic policy objective, a government must therefore be prepared to accept a certain level of inflation as a necessary evil.

4.3 The expectations augmented Phillips curve and NAIRU

The monetarist economist, Milton Friedman, developed an alternative approach to the Phillips curve to reflect the fact that expectations distort the inflationary process.

Friedman’s model is known as the expectations augmented Phillips curve, and indicates that, although in the short run there may be trade-offs between inflation and unemployment, in the long run an economy is faced with a vertical Phillips curve, and there are inflationary expectations which reflect the rates of inflation that workers expect in the future.

Figure 2 illustrates why. Each of the short run Phillips curves shown is for a given expectation of inflation.

The economy is initially in equilibrium with unemployment of U, and with very low inflation (PE). The government tries to boost aggregate demand in the economy, and this reduces unemployment to U. However, it also creates excess demand in the labour market which prompts wage inflation.
This wage inflation in turn becomes price inflation, so workers are no better off in real terms than they were before the wage rise. In this case, the labour supply and hence unemployment returns to its previous level: \( U \). Therefore the economy has found a new equilibrium with higher inflation (\( PE_1 \)) and the original unemployment (\( U \)), because the price expectations in the market mean that the equilibrium rate of unemployment can now only be achieved at this higher rate of (wage) inflation. In effect, the short-run Phillips curve has shifted outwards from \( PC \) to \( PC_1 \).

These shifts in the short-run Phillips curves illustrate that a cycle of wage inflation has been created, but note that long term unemployment rate remains the same (at \( U \) on Figure 2). This means that governments can no longer base policies on the simple trade off between inflation and unemployment as proposed in the original Phillips curve.

Moreover, instead of trying to eradicate unemployment, governments need to accept that there is a natural rate of unemployment in the economy, and that this will also be the level at which inflation rates will be stable.

This natural rate of unemployment is called the non-accelerating inflation rate of unemployment (NAIRU), the level at which the rate of inflation is stable. (The NAIRU is also sometimes referred to as the natural rate hypothesis.) Regulators could attempt to lower the NAIRU through various structural or policy changes, such as increased training in the labour market or improving matching efficiency.

NAIRU is a logical extension of the expectations-augmented Phillips curve, and is represented by \( U \) in Figure 2.

---

### 5 Money

#### Section overview
- There are four functions of money – it is a means of exchange, a unit of account, a standard of deferred payment and a store of value. Money needs to be universally accepted, durable, convenient and homogenous.

#### 5.1 Functions of money

Money as we know it today is a strange thing in that we value it highly but its intrinsic worth is, generally, negligible. We can solve this puzzle by understanding that the importance of money lies in what it does for us rather than what it actually is. Money enables a modern economy to function; most importantly, without
it we should be reduced to the time-consuming and inefficient process of barter. However, this is not all that money can do. What are the functions of money? We can identify four different functions of money.

- A means of exchange.
- A unit of account.
- A standard of deferred payment.
- A store of value.

5.1.1 Money as a means of exchange

This is arguably the most important function of money in an economy, because without money, the only way of exchanging goods and services would be by means of barter, i.e. by a direct exchange of goods or services. In other words, if a shoemaker wanted to buy a horse, he would have two alternative courses open to him:

(a) To find a horse-owner prepared to exchange a horse for a sufficient quantity of shoes of equal value to the horse.

(b) To find other people willing to exchange different goods such as clothes or food for shoes, and then trade these goods in exchange for a horse from the horse-owner.

A monetary economy is the only alternative to a barter economy, and it is a means of encouraging economic development and growth.

(a) People are prepared to organise and work for an employer, and in return, receive money wages.
(b) A business will exchange its goods or services for money in return.
(c) People will pay out money in order to obtain goods or services.

5.1.2 Money as a unit of account

This function of money is associated with the use of money as a means of exchange. Money should be able to measure exactly what something is worth. It should provide an agreed standard measure by which the value of different goods and services can be compared.

Case study

Suppose that only four products are traded in a market. These are pigs, sheep, hens and corn. The relative value of these products must be agreed before exchange can take place in the market. It might be decided that:

- 1 pig has the same value as 0.75 sheep, 3 hens or 1.5 bags of corn.
- 1 sheep is the same value as 1.33 pigs, 4 hens or 2 bags of corn.
- 1 hen is worth 0.33 pigs, 0.25 sheep or 0.5 bags of corn.
- 1 bag of corn has the same value as 0.67 pigs, 0.5 sheep or 2 hens.

In a market with more than four products, the relative values of each product compared with others could be worked out in the same way, although there would be many more value or price ratios to calculate.

The function of money in the economy would be to establish a common unit of value measurement or account by which the relative exchange values or prices of goods can be established.

5.1.3 Money as a standard for deferred payments

When a person buys a good or service, he might not want to pay for it straightaway, perhaps because he has not yet got the money. Instead, he might ask for credit. Selling goods on credit is not an essential feature of an economy, but it certainly helps to stimulate trade. The function of money in this respect is to establish, by agreement between buyer and seller, how much value will be given in return at some future date for goods provided now. Similarly, when a buyer and seller agree now to make a contract for the supply of certain goods in the future, the function of money is to establish the value of the contract, that is, how much the buyer will eventually pay the seller for the goods.
In order to provide an acceptable standard for deferred payments, it is important that money should maintain its value over a period of time. Suppose, for example, that a customer buys goods for an agreed sum of money, but on three months’ credit. Now if the value of money falls in the three-month credit period, the sum of money which the seller eventually receives will be worth less than it was at the time of sale. The seller will have lost value by allowing the credit.

When the value of money falls (or rises) over time, sellers (or buyers) will be reluctant to arrange credit, or to agree the price for future contracts. Money would then be failing to fulfil its function as a standard for deferred payments.

One major reason why money might lose value is because of price inflation.

(a) When inflation is high sellers will be reluctant to allow credit to buyers. For example, if a buyer asks for three months’ credit, and inflation is running at 20 per cent per annum, the ‘real’ value of the debt that the buyer owes will fall by about 5 per cent over the three-month credit period.

(b) Sellers will be reluctant to agree to a fixed price for long-term contracts. For example, a house-builder might refuse to quote a price for building a house over a 12 month period, and instead insist on asking a price which is ‘index-linked’ and rises in step with the general rate of inflation.

5.1.4 Money as a store of value

Money acts as a store of value, or wealth. So too do many other assets (e.g. land, buildings, art treasures, motorcars, machinery) some of which maintain or increase their money value over time, and some of which depreciate in value. This means of course that money is not the only asset which acts as a store of wealth, and we need to extend our definition of this function of money.

Money is more properly described as acting as a liquid store of value. This definition has two parts to it:

(a) Money is a store of value or wealth. A person can hold money for an extended period, for the purpose of exchanging it for services or goods or other assets.

(b) Money is a liquid asset.

5.2 Properties of money

In order to be able to perform the functions mentioned above, it is necessary that money should possess certain qualities or properties.

5.2.1 Acceptability

Money must be accepted as such without question if it is to be useful. This has important implications.

First, money should be scarce; it will not maintain its value if people can easily obtain more for themselves without economic activity. An episode in Douglas Adams’ Hitch-Hiker series in which a society of hairdressers and telephone sanitisers adopts leaves as its currency illustrates this point. Leaves grow on trees, while it is not merely proverbial but essential that money does not. Forgery is a serious threat to the scarcity and therefore the acceptability of money.

Second, money should be recognisable and establishing recognition should not damage or destroy it.

5.2.2 Durability

Money’s function as a store of value requires that it should be durable and not dwindle or deteriorate over time. When gold and silver coins were commonly used as money, wear eroded the amount of metal present in each coin, gold and silver both being soft metals. Similarly, even though cigarettes have been used as money, tobacco does not really make satisfactory money as it is an organic substance and subject to decay.

5.2.3 Convenience

Money must be easy to use. This implies portability and transferability: it should be possible to carry useful quantities of physical money with ease and to transfer ownership to another simply by handing it over.
Money that exists as credit balances need not be carried and may be transferred by means such as cheques and Internet bank transfers.

### 5.2.4 Homogeneity

Money should be available in units of consistent and measurable value. It should also be divisible, so that larger and smaller units are available. Both of these properties are present in notes and coin in their various denominations. A disadvantage of using commodities such as sheep or grain as money is the inevitable variation in quality. With homogeneity money should be of the same quality, so that equal weights will have exactly the same value.

### 6 Measuring the money supply

#### Section overview

- Money can be either broad or narrow and definitions of broad and narrow money vary greatly. Australia measures the money supply through $M_1$, $M_3$, broad money and money base.

#### Definition

The **money supply** is the total amount of money in the economy. It is also referred to as the money stock.

It is important that an economy should have enough money to enable economic activity to take place. As economic growth takes place, the money supply must increase so that there is no shortage. However, it is important that the growth in the money supply should not be too rapid if inflation is to be avoided. Measuring the money supply is therefore an important aspect of economic management.

#### 6.1 Narrow money and broad money

We are used to thinking of money in terms of notes and coin but as was mentioned earlier in this chapter, a wide range of assets, such as money market deposits and eligible bills are so liquid that, effectively, they are equivalent to money. It is not always easy to decide whether a particular financial asset is money or not and it is now considered appropriate to distinguish between **narrow money** and **broad money**.

Financial assets must have a high degree of liquidity to be regarded as **narrow money**.

#### Definitions

A definition of **narrow money** is 'money balances which are readily available to finance current spending, that is to say for transactions purposes'. *(UK Treasury Economic Progress Report)*

**Broad money** extends the range of assets that are regarded as money to include money held in the form of savings.

A definition of **broad money** is 'It provides an indicator of the private sector’s holdings of relatively liquid assets – assets which could be converted with relative ease and without capital loss into spending on goods and services'. *(UK Treasury Economic Progress Report)*

Narrow money can be defined in different ways, depending on how narrowly 'liquidity' is defined; similarly, broad money can be defined in a variety of ways. Even the broadest definition of money will exclude some financial assets. There will never be a clear dividing line between what is narrow money, what is broad money, and what is not money at all.
6.2 Australian measures of money supply

The Reserve Bank of Australia (RBA) is Australia’s central bank authority and its role in setting monetary policy is presented in Chapter 10 of the Study Manual. The RBA uses four main measures of money supply, \( M_1, M_3, \) broad money and the money base:

- \( M_1 \) is the narrowest definition of money, which is made up of bank currency and current deposits of the private, non-bank sector.
- \( M_3 \) is a broader definition consisting of all of \( M_1 \) plus all other bank deposits of the private, non-banking sector.
- Broad money consists of \( M_3 \) plus borrowings from the private sector by the non-bank sector less their currency and bank deposits.
- The money base consists of notes and coins in circulation, plus bank deposits placed with the RBA, plus other RBA liabilities to the non-bank sector.

Example

Statistics for the Australian money supply in April 2011, published by the RBA, were as follows:

<table>
<thead>
<tr>
<th></th>
<th>$bn</th>
</tr>
</thead>
<tbody>
<tr>
<td>Currency</td>
<td>47.7</td>
</tr>
<tr>
<td>Current deposits with banks</td>
<td>220.0</td>
</tr>
<tr>
<td>( M_1 )</td>
<td>267.7</td>
</tr>
<tr>
<td>Certificates of Deposit issued by banks</td>
<td>173.3</td>
</tr>
<tr>
<td>Time deposits with banks</td>
<td>452.4</td>
</tr>
<tr>
<td>Other deposits with banks</td>
<td>377.1</td>
</tr>
<tr>
<td>Deposits with non-bank ADIs</td>
<td>62.2</td>
</tr>
<tr>
<td>( M_3 )</td>
<td>1332.6</td>
</tr>
<tr>
<td>Other borrowings from private sector by AFIs</td>
<td>22.3</td>
</tr>
<tr>
<td>Broad money</td>
<td>1354.9</td>
</tr>
</tbody>
</table>

Notes:

1. DI stands for depositary institution and FI stands for financial institution
2. In April 2011, the money base was $56.4 billion.

7 Credit and interest rates

Section overview

- Credit is the borrowing and lending of money which are governed by interest rates. Interest rates can be short-term, medium-term and long-term; and expressed as nominal rates or real rates of interest. Interest rates will vary over time due to a variety or macroeconomic factors.

Credit for our purposes here concerns lending and borrowing money, rather than with buying goods on trade credit.

The functions of credit can be seen from the point of view either of the borrower or the lender:

(a) For the borrower, the reason for borrowing money is to be able to purchase goods or services now that he might not otherwise be able to afford. The borrower wants to buy now and pay later.

(b) For the lender, the reason for lending money is that there is nothing that he now particularly wants to spend his money on, and by lending it, he can earn some interest.

Credit involves the transfer of money from a lender to a borrower, in exchange for a promise to return it at some future time and to pay interest meanwhile.
Credit is a **scarce commodity**, priced through interest rates. Although there are many different interest rates in an economy, including building society mortgage rates, banks’ base rates and yields on gilt-edged securities, they tend to move up or down together.

(a) If some interest rates go up, for example the banks’ base rates, it is quite likely that other interest rates will move up too, if they have not gone up already.

(b) Similarly, if some interest rates go down, other interest rates will move down too.

### 7.1 Interest rates and the money supply

In the past, attempts have been made to prevent inflation by direct controls on the supply of money such as requiring the clearing banks to make large deposits of funds with the central bank. This approach assumes that the money supply can be controlled exogenously, or from outside the process of commercial lending and borrowing. Direct controls did not work very well and the alternative endogenous view of control now prevails. Under this approach, the creation of credit is controlled within the system, through the structure of interest rates, higher rates tending both to reduce demand for new credit and to reduce discretionary expenditure by increasing the cost of existing loans.

### 7.2 The term structure of interest rates

The various interest rates can be grouped into three broad classes, according to the length of the loan period.

- Short-term interest rates.
- Medium-term interest rates.
- Long-term interest rates.

In general, longer-term financial assets should offer a higher yield than short-term lending. There are several reasons for this:

(a) The investor must be compensated for tying up his money in the asset for a longer period of time. If the government were to make two issues of 9 per cent Treasury Bonds on the same date, one with a term of five years and one with a term of twenty years (and if there were no expectations of changes in interest rates in the future) then the **liquidity preference** of investors would make them prefer the five-year stock.

(b) The only way to overcome the liquidity preference of investors is to compensate them for the loss of liquidity; in other words, to offer a higher rate of interest on longer-dated stock.

(c) There is a greater risk in lending longer-term than shorter-term for two reasons:

   (i) **Inflation.** The longer the term of the asset, the greater is the possibility that the rate of inflation will increase, so that the fixed rate of interest paid on the asset will be overtaken by interest yields on new lending now that inflation is higher.

   (ii) **Uncertain economic prospects.** The future state of the economy cannot be predicted with certainty. If an organisation wishes to borrow money now for, say, fifteen years, there is no certainty about what might happen to that organisation during that time. It might thrive and prosper or it might run into economic difficulties for one reason or another. Investors will require a higher return to compensate them for the increased risk.

(d) Note, however, that two other factors also affect the cost of borrowing:

   (i) The risk associated with the perceived ability of the borrower to fulfil the terms of the loan.

   (ii) Whether or not the loan is secured by a mortgage on an asset.

### 7.3 Nominal and real rates of interest

**Nominal rates of interest** are rates expressed in money terms. If interest paid per annum on a loan of $1 000 is $150, the rate of interest would be 15 per cent. The nominal rate of interest might also be referred to as the money rate of interest, or the actual money yield on an investment.

**Real rates of interest** are the rates of return that investors get from their investment, adjusted for the rate of inflation. The real rate of interest is therefore a measure of the increase in the real wealth,
expressed in terms of buying power, of the investor or lender. Real rates of interest are lower than nominal rates when there is price inflation. For example, if the nominal rate of interest is 12 per cent per annum and the annual rate of inflation is 8 per cent per annum, the real rate of interest is the interest earned after allowing for the return needed just to keep pace with inflation.

The relationship between the inflation rate, the real rate of interest and the money rate of interest is:

\[(1 + \text{real rate of interest}) \times (1 + \text{inflation rate}) = 1 + \text{money rate of interest}\]

We may rearrange this to find the real rate of interest in the example above.

\[
\frac{1 + \text{money rate}}{1 + \text{inflation rate}} = 1 + \text{real rate}
\]

\[
\frac{1.12}{1.08} = 1.037
\]

The real rate of interest is thus 3.7 per cent.

The real rate of interest is commonly measured approximately, however, as the difference between the nominal rate of interest and the rate of inflation. In our example, this would be 12% – 8% = 4%.

### 7.4 Variations in the general level of interest rates over time

Interest rates on any one type of financial asset will vary over time. In other words, the general level of interest rates might go up or down. The general level of interest rates is affected by several factors:

(a) **The need for a real return.** It is generally accepted that investors will want to earn a 'real' rate of return on their investment, that is, a return which exceeds the rate of inflation. The suitable real rate of return will depend on factors such as investment risk.

(b) **Uncertainty about future rates of inflation.** When investors are uncertain about what future nominal and real interest rates will be, they are likely to require higher interest yields to persuade them to take the risk of investing, especially in the longer-term.

(c) **Changes in the level of government borrowing.** When the demand for credit increases, interest rates will go up. A high level of borrowing by the government is likely to result in upward pressure on interest rates.

(d) **Higher demand for borrowing from individuals.** If individuals want to borrow more, for example because they feel confident about their level of future earnings, then interest rates will tend to rise.

(e) **Monetary policy.** Governments control the level of interest rates in order to control inflation.

(f) **Interest rates abroad.** An appropriate real rate of interest in one country will be influenced by external factors, such as interest rates in other countries and expectations about the exchange rate.

### 8 Monetary theory

**Section overview**

- **Monetary theory** deals with the way changes in monetary variables affect the aggregate demand in the economy and its ultimate impact on prices and output. There are three theories of how the changes in the money supply are transmitted to the real economy: the classical quantity theory, the Keynesian theory, and the monetarist theory.

#### 8.1 The classical theory of money

The classical theory of money is based on the view that money is used only as a medium of exchange and people require it only in order to settle transactions in goods and services. This leads to the quantity theory of money, which holds that changes in the level of prices are caused predominantly by changes in the supply of money.
Monetarism is based on the quantity theory, following the 'Fisher' equation:

\[ MV = PT \]

where:
- \( M \) = supply of money
- \( V \) = velocity of circulation
- \( P \) = level of prices
- \( T \) = number of transactions.

Monetarists assume that \( V \) and \( T \) are constant and therefore prices are a function of the supply of money in an economy.

There is a logic behind the classical quantity theory and it relates to the basic transactions assumption. If it is true that money is used only for transactions relating to goods and services, it follows that an excess of money will lead to increased attempts to spend it. Similarly, a shortage of money will have the effect of reducing demand. If the economy is utilising its productive resources to the full (that is, if there is full employment) it will not be possible to increase output. Any increase in demand will therefore cause prices to rise by the action of market forces. Similarly, any reduction in demand will cause prices to fall.

This theory is very satisfactory in explaining past experience of price rises and falls over long periods, as during the nineteenth century. In the first half of the nineteenth century there was large scale economic expansion, but the money supply was based on the gold standard and expanded only slowly: prices generally fell. In the second half of the century there were extensive increases in the supply of gold as a result of mining in Australia and America. The growth of economic activity, and hence output, did not match the growth in the gold supply and prices rose.

Question 2: Money supply and inflation

According to the Fisher equation, if an economy’s money supply increases, what will happen to the level of prices in that economy?

(The answer is at the end of the chapter)

8.2 The Keynesian theory of money

The classical theory does not, in fact, fully explain the role of money in the economy. Keynes identified three reasons why people hold wealth as money rather than as interest-bearing securities.

Definitions

The transactions motive. Households need money to pay for their day-to-day purchases. The level of transactions demand for money depends on household incomes.

The precautionary motive. People choose to keep money on hand or in the bank as a precaution for when it might suddenly be needed.

The speculative motive. Some people choose to keep ready money to take advantage of a profitable opportunity to invest in bonds which may arise (or they may sell bonds for money when they fear a fall in the market prices of bonds).

There is an important contrast here with the classical quantity theory of money. The quantity theory assumes that the demand for money is governed by transactions only. By proposing two further reasons, Keynes strikes out into new territory.

The precautionary motive is really just an extension of the transactions motive. However, the speculative motive for holding money needs explaining a bit further.

(a) If individuals hold money for speculative purposes, this means that they are not using the money to invest in bonds. They are holding on to their savings for speculative reasons. Hence, savings and investment might not be in equilibrium, with consequences for changes in national income. (When we go on to examine national income accounting, we see that an economy would only be in equilibrium if savings equal investments.)
(b) The reason for holding money instead of investing in bonds is that interest rates are expected to go up. If interest rates go up, bond prices will fall. For example, if the current market price of bonds which pay 5 per cent interest on face value is $100, and interest rates doubled to 10 per cent, the market value of the bonds would fall, perhaps to $50. This is because the interest paid on a bond is fixed at a percentage of face value. The ratio between the income paid and the market value adjusts to the current prevailing interest rate by means of changes in the market value. So, if interest rates are expected to go up, any bonds held now will be expected to lose value, and bond holders would make a capital loss. Thus, it makes sense to hold on to money, for investing in bonds later, after interest rates have gone up. Keynes called such money holdings idle balances.

(c) What causes individuals to have expectations about interest rate changes in the future? Keynes argued that each individual has some expectation of a normal rate of interest. This concept of a normal interest rate reflects past levels and movements in the interest rate, and expectations of the future rate level, obtained from available market information.

Question 3: Interest rates

Following this Keynesian analysis, how would you expect an individual to act if:

(a) They think that the current level of interest is below the 'normal' rate?
(b) They think that the current level of interest is above the 'normal' rate?

(The answer is at the end of the chapter)

Definition

Liquidity preference describes people's preference for holding on to their savings as money (in liquid form) rather than investing it.

Keynes argued further that people will need money to satisfy the transactions motive and precautionary motive regardless of the level of interest. It is only the speculative motive which alters the demand for money as a result of interest rate changes.

(a) If interest rates are high, people will expect them to fall and will expect the price of bonds to rise. They will therefore purchase bonds in anticipation of a capital gain and will therefore have low liquidity preference. (That is, their demand for money is low because they prefer to hold bonds instead.)

(b) If interest rates are low but are expected to rise, this implies that bond prices are likely to fall. People will therefore hold liquid funds in order to be able to invest in bonds later on. Their liquidity preference will be high. (That is, their demand for money is high, because they would rather hold money than bonds.)

The conclusion is that the demand for money will be high (liquidity preference will be high) when interest rates are low. This is because the speculative demand for money will be high. Similarly, the demand for money will be low when interest rates are high, because the speculative demand for money will be low.

8.3 The new quantity theory of money

Friedman re-stated the quantity theory of money, as follows:

\[ MV = PQ \]

where:
- \( M \) is the money supply
- \( V \) is the velocity of circulation of money
- \( P \) is the average price level
- \( Q \) is the physical quantity of national output per period (i.e. the real volume of economic output)

Therefore, \( PQ \) is the money value of national output (i.e. national income \( Y \) at current prices). Remember that in the original Fisher equation, \( P \) was the general level of prices rather than a numerical average and \( T \) was the number of transactions. PT was therefore the money value of transactions and proportional to national income, not equal to it.
Monetarists argue that \( V \) and \( Q \) are independent of \( M \). Therefore, an increase in money supply \( M \) will tend to raise prices \( P \), via the direct transmission mechanism.

(a) Individuals will have more money than they want.
(b) They will spend this excess money, buying not just ‘bonds’ (as Keynes believed) but also equities and physical goods.
(c) The greater demand for physical goods will boost expenditure in the economy (and so the money value of national income).
(d) However, a rapid increase in the money supply will increase spending at a faster rate than the economy will be able to produce more physical output.
(e) A rapid increase in the money supply will therefore inevitably be inflationary.

In conclusion, for monetarists, changes in the money supply cause changes in the money value of national income. Remember that Keynes believed there was only a weak link between changes in the money supply and changes in aggregate demand.

8.3.1 The monetarist view of money supply and inflation in the economy

Whereas Keynes argued that an increase in the money supply would merely result in lower interest rates, with no immediate effect on national income, Friedman and other monetarists argue that an increase in the money supply will lead directly and quickly to changes in national income and ‘prices \( \times \) transactions’ (PT), with the velocity of circulation (\( V \)) remaining fairly constant.

In his analysis of the demand for money, Friedman argued that money is just one of five broad ways of holding wealth.

- Money
- Bonds
- Equities
- Physical goods
- Human wealth

(Human wealth here is a special concept and may be ignored for the purpose of our analysis.)

Each method of holding wealth brings some form of return or yield to the holder.

(a) The main yield from money is the convenience of having it when it is needed. This cannot be measured in money terms.
(b) The return on bonds is the interest plus any capital gain (or loss).
(c) Equities should provide dividends and capital growth which keep ahead of the rate of inflation.
(d) Physical assets in this analysis, do not waste away through use, because assets which are consumed cannot be a store of wealth. There might be an increase in their capital value but the yield also includes the non-monetary return, such as the use of furniture, the enjoyment from paintings and so on.

Friedman argued that the demand for money is related to the demand for holding wealth in its other forms. While Keynes believed that if people did not want to hold money, they would invest it to earn interest, monetarists believe that they might also use it instead to buy equities or physical assets.

Friedman argued that money gives a convenience yield but it is not an asset which is held for its own sake. It is a ‘temporary abode of purchasing power’ waiting to be spent on other types of financial or physical asset. The demand for money is therefore a function of the yield on money and the yield on other forms of holding wealth. Yield as defined here includes non-monetary yield such as convenience and enjoyment.

Monetarists would argue, further, that the demand for money is fairly interest-inelastic. The demand for money is related to a transactions motive, but not to any speculative motive. An expected rise in interest rates might persuade individuals to sell bonds and buy other assets, but not to hold speculative money.

Monetarists argue that, since money is a direct substitute for all other assets, an increase in the money supply, given a fairly stable velocity of circulation, will have a direct effect on demand for other assets because there will be more money to spend on those assets. If the total output of the economy is fixed, then an increase in the money supply will lead directly to higher prices.
Monetarists therefore reach the same basic conclusion as the old quantity theory of money. A rise in the money supply will lead to a rise in prices and probably also to a rise in money incomes. (It is also assumed by monetarists that the velocity of circulation remains fairly constant, again taking a view similar to the old quantity theory.) In the short run, monetarists argue that an increase in the money supply might cause some increase in real output and so an increase in employment. In the long run, however, all increases in the money supply will be reflected in higher prices unless there is longer term growth in the economy.

### 8.3.2 Weaknesses in monetarist theory

There are certain complications with the monetarist views, including the following:

(a) The velocity of circulation is known to fluctuate up and down by small amounts.

(b) Increases in prices will not affect all goods equally. Some goods will rise in price more than others and so the relative price of goods will change. For example, the price of houses might exceed the average rate of inflation but the price of electronic goods might rise more slowly.

(c) A higher rate of inflation in one country than another might affect the country’s balance of payments and currency value, thereby introducing complications for the economy from international trade movements.

(d) Prices in the economy might take some time to adjust to an increase in the money supply.

### 8.4 Comparison of theories

<table>
<thead>
<tr>
<th></th>
<th>Classical Quantity Theory</th>
<th>Keynesian Theory</th>
<th>New Quantity Theory</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Name</strong></td>
<td>Fisher</td>
<td>Keynes</td>
<td>Friedman</td>
</tr>
<tr>
<td><strong>Keyword/equation</strong></td>
<td>MV = PT</td>
<td>'liquidity preference'</td>
<td>MV = PQ</td>
</tr>
<tr>
<td><strong>Uses for money</strong></td>
<td>Transactions</td>
<td>Transaction motive</td>
<td>Cash</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Precautionary motive</td>
<td>Bonds</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Speculative motive</td>
<td>Equities</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Physical goods</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Human wealth</td>
</tr>
<tr>
<td><strong>Assumptions</strong></td>
<td>T, V constant;</td>
<td>Only speculative demand</td>
<td>Demand for money is interest-inelastic</td>
</tr>
<tr>
<td></td>
<td>M independently</td>
<td>varies;</td>
<td></td>
</tr>
<tr>
<td></td>
<td>determined</td>
<td>Money supply usually</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>fixed</td>
<td></td>
</tr>
<tr>
<td><strong>Effect if rise in money supply when economy at full employment</strong></td>
<td>AD up, price rises</td>
<td>Price of bonds up,</td>
<td>Spending on assets of all kinds up,</td>
</tr>
<tr>
<td></td>
<td></td>
<td>interest rate falls,</td>
<td>AD up, prices up</td>
</tr>
<tr>
<td></td>
<td></td>
<td>small rise in AD and</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>prices</td>
<td></td>
</tr>
<tr>
<td><strong>Transmission mechanism</strong></td>
<td>Direct</td>
<td>Indirect, via interest</td>
<td>Direct</td>
</tr>
<tr>
<td></td>
<td></td>
<td>rate</td>
<td></td>
</tr>
<tr>
<td><strong>Determination of interest rate</strong></td>
<td>Supply and demand for money</td>
<td>Since money supply is</td>
<td>Supply and demand of loanable funds</td>
</tr>
<tr>
<td></td>
<td></td>
<td>fixed, interest rate,</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>depends on speculative</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>demand for bonds</td>
<td></td>
</tr>
</tbody>
</table>
Key chapter points

- The monetarist concept of a stable equilibrium implies that with zero price inflation, there is a natural optimal level of unemployment, and a rate of economic growth and balance of trade position from which the economy will not deviate. Monetarism focuses on economic stability in the medium to long term, which can only be achieved by abandoning short-term demand management goals.

- The consequences of unemployment for the national economy are loss of output, loss of human capital, an increase in income inequalities in the population, social issues such as theft and suffering and increased social security payments.

- Types of unemployment are frictional, seasonal, technological, cyclical, real wage unemployment and voluntary unemployment.

- The causes of unemployment are manifold and can be summarised into three reasons – demand-deficient, structural change, and supply side unemployment.

- According to Keynes, full employment is when the country’s economic resources are fully employed and a percentage of the population may be unemployed at any time.

- High rates of inflation are harmful to an economy. Inflation redistributes income and wealth. Uncertainty about the value of money makes business planning more difficult. Constantly changing prices impose extra costs.

- Consumer price indices measure inflation and the underlying rate of inflation is the CPI adjusted to exclude mortgage costs.

- There are three causes of inflation – cost push inflation, demand pull inflation and the quantity theory of money.

- Cost push inflation arises from increases in the costs of production. Demand pull inflation arises from an excess of aggregate demand over the productive capacity of the economy. The quantity theory argues that inflation is caused by an increase in the supply of money and is calculated as $MV = PT$.

- There appears to be a connection between the rate of inflation and unemployment.

- The Phillips curve has been used to show the relationship between wage inflation and unemployment.

- The expectations augmented Phillips curve shows that they may be a short-term trade-off between inflation and unemployment but in the long-run the economy will have a vertical Phillips curve.

- There are four functions of money – it is a means of exchange, a unit of account, a standard of deferred payment and a store of value. Money needs to be universally accepted, durable, convenient and homogeneous.

- Money can be either broad or narrow and definitions of broad and narrow money vary greatly. Australia measures the money supply through $M_1$, $M_3$, broad money and money base.

- Credit is the borrowing and lending of money which are governed by interest rates. Interest rates can be short-term, medium-term and long-term; and expressed as nominal rates or real rates of interest. Interest rates will vary over time due to a variety or macroeconomic factors.

- Monetary theory deals with the way changes in monetary variables affect the aggregate demand in the economy and its ultimate impact on prices and output. There are three theories of how the changes in the money supply are transmitted to the real economy: the classical quantity theory, the Keynesian theory, and the monetarist theory.
Quick revision questions

1. What is the Phillips curve?

2. Name two types of short-term unemployment and two types of long-term unemployment.

3. Which one of the following is not an effect of inflation in an economy?
   A. prices convey less information so rational decision making becomes harder
   B. the purchasing power of people on fixed incomes is reduced
   C. the balance of trade weakens as imports become more expensive
   D. consumer behaviour will be distorted as consumers attempt to anticipate price changes

4. Structural unemployment is best defined as unemployment caused by
   A. lack of knowledge about job opportunities in the market.
   B. the long term decline of particular industries.
   C. minimum wage rates being set above the market clearing wage.
   D. the insufficient level of aggregate demand in the economy as a whole.

5. Complete the sentence:
   Money is a means of ................................., a unit of ................................., a standard of .................................,
   ........................................, and a store of ..................................

6. According to Keynes, what are the three motives for working to hold money?

7. According to Keynes, which one of the following is very sensitive to changes in interest rates?
   A. the money supply
   B. the speculative demand for money
   C. the precautionary demand for money
   D. transactions demand for money
Answers to quick revision questions

1. The Phillips curve is an illustration of the relationship between unemployment and inflation levels in an economy.

2. Short-term unemployment: seasonal, frictional
   Long-term unemployment: two from: – structural, technological, cyclical (demand-deficient), voluntary

3. C The balance of trade will weaken, but this will because exports become more expensive and imports become cheaper, leading to a decline in demand for exports and an increase in demand for imports.

4. B Structural unemployment only affects certain industries. A describes frictional unemployment; C describes real wage unemployment; and D describes demand deficient (or cyclical) unemployment, because the problem is lack of aggregate demand in the economy as a whole.

5. Money is a means of exchange, a unit of account, a standard of deferred payment and a store of value.

6. The transactions motive, the speculative motive and the precautionary motive.

7. B According to Keynes, the money supply would be fixed by the authorities. The demand for money depends on three motives (transactions, precautionary and speculative) but it is the speculative demand for money that is sensitive to changes in interest rates, and this explains the liquidity preference schedule.
Answers to chapter questions

1. The pairings are (1) C, (2) B and (3) A.

2. The increase in the money supply will lead to an increase in price levels in the future.

3. (a) If someone believes that the normal rate of interest is above the current level, he will expect the interest rate to rise and will therefore expect bond prices to fall. To avoid a capital loss the individual will sell bonds and hold money.

(b) Conversely, if an individual believes that the normal rate of interest is below the current market interest rate, he will expect the market interest rate to fall and bond prices to rise. Therefore, he will buy bonds, and run down speculative money holdings, in order to make a capital gain.
Chapter 10

Macroeconomic policy

<table>
<thead>
<tr>
<th>Learning objectives</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Macroeconomic policy</strong></td>
<td>LO10</td>
</tr>
<tr>
<td>Explain government policy to address the redistribution of income</td>
<td>LO10.1</td>
</tr>
<tr>
<td>Analyse the impact of interest rates on base employment</td>
<td>LO10.2</td>
</tr>
<tr>
<td>Explain the purpose of monetary policy and the implications of holding cash balances</td>
<td>LO10.3</td>
</tr>
<tr>
<td>Calculate the credit multiplier</td>
<td>LO10.3.1</td>
</tr>
<tr>
<td>Explain how fiscal policy relates to the stimulation of national income and rates of employment</td>
<td>LO10.4</td>
</tr>
<tr>
<td>Demonstrate how fiscal policy affects aggregate demand</td>
<td>LO10.4.1</td>
</tr>
<tr>
<td>Explain the relationship between interest rates, monetary policy, employment and national income</td>
<td>LO10.5</td>
</tr>
<tr>
<td>Prepare an expectations augmented Phillips curve</td>
<td>LO10.5.1</td>
</tr>
<tr>
<td>Analyse the role of the monetary authorities (Reserve Banks/Central Banks) in the control of money</td>
<td>LO10.6</td>
</tr>
</tbody>
</table>

**Topic list**

1. Government policies and objectives
2. Fiscal policy
3. Fiscal policy and taxation
4. Fiscal policy and aggregate demand
5. Fiscal policy and unemployment
6. Monetary policy
7. Effectiveness of macroeconomic policy
8. The role of the central bank
The previous chapter introduced the main macroeconomic concepts such as unemployment, inflation, money and interest rates. In this chapter we present an overview of the goals of macroeconomic policy in relation to these concepts, and concentrate on two broad types of policy: fiscal policy and monetary policy.

First we consider the role of fiscal policy in affecting aggregate demand, and we look at the different types of taxes which can be used as part of fiscal policy.

Next we present alternative theories of how the money supply and interest rates affect the aggregate demand and discuss the conduct of monetary policy.

Then we examine the effectiveness of macroeconomic policy in controlling inflation and unemployment.

The last section examines the role of central banks in controlling the supply of money. It examines the Reserve Bank of Australia (RBA) as an example of universal central bank functions.
Before you begin

If you have studied these topics before, you may wonder whether you need to study this chapter in full. If this is the case, please attempt the questions below, which cover some of the key subjects in the area.

If you answer all these questions successfully, you probably have a reasonably detailed knowledge of the subject matter, but you should still skim through the chapter to ensure that you are familiar with everything covered.

There are references in brackets indicating where in the chapter you can find the information, and you will also find a commentary at the back of the Study Manual.

1. What is the difference between fiscal policy and monetary policy? (Section 1)
2. What are the three types of taxation? (Section 3.2.1)
3. Distinguish between direct taxation and indirect taxation. (Section 3.5)
4. List the differences between quantitative and qualitative controls on bank lending. (Section 6.7.3)
5. What are inflationary expectations? (Section 7.3)
6. List four key functions of a central bank. (Section 8)
7. What are the central banks of:
   (a) Australia?
   (b) the United Kingdom?
   (c) the European Union?
1 Government policies and objectives

Section overview

- Macroeconomic policy objectives relate to economic growth, inflation, unemployment and the balance of payments.

1.1 Economic policy objectives

All modern governments are expected to manage their national economies to some extent. People generally suppose that government action can support or hinder the growth of prosperity in their country and look to their government for serviceable macroeconomic policies. There are four main objective of economic policy, though debate continues about their relative priority.

(a) **To achieve economic growth**, and growth in national income per head of the population. Growth implies an increase in national income in real terms. Increases caused by price inflation are not real increases at all.

(b) **To control price inflation** (to achieve stable prices). This has become a central objective of many Western countries' economic policy in recent years.

(c) **To achieve full employment**. Full employment does not mean that everyone who wants a job has one all the time, but it does mean that unemployment levels are low, and involuntary unemployment is short-term.

(d) **To achieve a balance between exports and imports** (on the country's balance of payments accounts) over a period of years. The wealth of a country relative to others, a country's creditworthiness as a borrower, and the goodwill between countries in international relations might all depend on the achievement of an external balance over time.

Definitions

Fiscal policy relates to government policy on taxation, public borrowing and public spending.

Monetary policy is concerned with government policy on the money supply, the monetary system, interest rates, exchange rates and the availability of credit.

Fiscal and monetary policy attempt to attain the macroeconomic policy objectives by influencing aggregate demand in an economy.

2 Fiscal policy

Section overview

- Fiscal policy provides a method of managing aggregate demand in the economy via taxation and government spending.

2.1 Fiscal policy and the Budget

A feature of fiscal policy is that a government must plan what it wants to spend, and so how much it needs to raise in income or by borrowing. It needs to make a plan in order to establish how much taxation there should be, what form the taxes should take and which sectors of the economy (firms or households, high income earners or low income earners) the money should come from. This formal planning of fiscal policy is usually done once a year and is set out in the Budget.
The two components of the budget which the government determines and through which it exercises its fiscal policy are:

(a) **Expenditure.** The government, at a national and local level, spends money to provide goods and services, such as a health service, public education, a police force, roads, public buildings and so on, and to pay its administrative work force. It may also, perhaps, provide finance to encourage investment by private industry, for example by means of grants.

(b) **Revenues.** Government expenditure must be financed, so the government must generate income. Most government income comes from **taxation**, although some income is obtained from **direct charges** to users of government services such as Medicare or other public health system charges.

A third element of the fiscal policy is:

(c) **Borrowing.** To the extent that a government’s expenditure exceeds its income it must borrow to make up the difference. Where the government borrows from has an impact on the effectiveness of fiscal policy.

### 2.2 Budget surplus and budget deficit

If a government decides to use fiscal policy to influence demand in the economy, it can choose either expenditure changes or tax changes as its policy instrument. Suppose, for example, that the government wants to stimulate demand in the economy.

(a) **It can increase demand directly by spending more itself** – for example, on the health service or education, and by employing more people itself.

   (i) This extra spending could be financed by higher taxes, but this would reduce spending by the private sector of the economy because the private sector’s after-tax income would be lower.

   (ii) The extra government spending could also be financed by extra government borrowing. Just as individuals can borrow money for spending, so too can a government.

(b) **It can increase demand indirectly by reducing taxation** and so allowing firms and individuals more after-tax income to spend (or save).

   (i) Cuts in taxation can be matched by cuts in government spending, in which case total demand in the economy will not be stimulated significantly, if at all.

   (ii) Alternatively, tax cuts can be financed by more government borrowing.

Just as aggregate demand in the economy can be boosted either by more government spending or by tax cuts, so too can demand in the economy be reduced by cutting government spending or by raising taxes, and using the savings or higher income to cut government borrowing.

Expenditure changes and tax changes are not mutually exclusive options, of course. A government has several options:

(a) Increase expenditure and reduce taxes, with these changes financed by a higher level of government borrowing.

(b) Reduce expenditure and increase taxes, with these changes reducing the level of government borrowing.

(c) Increase expenditure and partly or wholly finance this extra spending with higher taxes.

(d) Reduce expenditure and use these savings to reduce taxes.

When a government’s income exceeds its expenditure we say that the government is running a **budget surplus.** When a government’s expenditure exceeds its income, so that it must borrow to make up the difference we say that the government is running a **budget deficit.**
Definitions

**Budget deficit**: Government expenditure exceeds government revenue from taxation income.

**Budget surplus**: Government expenditure is less than government revenue from taxation income.

Case study: Budget deficit in the UK and Greece

A combination of government fiscal policies and the economic recession resulted in a large increase in the UK budget deficit (‘public sector net cash requirement’). Following a change of government in 2010, new fiscal policies were announced aimed at cutting the budget deficit and the size of the total public debt substantially, through a combination of severe government spending cuts and higher taxation (notably an increase in VAT, the sales tax).

Within the countries of the euro area (the eurozone) Greece, Ireland and Portugal all needed financial support from other countries of the European Union and the International Monetary Fund to deal with their fiscal deficits and public debt. These countries have agreed programs to cut the size of their budget deficits, but proposed measures to cut spending in Greece prompted strong political and social unrest.

Case study: Budget surplus in Australia

In contrast, Australia entered the global recession of 2008/09 with a strong budget surplus. The Federal Government chose to meet the threat of recession through a combination of taxpayer payments, spending programs and other bonuses in order to stimulate consumer spending and economic recovery. As a result, there was a budget deficit of about $55 billion in 2010 – 2011. The annual deficit has since been reduced, and the government stated its intention in 2011 of returning to a budget surplus by 2012 – 2013.

The size of Australia’s budget deficit has been relatively small compared with the deficits in other advanced economies. (Comparisons can be made by measuring the budget deficit as a percentage of GDP.)

3 Fiscal policy and taxation

Section overview

- **Direct taxes** have the quality of being **progressive** or **proportional**. Income tax is usually progressive, with higher rates of tax charged on higher bands of taxable income. **Indirect taxes** can be **regressive**, when the taxes are placed on essential commodities or commodities consumed by poorer people in greater quantities.
- A government must decide how it intends to raise tax revenues, from **direct** or **indirect taxes**, and in what proportions tax revenues will be raised from each source.

3.1 Functions of taxation

Taxation has several functions:

(a) **To raise revenues for the government** and to finance the provision of public and merit goods such as defence, health and education.

(b) **To manage aggregate demand**. Aggregate demand could be boosted by lowering taxes, or it could be reduced by increasing taxes.

(c) **To provide a stabilising effect on national income**. Taxation reduces the effect of the multiplier, and so can be used to dampen upswings in a trade cycle – i.e. higher taxation when the economy shows signs of a boom will slow down the growth of money GNP and so take some inflationary pressures out of the economy.
The size of the multiplier, remember, is \( \frac{1}{s + m + t} \) where \( t \) is the marginal rate of taxation.

(d) **To cause certain products to be priced to take into account their social costs.** For example, smoking entails certain social costs, including the cost of hospital care for those suffering from smoking-related diseases, and the government sees fit to make the price of tobacco reflect these social costs.

In a similar way, taxes could be used to discourage activities which are regarded as undesirable.

(e) **To redistribute income and wealth.** Higher rates of tax on higher incomes will serve to redistribute income across all segments of the population.

(f) **To protect industries from foreign competition.** If the government levies a duty on all imported goods much of the duty will be passed on to the consumer in the form of higher prices, making imported goods more expensive. This has the effect of transferring a certain amount of demand from imported goods to domestically produced goods.

### 3.2 Qualities of a good tax

*Adam Smith* (in his seminal work about the *Wealth of Nations*) ascribed **four features to a good tax system.**

(a) **Equity.** People should pay according to their ability.

(b) **Certainty.** The tax should be well-defined and easily understood by all concerned.

(c) **Convenience.** The payment of tax should ideally be related to how and when people receive and spend their income (e.g. pay as you go [PAYG] is deducted when wages are paid, and sales tax is charged when goods are bought).

(d) **Economy.** The cost of collection should be small relative to the yield.

Further features of a good tax can be identified.

- **Flexibility.** It should be adjustable so that rates may be altered up or down. For example, in the UK, the rate of sales tax (VAT, the equivalent of GST in Australia and New Zealand) was reduced from 17.5 per cent to 15 per cent in 2008 to try to boost aggregate demand in the economy. It was increased back to 17.5 per cent in January 2010, and subsequently raised still further to 20 per cent in January 2011.

- **Efficiency.** A tax needs to achieve its objective efficiently, and avoidance should be difficult. However, the tax should not undermine other aims or taxes.

- It should attain its purpose **without distorting economic behaviour.**

### 3.2.1 Types of taxation

Taxation can be classified into three categories on the basis of what is being taxed.

(a) **Income** – income tax, corporation tax, national insurance.

(b) **Expenditure** – sales tax (GST), duties and levies.

(c) **Capital** – inheritance tax, capital gains tax.

Taxes can also be categorised according to the percentage of income which is paid as tax by different groups in society.

(a) A **regressive tax** takes a higher proportion of a poor person's salary than of a rich person's. Television licences (the annual licence fee people have to pay in the UK to watch television) are an example of regressive taxes since they are the same for all people. Sales taxes (such as GST) are also regressive because they are the same for all people, regardless of a person's income. Therefore, as their income rises, the tax represents a smaller percentage of a person's income.

(b) A **proportional tax** takes the same proportion of income in tax from all levels of income. So an income tax with a basic rate of tax at 22 per cent is a proportional tax, (although it then becomes a progressive tax if higher income earners have to pay a higher rate than this basic rate).
(c) A **progressive tax** takes a higher proportion of income in tax as income rises. Income tax as a whole in Australia, the US and the UK is progressive. In Australia, the first part of an individual’s income is tax-free (up to the tax-free threshold which is currently $6,000). The rate of tax on additional income increases in four steps from 15 cents in $1, 30 cents in $1, 37 cents in $1 and 45 cents in $1 (the highest rate) (2010 – 2011 rates).

### 3.3 Advantages and disadvantages of progressive taxation

**Arguments in favour of progressive direct taxes**

(a) **They are levied according to the ability of individuals to pay.** Individuals with a higher income are more able to afford to give up more of their income in tax than low income earners, who need a greater proportion of their earnings for the basic necessities of life. If taxes are to be raised according to the ability of people to pay (which is one of the features of a good tax suggested by Adam Smith) then there must be some progressiveness in them.

(b) **Progressive taxes enable a government to redistribute wealth from the rich to the poor in society.** Such a redistribution of wealth will alter the consumption patterns in society since the poorer members of society will spend their earnings and social security benefits on different types of goods than if the income had remained in the hands of the richer people. Poorer people are also likely to have a higher marginal propensity to consume than richer people. Leaving more income in the hands of the poorer people is likely to increase aggregate demand in the economy as a whole.

(c) **Indirect taxes tend to be regressive and progressive taxes are needed as a counter-balance** to make the tax system as a whole more fair.

**Arguments against progressive taxes**

(a) **In an affluent society, there is less need for progressive taxes than in a poorer society.** Fewer people will live in poverty in such a society if taxes are not progressive than in a poorer society.

(b) **Higher taxes on extra corporate profits might deter entrepreneurs** from developing new companies because the potential increase in after-tax profits would not be worth the risks involved in undertaking new investments.

(c) **Individuals and firms that suffer from high taxes might try to avoid or evade paying tax** by transferring their wealth to other countries, or by setting up companies in tax havens where corporate tax rates are low. However, tax avoidance and evasion are practised whether tax rates are high or low. High taxes will simply raise the relative gains which can be made from avoidance or evasion.

(d) When progressive taxes are harsh, and either tax high income earners at very high marginal rates or tax the wealthy at high rates on their wealth, **they could act as a deterrent to initiative.** Skilled workers might leave the country and look for employment in countries where they can earn more money.

### 3.4 Proportional and regressive taxes

It is often argued that tax burdens should be **proportional** to income in order to be fair, although a proportional tax has the following disadvantages:

(a) A large administrative system is needed to calculate personal tax liabilities on a proportional basis. The costs of collecting income tax relative to tax revenues earned can be high, particularly in the case of lower income taxpayers.

(b) Such a tax does not contribute towards a redistribution of wealth among the population.

In the case of a **regressive tax**, a greater proportionate tax burden falls on those least able to afford it. The main disadvantage of a regressive tax is that it is not fair or equitable. The main advantage of a regressive tax is that it is often relatively easy to administer and collect. This is the case with the GST in Australia, for example. However, a regressive tax could still be expensive to collect.
**Question 1: Taxation types**

Below are details of three taxation systems, one of which is regressive, one proportional and one progressive. Which is which?

<table>
<thead>
<tr>
<th>System</th>
<th>Income before tax</th>
<th>Income after tax</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$</td>
<td>$</td>
</tr>
<tr>
<td>System 1</td>
<td>10 000</td>
<td>8 000</td>
</tr>
<tr>
<td></td>
<td>40 000</td>
<td>30 000</td>
</tr>
<tr>
<td>System 2</td>
<td>10 000</td>
<td>7 000</td>
</tr>
<tr>
<td></td>
<td>40 000</td>
<td>28 000</td>
</tr>
<tr>
<td>System 3</td>
<td>10 000</td>
<td>9 000</td>
</tr>
<tr>
<td></td>
<td>40 000</td>
<td>38 000</td>
</tr>
</tbody>
</table>

(The answer is at the end of the chapter)

### 3.5 Direct and indirect taxes

A **direct tax** is paid directly by a person to the relevant taxation or revenue authority (in Australia this is the Australian Taxation Office or ATO). Examples of direct taxes in Australia are income tax, corporation tax, and capital gains tax. A direct tax can be levied on income and profits, or on wealth. Direct taxes tend to be progressive or proportional taxes. They are also usually unavoidable, which means that they must be paid by everyone.

An **indirect tax** is collected by the relevant revenue authority from an intermediary (a supplier) who then attempts to pass on the tax to consumers in the price of goods they sell, for example, as with the GST in Australia. Indirect taxes are of two types:

- **A specific tax** is charged as a fixed sum per unit sold.
- **An ad valorem tax** is charged as a fixed percentage of the price of the good.

#### 3.5.1 Incidence of tax

We can also distinguish between the **formal** and **real incidence of a tax**. The formal incidence relates to the person administering the tax, the real incidence relates to the person who bears the burden of the tax. So, for GST, the formal incidence lies with retailers, the real incidence passes to the consumer. (This topic has already been explained, in Chapter 6.)

We can illustrate the impact of the incidence of taxation graphically.

In Figure 1 the imposition of a sales tax causes the supply curve to shift to the left (S to S') because producers will reduce their supply at any given price due to part of the sales price being paid to the government. A new equilibrium will be reached, but this will be at a higher price and lower quantity than the pre-tax equilibrium.

The tax per unit is represented by the vertical distance between the supply curves (‘ab’ on Figure 1). However, the burden of tax is split between the producer and the consumer.

The distribution of the tax burden between producer and consumer will depend on both the **elasticity of demand** and the **elasticity of supply**. For elastic demand curves, the burden tends to the producer; for inelastic demand curves, the burden tends to the consumer.

In Figure 1, demand is relatively inelastic, and we can see that the consumer bears a greater proportion of the burden of the tax than the producer.
3.6 Advantages of direct taxation on income

The main advantages of direct taxes on income are that they can be made fair and equitable by being designed as progressive or proportional to the degree desired. Because of their generally progressive nature they also tend to stabilise the economy, automatically taking more money out of the system during a boom and less during depression. Moreover, because they are more difficult to pass on, they are less inflationary than indirect taxes. Finally, taxpayers know what their tax liability is.

3.7 Incentive effects

A direct tax on profits might act as a disincentive to risk-taking and enterprise. The tax will reduce the net return from a new investment and any disincentive effects will be greater when the tax is progressive. In addition, a tax on profits will reduce the ability to invest. A considerable part of the finance for new investment comes from retained profits so any tax on corporate profits will reduce the ability of firms to save and therefore limit the sources of funds for investment.

High taxation acts as a disincentive to work because if marginal tax rates (i.e. the proportion of additional income taken as tax) are high, individuals are likely to behave in one of two ways.

(a) They may forgo opportunities to increase income through additional effort on the basis that the increase in net income does not adequately reward the effort or risk.

(b) They may resort to working in the parallel 'black' economy to avoid paying the tax.

3.8 The Laffer curve and tax yields

Definition

Laffer curve: a curve depicting the relationship between tax revenue and the average tax rate, designed to illustrate the thesis that there is an optimal tax rate at which tax revenues are maximised.

The Laffer curve (named after Professor Arthur Laffer) illustrates the effect of tax rates upon government revenue and national income.

In the hypothetical economy depicted in Figure 2 a tax rate of 0 per cent results in the government receiving no tax revenue irrespective of the level of national income. If the rate is 100 per cent then nobody will work because they keep none of their earnings and so once again total tax revenue is zero. In our example, at 25 per cent tax rates the government will achieve a total tax take of $30bn; the same as the revenue they achieve from tax at rates of 75 per cent. By deduction, the level of national income when...
taxes are 25 per cent must be $120bn compared with only $40bn if taxes are 75 per cent. High taxation appears to operate as a disincentive and reduce national income.

At low tax rates, the income effect dominates the substitution effect leading to increases in hours of work and increasing tax revenue. However, at higher tax rates, the substitution effect dominates and reduces the desired hours of work.

The government will be keen to identify the tax rate 'T,' which maximises revenue, and it will not want to set taxes higher than that with the effect that they become a disincentive to work. The disincentive effect is arising from the substitution effect. Labour taxes are inefficient because leisure (and possibly non-labour income) is not taxed so the tax can be avoided by non-workers.

![Laffer curve for a hypothetical economy](image)

**Figure 2** Laffer curve for a hypothetical economy

Three consequences flow from this Laffer curve analysis:

(a) **High rates of taxation act as a disincentive to work** and accordingly reduce output and employment, because people will substitute leisure for work. This is known as the **disincentive effect**.

(b) Governments cannot always expect to increase tax revenue by increasing tax rates. There appears to be a crucial tax rate beyond which the fall in national income resulting from the erosion of incentives and effort outweighs the increased tax rate. In Figure 2 the maximum tax revenue is $T_x$, at average tax rate $T_r$. If tax rates are above $T_r$, the government can increase tax revenues by cutting tax rates.

(c) There will always be two tax rates available which can yield the same total tax revenue: one associated with a high level of national income and another associated with a lower level. In consequence, governments committed to high government expenditure need not necessarily be associated with high rates of tax. Taxes could be set at the lower of the two rates and earn just the same amount of revenue as if they had been set at the higher amount.

In addition, the following arguments can be made against high tax rates.

When income tax is levied at a high rate, it could act as a disincentive to work.

The idea of the **marginal tax rate** is important here. This is the rate of tax paid on an extra unit of income.

The marginal tax rates facing economic agents are often important in determining how tax affects a decision to work or not. For example, is it worth taking a higher paid job or working overtime if the majority of the additional income you earn will have to be paid over as tax?

The importance of this from a government's perspective is that if they set the tax rate too high, tax yield will fall because workers will choose to have leisure time instead of working. In this case, the tax is no longer efficient.

Another danger of setting income tax too high is that it could encourage a **brain drain**. If a country has high marginal rates of taxation, this might cause a migration of highly skilled – and therefore highly paid –
workers to countries with more favourable tax regimes. This means that the country with the high tax rates not only loses the potential tax revenue from these high earners, but it also loses their valuable skills.

High marginal tax rates, by narrowing the differentials in the after-tax pay of skilled and unskilled labour, may reduce the incentive to train and thereby cause a shortage of skilled labour.

High marginal tax rates may also be seen as creating a 'poverty trap'. For example, if people are not working, they receive unemployment benefit. Once they find a job they will no longer receive their unemployment benefit, but will have to pay tax on their income instead. If tax rates are felt to be too high there would be little incentive for an unemployed worker to accept a low paid job. However, if direct taxes are lower, then workers will be more likely to accept the job because their post tax income will be higher.

High marginal rates of tax could encourage tax avoidance (finding legal loopholes in the tax rules so as to avoid paying tax). In many countries with relatively high taxation, such as Australia and the UK, high tax rates led to a growth of making income payments-in-kind by way of fringe benefits (benefits-in-kind), for example, free medical and life insurance, preferential loans and favourable pension rights. The tax authorities in both countries responded to this by bringing an increasing number of fringe benefits into the tax net – for example, the private use of a company car.

In some cases individuals and companies may resort to tax evasion, which is the illegal non-payment of tax. Employed people have limited scope for tax evasion, because of the pay as you go (PAYG) tax system in Australia. This is referred to as the pay as you earn (PAYE) system in the UK. Undoubtedly, the self-employed have a greater ability to evade tax by failing to declare earnings. If evasion becomes widespread the cost of enforcing the tax laws may become very expensive.

Reductions in tax liability may encourage workers to accept lower wage increases, and this will mean there is less inflationary pressure from high wage claims.

### 3.9 Indirect taxation: advantages and disadvantages

Indirect taxes are hidden in the sense that the taxpayer is frequently unaware of the amount of tax he is paying, so that it is almost painlessly extracted (for example, in the high rate of tax on beer and spirits). This has considerable advantages from the government’s point of view.

Indirect taxation can be used to encourage or discourage the production or consumption of particular goods and services to affect the allocation of resources. For example, the production of goods that produce environmental pollution may be taxed as a means of raising the price in order to reduce demand and output. Similarly, the consumption of cigarettes can be discouraged by high indirect taxation.

#### Question 2: Indirect tax

The burden of an indirect tax must either be borne by the producer or passed on by the producer to consumers. If a producer feels able to pass on the whole of the burden, what can you deduce about the elasticity of demand for his product?

(The answer is at the end of the chapter)

The effectiveness of indirect taxation is likely to depend on price elasticity of demand (and its corresponding elasticity of supply). If a good is price inelastic, then sales will not fall much when the tax is included in the price. By contrast, if a sales tax is added onto a good with elastic demand, demand will fall significantly following the increase in price.

Consequently, the amount of indirect tax revenue generated is likely to be higher when imposed on a good with inelastic demand than on one with elastic demand. This is another reason (alongside health issues) why indirect taxes are imposed on alcohol and cigarettes – because they have price inelastic demand.

Indirect taxation is a relatively flexible instrument of economic policy. The rates of indirect taxes may be changed to take effect immediately. The example of the UK government changing the rate of sales tax several times between 2008 and 2011 was mentioned previously.

Indirect taxes can be cheap to collect. Traders and companies are required to act as collectors of GST in Australia, so reducing the administrative burden on government.
Indirect taxes do have disadvantages, however:

(a) **They can be inflationary.** The introduction of a flat 10 per cent GST in 2000 resulted in a large increase in the reported rate of inflation. When prices are rising the burden of ad valorem indirect taxes will also naturally rise. However, a specific indirect tax will only be inflationary if the rate at which it is applied is changed.

(b) **Indirect taxes also tend to be regressive.** A broadly levied indirect tax like GST is likely to be quite regressive because the poorer members of the community spend a very much larger fraction of their income than very rich people. A system of indirect taxes on luxury goods would not be regressive, however.

(c) **Indirect taxes are not completely impartial in their application in other ways.** For example, someone who seeks to relax with a cigarette in a beer garden is going to be much more heavily hit by indirect taxes than someone who likes walking. The differential taxes, moreover, prevent resources from being distributed optimally according to consumer preference. Unlike an income tax, indirect taxes change the relative price of goods. This means that consumers have to arrange their patterns of expenditure accordingly. This substitution may involve loss of satisfaction.

(d) Like taxes on incomes, **indirect taxes may be evaded by some.** The so-called 'black economy', in which cash payments are made and income is not declared for tax purposes, is undoubtedly large and widespread in many countries (including Australia) particularly in the self-employed sector.

4 Fiscal policy and aggregate demand

**Section overview**

- A fiscal stance can be neutral, expansionary or contractionary. This depends on its effect on national income.

**Fiscal policy** is concerned with government spending (an injection into the circular flow of income) and taxation (a withdrawal from the circular flow).

(a) If government spending is increased, there will be an increase in the amount of injections, expenditure in the economy will rise and so national income will rise (either in real terms, or in terms of price levels only; that is, the increase in national income might be real or inflationary).

(b) If government taxation is increased, there will be an increase in withdrawals from the economy, and expenditure and national income will fall. A government might deliberately raise taxation to take inflationary pressures out of the economy.

A government's 'fiscal stance' may be neutral, expansionary or contractionary, according to its effect on national income.

(a) Spending more money and financing this expenditure by borrowing would indicate an expansionary fiscal stance.

(b) Collecting more in taxes without increasing spending would indicate a contractionary fiscal stance.

(c) Collecting more in taxes in order to increase spending, thus diverting income from one part of the economy to another would indicate a broadly neutral fiscal stance.
Figure 3 Increase in taxation

An increase in taxation by $T$ in Figure 3, without any matching increase in government expenditure, would reduce the aggregate expenditure in the economy from $AD_1$ to $AD_2$, and so the money value of national income would fall from $Y_1$ to $Y_2$. This would result in either a fall in real output or it would dampen inflationary pressures.

Similarly, a reduction in taxation without any reduction in government spending would increase the money value of national income. This would either cause real output to increase, or it would give a boost to price rises and inflation.

5 Fiscal policy and unemployment

Section overview

- Fiscal policy is used to create jobs and reduce unemployment. The balanced budget multiplier is an increase in aggregate monetary demand.

Fiscal policy can be used to reduce unemployment and provide jobs.

(a) More government spending on capital projects would create jobs in the construction industries.

(b) Government-funded training schemes are a means of spending by government to improve training, so as to make people more qualified for jobs in private industry.

(c) A government might tax companies on the basis of the numbers and pay levels of people they employ (as with employers’ National Insurance Contributions). Lower employment taxes would possibly make employers more willing to take on extra numbers of employees.

If government raises taxes and spending by the same amount, so that the budget remains in balance, there will be an increase in aggregate monetary demand. This is because tax payers would have saved some of the money they pay in increased tax and the government spends all of it within the economy. This effect is called the balanced budget multiplier.

Government spending, however, might create inflationary pressures. Fiscal policy must therefore be used with care, even to create new jobs.
6 Monetary policy

Section overview
- Monetary policy focuses on the relationship between interest rates and the supply of money in an economy, and how the two of them together can influence aggregate demand.

6.1 Objectives of monetary policy

Questions on monetary policy will often focus on the relationship between monetary policy and interest rates.

Monetary policy can be used as a means towards achieving ultimate economic objectives for inflation, the balance of trade, full employment and real economic growth. To achieve these ultimate objectives, the authorities will set intermediate objectives for monetary policy.

In Australia, the ultimate objective of monetary policy is principally to control inflation through maintaining a stable CPI of between two to three per cent. Controlling inflation over the medium term allows the three long-term aims of monetary policy to be achieved: a stable currency, maintenance of full employment and ensuring economic prosperity and welfare.

6.2 The money supply as a target of monetary policy

To monetarist economists, the money supply is an obvious intermediate target of economic policy. This is because they claim that an increase in the money supply will raise prices and incomes and this in turn will raise the demand for money to spend.

When such a policy is first introduced, the short-term effect would be unpredictable for three reasons:

(a) The effect on interest rates might be erratic.

(b) There might be a time lag before anything can be done. For example, it takes time to cut government spending and hence to use reduction in government borrowing as an instrument of monetary policy to control the growth in $M_1$ and $M_3$.

(c) There might be a time lag before control of the money supply alters expectations about inflation and wage demands.

Growth in the money supply, if it is a monetary policy target, should therefore be a medium-term target. For example, when the UK Government set targets for the growth of the money supply as a main feature of its economic policy strategy from 1980, it was consequently prepared to wait for some years to see any benefits from its policies. It therefore set out its policy targets in a medium-term financial strategy.

6.2.1 The interrelationship between targets: the money supply and interest rate targets

The authorities can set intermediate targets for the growth of the money supply, but to achieve their targets of growth it will be necessary to allow interest rates to adjust to a level at which the demand for money matches the size of the money supply. For example, a policy to cut the growth of the money supply might result in higher real interest rates.

On the other hand, the authorities might set targets for the level of interest rates. If they do so, they must allow whatever demand for money there is to be met at that rate of interest by allowing the money supply to meet the demand. If they did not, interest rates would then rise above or fall below the target level.

This means that the authorities can set a target for the money supply or a target for interest rates, but they cannot set independent targets for both at the same time.
6.3 The effects of changing interest rates

While controlling the money supply is a key target of economic policy, the main feature of monetary policy is that of **controlling interest rates**. Governments will be particularly keen to use an interest rate policy if they consider there is a direct relationship between interest rates and the level of expenditure in the economy, or between interest rates and the rate of inflation. In other words, they will use interest rates to control the level of aggregate demand in an economy.

A rise in interest rates will raise the price of borrowing in the internal economy for both companies and individuals. If companies see the rise as relatively permanent, rates of return on investments will become less attractive and **investment plans may be curtailed**. Corporate profits will fall as a result of higher interest payments. Companies will reduce inventory levels as the cost of having money tied up in inventory rises. Individuals should be expected to reduce or postpone consumption in order to reduce borrowings, and should become less willing to borrow for major purchases, such as house purchases.

Although it is generally accepted that there is likely to be a connection between interest rates and investment (by companies) and consumer expenditure, **the connection is not a stable and predictable one**, and interest rate changes are only likely to affect the level of expenditure after a **considerable time lag**.

**Other effects of raising interest rates**

(a) High interest rates will keep the value of a country’s currency higher than it would otherwise be. This will keep the cost of exports high, and so discourage the purchase of exports. This may be necessary to protect the balance of payments and to prevent ‘import-cost-push’ inflation.

(b) High interest rates will attract foreign investors into capital investments, and so high interest rates in Australia could be used to generate capital inflows to help finance any future balance of payments deficit.

**The impacts of a rise in interest rates**

<table>
<thead>
<tr>
<th>Impact</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Spending falls</td>
<td>Higher interest rates increase the cost of credit and thereby deter spending. The higher interest rates are the more attractive it is to hold money than to spend it.</td>
</tr>
<tr>
<td>Investment falls</td>
<td>The increased rate will increase the opportunity cost of investment and reduce the net present value of the investment. This will discourage firms from investing. The increased interest rates will make borrowing more expensive.</td>
</tr>
<tr>
<td>Foreign funds are attracted into the country</td>
<td>Interest rates are the reward for capital, so a rise in interest rates will encourage overseas investors because of the increased rate of return relative to other countries.</td>
</tr>
<tr>
<td>Exchange rate rises</td>
<td>The inflow of foreign funds (above) increases the demand for currency and therefore increases the exchange rate. A rise in exchange rates will make exports more expensive, and imports cheaper. The impact this will have on the balance of payments current account will depend on the relative price elasticities of imports and exports.</td>
</tr>
<tr>
<td>Inflation rate falls</td>
<td>This is the overall goal of an interest rate rise. The reduction in spending and investment will reduce aggregate demand in the economy. Higher exchange rates will force producers to make prices more competitive by cutting costs.</td>
</tr>
<tr>
<td>Bond prices fall</td>
<td>There is an inverse relationship between bond prices and the rate of interest.</td>
</tr>
</tbody>
</table>

An increase in interest rates will have a **deflationary** impact on the economy.

Note, however, the potential conflicting objectives which monetary policy may face. A change in interest rates will have effects on both the domestic economy and on a country’s international trade position (for example, through exchange rate movements). It may be the case that the interest rate movement required for the domestic economy conflicts with that required to achieve a balance on the external current account.
An important reason for pursuing an interest rate policy is that the authorities are able to influence interest rates much more effectively and rapidly than they can influence other policy targets, such as the money supply or the volume of credit.

6.4 The exchange rate as a target of monetary policy

Why the exchange rate is a target

(a) If the exchange rate falls, exports become cheaper to overseas buyers and so more competitive in export markets. Imports will become more expensive and so less competitive against goods produced by manufacturers at home. A fall in the exchange rate might therefore be good for a domestic economy, by giving a stimulus to exports and reducing demand for imports.

(b) An increase in the exchange rate will have the opposite effect, with dearer exports and cheaper imports. If the exchange rate rises and imports become cheaper, there should be a reduction in the rate of domestic inflation. A fall in the exchange rate, on the other hand, tends to increase the cost of imports and adds to the rate of domestic inflation.

When a country's economy is heavily dependent on overseas trade it might be appropriate for government policy to establish a target exchange value for the domestic currency. However, the exchange rate is dependent on both the domestic rate of inflation and the level of interest rates. Targets for the exchange rate cannot be achieved unless the rate of inflation at home is first brought under control. Governments also require large international reserves to be held by the central bank in order to maintain the targeted price.

6.5 Growth in money national income as a target of monetary policy

The authorities might set targets for the level of national income in the economy. For example, the policy might be for the growth in the national income (or GNP or GDP as defined in Chapter 7) to be X per cent per annum for Y years. However, it takes time to collect information about national income whereas targets of monetary policy should be items for which statistical data can be collected regularly and easily.

For this reason, although a target growth rate in national income itself is, in theory, probably the most suitable target of monetary policy, it is the least practical because the authorities would always be working with out-of-date information.

6.6 Targets and indicators

An economic indicator provides information about economic conditions and might be used as a way of judging the performance of government.

(a) A leading indicator is one which gives an advance indication of what will happen to the economy in the future. It can therefore be used to predict future conditions. For example, a fall in the value of the Australian dollar of, say, 2 per cent might be used to predict what will happen to the balance of payments and to the rate of inflation.

(b) A coincident indicator is one which gives an indication of changes in economic conditions at the same time that these changes are occurring. For example, if the narrow money supply rises by 5 per cent, this might ‘confirm’ that the rate of increase in GDP over the same period of time has been about the same, 5 per cent in ‘money’ terms.

(c) A lagging indicator, not surprisingly, is one which ‘lags behind’ the economic cycle. Unemployment, to take an example, often continues to rise until after a recession has ended and only starts to fall again after recovery has begun.

Items which are selected as monetary targets will also be indicators, but not all indicators are selected by the authorities as targets. There are a number of monetary indicators:

(a) The size of the money stock.

(b) Interest rates such as the banks’ base rate of interest, the Treasury bill rate and the yield on long-dated government securities.

(c) The exchange rate against the US dollar, or the trade-weighted exchange rate index.

(d) The size of the government’s borrowing.

(e) Government borrowing as a percentage of Gross Domestic Product.
6.7 Instruments of monetary policy

There are a number of techniques or instruments which are available to the authorities to achieve their targets for monetary policies:

- Changing the level and/or structure of interest rates through open market operations.
- Reserve requirements.
- Direct controls, which might be either quantitative or qualitative.
- Intervention to influence the exchange rate.

6.7.1 Control over the level and structure of interest rates

When a government uses interest rates as an instrument of policy, it can try to influence either the general level of interest rates or the term structure of interest rates. It could do this by influencing either short-term interest rates or long-term interest rates. It is a function of the central bank to set interest rate policy as discussed later in the chapter.

6.7.2 Reserve requirements on banks as a means of controlling the money supply

As another technique for controlling money supply growth, the government might impose reserve requirements on banks. A reserve requirement might be a compulsory minimum cash reserve ratio (i.e. ratio of cash to total assets) or a minimum liquid asset ratio.

Any initial increase in bank deposits or building society deposits will result in a much greater eventual increase in deposits, because of the credit multiplier.

Ignoring leakages, the formula for the credit multiplier is:

\[ D = \frac{C}{r} \]

where:
- \( C \) is the initial increase in deposits
- \( r \) is the liquid assets ratio or reserve assets ratio
- \( D \) is the eventual total increase in deposits

If the authorities wished to control the rate of increase in bank lending and building society lending, they could impose minimum reserve requirements — i.e. a minimum value for \( r \). The bigger the value of \( r \), the lower size of the credit multiplier would be.

There are drawbacks to reserve requirements as a monetary policy instrument:

(a) Unless the same requirements apply to all financial institutions in the country, some institutions will simply take business from others. For example, reserve requirements placed on Australian banks but not on building societies or credit unions would give the building societies a competitive advantage over the banks, without having any effect on the control of total credit/money supply growth.

(b) Similarly, restrictions on domestic financial institutions which do not apply to foreign banks would put the domestic financial institutions at a competitive disadvantage in international markets. This is one reason why international co-operation on the capital adequacy of banks (the international Basel accords) is an important step towards better regulation of financial markets.

6.7.3 Direct controls as a technique of monetary control

Another way of controlling the growth of the money supply is to impose direct controls on bank lending. Direct controls may be either quantitative or qualitative.

(a) Quantitative controls might be imposed on either bank lending (assets), for example a ‘lending ceiling’ limiting annual lending growth, or bank deposits (liabilities). The purpose of quantitative controls might be seen as a means of keeping bank lending in check without having to resort to higher interest rates.

(b) Qualitative controls might be used to alter the type of lending by banks. For example, the government (via the bank) can ask the banks to limit their lending to the personal sector, and lend more to industry, or to lend less to a particular type of firm (such as, for example, property companies) and more to manufacturing businesses.
Quantitative controls

Controls might be temporary, in which case, in time, interest rates would still tend to rise if the money supply growth is to be kept under control. However, the advantage of a temporary scheme of direct quantitative controls is that it gives the authorities time to implement longer term policy. Quantitative controls are therefore a way of bridging the time-lag before these other policies take effect.

Quantitative controls might be more permanent. If they are, they will probably be unsuccessful because there will be financial institutions that manage to escape the control regulations, and so thrive at the expense of controlled institutions.

Direct controls on banks, for example, might succeed in reducing bank deposits but they will not succeed in controlling the level of demand and expenditure in the economy if lending is re-directed into other non-controlled financial instruments of non-controlled financial institutions. For example, large companies might use their own bank deposits to set up a scheme of lending themselves.

Direct controls are therefore rarely effective in dealing with the source rather than the symptom of the problem. Direct controls tend to divert financial flows into other, often less efficient, channels, rather than to stop the financial flows altogether, i.e. 'leakages' are inevitable.

Qualitative controls

Qualitative controls might be mandatory or they might be applied through moral suasion. Mandatory directives of a qualitative nature are unlikely in practice, because they are difficult to enforce without the co-operation of banks and other financial institutions. Moral suasion, on the other hand, might be used frequently. This is a process whereby the central bank appeals to the banks to do one or more things:

- To restrain lending.
- To give priority to certain types of lending such as finance for exports or for investment.
- Refuse other types of lending such as loans to private individuals.

Moral suasion might therefore be a temporary form of control. As just one example, prior to deregulation in the 1970s the Governor of the Reserve Bank of Australia (RBA) 'advised' banks concerning the industries to which loans should be made. This often led to a lack of innovation in industry and lending practices and this moral suasion was lifted as deregulation continued.

6.8 Monetary policy and fiscal policy

Monetary policy can be made to act as a subsidiary support to fiscal policy and demand management. Since budgets are once-a-year events, a government must use non-fiscal measures in between budgets to make adjustments to its control of the economy:

(a) A policy of low interest rates or the absence of any form of credit control might stimulate bank lending, which in turn would increase expenditure (demand) in the economy.

(b) High interest rates might act as a deterrent to borrowing and so reduce spending in the economy.

(c) Strict credit controls (for example, restrictions on bank lending) might be introduced to reduce lending and so reduce demand in the economy.

Alternatively, monetary policy might be given prominence over fiscal policy as the most effective approach by a government to achieving its main economic policy objectives.

6.9 Monetary policy, inflation control and economic growth

Monetarists argue that monetary control will put the brake on inflation, but how does this help the economy? We have already suggested that inflation seems to hinder economic growth, and so we could argue like this:

(a) High inflation increases economic uncertainty. Bringing inflation under control will restore business confidence and help international trade by stabilising the exchange rate.

(b) A resurgence of business confidence through lower interest rates (due to less uncertainty and lower inflation) will stimulate investment and real output.

(c) A controlled growth in the money supply will provide higher incomes for individuals to purchase the higher output.
7 Effectiveness of macroeconomic policy

Section overview

- The effectiveness of monetary policy in influencing aggregate demand and unemployment is limited in the long-run.

Having discussed the way fiscal and monetary policies operate on aggregate demand, here we review their role in controlling inflation and affecting the level of unemployment.

7.1 The control of inflation

The best way of controlling inflation will depend on the causes of it. In practice, it may be difficult to know which cause is most significant. The table below sets out various policies designed to control inflation.

<table>
<thead>
<tr>
<th>Cause of inflation</th>
<th>Policy to control inflation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Demand pull (high consumer demand)</td>
<td>Take steps to reduce demand in the economy:</td>
</tr>
<tr>
<td></td>
<td>• Higher taxation, to cut consumer spending</td>
</tr>
<tr>
<td></td>
<td>• Lower government expenditure (and lower government borrowing to finance its expenditure)</td>
</tr>
<tr>
<td></td>
<td>• Higher interest rates</td>
</tr>
<tr>
<td>Cost push factors (higher wage costs and other costs working through to higher prices)</td>
<td>Take steps to reduce production costs and price rises:</td>
</tr>
<tr>
<td></td>
<td>• De-regulate labour markets</td>
</tr>
<tr>
<td></td>
<td>• Encourage greater productivity in industry</td>
</tr>
<tr>
<td></td>
<td>• Apply controls over wage and price rises (prices and incomes policy)</td>
</tr>
<tr>
<td>Import cost push factors</td>
<td>Take steps to reduce the quantities or the price of imports. Such a policy might involve trying to achieve either an appreciation or depreciation of the domestic currency</td>
</tr>
<tr>
<td>Excessively fast growth in the money supply</td>
<td>Take steps to try to reduce the rate of money supply growth:</td>
</tr>
<tr>
<td></td>
<td>• Reduced government borrowing and borrow from the non-bank private sector</td>
</tr>
<tr>
<td></td>
<td>• Try to control or reduce bank lending</td>
</tr>
<tr>
<td></td>
<td>• Try to achieve a balance of trade surplus</td>
</tr>
<tr>
<td></td>
<td>• Maintain interest rates at a level that might deter money supply growth</td>
</tr>
<tr>
<td>Expectations of inflation</td>
<td>Pursue clear policies which indicate the government’s determination to reduce the rate of inflation</td>
</tr>
</tbody>
</table>

7.2 High interest rates and inflation

A government (or a central bank acting on behalf of the government) may adopt a policy of raising interest rates as a means of trying to reduce the rate of inflation, when inflation is being caused by a boom in consumer demand (with demand rising faster than the ability of industry to increase its output to meet the demand).

If interest rates are high enough, there should eventually be a reduction in the rate of growth in consumer spending.

(a) People who borrow must pay more in interest out of their income. This will leave them less income, after paying the interest, to spend on other things. (The government would not want wages to rise, though, because if people build up their income again with high wage settlements, the consumer spending boom could continue and a wage-price spiral could be instigated.)

(b) High interest rates might deter people from borrowing, and so there would be less spending with borrowed funds.

(c) High interest rates should encourage more saving, with individuals therefore spending less of their income on consumption.

(d) High interest rates will tend to depress the values of non-monetary assets, such as houses, and the reduction in people’s perceived wealth may make people feel ‘poorer’ and consequently reduce the amounts they spend on consumer goods.
7.3 Inflationary expectations

An explanation of rising inflation rates combined with rising unemployment was put forward, based on inflationary expectations. This natural rate hypothesis is supported by monetarist economists.

Inflationary expectations reflect the rates of inflation that are expected in the future. The inflationary expectations of the work force will be reflected in the level of wage rises that is demanded in the annual round of pay negotiations between employers and workers. This is the logic behind the expectations augmented Phillips curve which we looked at in connection with inflation and unemployment levels.

In the long run, unemployment will revert towards its natural level. The rate of inflation, however, will be determined by the short-run Phillips curve, which will shift upwards as inflationary expectations increase. The distinction between short and long-run Phillips curves can help explain the observation that, in Australia and other countries, unemployment and inflation have often both risen at the same time.

You should refer back to Chapter 9 for a more detailed explanation of inflationary expectations.

Definition

The expectations augmented Phillips curve shows that the expansion of aggregate demand to reduce unemployment below its natural rate will only produce inflation.

There is a non-accelerating rate of unemployment (NAIRU) at which the rate of inflation is stable.

Question 3: Natural rate hypothesis

Suppose that the short-run Phillips curve is currently at PC2, with unemployment at 5 per cent and annual inflation at 4 per cent. What would happen if the government now took measures to reduce unemployment to 3 per cent?

(The answer is at the end of the chapter)

Monetarist economists argue that the only way to reduce the rate of inflation is to get inflationary expectations out of the system. In doing so, excessive demands for wage rises should be resisted by employers. However, a firm approach to reducing the rate of inflation could mean having to accept high levels of unemployment for a while.
7.4 New classical school

The new classical school of monetarists believe that the aggregate supply and Phillips curves are vertical in the short run. Therefore, they condemn any policy to expand demand as leading only to increased inflation. They suggest that human behaviour is governed by rational expectations; that is, by a rational assessment of all the information currently available. The implication is that the public will recognise inflation-producing policies as soon as they are introduced and adjust their expectations of inflation immediately. This will lead to increased wage demands and the expected inflation will ensue.

8 The role of the central bank

Section overview

- The central bank has various functions. These include acting as a banker to the central government and to the commercial banks.
- The Reserve Bank of Australia (RBA) has responsibility for monetary policy chiefly through controlling inflation and maintaining a constant CPI by setting interest rates.
- The RBA is an example of an independent central bank.

Exam comments

The examiner has stated that, where possible, references to institutions will be to generic types (e.g., central banks) rather than to particular national institutions. You might be able to make use of your knowledge of institutions in any country in answering questions, but the questions will not be about the specific institutions of any individual country.

A central bank is a bank which acts on behalf of the government. The central bank for Australia is the Reserve Bank of Australia commonly referred to as the RBA. The RBA (not the central government) is responsible for formulating and implementing Australia’s monetary policy.

The RBA operates as a statutory authority and is governed by two boards. The Reserve Bank Board has responsibility for monetary policy and financial stability and the Payments System Board has responsibility for the payment system. The Reserve Bank Board has an Audit Committee.

Functions of the RBA (as Australia’s central bank)

The RBA’s broad responsibility is to conduct monetary policy in a way that provides the greatest benefits to the Australian people, and it should use its powers to contribute to the stability of the currency, the maintenance of full employment in Australia and the economic prosperity and well-being of the Australian people. This general mandate has taken the form of control or influence over short-term interest rates (see below).

(a) The RBA’s most important function is maintaining monetary stability in the economy. The Reserve Bank Board has operational responsibility for setting interest rates at the level it considers appropriate in order to meet their stated annual CPI target of between two to three per cent on average over the economic cycle. This target was announced by the Governor of the Reserve Bank in 1993. (Higher interest rates should be expected to have the effect of slowing the rate of economic growth and so reducing inflationary pressures. Lower interest rates might be expected to stimulate economic activity, although this could also result in some increase in the rate of inflation.) The central bank can control or influence short-term interest rates by buying or selling government stock in the money market. Such activity is known as open market operations.

(b) The RBA also has a key role in maintaining the stability of the financial system. In exceptional circumstances, this may require it to act as a lender to the banking system (lender of last resort). When the banking system is short of money and this lack of money is perceived to have serious implications for the rest of the banking system, the RBA will provide the money the banks need – at a suitable rate of interest. The function of lender of last resort was famously demonstrated...
10: Macroeconomic policy

by the UK’s central bank – the **Bank of England** – in Autumn 2007 when the Northern Rock building society applied to the Bank for financial assistance and the funds needed to cover its immediate liabilities. The saving of Northern Rock was one of the first episodes of the 2008/2009 global downturn.

(c) It acts as a **banker to the commercial banks**. All commercial banks keep a bank account with the RBA. This enables the cheque clearing system to operate. At the end of each day the net balances on each bank’s accounts with all the other banks are settled through their clearing accounts at the central bank. The funds which banks hold with the central bank act as a liquid reserve for the commercial bank, and are controlled by the fractional reserve ratio. This is controlled by the RBA’s Payments System Board.

(d) It acts as **banker to the government** and holds the ‘public deposits’.

(e) It is the **central note-issuing authority** in Australia – it is responsible for issuing bank notes in Australia.

(f) It is the **manager of the National Debt** – i.e. it deals with long-term and short-term borrowing by the central government and the repayment of central government debt.

(g) It is the manager of the **Foreign Currency Reserves**.

Supervision of the banking system is the responsibility of the Australian Prudential Regulation Authority (APRA). This supervision includes ensuring the banks have sufficient capital to cover any business losses (for example, bad debts) and sufficient liquidity to meet customers’ day-to-day requirements for cash.

### 8.1 The central bank as lender of last resort

In Australia, the short-term money market provides a link between the banking system and the government (the RBA) whereby the RBA lends money to the banking system, when banks which need cash cannot get it from anywhere else.

(a) The RBA will supply cash to the banking system on days when the banks have a cash shortage. It does this by buying eligible bills and other short-term financial investments from approved financial institutions in exchange for cash.

(b) The RBA will remove excess cash from the banking system on days when the banks have a cash surplus. It does this by selling bills to institutions, so that the short-term money markets obtain interest-bearing bills in place of the cash that they do not want.

The process whereby this is done currently is known as **open market operations** by the RBA. This simply describes the buying and selling of eligible bills and other short-term assets between the RBA and the short-term money market.

### 8.2 Open market operations and short-term interest rates

**Definition**

**Open market operations**: the RBA’s dealings in the capital market. The bank uses open market operations to control interest rates.

Open market operations provide the RBA (or other central banks) with a method of control over **short-term interest rates** through the control of broad and narrow money. They are therefore an important feature of a country’s monetary policy. Note that in Australia the RBA has been delegated the responsibility for the formulation of monetary policy. This is not the case in other countries. For example, in the UK, the central government is still responsible for formulating monetary policy and the central bank’s role is to administer the policy on its behalf.

When bills are bought and sold, they are traded at a discount to their face value, and there is an implied interest rate in the rate of discount obtained. Discounts on bills traded in open market operations have an immediate influence on other money market interest rates, such as the Australian and New Zealand **Official Cash Rate** (OCR) and these in turn influence the ‘benchmark’ base rates of the major banks.
Because the eligible bills and other assets which the RBA acquires in its money market operations are short-term assets, a proportion mature each day. The market is then obliged to redeem these claims and must seek further refinancing from the bank. This continual turnover of assets gives the RBA the opportunity to determine the level of interest rates day by day.

The RBA reviews the cash rate (OCR) regularly and announces its decision to raise or lower the rate, or keep the rate at its current level, giving reasons for its decision. For example in June 2011, with the CPI rate at 3.2 per cent, the Bank announced its decision to maintain the cash rate at 4.7 per cent. It gave as its reasons that:

(a) in spite of a temporary fall in GDP due to flooding and cyclones, there was strong underlying growth in the economy and investment by businesses
(b) the CPI rate was above the target level, but was expected to fall back when the temporary effects of extreme weather on prices had ended.

Consequently, the Bank had decided to continue with its ‘mildly restrictive’ monetary policy, with the interest rates (hopefully) applying some restrictions on the rate of inflation.

8.3 Controlling the supply of money

Our examination of macroeconomic policy throughout the Study Manual demonstrates how governments attempt to manage both economic growth and inflation. When we do so, we will see how interest rates are a vital part of economic policy.

However, for now, and in conjunction with looking at the open market operations of central banks, we should consider briefly what interest rates are within this context. In essence, interest rates reflect the price of money, or credit. And money is priced, like everything else, through the supply and demand for it. The key aspect to consider in connection with open market operations is how to control the supply of money.

Through its open market operations, a central bank will either buy government stock from commercial banks or sell government stock to them.

If the central bank buys government stock, the commercial banks will get extra money in return for the stocks they have sold to the central bank. Therefore, the supply of money will increase, and the bank’s ability to create credit through the credit multiplier will also increase. In effect, this increase in the supply of money should make money cheaper (reduce interest rates).

Conversely, if the central bank wants to reduce the supply of money through its open market operations (and thereby raise interest rates), it will look to sell government stocks. In buying the stocks, the commercial banks’ money base will be reduced, as will the level of credit they can subsequently create through the credit multiplier.

8.4 The RBA as an independent central bank

The RBA operates as an independent central bank and is solely responsible for the management of monetary policy.

Proponents of independence for central banks argue that independence can prevent the worst government monetary excesses, which in some cases result in hyperinflation. High levels of existing public expenditure commitments combined with electoral pressures (along with other factors) build in strong underlying inflationary pressures. An independent central bank is seen as an essential counterweight to the potentially reckless decisions of politicians. As well as avoiding the worst excesses, a strong central bank is regarded as vital for the shorter term stability of domestic prices and of the currency, and so is important to overseas trade. Any government wishing to reduce an already high rate of inflation will, however, have to listen carefully to the advice of its central bank if it is to have any real success.

Those arguing against independence point out that the central bank is an unelected body and therefore does not have the open responsibility of politicians. However, the danger of having an unelected and unaccountable body is minimised by the formal publication of decisions and recommendations of the central bank. Further, it is claimed that central bank views on monetary policy could be in conflict with other
economic objectives of the government. For example, excessively strict pursuit of monetary policy in order to pursue an inflation target might result in prolonged recession and heavy under-utilisation of resources.

Another example of an independent central bank is the **European Central Bank** which is designed to be totally free of political interference. It came into existence at the end of 1998, ready for the European single currency. The Bank of England is an example of a non-independent central bank. Its decisions can still be overridden by the UK Government in an emergency, and it does not have responsibility for setting inflation or monetary targets. That still rests with the government.
Key chapter points

- Macroeconomic policy objectives relate to economic growth, inflation, unemployment and the balance of payments.
- Fiscal policy provides a method of managing aggregate demand in the economy via taxation and government spending.
- Direct taxes have the quality of being progressive or proportional. Income tax is usually progressive, with higher rates of tax charged on higher bands of taxable income. Indirect taxes can be regressive, when the taxes are placed on essential commodities or commodities consumed by poorer people in greater quantities.
- A government must decide how it intends to raise tax revenues, from direct or indirect taxes, and in what proportions tax revenues will be raised from each source.
- A fiscal stance can be neutral, expansionary or contractionary. This depends on its effect on national income.
- Fiscal policy is used to create jobs and reduce unemployment. The balanced budget multiplier is an increase in aggregate monetary demand.
- Monetary policy focuses on the relationship between interest rates and the supply of money in an economy, and how the two of them together can influence aggregate demand.
- The effectiveness of monetary policy in influencing aggregate demand and unemployment is limited in the long-run.
- The central bank has various functions. These include acting as a banker to the central government and to the commercial banks.
- In Australia, the Reserve Bank of Australia (RBA) has responsibility for monetary policy chiefly through controlling inflation and maintaining a constant CPI by setting interest rates.
- The RBA is an example of an independent central bank.
Quick revision questions

1. Outline how the government may use fiscal policy to influence aggregate demand.

2. What is
   (a) a regressive tax?
   (b) a proportional tax?
   (c) a progressive tax?

3. The government of a certain country decides to introduce a poll tax, which will involve a flat rate levy of $200 on every adult member of the population. This new tax could be described as
   A regressive.
   B proportional.
   C progressive.
   D ad valorem.

4. High rates of personal income tax are thought to have a disincentive effect. This refers to the likelihood that the high rates of tax will
   A encourage illegal tax evasion by individuals.
   B lead to a reduction in the supply of labour.
   C lead to a reduction in savings by individuals.
   D discourage consumer spending and company investments.

5. The total yield from an indirect tax levied on a good is likely to be greatest when
   A demand is inelastic, supply is elastic.
   B demand is inelastic, supply is inelastic.
   C demand in elastic, supply is elastic.
   D demand is elastic, supply is inelastic.

6. Which of the following will not be the immediate purpose of a tax measure by the government?
   A to discourage an activity regarded as socially undesirable
   B to influence interest rates
   C to influence the level of aggregate demand
   D to raise revenue to spend on social or merit goods

7. Which of the following government aims might be achieved by means of fiscal policy?
   I A redistribution of income between firms and households.
   II A reduction in aggregate monetary demand.
   III A change in the pattern of consumer demand.
   A I and II only
   B I and III only
   C II and III only
   D I, II and III

8. What effect does an increase in interest rates have on the exchange rate?

9. List the likely functions of a central bank.
1. A government can increase demand by spending more itself or by reducing taxation so that firms and households have more after-tax income to spend.

2. A regressive tax takes a higher proportion of a poor person’s income than a rich person’s. A progressive tax takes a higher proportion of a rich person’s income and a lower proportion of a poor person’s. A proportional tax takes the same proportion of all incomes.

3. A flat-rate poll tax, with no concession for the lower-paid, would take a higher proportion of the income of lower-income earners than of higher income earners. This is a regressive tax system.

4. The disincentive effect refers specifically to the disincentive of individuals to work.

5. The total yield from an indirect tax is likely to be greatest when (a) demand for the good is relatively unaffected by the addition of a tax on to the price and (b) supply is relatively unaffected, even though suppliers will be receiving the price net of the tax.

6. The main purpose of taxation will be to raise revenue for the government. Other aims might be to redistribute wealth or affect demand in the economy. Changes in rate of tax do not have a direct influence on interest rates, which can be influenced by a government’s monetary policies.

7. Objective I could be achieved by raising (or lowering) taxes on firms and lowering (or raising) taxes on households. Objective II could be achieved by raising taxation in order to reduce consumers’ disposable income and so to reduce aggregate expenditure in the economy: these consequences should lead to a fall in the demand for money. Objective III can be achieved either by taxing income or by means of selective indirect taxes on certain goods.

8. A rise in interest rates attracts foreign investment, thus increasing the demand for the currency. The currency typically strengthens as a result.

   Banker to the government.
   Central issuer of banknotes.
   Manager of the national debt.
   Manager of the nation’s foreign currency reserves.
   Banker to the clearing banks.
   Lender to the clearing banks (lender of last resort).
   Supervision of the banking system.
### Answers to chapter questions

<table>
<thead>
<tr>
<th></th>
<th><strong>System 1</strong></th>
<th><strong>System 2</strong></th>
<th><strong>System 3</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Tax paid on low income</strong></td>
<td>20%</td>
<td>30%</td>
<td>10%</td>
</tr>
<tr>
<td><strong>Tax paid on high income</strong></td>
<td>25%</td>
<td>30%</td>
<td>5%</td>
</tr>
<tr>
<td><strong>Nature of tax</strong></td>
<td>Progressive</td>
<td>Proportional</td>
<td>Regressive</td>
</tr>
</tbody>
</table>

1. Demand for the product must presumably be inelastic, at least in the opinion of the producer. Otherwise he would fear to pass on the tax burden by increasing his prices as this would lead to a fall in demand.

2. Inflation would rise to about 10 per cent, which is the rate of inflation on $PC_2$ associated with 3 per cent unemployment. (Inflation and unemployment levels still follow $PC_2$, rather than shifting inwards to $PC_1$.) However, according to the natural rate hypothesis, in the longer run, unemployment would move back to 5 per cent. A new short-run Phillips curve would be established, according to which an unemployment rate of 5 per cent would be associated with 10 per cent inflation.
Chapter 11

Government intervention and income distribution

### Topic list

1. Market failure and regulation
2. Privatisation and denationalisation
3. Public policy towards monopolies
4. Income distribution and inequality

### Learning objectives

<table>
<thead>
<tr>
<th>Learning objectives</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Government intervention and income distribution</strong></td>
<td>LO11</td>
</tr>
<tr>
<td>Explain how the government may intervene to reduce misallocation of resources</td>
<td>LO11.1</td>
</tr>
<tr>
<td>Analyse ways to redress income inequalities</td>
<td>LO11.2</td>
</tr>
<tr>
<td>Explain the concept of income distribution and describe the Lorenz curve</td>
<td>LO11.3</td>
</tr>
<tr>
<td>Measure income inequality</td>
<td>LO11.4</td>
</tr>
</tbody>
</table>
Introduction

There is a role for government in the regulation of private markets where these fail to bring about an efficient use of resources, potentially resulting in market failure.

The first section of this chapter examines this role in terms of government regulation of private markets, privatisation and competitive practices.

The last section of the chapter looks at the role of government in measuring income and redressing income inequalities. Also explained in the last section of the chapter is income distribution across the population, specifically using the Lorenz curve.
Before you begin

If you have studied these topics before, you may wonder whether you need to study this chapter in full. If this is the case, please attempt the questions below, which cover some of the key subjects in the area.

If you answer all these questions successfully, you probably have a reasonably detailed knowledge of the subject matter, but you should still skim through the chapter to ensure that you are familiar with everything covered.

There are references in brackets indicating where in the chapter you can find the information, and you will also find a commentary at the back of the Study Manual.

1. What is deregulation? (Section 1.3)
2. Identify two advantages of privatisation. (Section 2.1)
3. Identify two disadvantages of privatisation. (Section 2.1)
4. What are the three types of inefficiencies associated with State monopolies? (Section 2.2)
5. What is the primary function of the ACCC? (Section 3.4.1)
6. What is the Gini coefficient? (Section 4.2)
7. What was Australia’s Gini coefficient in 2007-08? (Section 4.3)
1 Market failure and regulation

Section overview

- **Market failure** is said to occur when the market mechanism fails to result in economic efficiency, and therefore the outcome is sub-optimal.
- This results in the government intervening through the regulation of private economic activity. Markets can be self regulated or subject to much government regulation.
- Many governments, including those of Australia and the UK, have been pursuing a policy of deregulation over the last few decades.

<table>
<thead>
<tr>
<th>Market failure</th>
<th>Regulation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Imperfect competition</td>
<td>Where monopoly power is leading to inefficiency, the State will intervene through controls on, say, prices or profits in order to try to reduce the effects of the monopoly.</td>
</tr>
<tr>
<td>Externalities</td>
<td>A possible means of dealing with the problem of external costs and benefits is via some form of regulation. Regulations might include, for example, controls on emissions of pollutants, restrictions on car use in urban areas, the banning of smoking in public buildings, compulsory car insurance and compulsory education.</td>
</tr>
<tr>
<td>Imperfect information</td>
<td>Regulation is often the best form of government action whenever informational inadequacies are undermining the efficient operation of private markets. This is particularly so when consumer choice is being distorted. Examples here would include legally enforced product quality/safety standards, consumer protection legislation, and the provision of job centres and other means of improving information flows in the labour market.</td>
</tr>
<tr>
<td>Equity</td>
<td>The government may also resort to regulation for social reasons. For example, legislation to prevent racial, sexual or age discrimination in the labour market; regulation to ensure equal access to goods such as health care, education and housing; minimum wage regulations and equal pay legislation.</td>
</tr>
</tbody>
</table>

Regulation can be defined as any form of State interference in the operation of the free market. This could involve regulating demand, supply, price, profit, quantity, quality, entry, exit, information, technology, or any other aspect of production and consumption in the market.

With specific reference to competition and free market economics, there are two aspects which government will regulate:

- Mergers which would create monopolies.
- Restrictive trade practices which reduce competition in a market, for example price fixing.

1.1 Self-regulation

Bear in mind that in many markets the participants (especially the firms) may decide to maintain a system of voluntary self-regulation, possibly in order to try to avert the imposition of government controls. Self-regulation often exists in the professions, for example, the Certified Practising Accountants (CPA), Financial Planning Association (FPA), the Law Society and other professional bodies.

1.2 Costs of regulation

1.2.1 Potential costs of regulation

(a) **Enforcement costs.** Direct costs of enforcement include the setting up and running of the regulatory agencies. Indirect costs are those incurred by the firm being regulated in conforming to the restrictions. These requirements will add to the firms’ costs of production and ultimately to their prices.
(b) **Regulatory capture** refers to the process by which the regulator becomes dominated and controlled by the regulated firms, such that it acts increasingly in the latter's interests, rather than those of consumers. This is a phenomenon which has been observed in the US.

(c) **Unintended consequences of regulation.** Firms will not react passively to regulatory constraints on their behaviour; they will instead try to limit the effectiveness of the constraints. In general, theory and observation suggest that firms will substitute away from the regulated activity towards those which are less constrained or completely unregulated.

### 1.3 Deregulation

Deregulation can be defined as the removal or weakening of any form of statutory (or voluntary) regulation of free market activity. Deregulation allows free market forces more scope to determine the outcome.

There was a shift in policy in the 1980s in Australia and the UK towards greater deregulation of markets, in the belief that this would improve efficiency. For example, many politicians and commentators in the UK believed that it was over-regulation of British industry that was largely responsible for Britain's uncompetitive and inefficient performance in comparison to other developed nations in the preceding years.

A rational assessment of deregulation should weigh the potential social benefits against the social costs. If there will be a net gain to society, we can say that the deregulation should proceed. It would be simplistic to contend that all regulation is detrimental to the economy. As we have seen, where there is a clear case of market failure, then State regulation may be required.

#### 1.3.1 Advantages and disadvantages of deregulation

Deregulation measures, whose main aim is to introduce more competition into an industry by removing statutory or other entry barriers, are also known as **liberalisation**. The benefits of liberalising an industry include the following:

(a) **Improved incentives for internal/cost efficiency.** Greater competition compels managers to try harder to keep down costs.

(b) **Improved allocative efficiency.** Competition keeps prices closer to marginal cost, and firms therefore produce closer to the socially optimal output level.

In some industries, liberalisation could have certain disadvantages:

(a) **Loss of economies of scale.** If increased competition means that each firm produces less output on a smaller scale, unit costs will be higher.

(b) **Lower quality or quantity of service.** The need to reduce costs may lead firms to reduce quality or eliminate unprofitable but socially valuable services.

(c) **Need to protect competition.** It may be necessary to implement a regulatory regime to protect competition where inherent forces have a tendency to eliminate it, for example if there is a dominant firm already in the industry, such as Telstra in telecommunications in Australia. In this type of situation, effective regulation for competition will be required, i.e. there will need to be regulatory measures aimed at maintaining competitive pressures, whether existing or potential.

#### 1.3.2 Deregulation and liberalisation in Australia

In Australia deregulation and liberalisation have taken place in the areas of electricity supply, the labour market, banking and the media:

(a) The deregulation of the **electricity supply industry** by the various State and Territory governments throughout the 1990s.

(b) The Hawke/Keating Labour Government extensively deregulated the **labour market** in the late 1980s as part of a series of large scale monetary reforms following up from **deregulation in the money market** in the early part of the decade.

(c) The **banking system** was deregulated under the 1980s reforms referred to above.

(d) There has been much call for further deregulation of the media which is governed by the Federal Government’s Broadcasting Services Act 1992. To date, further deregulation measures in this area have not been successful.
2 Privatisation and denationalisation

Section overview

- **Privatisation** occurs when the government transfers State-operated industries to the private sector. This can be through contracting out services, deregulation or transferring asset ownership.
- The Australian telecommunications industry was opened to competition in 1997 and **Telstra** is still obligated to provide service to all residents.

Definition

**Privatisation** is the transfer by government of state owned activities to the private sector.

Privatisation as originally envisaged takes three broad forms:

(a) The **deregulation of industries**, to allow private firms to compete against State-owned businesses where they were not allowed to compete before (for example, deregulation of bus and coach services; deregulation of telecommunications).

(b) **Contracting out** work to private firms, where the work was previously done by government employees — for example, refuse collection or hospital laundry work.

(c) **Transferring the ownership of assets** from the State to private shareholders, for example in Australia with the denationalisation of Telstra, Qantas and many other enterprises.

The Australian Government, like many other government of developed countries, has been carrying out a policy of denationalisation since the 1970s. Telecoms, electricity and gas suppliers and water authorities are among the enterprises which have been privatised. Many of the utility industries which have been privatised are still subject to regulations though, for example, the price increases pegged to the Consumer Prices Index (CPI) minus a specified percentage.

2.1 For and against privatisation generally

The following are **possible advantages of privatisation**:

(a) Privatised companies may be **more efficient** than State monopolies and private sector managers are likely to try to reduce costs and strip out unproductive labour. Private companies may also provide **better quality** because they will have to compete to survive. The threat of competition may also lead to innovation.

(b) Denationalisation provides an immediate **source of money** for the government, through the sale of assets or businesses.

(c) Privatisation **reduces bureaucratic and political meddling** in the industries concerned.

(d) Privatised companies may have a **more flexible** and **profit-oriented management culture**.

(e) There is a view that **wider share ownership** should be encouraged. Denationalisation is one method of creating wider share ownership, as the sale of Telstra and NRMA shares have demonstrated in Australia. If workers own shares in their company, they are more likely to want it to be successful.

There are arguments against privatisation too:

(a) State-owned industries are more likely to respond to the **public interest**, ahead of the profit motive. For example, State-owned industries are more likely to cross-subsidise unprofitable operations from profitable ones. For example, Australia Post will continue to deliver letters to the remote outback stations even though the service might be very unprofitable. But privatisation may lead to fewer deliveries and higher prices.

(b) Encouraging private competition to State-run industries might be inadvisable where **significant economies of scale** can be achieved by monopoly operations.
Government can provide capital more cheaply than the market to industries whose earning potential is low, but which are deemed to be of strategic importance, such as aircraft manufacture. Opponents of privatisation suggest that the very idea of privatising a strategic industry is spurious.

State-owned industries can be run in a way that protects employment, as in the case in China. The problem with this is that the taxpayer is effectively subsidising technical inefficiency.

Surpluses from State-run industries can be used for public welfare rather than private wealth. However, the problem here is that points (a) and (d) above tend to preclude the creation of surpluses.

### Privatisation, efficiency and competition

The inefficiency associated with state monopolies was discussed earlier and could be said to be of three types:

1. **Technical inefficiency.** A firm is only technically efficient when it uses the least-cost combination of productive resources for a given level of output. Failure to do this leads to its not operating at the lowest possible cost per unit of output. (This is also known as productive inefficiency.)

2. **Allocative inefficiency.** Second, there is allocative inefficiency where price is higher than marginal cost such that the good or service is under-produced and under-consumed.

3. **X-inefficiency.** The monopolist's privileged position relieves it of the need to exert constant effort to keep costs down. The resulting rise in costs is called x-inefficiency.

The contention that privatisation will lead to efficiency improvements is based on a number of assumptions. The most important assumption is that privatisation will place the industry in a market subject to competitive pressures approaching the perfectly competitive ideal. In practice, however, there are few examples of perfectly competitive markets and the likelihood is of restricted competition between a small number of large firms, with a degree of inefficiency persisting. If there are no legal restrictions, it is also possible that State monopolies will merely become private monopolies with even less regard than before for the consumer.

Another assumption of the privatisation analysis is that there are no economies of scale. Where there are significant economies of scale there may be a natural monopoly, with room in the market for one firm only, and monopoly may be productively more efficient than a market with competing firms.

It is believed that management and working practices will be changed under private sector control. Management will become largely free of interference from the government, working practices should become more efficient and trade union power may be reduced.

It is claimed that with the central objective of profit maximisation, privatised firms will be responsive to the wants of consumers, as revealed by market research and signalled by the operation of the price mechanism. The profit objective induces firms to innovate and to seek out new markets, a process which promotes the efficient allocation of resources.

Where privatisation increases competition, the greater competition is likely to make firms produce output more cheaply and to sell it at a lower price. Nationalised organisations have often acted as monopolists, with the consequences of higher prices and lower output characteristic of non-competitive markets.

### Privatisation in practice

In Australia, there has been only a limited increase in competition following some privatisations. Indeed, some organisations have been sold as monopolies to increase their attractiveness to shareholders in effect transferring a public monopoly to a private monopoly. As monopolies, they would aim to produce at the profit maximising level of output, which leads to higher prices than perfect competition, and does not create technical or allocative efficiencies like perfect competition does.

To try to avoid creating private sector monopolies, the government has tried to create competition alongside privatisation.
Privatisation alone is often not sufficient to improve the performance of a monopoly and other steps may need to be taken to increase competition. On the other hand, improvements to the competitive environment might equally be achieved without privatisation.

**Case study: Australian telecommunications industry**

The example of Telstra in Australia can be used to illustrate the attempt to encourage competition along with privatisation, while also ensuring core services are provided to the public. In 1997 the Government-owned telecommunications monopoly, Telstra, was partially privatised. In the same year the telecommunications industry was opened up to full competition. By 2007 the Government had sold or transferred its remaining shares in Telstra.

The introduction of full competition in the telecommunications industry led to a large number of small to medium competitors coming into the market relatively quickly, particularly in the mobile phone market. The largest competitor in the home phone, mobile and Internet market is Optus.

A difficulty for competitors entering the market was the fact that Telstra owned the existing telecommunications infrastructure, a difficulty exacerbated by the size of the country and location of remote communities. Without access to this infrastructure, competition could not prosper. In response, there were regulations placed upon privatisation on Telstra concerning competitors’ rights to access existing infrastructure. These have not always been fully effective as indicated by the fact that Telstra still owns a large majority of the home phone/fixed line market.

The privatisation of Telstra raised many concerns about ensuring continuity of service throughout the country. Telstra is still subject to much government regulation in regards to services, and the Commonwealth government ensures that Telstra provides services to all Australians under the Universal Services Obligation. It also sets provisions regarding service standards, faults and new connections.

### 2.4 Criticisms of privatisation in practice

There are also a number of criticisms of privatisation in practice:

(a) Critics argue that privatisation has not enhanced competition, and in some cases it has merely transferred a public monopoly to a private monopoly.

(b) Quality of service has diminished where privately-owned companies have tried to cut costs on services previously provided centrally.

(c) In some cases the level of service has been reduced and prices raised as private companies do not want to operate loss-making services (for example, on air routes in regional Australia.) Whereas nationalised organisations were prepared to cross-subsidise loss-making elements of their business, private sector companies are less likely to be prepared to do this. They are more likely to focus on the profitable elements of their business and discontinue the less profitable elements.

(d) The assets sold by governments have been undervalued and this allowed private investors to make large capital gains by acquiring the assets.

(e) In some industry sectors, privatisation has been selective, and only the profitable parts of the sector have been sold off. This means the unprofitable areas remain in the public sector, and are a drain on public funds.

(f) Top executives of privatised companies have been granted very large salaries and share options, which looks insensitive in the context of trying to improve employee efficiency and competitiveness.

### Question 1: Benefits of privatisation

Can you identify three possible benefits of privatisation?  

(The answer is at the end of the chapter)

---

Economics and Markets
3 Public policy towards monopolies

Section overview

- There are various arguments in favour of public (or 'nationalised') ownership of production concerning public goods, merit goods and natural monopoly. It is important to also be aware of the efficiency arguments against public ownership.
- The chief instrument of Australian government regulation in the area is now the Competition and Consumer Act 2010 which is by the Australian Competition and Consumer Commission (the ACCC), jointly with the respective State and Territory regulators for the legislation on consumer protection.

Public sector organisations are all ultimately responsible (accountable) to government, and their purposes are defined in the laws that established them. Public sector bodies' objectives will usually be defined in terms of the provision of goods or services which are beneficial to society, rather than in terms of the profit maximising objectives of private sector firms.

3.1 Public policy towards monopolies

Monopolies might be harmful or beneficial to the public interest.

(a) A beneficial monopoly is one that succeeds in achieving economies of scale in an industry where the minimum efficiency scale is at a level of production that would mean having to achieve a large share of the total market supply.

(b) A monopoly would be detrimental to the public interest if cost efficiencies are not achieved. Oliver Williamson suggested that monopolies might be inefficient if 'market power provides the firm with the opportunity to pursue a variety of other-than-profit objectives'. For example, managers might instead try to maximise sales, or try to maximise their own prestige.

3.2 Methods of government control

There are several different ways in which a government can attempt to control monopolies.

(a) It can stop them from developing, or it can break them up once they have been created. Preventing monopolies from being created is the reason why a government might have a public policy on mergers.

(b) It can take them over. Nationalised industries are often government-run monopolies, and central and/or local government also have virtual monopolies in the supply of other services, such as health, the police, education and social services. Government-run monopolies or nationalised industries are potentially advantageous.

(i) They need not have a profit-maximising objective so that the government can decide whether or not to supply a good or service to a household on grounds other than cost or profit. In this way, services which provide a social benefit rather than an economic profit will be provided, whereas they would not be under free market forces.

(ii) The government can regulate the quality of the good or service provided more easily than if the industry were operated by private firms. The government can also keep strategic control over provision of the goods and services, and can maintain strategic control over key resources.

(iii) Key services can be protected (for example, health and education), and capital can be made available for investment where it might not be in the private sector.

(iv) Nationalised industries may benefit from economies of scale, particularly if they require significant investment in infrastructure.

(v) The government can protect employment and reduce social costs related to unemployment. For example, this argument was used to justify keeping 'uneconomic' coal mines open in the UK. If the mines were closed, as free market economics would dictate, this would reduce production, but it would also increase unemployment and the related costs of unemployment benefits and social unease.
Nationalised industries may also promote a fairer distribution of wealth, since economic surpluses can be reinvested for the benefit of society rather than being related as profit by capitalist entrepreneurs.

It can allow monopolies or oligopolies to operate, but try to control their activities in order to protect the consumer. For example, it can try to prohibit the worst forms of restrictive practice, such as price cartels. Or it may set up regulatory ‘consumer watchdog’ bodies to protect consumers’ interests where conditions of natural monopoly apply.

Control over markets can arise by firms eliminating the opposition, either by merging with or taking over rivals or stopping other firms from entering the market. When a single firm controls a big enough share of the market it can begin to behave as a monopolist even though its market share is below 100 per cent.

Several firms could behave as monopolists by agreeing with each other not to compete. This could be done in a variety of ways – for example, by exchanging information, by setting common prices or by splitting up the market into geographical areas and operating only within allocated boundaries.

In a perfect monopoly, there is only one firm that is the sole producer of a good that has no closely competing substitutes, so that the firm controls the supply of the good to the market. The definition of a monopoly in practice is wider than this, because governments seeking to control the growth of monopoly firms will probably choose to regard any firm that acquires a certain share of the market as a potential monopolist.

### 3.3 Pricing in nationalised industries

Nationalised industries are monopolies, and so if they operated like private-sector monopolies they would earn supernormal profits by producing where $MC = MR$.

However, the logic of public sector organisations is different to the private sector; an alternative requirement of public sector monopolies can be that they produce at breakeven levels. (Remember, breakeven occurs where total cost equals total revenue, and therefore also where average cost equals average revenue.)

Figure 1 below illustrates monopoly output under breakeven pricing.

![Figure 1 Breakeven pricing in a monopoly](image)

The breakeven output ($Q_{BE}$) is greater than it would have been under a profit maximising monopolist ($Q_{PM}$). However, the public sector monopoly is neither technically nor allocatively efficient.

For technical efficiency to be achieved, output quantity would need to be $Q_{T}$, so that the firm is producing at the lowest point on its average cost curve.

A second option would be for a nationalised industry to produce at the level of allocative efficiency (Price = MC) but this could cause it to make a loss, if marginal cost (and therefore, price) are less than average cost.

However, nationalised monopolies do still have the option to discriminate on price to increase profits. In the Australian State of NSW, where railway services are run by a State-owned company – State Rail – price discrimination is applied to charge commuters different (higher) fares to leisure travellers.
3.4 Australian Regulation and Competition Policy

There are three main aspects of government policy:

• Investigating mergers and acquisitions which might lead to the creation of a monopoly.

• Investigating restrictive trade practices in a market (for example, collusion by suppliers) which undermine competition and consumer sovereignty in the market.

• Consumer protection.

Under the Australian Constitution, the 'corporations' power' means that the regulation of corporations (including commonwealth monopolies, public and private companies) is a responsibility of the Federal Government. Consumer protection has a responsibility of both Federal and State and Territory Governments.

The chief piece of Federal legislation delivering government policy is the Competition and Consumer Act 2010 (CCA), which replaced the Trade Practices Act 1974 (the TPA). The CCA contains much that is the same as or similar to the TPA in matters of competition policy but introduced new regulations relating to consumer protection. The Act deals with anti-competitive conduct, unfair market practices, industry codes, mergers and acquisitions of companies, product safety, product labelling, price monitoring, and regulated industries. The main anti-competitive conduct regulations are described below:

(a) **Prohibited outright**: Certain types of anti-competitive conduct are prohibited outright and these are price agreements or 'fixing', exclusionary provisions, and primary boycotts. Primary boycotts are a combined refusal to deal with a third party.

(b) **Mergers**: These are prohibited if they will have a likely effect of substantially lessening competition.

(c) **Unconscionable conduct**: This is where a stronger party exploits the position of a weaker party to gain advantage. This is prohibited in small business dealing and trade and commerce generally.

(d) **Secondary boycotts**: They are prohibited if the action will call a substantial loss of competition to the market, or a substantial loss to a business. A secondary boycott is where two persons engage in conduct that prevents a third person supplying goods to a fourth person.

(e) **Misuse of market power**: If a corporation has a substantial degree of market power they cannot take advantage of this power to eliminate or damage any actual or potential competitor. For example, Telstra must comply with provisions to allow competitor access to existing telecommunications infrastructure which was created under monopolistic conditions.

(f) **Enforcement**: The ACCC has a range of enforcement measures that it can use for dealing with anti-competitive conduct by corporations. These include the power to take civil action against corporations for anti-competitive behaviour or unfair commercial practices against consumers. An offending corporation can also be ordered to pay compensation to third parties.

The CCA contains extensive regulations to protect consumers from unfair practices, known as Australian Consumer Law (ACL). These are contained in Schedule 2 of the CCA. They cover issues such as unfair terms in standard form consumer contracts, a national product safety regime, national consumer guarantee provisions, rules dealing with unsolicited consumer agreements, and provisions dealing with information standards for goods and services.

3.4.1 The ACCC

The Australian Competition and Consumer Commission – the ACCC – is the Federal Government’s independent statutory authority. It was established by the Federal Government in 1995 and its role is now to administer the CCA. It is the only national agency in this area and is therefore the number one authority in the country. Its main responsibility is to ensure that businesses and individuals comply with competition, consumer and fair trading laws. As part of this role it is responsible for monitoring any anti-competitive practices in business and industries as well as monitoring regulated industries (such as electricity services, water authorities and telecommunications). For example, it is the ACCC which monitors the behaviour of Telstra to ensure it meets its Universal Services Obligation referred to earlier in the Telstra Case study.

Although there are similar statutory bodies in each State and Territory, their focus is more on consumer protection issues rather than anti-competitive practice. The CCA provides for joint enforcement of the consumer protection regulations by the ACCC and the respective State or Territory regulator.
Exam comments

Although the ACCC is specific to Australia so you won’t be asked questions specifically about the details of the ACCC, you could still be set questions about public control of monopolies and mergers as this is part of its measures in reducing private misallocation of resources.

4 Income distribution and inequality

Section overview

- The Lorenz curve is the most common tool used to show the distribution of income within a population. The Gini coefficient measures income inequality within a population using a ratio based on the Lorenz curve.
- Government measures to address income inequalities are through progressive taxation systems, provision of national services including education and ensuring a minimum wage.

4.1 The Lorenz curve

The Lorenz curve is used to measure income distribution within a population.

A Lorenz curve of income distribution would be drawn on a horizontal axis representing the percentage of households in the population starting from 0 to 100 per cent of households. The vertical axis shows the percentage of total income.

![Figure 2 Lorenz curve](image)

A straight line at 45 degrees from the origin would show a perfectly equal income distribution in the population, meaning each person has the same income. This line is called the ‘line of perfect equality’. A perfectly equal income distribution is one where each person has the same income. A perfectly unequal distribution is where one person has all the income and everyone else has none. The curve in the Lorenz curve in Figure 2 shows a possible actual income distribution.

The x axis for the curve starts on the left with the lowest income earning households, and the highest income households are included on the right. The x axis therefore moves from poorest to richest. Since the poorest households enjoy the lowest percentage of total national income, the curve has a convex shape.

The steep rise in the curve on the right shows that the highest-earning proportion of the population earn the highest percentage share of national income.
Both situations are, however, unrealistic: the Lorenz curves typically display increasing gradients as the horizontal coefficient increases towards one hundred per cent. Every point on the Lorenz curve has a representative income such as 'the top 20 per cent of all households have 5 per cent of total income'. The Lorenz curve for each population will vary.

4.2 The Gini coefficient

The construction of the Lorenz curve is such that the greater the area between the curve and the 45 degree line, the more unequal the distribution of income. The Gini coefficient measures the deviation of the Lorenz curve from the 45 degree line. It is the ratio of the area between the curve and the 45 degree line to the whole area below the 45 degree line. In Figure 2, this is the ratio of area A to (area A + area B). The Gini coefficient of perfect income equality would therefore be zero: that is to say, the Lorenz curve would not deviate from the 45 degree line, as explained above. The Gini coefficient of perfectly unequal income distribution would be unity: area B would disappear completely, since the Lorenz curve would run along the horizontal axis until the one hundred per cent of population point was reached and then would rise vertically. Thus, any Gini coefficient will be between zero and one: the higher it is, the more unequal the distribution of income.

4.3 Addressing income inequalities within a population

Income inequality is present in all free market and mixed economies. In a centrally planned (or command) economy, the decisions and choices about resource allocation are made by the government. The command economy ideology is based on the theory that only the government can make fair and proper provision for all members of society. However, all command economies have failed in their quest for perfect equality of income due to the black market, corruption, government incompetence and/or repressive political regimes. These are all some of the reasons for the failure of the Soviet Union and the end of the Cold War in the early 1990s.

Unequal distribution of income has been growing in many free market and mixed economies countries since the 1990s, including Australia, New Zealand, the US and the UK. Australia’s Gini co-efficient was 0.319 in 2007-08, up 5.6 per cent from the 1994-95 measure of 0.302.

Incomes can be influenced by governments chiefly through the operation of tax and social security systems. Specific measures of addressing income inequality in a population are as follows:

(a) Progressive taxation systems are used by governments to redistribute wealth from the rich to the poor. The advantages and disadvantages of progressive taxation as a means of addressing income inequality are discussed in Chapter 10, Sections 3.1 – 3.3.

(b) Setting a universal minimum wage in the labour market. Over 90 per cent of countries have minimum wage legislation in place, and its purpose is to ensure low-paid workers earn enough income to maintain an acceptable standard of living. The disadvantage of minimum wage legislation may be that job losses are incurred as low paid jobs are taken offshore or made redundant through technological advantage. This may actually reduce the incomes of the lowest skilled members of society. Minimum wages are discussed in Chapter 2, Section 5.4.

(c) Nationalisation and subsidisation of the supply of goods and services. Governments can choose to supply essential goods and services to the market to lessen inequality of access across society. The nationalisation of healthcare (for example, Australia’s Medicare system) is an example of a nationalised service. Another example is the provision of subsidised housing to those on lower incomes.

(d) Public education is an example of a nationalised service. The provision of good quality public education is integral to increasing the number of skilled workers and reducing income inequality in a society.
• Market failure is said to occur when the market mechanism fails to result in economic efficiency, and therefore the outcome is sub-optimal.

• This results in the government intervening through the regulation of private economic activity. Markets can be self regulated or subject to much government regulation.

• Many governments, including those of Australia and the UK, have been pursuing a policy of deregulation over the last few decades.

• Privatisation occurs when the government transfers State-operated industries to the private sector. This can be through contracting out services, deregulation or transferring asset ownership.

• The Australian telecommunications industry was opened to competition in 1997 and Telstra is still obligated to provide service to all residents.

• There are various arguments in favour of public (or 'nationalised') ownership of production concerning public goods, merit goods and natural monopoly. It is important to also be aware of the efficiency arguments against public ownership.

• The chief instrument of Australian government regulation in the area is the Competition and Consumer Act (CCA) 2010 which is administrated by the Australian Competition and Consumer Commission (the ACCC).

• The Lorenz curve is the most common tool used to show the distribution of income within a population. The Gini coefficient measures income inequality within a population using a ratio based on the Lorenz curve.

• Government measures to address income inequalities are through progressive taxation systems, provision of national services including education and ensuring a minimum wage.
Quick revision questions

1. In what circumstances might government regulation of markets have an economic justification?
2. What different forms can privatisation take?
3. Why might a government wish to control monopolies?
4. How might a government be able to control monopolies?
5. Which of the following are arguments in favour of a policy of privatisation?
   I. It will reduce X-inefficiency because competition is always introduced.
   II. It raises useful short term funds for the government.
   III. It encourages a more profit-oriented management culture.
   A. I and II only
   B. I and III only
   C. II and III only
   D. I, II and III
6. Explain the concept of perfectly equal income distribution.
7. Which one of the following is not a measure generally used to address income equality in a population?
   A. minimum wage legislation
   B. progressive taxation
   C. regressive taxation
   D. nationalisation of services
Answers to quick revision questions

1. The undesirable effects of various forms of market failure can be reduced by government action. Monopoly power can be attacked by regulation of price or profit or even the break-up of the monopoly firm. Externalities can be reduced by bans on some forms of behaviour, such as dangerous pollution, and by levying taxes on others such as the consumption of alcohol. Where imperfect information distorts consumer choice, legally enforceable product standards and disclosure requirements can improve the operation of the market. Finally, governments may intervene for social and political reasons, banning racial discrimination and enforcing a minimum wage, for example.

2. Privatisation can take three forms:
   - Deregulation allows private firms to compete against State-owned organisations.
   - Work done by government employees may be contracted out.
   - State-owned businesses may be sold to private shareholders.

3. There are arguments both for and against monopolies. Monopolies are detrimental to the public interest when they are inefficient in their allocation of resources and their operations generally. They may also be objected to on the grounds that they charge higher prices, and hence earn higher profits, than would be possible in a competitive market; and they restrict choice.

4. Governments can regulate monopolies, particularly their prices and the quality of their goods and services; prevent them from developing; break them up; or take them into public ownership.

5. C Selling off nationalised industries can raise useful funds for the government, and private companies are likely to have a more profit-oriented management culture than State-owned monopolies. However, competition is not always introduced following privatisation. A firm could simply be transferred from being a State-owned monopoly to being a private sector monopoly; this would not reduce X-inefficiency.

6. Perfectly equal income distribution is where each person in a population has the same income. It is represented on the Lorenz curve by a straight line of a 45 degree angle – the line of perfect equality.

7. C Regressive taxation actually serves to punish those on lower incomes as it is a flat tax on goods and services across the population.
Answers to chapter questions

1 Any three of:
   – improved efficiency of production in privatised companies
   – reduces bureaucracy and political interference in the industries concerned, so producers have greater economic freedom
   – private companies have a more profit-oriented management culture
   – encourages wider share ownership
Part 2:

Statistics
Chapter 12

Statistical analysis, data, and methods of describing data

<table>
<thead>
<tr>
<th>Learning objectives</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Statistical analysis, data, and methods of describing data</td>
<td>LO12</td>
</tr>
<tr>
<td>Explain the role of statistical analysis in decision making</td>
<td>LO12.1</td>
</tr>
<tr>
<td>Distinguish between quantitative and qualitative data</td>
<td>LO12.2</td>
</tr>
<tr>
<td>Explain and apply the different sampling methods</td>
<td>LO12.3</td>
</tr>
<tr>
<td>random sampling</td>
<td>LO12.3.1</td>
</tr>
<tr>
<td>cluster sampling</td>
<td>LO12.3.2</td>
</tr>
<tr>
<td>stratified sampling</td>
<td>LO12.3.3</td>
</tr>
<tr>
<td>Describe the different methods of collecting data and statistical information</td>
<td>LO12.4</td>
</tr>
<tr>
<td>survey</td>
<td>LO12.4.1</td>
</tr>
<tr>
<td>published source</td>
<td>LO12.4.2</td>
</tr>
<tr>
<td>Explain the different levels of data measurement</td>
<td>LO12.5</td>
</tr>
<tr>
<td>nominal-level data</td>
<td>LO12.5.1</td>
</tr>
<tr>
<td>ordinal-level data</td>
<td>LO12.5.2</td>
</tr>
<tr>
<td>interval-level data</td>
<td>LO12.5.3</td>
</tr>
<tr>
<td>ratio-level data</td>
<td>LO12.5.4</td>
</tr>
<tr>
<td>Describe different ways of presenting data</td>
<td>LO12.6</td>
</tr>
<tr>
<td>Construct a bar graph, a pie chart, a histogram and a scatter diagram from a given set of data</td>
<td>LO12.6.1</td>
</tr>
<tr>
<td>Interpret data presented in a bar graph, a pie chart, a histogram and a scatter diagram</td>
<td>LO12.6.2</td>
</tr>
</tbody>
</table>

Topic list

1. Statistical analysis
2. Sampling
3. Survey methods for collecting statistical information
4. Published sources of statistical information
5. Levels of data measurement
6. Data presentation
7. Graphical representation of data
Introduction

Organisational data is a collection of raw facts relating to the entity and its environment. It can be classified in a number of ways e.g. quantitative/qualitative, discrete/continuous, internal/external, formal/informal, primary/secondary.

Data must be processed or analysed in some way to form information that is useful in the decision-making process of the organisation. Much of a manager’s work will involve the use of data and information, collected internally or externally. Decisions regarding the future plans and operations of the organisation will incorporate information about past performance, future market potential, industry and company statistics etc, all of which will need to be gathered, processed and analysed. You have to analyse and present the data you have collected so that they can be of use and in this chapter we look at how data can be presented in tables and charts. Such methods are helpful in presenting key data concisely and in a way that is easy to understand. However, a note of caution, they are purely descriptive and offer little opportunity for further detailed numerical analysis of a situation.

Data that is a mass of numbers can usefully be summarised into a frequency distribution (effectively a table which details the frequency with which a particular value occurs). Histograms and scatter diagrams (scattergraphs) are the pictorial representation of frequency distributions and provide the link between the purely descriptive approach to data analysis and the numerical approach.
Before you begin

If you have studied these topics before, you may wonder whether you need to study this chapter in full. If this is the case, please attempt the questions below, which cover some of the key subjects in the area.

If you answer all these questions successfully, you probably have a reasonably detailed knowledge of the subject matter, but you should still skim through the chapter to ensure that you are familiar with everything covered.

There are references in brackets indicating where in the chapter you can find the information, and you will also find a commentary at the back of the Study Manual.

1. Explain the difference between objective and subjective decision-making. (Section 1.1)
2. Look through the following list of surveys and decide whether each is collecting data on attributes, discrete variables or continuous variables:
   a. a survey of maths students, to determine how many have red hair
   b. a survey of tinned peas in a supermarket, to determine how many are damaged
   c. a survey of swimmers to find out how long they take to swim 50 metres
   d. a survey of the number of pensioners over 75 in Australia
3. List the arguments in favour of using a sample. (Section 2.1)
4. What is stratified sampling? (Section 2.2)
5. Retail prices are very important to a wide variety of users. Fill in the gaps below: (Section 4.1.2)
   a. for .................................. the Consumer Price index (CPI) indicates the degree of success there has been in fighting inflation.
   b. for .................................. the CPI may give an indication of how much wages need to rise to keep pace with inflation.
   c. for .................................. the CPI indicates the increases to be expected in the prices of goods in shops.
   d. for .................................. the CPI may give a broad indication of how much costs should have been expected to rise over recent years and months.
   e. for .................................. and ................................ recipients, the movement in the CPI is used to update benefit levels............
6. ‘Tiny’ < ‘little’ < ‘medium-sized’ < ‘huge’ < ‘infinite’ is an example of which scale level? (Section 5.1)
   A. ordinal
   B. ratio
   C. nominal
   D. interval
7. Which of the following graphical methods would be most suitable to illustrate the occupations of 50 female adults? (Section 6.2)
   A. line graph
   B. pie chart
   C. histogram
   D. component bar chart
1 Statistical analysis

Section overview

Today's good decisions are driven by data. In all aspects of our lives, and importantly in the business context, an amazing diversity of data is available for inspection and analysis. Business managers and professionals are increasingly required to justify decisions on the basis of data.

1.1 Statistical analysis in decision-making

Definition

Decision-making is the selection of a course of action from among alternatives. A decision is a choice whereby a person forms a conclusion about a situation.

Decision-making can be subjective or objective. In objective decision-making, decision models are used in an attempt to eliminate bias or hunch, and to ensure a decision meets the objectives of the organisation. Subjective decision-making involves choosing an action that produces the best possible outcome based on the individual's preferences, prejudices, and other subjective factors. For example, in a business an individual might decide to produce a new product in red because they like the colour. An objective decision would involve the statistical analysis of relevant market research and other investigations as to the relative cost and benefits of different colours.

The purpose of statistical analysis is to provide estimates or comparisons on which sound decisions can be made. For example:

(a) Estimates – when formulating plans for the future, estimates based on present knowledge are essential to government and industry alike. A local authority planning to build new schools must estimate the number of children likely to require admission during the next few years. A manufacturing firm deciding whether to build a new plant, or launch an export drive or an advertising campaign, requires estimates of future demand as well as knowledge of current trade and production statistics. Such estimates may be based partly on statistics of population, incomes and regional employment and partly on a mathematical analysis of past records projecting into the future.

(b) Comparisons – in a current situation, if a management team is told that factory A produced 500 units of a certain product in a week, they must have some standard of comparison to know whether that is satisfactory. They may judge that the capacity of the factory is 650 units a week, or they may be given a series of figures showing that it normally produces about 600 units a week, or they may be told that factory B, with half as many workers but otherwise comparable with factory A, produced 350 units in the same week. Such comparisons provide information on which action can be taken.

Statistical skills enable managers and professionals to collect, analyse and interpret data relevant to their decision-making. Statistical concepts and statistical thinking enable them to:

(a) solve problems in a diversity of contexts.

(b) add substance to decisions.

(c) reduce guesswork.

1.2 Data and information

Information is different from data, and although the two terms are often used interchangeably in everyday language, it is important to make a clear distinction between them, as follows.
We think of data as a ‘scientific’ term for facts, figures, information and measurements. They are the raw materials for data processing.

Data consists of numbers, letters, symbols, raw facts, events and transactions which have been recorded but not yet processed or analysed into a form which is suitable for making decisions.

**Statistics** represent a set of methods that are used to collect, analyse, present, and interpret data. They are the means by which data is interpreted to give meaningful information.

**Information** is data, which has been processed or analysed in such a way that it has a meaning to the person who receives it, who may then use it to improve the quality of decision-making.

For example, in cost accounting the accounting system records a large number of facts (data) about materials, times, expenses and other transactions. These facts are then classified and summarised to produce accounts that are organised into reports, which are designed to help management to plan and control the firm’s activities. Note that as data is converted into information, some of the detail is eliminated and replaced by summaries, which are easier to interpret.

**1.3 Classification of data**

**1.3.1 Attributes and variables**

Data may be of several types. The first distinction to make is between attributes and variables.

**Definition**

An **attribute** is a feature that an object either has or does not have.

An attribute cannot be measured. For example, an individual is either male or female. There is no measure of *how* male or *how* female somebody is: the sex of a person is an attribute. This type of data is qualitative.

**Definition**

A **variable** is something which can be measured.

For example, the height of a person is a variable which can be measured according to some scale (such as centimetres). This type of data is quantitative.

**1.3.2 Quantitative and qualitative data**

**Definitions**

**Quantitative data** is data that can be measured.

Examples of quantitative data include the following.

(a) The maximum temperature on each day of January in Singapore. This can be **measured** in degrees Fahrenheit or Celsius.

(b) The time it takes you to swim 50 metres in a swimming pool. This can be **measured** in minutes and seconds.

**Qualitative data** cannot be measured numerically but may reflect distinguishing characteristics or attributes, e.g. town where a product is manufactured.
1.3.3 Discrete and continuous data

Quantitative data can be further classified as discrete or continuous.

(a) Data is said to be discrete when they can only take on specific fixed values, e.g. the actual number of vehicles through a car wash per day could be 35 but not 35.3, a shoe size could be 5½ but not 5.193 and 9 people could enter a supermarket between 2.30pm and 3.00pm but not 9.999.

(b) Continuous data may take on any value that is measured rather than counted. For example, we could, in an eight hour day, measure the throughput of cars as 4.375 per hour i.e. 35 cars / 8 hours. It may be considered sufficient to measure the heights of a number of people to the nearest centimetre but there is no reason why the measurements should not be made to the nearest 1/100 cm. Two people who are found to have the same height to the nearest centimetre could almost certainly be distinguished if more precise measurements were taken. Much physical data is continuous e.g. length, weight, time.

1.3.4 Primary and secondary data

Data needs to be collected and summarised to the form required by the user.

Primary data is collected especially for a particular enquiry. Raw data is primary data which has not been processed at all, and which is still just a list of numbers.

The main sources of primary data are personal investigation, teams of investigators, interviews, questionnaires and telephone surveys.

Worked Example: Quantitative and qualitative data

Look through the following list of surveys and decide whether each is collecting qualitative data or quantitative data. If you think the data is quantitative, indicate whether it is discrete or continuous.

(a) A survey of accountancy textbooks, to determine how many diagrams they contain.

(b) A survey of greeting cards on a newsagent’s shelf, to determine whether or not each has a price sticker on it.

(c) A survey of the results in a cost accounting assessment, to determine what percentage of marks the students obtained.

(d) A survey of heights of telegraph poles in Papua New Guinea, to find out if there is any variation across the country.

(e) A survey of swimmers to find out how long they take to swim a kilometre.

Solution

(a) The number of diagrams in an accountancy text book is an example of quantitative data, because it can be measured. Because the number of diagrams can only be counted in whole number steps, the resulting data is discrete. For example you cannot have 42½ diagrams, but you can have 42 or 43 diagrams.

(b) Whether or not a greeting card has a price sticker on it is not something that can be measured. This is therefore an example of qualitative data, as a greeting card either has a price sticker on it, or it does not have a price sticker on it.

(c) The results of an assessment can be measured, and are therefore an example of quantitative data. The assessment results can only take on whole number values between 0 per cent and 100 per cent, and the data is therefore discrete. (It may be possible to score 62½ per cent, or 64½ per cent, but it is not possible to score 62.41 per cent, so the variable is not continuous.)

(d) The heights of telegraph poles is an example of quantitative data as they can be measured. Since the telegraph poles may take on any height, the data are said to be continuous.

(e) The time taken to swim a kilometre may be measured and is therefore quantitative data. Because the time recorded can take on any value, in theory, the data are said to be continuous.
Secondary data is data which have already been collected elsewhere, for some other purpose, but which can be used or adapted for the survey being conducted. For example from government, banks, newspapers, the Internet.

An advantage of using primary data is that the investigator knows where the data came from, the circumstances under which it was collected, and any limitations or inadequacies in the data. However, it can take time to collect and is expensive.

In contrast, note the following inadequacies of secondary data:

(a) Any limitations in the data might not be known to the investigator, because he or she did not collect it.
(b) The data might not be entirely suitable for the purpose it is being used for.

Secondary data is sometimes used despite its inadequacies, simply because it is available cheaply whereas the extra cost of collecting primary data would far outweigh its extra value.

1.3.5 Internal and external data

Data collected by an organisation may be internal (from the organisation itself) or external (from outside the organisation).

Internal data relates to activities or transactions performed within the organisation, e.g. administrative tasks such as correspondence or payroll calculations, the production of products and services, or the sale of those products. Often these activities generate costs and revenues, so much of the internal data collected will be quantitative.

Internal sources of data can be classified according to the department of the organisation to which it relates: e.g. Purchasing, Production, Sales, and Marketing.

Gathering data/information from inside the organisation involves:

(a) Establishing a system for collecting or measuring data, in other words, what data is collected, how frequently, by whom and by what method – and how it is processed, filed (or stored) and communicated;
(b) Relying to some extent on the informal communication lines between managers and staff e.g. word of mouth, conversations at meetings and email.

External data relates to data collected from outside or the 'environment' of the organisation.

Data relating to the environment of an organisation might be classified under the following headings:

(a) Political (such as government policy).
(b) Economic (such as inflation or exchange rates).
(c) Social (such as buying-patterns or fashion).
(d) Technological (such as materials and production methods).
(e) Competitive (such as the behaviour of customers, suppliers and rivals).

1.3.6 Formal and informal data

There are usually several information systems collecting internal data. These include formal systems for producing accounts, production statistics, sales analysis and personnel records. Data is also collected from informal sources, which include meetings, face-to-face exchanges, telephone conversations, ad hoc memoranda and 'grapevine' rumours. Data from outside the organisation may also be formal or informal. Informal gathering of data from outside sources goes on all the time, consciously or unconsciously, because we all learn what is going on in the world from newspapers, Internet, television or radio. Examples of formal systems for external information include:

(a) legal and regulatory update information: changes to company law, tax law, employment law, accounting standards, environmental protection.
(b) research intelligence: information about technology changes or new discoveries which may have an impact on the firm's operations.
other forms of market intelligence: for example, the formal collection of feedback forms from customers, salespeople and others 'in the field', such as maintenance staff.

For some types of data, the formal collection from outside sources needs to be the responsibility of particular individuals within the organisation. For example:

(a) A tax specialist within the organisation will be expected to gather facts about changes in tax law and determine whether it will affect the organisation.

(b) Companies registering under the Privacy Act 1988 should appoint a data registration officer with the responsibility for finding out the procedure for keeping personal details.

(c) Research and development work needs someone to co-ordinate data on similar work being done by other companies.

(d) Market research is undertaken by specialists to find out about the opinions and buying attitudes of potential customers.

**Question 1: Data types**

Look through the following list of surveys and decide whether each is collecting data on attributes, discrete variables or continuous variables.

(a) A survey of statistics textbooks, to determine how many diagrams they contain.

(b) A survey of cans in a shop, to determine whether or not each has a price sticker.

(c) A survey of athletes to find out how long they take to run a mile.

(d) A survey of the heights of telegraph poles in Australia.

(The answer is at the end of the chapter)

**1.4 Types of data collection**

Once you decide on the type of data – quantitative or qualitative – appropriate for the problem at hand, you will need to collect the data. You can obtain the data in different ways:

(a) **Data from a published source.** Sometimes, the data set of interest has already been collected for you and is available in a published source, such as a book, journal, newspaper or website.

(b) **Administrative by-product data.** Collected as a by-product of an organisation’s day-to-day operations, is another published source. Examples include data on births, deaths, marriages, divorces, airport arrivals, and motor vehicle registrations. For example, prior to a marriage licence being issued, a couple must provide the registrar with information about their age, sex, birthplace, whether previously married, and where they live.

(c) **Data from a designed experiment.** In which the researcher exerts strict control over the units (people, objects, or events) in the study.

(d) **Surveys** – the researcher samples a group of people, asks one or more questions, and records the responses.

(e) **Observational studies** can also be employed to collect data. In an observational study, the researcher observes the experimental units in their natural setting and records the variable(s) of interest. For example, a finance researcher may observe and record the closing stock prices of companies that are acquired by other firms on the day prior to the buyout and compare them to the closing prices on the day the acquisition is announced. Unlike a designed experiment, an observational study is one in which the researcher makes no attempt to control any aspect of the experimental units.

Regardless of the data collection method employed, it is likely that the data will be a sample from some population. And if we wish to apply inferential statistics, we must obtain a representative sample.
2 Sampling

Section overview

- Successful statistical practice is based on focused problem definition. In sampling, this includes defining the population from which our sample is drawn. Because there is very rarely enough time or money to gather information from everyone or everything in a population, the goal becomes finding a representative sample (or subset) of that population.

2.1 Samples and populations

Data is often collected from a sample rather than from a population. In many situations, it will not be practical to carry out a survey which considers every item of the population. For example, if a poll is taken to try to predict the results of an election, it would not be possible to ask all eligible voters how they are going to vote. It would take far too long and cost too much money.

Definitions

The population is the entire group of individuals that we want information about.

The sample is the part of the population that we actually examine in order to gather information.

The critical difference between a population and a sample is that with a population our interest is to identify its characteristics whereas with a sample, our interest is to make inferences about the characteristics of the population from which the sample was drawn.

For example, if we wanted to find out the percentage of students at a certain High School who enjoy reading during the 15 minutes block of time provided for it each day, we might select 20 per cent of the population. This selection would be the sample in this experiment. The population would be all of the students who attend that particular High School.

If the whole population is examined, the survey is called a census.

Definition

A census is a survey of an entire population, as opposed to a sample survey.

This method gives a high degree of accuracy and allows further detailed information about small subgroups to be made but it has a number of disadvantages:

(a) The high cost of a census may exceed the value of the results obtained.

(b) It might be out of date by the time it is completed.

The term ‘census’ can be applied to a survey of 100% of items in any population, but the word is commonly associated with periodic surveys of their entire population by the government of a country. Every household is required to complete a set of questions about the individuals in their household, and the completed census forms are returned to the government for statistical analysis.

The principle underlying sampling is that a set of objects taken at random from a larger group tends to reproduce the characteristics of that larger group: this is called the law of statistical regularity. There are exceptions to this rule and a certain amount of judgment must be exercised, especially when there are a few abnormally large items in the larger group. With inconsistent data, the accuracy of sampling can often be tested by comparing several samples.
There are three main reasons why sampling is necessary:

(a) The whole population may not be known.

(b) Even if the population is known, the process of testing every item can be extremely costly in time and money. For example, checking every packet of biscuits coming off a production line would be a lengthy process.

(c) The item being tested may be completely destroyed in the process. For example, to check the lifetime of a light bulb it would be necessary to leave it switched on until it breaks and is of no further use.

The advantages of a sample are as follows:

(a) It can be shown mathematically that once a certain sample size has been reached, very little accuracy is gained by examining more items. However, the larger the size of the sample, the more accurate the results.

(b) It is possible to ask more questions with a sample e.g. when conducting market research.

2.2 Sampling methods

2.2.1 Sample design and sampling frames

In sample studies, we have to make a plan regarding the size and selection of the sample, collection of the sample data and preparation of the final results based on the sample study. The whole procedure involved is called the sample design.

A complete list of all the units of the population is called the sampling frame. A unit of population is a relative term. If all the workers in a factory make a population, a single worker is a unit of the population. If all the factories in a country are being studied for some purpose, a single factory is a unit of the population of factories. The sampling frame contains all the units of the population. It is to be defined clearly as to which units are to be included in the frame. The frame provides a base for the selection of the sample.

If random sampling is used then it is necessary to construct a sampling frame. Once this has been made, it is easy to select a random sample, simply by generating a list of random numbers.

Questions to ask about a sampling frame should include:

(a) Are all members of the population included on the list?
(b) Is the information correct?
(c) Does it cover the entire population?
(d) Is the list up-to-date?
(e) Is the sampling frame readily accessible?
(f) Does each member of the population appear on the list only once?

A readily available sampling frame is the electoral register (list of individuals).

2.2.2 Random number tables

Assuming that a sampling frame can be drawn up, then a random sample can be picked from it by one of the following methods.

One of the most important requirements of sample data is that it should be representative. That is, the data should cover all areas of the population to be examined. If this requirement is not met, then the sample will be biased.

2.2.3 Random sampling

Random sampling simply means choosing a sample in such a way that at each step all members of the population not already selected stand an equal chance of being chosen. For example, a hand of cards dealt from a properly shuffled pack is a random sample. Winning tickets in a lottery are chosen at random by taking numbers out of a revolving drum. By a simple extension of this idea, a random sample can be taken from a group or population by numbering all items in the group and picking numbers by means of a table of
"random numbers" i.e. numbers of, for example, two or four digits compiled by some random process. Note that random does not mean 'haphazard'. The method must be completely objective.

Provided the sampling frame is sound, if a sample is selected using random sampling, it will be **free from bias** (since every item will have an equal chance of being selected). This cannot be said of any other sampling method. Once the sample has been selected, valid inferences about the population being sampled can be made.

Technology can sometimes help us choose a random sample (e.g. of customers or invoices). Computer software can choose rows randomly from a spreadsheet or database. Then those units are interviewed or inspected.

### Worked Example: Random number tables

Set out below is part of a typical random number table.

<table>
<thead>
<tr>
<th>32886</th>
<th>59780</th>
<th>09958</th>
<th>18065</th>
</tr>
</thead>
<tbody>
<tr>
<td>93716</td>
<td>16894</td>
<td>98953</td>
<td>73231</td>
</tr>
<tr>
<td>39510</td>
<td>35905</td>
<td>85244</td>
<td>35159</td>
</tr>
<tr>
<td>27699</td>
<td>06494</td>
<td>03152</td>
<td>19121</td>
</tr>
<tr>
<td>92962</td>
<td>61773</td>
<td>22109</td>
<td>78508</td>
</tr>
</tbody>
</table>

An investigator wishes to select a random sample from a population of 800 people, who have been numbered 000, 001, ..., 799. As there are three digits in 799 the random numbers will be selected in groups of three. Randomly starting on the second line and working along the table shown above, the first few groups are as follows:

937 161 689 498 953 732

Numbers over 799 are discarded. The first four people in the sample will therefore be those numbered 161, 689, 498 and 732.

### 2.2.4 Stratified sampling

A variation on the random sampling method is **stratified sampling**. This is the best method of choosing a sample in many situations. The sample frame must be divided into strata or categories. The strata may be regions, towns, streets, sexes, age groups, social classes, occupations and so on, depending on the nature of the enquiry.

If we took a random sample of all cost and management accountants in the country, it is conceivable that the entire sample might consist of members of CPA Australia working in public companies. Stratified sampling removes this possibility as random samples could be taken from each type of employment, the number in each sample being proportional to the total number of cost and management accountants in each type (for example, those in partnerships, those in public companies and those in private companies).

In many situations, stratified sampling is the best method of choosing a sample. It takes **more time** than simple random sampling but samples should be **more representative** and so **sample error** should be reduced.

Stratified sampling is best demonstrated by means of an example.

### Worked Example: Stratified sampling

The number of accountants in each type of work in a particular country are as follows:

- Partnerships: 500
- Public companies: 500
- Private companies: 700
- Public practice: 800

Total: 2500
If a sample of 20 was required the sample would be made up as follows:

<table>
<thead>
<tr>
<th>Sample</th>
<th>Partnerships</th>
<th>Public companies</th>
<th>Private companies</th>
<th>Public practice</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>500</td>
<td>500</td>
<td>700</td>
<td>800</td>
</tr>
<tr>
<td></td>
<td>2 500</td>
<td>2 500</td>
<td>2500</td>
<td>2 500</td>
</tr>
<tr>
<td></td>
<td>× 20</td>
<td>× 20</td>
<td>× 20</td>
<td>× 20</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>4</td>
<td>6</td>
<td>6</td>
</tr>
</tbody>
</table>

The strata frequently involve multiple classifications. In social surveys, for example, there is usually stratification by age, sex and social class. This implies that the sampling frame must contain information on these three variables before the threefold stratification of the population can be made.

**Advantages** of stratification are as follows:

(a) It ensures a representative sample since it guarantees that every important category will have elements in the final sample.

(b) The structure of the sample will reflect that of the population if the same proportion of individuals is chosen from each stratum.

(c) Each stratum is represented by a randomly chosen sample and therefore inferences can be made about each stratum.

(d) Precision is increased. Sampling takes place within strata and, because the range of variation is less in each stratum than in the population as a whole and variation between strata does not enter as a chance effect, higher precision is obtainable. (For this to occur, the items in each stratum must be as similar as possible and the difference between the individual strata must be as great as possible.)

Note, however, that stratification requires prior knowledge of each item in the population. Sampling frames do not always contain this information. Stratification from the electoral register as to age structure would not be possible because the electoral register does not contain information about age.

### 2.2.5 Cluster sampling

**Definition**

Cluster sampling is a non-random sampling method that involves selecting one definable subsection of the population as the sample, that subsection taken to be representative of the population in question.

Cluster sampling is typically used when the researcher cannot get a complete list of the members of a population they wish to study but can get a complete list of groups or 'clusters' of the population. It is also used when a random sample would produce a list of subjects so widely scattered that surveying them would prove to be far too expensive, for example, people who live in different postal districts in Australia.

For example, a brewery could employ cluster sampling to test the market for a new canned beer. The population could be divided geographically into beer drinkers at pubs in different regions of the country and beer purchasers at bottle shops within these regions. A sample of regions could be selected, and a sample of pubs and bottle shops in each selected region could be chosen. All consumers at the chosen pubs and bottle shops would then form the sample.

The main advantage of cluster sampling is its relative cheapness. The main disadvantage is that the sample obtained will not be truly random, so some forms of statistical analysis will not be possible.

This sampling technique may well be more practical and/or economical than simple random sampling or stratified sampling.
2.2.6 The size of a sample

As well as deciding on the appropriateness of a particular sampling method for a given situation, the size of the sample actually selected must also be given consideration.

Although, in certain circumstances, statistical processes can be used to calculate sample sizes, there is no universal law for determining the size of the sample. Two general considerations should, however, be borne in mind:

(a) The larger the size of the sample, the more accurate the results.
(b) There is a point after which there is little to be gained from increasing the size of the sample.

Despite these principles other, mainly administrative factors, play a role in determining sample size.

(a) **Money and time available.**
(b) **Degree of precision required.** A survey may have the aim of discovering residents' reaction to a road widening scheme and hence a fairly small sample, producing imprecise results, would be acceptable. An enquiry into the safety of a new drug would, on the other hand, require an extremely large sample so that the information gained was as precise as possible.
(c) **Number of sub samples required.** If a complicated sampling method such as stratified sampling is to be used, the overall sample size will need to be large so as to ensure adequate representation of each subgroup (in this case, each stratum).

**Question 2: Sampling methods**

Sampling methods are frequently used for the collection of data. Three commonly used types of samples are (A) simple random, (B) stratified random and (C) cluster. State which of these sample types is being described in the following situations.

(a) One school in an area is selected at random and then all pupils in that school are surveyed.

Type of sample is □ □

(b) The local authority has a list of all pupils in the area and the sample is selected in such a way that all pupils have an equal probability of selection.

Type of sample is □ □

(c) The local authority has a list of all pupils in the selected area, categorised according to their gender and age. The sample selected is chosen randomly from the various categories, in proportion to their sizes.

Type of sample is □ □

(The answer is at the end of the chapter)

3 Survey methods for collecting statistical information

**Section overview**

- It has been said that we live in an 'information society' and our major problems and tasks no longer centre on the production of the goods and services necessary for survival and comfort. Because of this, our society requires a prompt and accurate flow of information on preferences, needs, and behaviour. It is in response to this critical need for information on the part of the government, business, and social institutions that so much reliance is placed on surveys.
Definition

A survey is a method of descriptive research used for collecting primary data based on verbal or written communication with a representative sample of individuals or respondents from the target population.

The following are all types of survey:

- A sample of voters is questioned in advance of an election to determine how the public perceives the candidates and the issues.
- A manufacturer does a survey of the potential market before introducing a new product.
- A government entity commissions a survey to gather the factual information it needs to evaluate existing legislation or to draft proposed new legislation.

3.1 Who conducts surveys?

Probably the most familiar type of survey is the political polls conducted by any one of a number of organisations (e.g. Galaxy, Gallup and Newspoll) and designed to predict the outcome of a political election. State polls and area polls, often supported by a local newspaper or TV station, are reported regularly in many localities. The major broadcasting networks and national news magazines also conduct polls and report their findings. The Nielsen survey provides the major television networks with information on the most watched TV programs.

Surveys also provide an important source of basic scientific knowledge. Economists, psychologists, health professionals, political scientists, and sociologists conduct surveys to study such matters as income and expenditure patterns among households, the roots of ethnic or racial prejudice, the implications of health problems on people’s lives, comparative voting behaviour, and the effects on family life of women working outside the home.

Most surveys are directed to a specific administrative, commercial, or scientific purpose. The wide variety of issues with which surveys deal is illustrated by the following listing of actual uses:

- Major TV networks rely on surveys to tell them how many and what types of people are watching their programs.
- Some countries conduct continuing panel surveys of children (and their families) to study educational and other needs.
- Auto manufacturers use surveys to find out how satisfied people are with their cars.
- Government officials conduct a survey each month to obtain information on employment and unemployment in the nation.
- Local transportation authorities conduct surveys to acquire information on commuting and travel habits.
- Magazine and trade journals use surveys to find out what their subscribers are reading.
- Surveys are conducted to ascertain who uses our national parks and other recreation facilities.

3.2 Common survey methods

Surveys can be classified in many ways, for example by size and type of sample. Surveys can be of human or non-human populations (for example, animals, soils, housing, and so on). While many of the principles are the same for all surveys, the focus here will be on methods for surveying individuals. Many surveys study all persons living in a defined area, but others might focus on special population groups, children, physicians, community leaders, the unemployed, or users of a particular product or service. Surveys may also be conducted with national, state, or local samples.

Surveys can be classified by their method of data collection.
3.2.1 Questionnaires

Many surveys require asking the respondents for information either face-to-face, using the telephone to interview, or through email, fax or Internet.

Questionnaires can be used in a variety of ways as follows:

(a) **Telephone interviews** are an efficient method of collecting some types of data and are commonly used. They lend themselves particularly well to situations where timeliness is a factor and the length of the survey is limited.

(b) **Personal interviews** may be conducted by market researchers in the interviewee's home. In-person interviews in a respondent's home or office are much more expensive than mail or telephone surveys. They may be necessary, however, especially when complex information is to be collected.

(c) **Mail surveys** can be relatively low in cost. As with any other survey, problems exist in their use when insufficient attention is given to getting high levels of co-operation. Mail surveys can be most effective when directed at particular groups, such as subscribers to a specialised magazine or members of a professional association.

(d) Questionnaires for self-completion, perhaps at the place of purchase.

You can further classify surveys by their content. Some surveys focus on opinions and attitudes (such as a pre-election survey of voters), while others are concerned with factual characteristics or behaviours (such as people's health, housing, consumer spending, or transportation habits).

Many surveys combine questions of both types. Respondents may be asked if they have heard or read about an issue, what they know about it, their opinion, how strongly they feel and why, their interest in the issue, past experience with it and certain factual information that will help the survey analyst classify their responses (such as age, gender, marital status, occupation, and place of residence).

Questions may be open-ended (‘Why do you feel that way?’) or closed (‘Do you approve or disapprove?’). Survey takers may ask respondents to rate a product on some type of scale, or they may ask for a ranking of various alternatives.

Because changes in attitudes or behaviour cannot be reliably ascertained from a single interview, some surveys employ a 'panel design', in which the same respondents are interviewed on two or more occasions. Such surveys are often used during an election campaign or to chart a family's health or purchasing pattern over a period of time.

3.2.2 Observation

Observation can be used as a means of obtaining sample data where quantitative data are required. For example, if data is needed about the volume of traffic passing along a road at a certain time of day, observers (either people or recording equipment) can be placed so as to count the traffic as it passes by. Observation can also be used to study consumer behaviour, although this is usually within a controlled experiment.

3.2.3 Consumer panels

Consumer panels (test panels) – some research firms have created consumer panels consisting of a representative cross-section of consumers who have agreed to give regular information about their attitudes or buying habits through personal visits or mail questionnaires. Consumer panels with personal visits are called home audit panels and panels which send data by post are called diary panels. For example, a panel of households might keep a purchase diary of the goods they have bought, and submit this diary regularly to the market research company. Panels might be established for a long-term or short-term period.

3.2.4 Trade and retail audits

Trade audits are carried out among panels of wholesalers and retailers, and the term 'retail audits' refers to panels of retailers only. A research firm sends auditors to selected outlets at regular intervals to count stock and deliveries, thus enabling an estimate of throughput to be made.

Retail audits, because they provide continuous monitoring of retail activity, may be of value to a manufacturing firm for the following reasons:
(a) Changes in retail sales provide an early warning of problems the manufacturer may have in sales directly from the factory.

(b) They indicate long-term trends in the market place, thus providing helpful information for strategic marketing planning.

(c) In the shorter term, they may indicate the need for changes in pricing policy, sales promotion or advertising, distribution policy, package design or product design.

4 Published sources of statistical information

4.1 Internal and external sources

Internal sources of published information include the following:

(a) Procedures manual
(b) Financial data
(c) Sales reports by region, customer and product
(d) Competitor intelligence
(e) Market prospects and reports
(f) Customer complaints
(g) Marketing research reports
(h) Number of employees
(i) Training programs
(j) Staff turnover details
(k) Details of pay

External sources of published information include the following:

(a) Books
(b) The Internet is an excellent source of secondary data if used with care. For example, you can use a search engine that will bring up websites of interest. Your search engine may also refer you to magazines and online newspapers.
(c) Government agencies are good sources of economic and other statistical information. Most countries have an agency that provides national statistics. For example, you may want to examine and summarise the unemployment rates (i.e., percentages of eligible workers who are unemployed) in Australia. You can find this data set (as well as numerous other data sets) on the Australian Bureau of Statistics website.
(d) Regulatory bodies and industry associations – there are many quasi-government and other public sector bodies that can provide data on particular industry sectors.
(e) Research reports, research papers, textbooks – academic research and publication forms a major body of secondary data sources.
(f) Opinion polls – the process of collecting public opinion via surveys and questionnaires has been developed to a high level of sophistication in recent decades, and the results of such surveys (if made public) can be an important contextual source.
(g) Market research – like opinion polls, most market research is carried out by private organisations on behalf of specific clients. Depending on the requirements of the client, the results of such surveys may or may not be made public. The commercial orientation of most surveys tends to limit their general applicability.
(h) Online databases – may range from bibliographic information to census data.

Statistics compiled from secondary data are termed secondary statistics. For example, the government publish tables of unemployment figures (these are secondary data): when this data are used for calculations, they are termed secondary statistics.

Government statistics are widely available and are produced to measure the effects of their policies and the effect external factors have on them. Trends can also be noted and this assists future planning.

4.1.1 Internet research

The Internet is a global interconnection of networks. Using browsers such as Microsoft’s Internet Explorer or Mozilla Firefox, it is possible to access businesses, academic institutions, trade associations, government agencies, medical facilities, scientific establishments and private individuals through the Internet.
4.1.2 Government data and official publications

Government departments’ websites are an excellent source of information and data. A gateway to all government sites is www.gov.au, giving access to the information and services of the Australian state, territory and local governments with an A to Z of government sites.

Retail prices are very important to a wide variety of users.

(a) For the government, the Consumer Price index (CPI) indicates the degree of success there has been in fighting inflation.

(b) For employees, the CPI may give an indication of how much wages need to rise to keep pace with inflation.

(c) For consumers, the CPI indicates the increases to be expected in the prices of goods in shops.

(d) For businesses, the CPI may give a broad indication of how much costs should have been expected to rise over recent years and months.

(e) For pensioners and social security recipients, the movement in the CPI is used to update benefit levels.

One of the cornerstones of the government’s statistical service and a massive source of data is the Census of Population. The Australian census is administered once every five years by the Australian Bureau of Statistics. The most recent censuses have been in 2006 and 2011. Researchers use the data from a census for segmentation by demographics and survey planning (e.g. setting quota samples).

http://business.gov.au/ is a national network of advice agencies for business. They provide training and support with a variety of business issues including start-up, business planning, marketing and finance.

In addition to the free basic company details, certain company documents and reports can be purchased for very modest sums by credit card and delivered electronically. These include scanned image documents of the latest company accounts, annual returns, current appointments and outstanding mortgages.

There are many other official publications that can be accessed — some more examples follow:

(a) The Reserve Bank of Australia (www.rba.gov.au) issues a quarterly publication, which includes data on the money supply and government borrowing and financial transactions.

(b) The individual Economic Development Councils of the National Economic Development Office publish many data about their industries.

(c) Other bodies such as the World Bureau of Metal Statistics can provide data on an international basis.

(d) International bodies include: the United Nations’ Yearbook of International Statistics (details of trade by commodity); the Organisation for Economic Co-operation and Development collects data on the foreign trade of member countries (the OECD Monthly Statistics of Foreign Trade) and produces macroeconomic forecasts for member countries (the OECD Economic Outlook); and the IMF’s International Financial Statistics show the balance of payments and related items.

4.1.3 Internal and by-product data

Your employer probably has large amounts of information, not designed for your research, but may be useful to you, and this will be held on a departmental basis.

(a) The accounts department will have financial data, a procedures manual with accounting policies, tax details, management accounts and financial statements.

(b) The sales and marketing department will keep sales reports by region, details of sales by customer and by product, competitor intelligence, customer complaints, marketing research and prospects reports, brand strategy and values and details of distribution chains.

(c) Production and operations will have operations data, process flow charts, input prices and product costings, transport costs and details on efficiency and capacity.

(d) The human resources department will have data on the number of employees, recruitment procedures, training programs, staff turnover details and details of pay.
5 Levels of data measurement

5.1 Types of scales

Before we can conduct a statistical analysis, we need to measure our dependent variable. Exactly how the measurement is carried out depends on the type of variable involved in the analysis. Different types are measured differently. To measure the time taken to respond to a stimulus, you might use a stop watch. Stop watches are of no use, of course, when it comes to measuring someone’s attitude towards a piece of cake. A rating scale is more appropriate in this case (with labels like ‘very favourable’, ‘somewhat favourable’ etc.). For a dependent variable such as ‘favourite colour’, you can simply note the colour (like ‘yellow’) that the subject offers.

Although procedures for measurement differ in many ways, they can be classified using a few fundamental categories. They are:

(a) Nominal.
(b) Ordinal.
(c) Interval.
(d) Ratio.

It is important for the researcher to understand the different levels of measurement, as these levels of measurement play a part in determining the arithmetic and the statistical operations that are carried out on the data. The nominal level is simplest, while ratio measures are the most sophisticated.

5.1.1 Nominal level data

A nominal measurement scale is used for variables in which participants or observations are put into mutually exclusive categories. For example, categorising study participants into ‘male’ and ‘female’ categories demonstrates that ‘sex’ is measured on a nominal scale. Every observation in the study falls into one, and only one, nominal category. The essential point about nominal scales is that they do not imply any value or ranking among the responses. For example, when classifying people according to their favourite colour, there is no sense in which green is placed ‘ahead of’ blue. Responses are merely categorised. Nominal scales embody the lowest level of measurement.

5.1.2 Ordinal level data

The second level of data is called ordinal level data. This level of measurement does indicate something about the rank ordering of study participants. For example, if you think of some type of competition or race (swimming, running), it is possible to rank order the finishers from first place to last place. If someone tells you they finished second, you know that one person finished ahead of them, and all other participants finished behind them.

Although ordinal variables provide information concerning the relative position of participants or observations in our research study, ordinal variables do not tell us anything about the absolute magnitude of the difference between first and second or between second and third.

5.1.3 Interval level data

Interval level data is similar to ordinal level data in that it has a definite ordering scheme, but differences between data are meaningful and can be measured.

For example, in a list of boiling points of liquids, water boils at 100 degrees Celsius and alcohol at 65 degrees Celsius. The items can be put in a definite order and the differences between the values are measurable and meaningful. Also, different sections of the scale measure the data in the same way, so an increase of temperature from, for example, 20 degrees to 30 degrees requires the same amount of heat as for 50 degrees to 60 degrees.

5.1.4 Ratio level data

With interval scales, there is no absolute zero point. For this reason, it is inappropriate to express interval level measurements as ratios; it would not be appropriate to say that 60 degrees is twice as hot as 30 degrees. Ratio scales do have a meaningful and fixed zero point, which allows for such ratio comparisons.

Weight, for example, has a definite zero (no weight) and 10 kg is twice as much as 5 kg.
Worked Example: Ratio level data

Four people are randomly selected and asked how much money they have with them. Here are the results: $21, $50, $65, and $300.

Is there an order to this data? Yes, $21 < $50 < $65 < $300.

Are the differences between the data values meaningful? Yes, the person who has $50 has $29 more than the person with $21.

Can we calculate ratios based on this data? Yes, because $0 is the absolute minimum amount of money a person could have with them. The person with $300 has six times as much as the person with $50.

Other examples of ratio level data would be ages of people, scores on exams (graded from 0 to 100), and hours of study for a test.

6 Data presentation

Section overview

- We now have to present the data we have collected so that it can be of use. This section begins by looking at how data can be presented in tables and charts. Such methods are helpful in presenting key data in a concise and easy to understand way.
- Data is a mass of numbers can usefully be summarised into a frequency distribution. Histograms and scattergraphs are pictorial representations of frequency distributions.

6.1 Tables and tabulation

Definitions

Tables are a simple way of presenting information about two variables. A table is a matrix of data in rows and columns, with the rows and the columns having titles.

Raw data (for example, a list of results from a survey) need to be summarised and analysed, to give them meaning. One of the most basic ways is the preparation of a table.

Tabulation means putting data into tables.

Since a table is two-dimensional, it can only show two variables. To tabulate data, you need to recognise what the two dimensions should represent, prepare rows and columns accordingly with suitable titles, and then insert the data into the appropriate places in the table.

Worked Example: Tables

The total number of employees in a certain trading company is 1 000. They are employed in three departments: production, administration and sales. There are 600 people employed in the production department and 300 in administration. There are 110 male juveniles, 110 female juveniles, and 290 adult females. The remaining employees are adult males.

In the production department there are 350 adult males, 150 adult females and 50 male juveniles, while in the administration department there are 100 adult males, 110 adult females and 50 juvenile males.

Required

Draw up a table to show all the details of employment in the company and its departments and provide suitable secondary statistics to describe the distribution of people in departments.
Solution

The basic table required has the following two dimensions:

(a) Departments
(b) Age/sex analysis

Secondary statistics (not the same thing as secondary data) are supporting figures that are supplementary to the main items of data, and which clarify or amplify the main data. A major example of secondary statistics is percentages. In this example, we could show one of the following:

(a) The percentage of the total work force in each department belonging to each age/sex group.
(b) The percentage of the total of each age/sex group employed in each department.

In this example, (a) has been selected but you might consider that (b) would be more suitable. Either could be suitable, depending of course on what purposes the data are being collected and presented for.

Analysis of employees

<table>
<thead>
<tr>
<th>Department</th>
<th>Production</th>
<th>Administration</th>
<th>Sales</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>No</td>
<td>%</td>
<td>No</td>
<td>%</td>
</tr>
<tr>
<td>Adult males</td>
<td>350</td>
<td>58.4</td>
<td>100</td>
<td>33.3</td>
</tr>
<tr>
<td>Adult females</td>
<td>150</td>
<td>25.0</td>
<td>110</td>
<td>36.7</td>
</tr>
<tr>
<td>Male juveniles</td>
<td>50</td>
<td>8.3</td>
<td>50</td>
<td>16.7</td>
</tr>
<tr>
<td>Female juveniles</td>
<td>*50</td>
<td>8.3</td>
<td>*40</td>
<td>13.3</td>
</tr>
<tr>
<td>Total</td>
<td>600</td>
<td>100.0</td>
<td>300</td>
<td>100.0</td>
</tr>
</tbody>
</table>

* Balancing figure to make up the column total
** Balancing figure then needed to make up the row total

Rounding errors may become apparent when, for example, a percentages column does not add up to 100 per cent. Any rounding should therefore be to the nearest unit and the potential size of errors should be kept to a tolerable level by rounding to a small enough unit (for example, to the nearest $10, rather than to the nearest $1 000).

6.2 Presenting data in a visual form

There are several advantages of presenting a mass of data in tabular form. The figures can easily be located, comparisons between classes can be made at a glance, patterns of figures are highlighted and tables are easily understood by non-statisticians. However, charts, diagrams and graphs are more popular ways of displaying data simply. The purpose of a chart is to convey the data in a way that will demonstrate its meaning more clearly than a table of data would.

Charts are not always more appropriate than tables, and the most suitable way of presenting data will depend on the following:

(a) **What the data are intended to show.** Visual displays usually make one or two points quite forcefully, whereas tables usually give more detailed information.

(b) **Who is going to use the data.** Some individuals might understand visual displays more readily than tabulated data.

6.2.1 Bar graphs

The **bar graph**, also known as the **bar chart**, is one of the most common methods of presenting data in a visual form. Bar graphs are a family of charts that display quantitative information by means of a series of rectangles (‘bars’) that can be displayed horizontally or vertically. Each bar represents a data element in a data series and a complete set of bars represents a data series. The end of each bar is located at the value it represents. Because the ends of the bars are so pronounced, this type of graph is considered good for showing specific values. Bar graphs have a quantitative scale on the vertical axis which is normally linear and typically starts at zero.
There are three main types of bar graph: simple, component (including percentage component) and multiple (or compound) bar graphs.

### 6.2.2 Simple bar graph

**Definition**

A **simple bar graph** is a graph consisting of one or more bars, in which the length of each bar indicates the magnitude of the corresponding data item.

When a bar graph displays a single data series it is generally referred to as a simple bar graph. The bars can be any width; however, with few exceptions, bars are uniform in width throughout a given graph. The same is true for the spaces between the bars.

Simple bar graphs serve two purposes:

- The actual magnitude of each item is shown.
- The lengths of bars on the graph allow magnitudes to be compared.

This means that a simple bar graph presents in visual form the total size of a number of different items, so that their sizes or amounts can be compared.

Each bar represents an item that has a measurable value. For example, a bar might represent a section of the country’s population (such as people in full-time education, working people and retired people) or different groups within a work force (administrators, accountants, IT staff, operations staff and so on). In the example below, each bar represents a year. Remember that the purpose of a simple bar graph is to compare the sizes of each item represented by a bar.

**Worked Example: A simple bar graph**

A company’s total sales for the years from 20X4 to 20X9 are as follows:

<table>
<thead>
<tr>
<th>Year</th>
<th>Sales $ 000</th>
</tr>
</thead>
<tbody>
<tr>
<td>20X4</td>
<td>800</td>
</tr>
<tr>
<td>20X5</td>
<td>1 200</td>
</tr>
<tr>
<td>20X6</td>
<td>1 100</td>
</tr>
<tr>
<td>20X7</td>
<td>1 400</td>
</tr>
<tr>
<td>20X8</td>
<td>1 600</td>
</tr>
<tr>
<td>20X9</td>
<td>1 700</td>
</tr>
</tbody>
</table>

In this bar graph, each bar will represent a year, and for each year we are measuring sales income. The length of the bar for each year simply represents the amount of sales income.

The data could be shown on a simple bar graph as follows:
Each axis of the chart must be clearly labelled, and there must be a scale to indicate the magnitude of the data. Here, the y axis has a scale for the amount of sales, and so readers of the bar chart can see the comparative total sales in each year, and also, since the bar graph represents a time series, it can be used to see whether there is any noticeable upward or downward trend in total annual sales over time.

Here we can see in visual form what sales income was in each year. In addition, we can see that annual sales were much higher in 20X9 than in 20X4, and that with the exception of 20X6 when there was a drop in sales, there has been a continual rising trend in annual sales over the period.

Note that different forms of visual presentation can be used for the same data. In the example above, a bar chart is used to show annual sales over a period of time. This information could also be shown in a simple scatter graph, with the year on the x axis and sales on the y axis, and with the dots joined up to show a time line. Scatter graphs are described later.

### 6.2.3 Component bar graph

A **component bar graph** is a bar graph in which each bar contains a breakdown of its total amount into its component parts. The total length of each bar and each component on a component bar graph indicates magnitude (a bigger amount is shown by a longer bar). It is sometimes called a stacked bar graph.

The purpose of a component bar chart is to show, in visual form:

- the total amount of each item represented by a bar
- how this total amount is made up of component elements.

### Worked Example: A component bar graph

Charbart's sales for the years from 20X7 to 20X9 are as follows:

<table>
<thead>
<tr>
<th>Year</th>
<th>Product A</th>
<th>Product B</th>
<th>Product C</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>20X7</td>
<td>1 000</td>
<td>900</td>
<td>500</td>
<td>2 400</td>
</tr>
<tr>
<td>20X8</td>
<td>1 200</td>
<td>1 000</td>
<td>600</td>
<td>2 800</td>
</tr>
<tr>
<td>20X9</td>
<td>1 700</td>
<td>1 000</td>
<td>700</td>
<td>3 400</td>
</tr>
</tbody>
</table>

A component bar graph would show the following:

- the total sales in each year
- the components of each year’s total, which in this example are the sales of each of the three products.
Notice that the components in each bar are stacked on top of each other in the same order in each bar. It does not matter which order they are shown in, but often you will find the largest component at the bottom of each bar and the smallest at the top. (The components are shown in the example below in alphabetical order, from A at the top to C at the bottom.)

A component bar chart allows the reader to look at the total amounts and also at the size of each component. In this diagram the total sales can be seen to rise each year and also that the growth in sales of Product A seems to be the main reason for the total sales growth.

### 6.2.4 Percentage component bar graph

A **percentage component bar graph** is a component bar graph which does not show **total magnitudes** – instead, the length of each bar represents 100% of the item that is being measured, and the component elements are shown as a percentage of the total.

A percentage component bar graph may consist of just one bar, or there may be several bars, each with its components shown as percentage amounts. The lengths of the component sections of the bar vary according to the **relative sizes** of the components. For example, if a component makes up 50% of the total item, it will be shown as one half (50%) of the length of the bar. Similarly, a component that makes up 20% of the item will be shown as one fifth of the length of the bar.

A percentage component bar graph is an **alternative to a pie chart** for presenting the proportions of components in a total. (Pie charts are described later.) Two or more component bar charts side by side can be used to compare the proportionate make-up of several different items.

To draw a percentage component bar chart:

- Every bar in the chart is exactly the same length
- The value for each component must be calculated as a percentage of the total, which should then be shown in the bar in the graph.
Worked Example: Percentage component bar graph

The information in the previous example of sales of Charbart could have been shown in a percentage component bar graph as follows:

![Charbart sales analysis 20X7 - 20X9](image)

### Working

<table>
<thead>
<tr>
<th>Year</th>
<th>Product A</th>
<th>Product B</th>
<th>Product C</th>
</tr>
</thead>
<tbody>
<tr>
<td>20X7</td>
<td>1 000</td>
<td>900</td>
<td>500</td>
</tr>
<tr>
<td>20X8</td>
<td>1 200</td>
<td>1 000</td>
<td>600</td>
</tr>
<tr>
<td>20X9</td>
<td>1 700</td>
<td>1 000</td>
<td>700</td>
</tr>
</tbody>
</table>

A percentage component bar graph does not show or compare totals, so that in this example we cannot see from the graph what total sales were in each year, or what total sales of each product were. The bar graph does provide a fairly clear picture, however, of proportionate amounts and whether these are considerably different in each year.

This graph shows that sales of C have remained a steady proportion of total sales, but the proportion of A in total sales has gone up quite considerably, while the proportion of B has fallen correspondingly.

6.2.5 Multiple bar graph

**Definition**

A **multiple bar graph** (or **compound bar graph**) is a bar graph in which two or more separate bars are used to present sub-divisions of data. This is sometimes called a side-by-side bar graph.

### Worked Example: Multiple bar graph

The data on Charbart's sales could be shown in a multiple bar graph as follows:
A multiple bar graph uses several bars for each total. In this multiple bar graph, the sales in each year are shown as three separate bars, one for each product, A, B and C.

Multiple bar graphs present similar information to component bar graphs, except for the following.
(a) Multiple bar graphs do not show the grand total whereas component bar graphs do.
(b) Multiple bar graphs illustrate the comparative magnitudes of the components more clearly than component bar graphs.

Multiple bar graphs are sometimes drawn with the bars horizontal instead of vertical.

Question 3: Bar graph

Income for Canary Bank in 20X7, 20X8 and 20X9 is made up as follows:

<table>
<thead>
<tr>
<th>Year</th>
<th>Interest income</th>
<th>Commission income</th>
<th>Other income</th>
</tr>
</thead>
<tbody>
<tr>
<td>20X7</td>
<td>$ 3,579</td>
<td>857</td>
<td>$ 62</td>
</tr>
<tr>
<td>20X8</td>
<td>$ 2,961</td>
<td>893</td>
<td>$ 59</td>
</tr>
<tr>
<td>20X9</td>
<td>$ 2,192</td>
<td>917</td>
<td>$ 70</td>
</tr>
</tbody>
</table>

Using the above data complete the following graphs:
(a) A simple bar graph
(b) A multiple bar graph

What do these graphs show you?

(The answer is at the end of the chapter)

6.2.6 Pie charts

Definition

A pie chart is a chart which is used to show pictorially the relative size of component elements of a total.

It is called a pie chart because it is circular, and so has the shape of a pie in a round pie dish. The 'pie' is then cut into slices with each slice representing part of the total.
A pie chart is similar in purpose to a percentage component bar graph. In a percentage component bar graph, each bar represents 100 per cent of the total, and in a pie chart the total pie represents 360 degrees.

Pie charts have sectors of varying sizes, and you need to be able to draw sectors fairly accurately. To do this, you need a protractor. Working out sector sizes involves converting parts of the total into equivalent degrees of a circle. A complete 'pie' = 360 degrees: the number of degrees in a circle = 100 per cent of whatever you are showing. An element which is 50 per cent of your total will therefore occupy a segment of 180 degrees, and so on.

(Note: Standard computer software such as Excel may be used to convert data into a pie chart without the need to draw a chart manually.)

Using shading and colour

Two pie charts are shown as follows. 

![Air and Noise Pollution Complaints Pie Charts](image)

Figure 1 Pie charts of air and noise pollution complaints

Note how shading distinguishes the segments from each other and colour may be used to distinguish segments.

Worked Example: Pie charts

The costs of production at Factory A and Factory B during March 20X1 were as follows:

<table>
<thead>
<tr>
<th></th>
<th>Factory A</th>
<th></th>
<th>Factory B</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$ 000</td>
<td>%</td>
<td>$ 000</td>
<td>%</td>
</tr>
<tr>
<td>Direct materials</td>
<td>70</td>
<td>35</td>
<td>50</td>
<td>20</td>
</tr>
<tr>
<td>Direct labour</td>
<td>30</td>
<td>15</td>
<td>125</td>
<td>50</td>
</tr>
<tr>
<td>Production overhead</td>
<td>90</td>
<td>45</td>
<td>50</td>
<td>20</td>
</tr>
<tr>
<td>Office costs</td>
<td>10</td>
<td>5</td>
<td>25</td>
<td>10</td>
</tr>
<tr>
<td></td>
<td><strong>200</strong></td>
<td><strong>100</strong></td>
<td><strong>250</strong></td>
<td><strong>100</strong></td>
</tr>
</tbody>
</table>

Required

Show the costs for the factories in pie charts.

Solution

To convert the components into degrees of a circle, we can use either the percentage figures or the actual cost figures.
(a) Using the percentage figures, the total percentage is 100 per cent, and the total number of degrees in a circle is 360 degrees. To convert from one to the other, we multiply each percentage value by $\frac{360}{100} = 3.6$.

<table>
<thead>
<tr>
<th></th>
<th>Factory A</th>
<th></th>
<th>Factory B</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Percentage</td>
<td>Degrees</td>
<td>Percentage</td>
<td>Degrees</td>
</tr>
<tr>
<td>Direct materials</td>
<td>35</td>
<td>126</td>
<td>20</td>
<td>72</td>
</tr>
<tr>
<td>Direct labour</td>
<td>15</td>
<td>54</td>
<td>50</td>
<td>180</td>
</tr>
<tr>
<td>Production overhead</td>
<td>45</td>
<td>162</td>
<td>20</td>
<td>72</td>
</tr>
<tr>
<td>Office costs</td>
<td>5</td>
<td>18</td>
<td>10</td>
<td>36</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>360</td>
<td>100</td>
<td>360</td>
</tr>
</tbody>
</table>

(b) Using the actual cost figures, we would multiply each cost by:

\[
\frac{360}{200} = 1.8 \quad \frac{360}{250} = 1.44
\]

<table>
<thead>
<tr>
<th></th>
<th>Factory A</th>
<th>$000</th>
<th>Degrees</th>
<th>Factory B</th>
<th>$000</th>
<th>Degrees</th>
</tr>
</thead>
<tbody>
<tr>
<td>Direct materials</td>
<td>70</td>
<td>126</td>
<td>50</td>
<td>72</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Direct labour</td>
<td>30</td>
<td>54</td>
<td>125</td>
<td>180</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Production overhead</td>
<td>90</td>
<td>162</td>
<td>50</td>
<td>72</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Office costs</td>
<td>10</td>
<td>18</td>
<td>25</td>
<td>36</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>200</td>
<td>360</td>
<td>250</td>
<td>360</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

A pie chart could be drawn for each factory. A protractor is used to measure the degrees accurately to obtain the correct sector sizes, although most people use a spreadsheet program to do it.

Figure 2 Pie charts of costs of Factories A and B

The advantages of pie charts are as follows:

(a) They give a simple pictorial display of the relative sizes of elements of a total.
(b) They show clearly when one element is much bigger than others.
(c) They can sometimes clearly show differences in the elements of two different totals. In the example above, the pie charts for factories A and B show how factory A’s costs mostly consist of production overhead and direct materials, whereas at factory B, direct labour is the largest cost element.

The disadvantages of pie charts are as follows:

(a) They show only the relative sizes of elements. In the example of the two factories, for instance, the pie charts do not show that costs at factory B were $50 000 higher in total than at factory A.
(b) They involve calculating degrees of a circle and drawing sectors accurately, and this can be time consuming.
(c) It is sometimes difficult to compare sector sizes accurately by eye.
Exam comments
A computer-based assessment cannot require you to draw charts so questions will focus on labelling, calculating values, choosing an appropriate chart and interpreting the data presented in one or more pie charts.

7 Graphical representation of data
7.1 Introduction to frequency distributions

Frequency distributions are used if values of particular variables could occur more than once.

Definition
A frequency distribution is a summary of the values obtained in a survey and the frequencies with which these values have occurred.

Frequently the data collected from a statistical survey or an investigation is simply a mass of numbers.

<table>
<thead>
<tr>
<th>Units</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>65</td>
<td>1</td>
</tr>
<tr>
<td>66</td>
<td>0</td>
</tr>
<tr>
<td>67</td>
<td>2</td>
</tr>
<tr>
<td>68</td>
<td>2</td>
</tr>
<tr>
<td>69</td>
<td>4</td>
</tr>
<tr>
<td>70</td>
<td>5</td>
</tr>
<tr>
<td>71</td>
<td>3</td>
</tr>
<tr>
<td>72</td>
<td>1</td>
</tr>
<tr>
<td>73</td>
<td>1</td>
</tr>
<tr>
<td>74</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>20</td>
</tr>
</tbody>
</table>

When the data is arranged in a table in this way it is immediately obvious that 69 and 70 units are the most common volumes of output per employee per week.
7.2 Grouped frequency distributions

If there is a large set of data or if every (or nearly every) data item is different, it is often convenient to group frequencies together into bands or classes. For example, suppose that the output produced by another group of 20 employees during one week was as follows, in units.

<table>
<thead>
<tr>
<th>Units</th>
<th>Number of employees (frequency)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 087 – 850</td>
<td>1</td>
</tr>
<tr>
<td>924 – 1 226</td>
<td>1</td>
</tr>
<tr>
<td>1 265 – 1 028</td>
<td>1 230 – 1 129</td>
</tr>
<tr>
<td>1 086 – 1 130</td>
<td>989 – 1 129</td>
</tr>
<tr>
<td>1 134 – 1 166</td>
<td>1 129 – 1 160</td>
</tr>
</tbody>
</table>

The range of output from the lowest to the highest producer is 792 to 1 265, a range of 473 units. This range could be divided into classes of, for example, 100 units (the class width or class interval), and the number of employees producing output within each class could then be grouped into a single frequency, as follows:

<table>
<thead>
<tr>
<th>Output</th>
<th>Number of employees (frequency)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Units 700 – 799</td>
<td>1</td>
</tr>
<tr>
<td>800 – 899</td>
<td>1</td>
</tr>
<tr>
<td>900 – 999</td>
<td>2</td>
</tr>
<tr>
<td>1 000 – 1 099</td>
<td>5</td>
</tr>
<tr>
<td>1 100 – 1 199</td>
<td>7</td>
</tr>
<tr>
<td>1 200 – 1 299</td>
<td>4</td>
</tr>
<tr>
<td>20</td>
<td></td>
</tr>
</tbody>
</table>

Note, however, that once items have been 'grouped' in this way their individual values are lost.

7.2.1 Grouped frequency distributions of continuous variables

As well as being used for discrete variables (as above), grouped frequency distributions (or grouped frequency tables) can be used to present data for continuous variables.

Worked Example: Distributions of continuous variables

Suppose we wish to record the heights of 50 different individuals. The information might be presented as a grouped frequency distribution, as follows:

<table>
<thead>
<tr>
<th>Height (cm)</th>
<th>Number of individuals (frequency)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Up to and including 154</td>
<td>1</td>
</tr>
<tr>
<td>Over 154, up to and including 163</td>
<td>3</td>
</tr>
<tr>
<td>Over 163, up to and including 172</td>
<td>8</td>
</tr>
<tr>
<td>Over 172, up to and including 181</td>
<td>16</td>
</tr>
<tr>
<td>Over 181, up to and including 190</td>
<td>18</td>
</tr>
<tr>
<td>Over 190</td>
<td>4</td>
</tr>
<tr>
<td>50</td>
<td></td>
</tr>
</tbody>
</table>

Note the following points.

(a) It would be wrong to show the ranges as 0–154, 154–163, 163–172 and so on, because 154 cm and 163 cm would then be values in two classes, which is not permissible. Although each value should only be in one class, we have to make sure that each possible value can be included. Classes such as 154–162, 163–172 would not be suitable since a height of 162.5 cm would not belong in either class. Such classes could be used for discrete variables, however.

(b) There is an open-ended class at each end of the range. This is because heights up to 154 cm and over 190 cm are thought to be uncommon, so that a single 'open-ended' class is used to group all the frequencies together.

7.2.2 Preparing grouped frequency distributions

To prepare a table of a grouped frequency distribution, a decision must be made about how wide each class should be. You should, however, generally observe the following guidelines:
(a) The size of each class should be appropriate to the nature of the data being recorded, and the most appropriate class interval varied according to circumstances.

(b) The upper and lower limits of each class interval should be suitable 'round' numbers for class intervals which are in multiples of 5, 10, 100, 1000 and so on. For example, if the class interval is 10, and data items range in value from 23 to 62 (discrete values), the class intervals should be 20–29, 30–39, 40–49, 50–59 and 60–69, rather than 23–32, 33–42, 43–52 and 53–62.

(c) With continuous variables, either:

(i) the upper limit of a class should be 'up to and including ...' and the lower limit of the next class should be 'over ...'; or

(ii) the upper limit of a class should be 'less than...', and the lower limit of the next class should be 'at least ...'.

**Question 4: Grouped frequency distribution**

The commission earnings for May 20X9 of the assistants in a department store were as follows (in $):

| 60 | 35 | 53 | 47 | 25 | 44 | 55 | 58 | 47 | 71 |
| 63 | 67 | 57 | 44 | 48 | 50 | 56 | 61 | 42 |
| 43 | 38 | 41 | 39 | 61 | 51 | 27 | 56 | 57 | 50 |
| 55 | 68 | 55 | 50 | 25 | 48 | 44 | 43 | 49 | 73 |
| 53 | 35 | 36 | 41 | 45 | 71 | 56 | 40 | 69 | 52 |
| 36 | 47 | 66 | 52 | 32 | 46 | 44 | 32 | 52 | 58 |
| 49 | 41 | 45 | 45 | 48 | 36 | 46 | 42 | 52 | 33 |
| 31 | 36 | 40 | 66 | 53 | 58 | 60 | 52 | 66 | 51 |
| 51 | 44 | 59 | 53 | 51 | 57 | 35 | 45 | 46 | 54 |
| 46 | 54 | 51 | 39 | 64 | 43 | 54 | 47 | 60 | 45 |

Required

Prepare a grouped frequency distribution, in the form of a table, classifying the commission earnings into categories of $5 commencing with '$25 and under $30'.

*(The answer is at the end of the chapter)*

You should be able to interpret a grouped frequency distribution and express an interpretation in writing. In the previous worked example on heights, an interpretation of the data are fairly straightforward.

(a) Most heights fell between 154 cm and 190 cm.

(b) Most heights were in the middle of this range, with few people having heights in the lower and upper ends of the range.

### 7.3 Histograms

#### Definition

A **histogram** is a pictorial representation of a frequency distribution.

- Like a bar chart, a histogram is made up of columns plotted on a graph.
- The horizontal axis of the histogram represents the values whose frequencies are being measured.
- The histogram is divided into ranges of values (class intervals), the same as for a frequency distribution in tabular form.
Statistical analysis, data, and methods of describing data

- For each class interval, the histogram will show the frequency of the value in the form of a column or rectangle.
- The width of each column (class interval) is proportional to the size of the value range.
- There is no space between adjacent columns.

The method of drawing a histogram differs according to whether each class interval has the same length or range, or whether the class intervals have differing lengths/ranges.

7.3.1 Histograms of frequency distributions with equal class intervals

If all the class intervals are the same, as shown below, the columns of the histogram all have the same width and the heights of the columns are proportional to the frequencies.

<table>
<thead>
<tr>
<th>Output Units</th>
<th>Number of employees (frequency)</th>
</tr>
</thead>
<tbody>
<tr>
<td>700 – 799</td>
<td>1</td>
</tr>
<tr>
<td>800 – 899</td>
<td>1</td>
</tr>
<tr>
<td>900 – 999</td>
<td>2</td>
</tr>
<tr>
<td>1000 – 1099</td>
<td>5</td>
</tr>
<tr>
<td>1100 – 1199</td>
<td>7</td>
</tr>
<tr>
<td>1200 – 1299</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>20</td>
</tr>
</tbody>
</table>

The histogram looks almost identical to a simple bar graph except that the bars are joined together.

Note that the discrete data have been treated as continuous, the intervals being changed to >700 but ≤ 800, >800 but ≤ 900 and so on.

Because the bars are joined together, when presenting discrete data the data must be treated as continuous so that there are no gaps between class intervals. For example, for a cricketer's scores in various games the classes would have to start ≥0 but <10, ≥10 but <20, instead of 0–9, 10–19 and so on.

7.3.2 Histograms of frequency distributions with unequal class intervals

If a frequency distribution has unequal class intervals, the heights of the bars have to be adjusted for the fact that the bars do not have the same width.

For example, if one class interval is twice as large as another, its height should be adjusted by a factor of ½.
Worked Example: A histogram with unequal class intervals

The daily wages of employees of Salt Lake Ltd are as follows:

<table>
<thead>
<tr>
<th>Wages per employee</th>
<th>Number of employees</th>
<th>Size of class interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>≤ $60</td>
<td>4</td>
<td>open-ended</td>
</tr>
<tr>
<td>&gt; $60 ≤ $80</td>
<td>6</td>
<td>$20</td>
</tr>
<tr>
<td>&gt; $80 ≤ $90</td>
<td>6</td>
<td>$10</td>
</tr>
<tr>
<td>&gt; $90 ≤ $120</td>
<td>6</td>
<td>$30</td>
</tr>
<tr>
<td>&gt; $120</td>
<td>3</td>
<td>open-ended</td>
</tr>
</tbody>
</table>

The class intervals for wages per employee are not all the same, and range from $10 to $30, with two open-ended class intervals at the bottom and top of the range.

Solution

A histogram is drawn as follows:

(a) The width of each bar on the chart must be proportionate to the corresponding class interval. In other words, the bar representing wages of > $60 ≤ $80, a range of $20, will be twice as wide as the bar representing wages of > $80 ≤ $90, a range of only $10.

(b) A standard width of bar must be selected. This should be the size of class interval which occurs most frequently. In our example, class intervals $10, $20 and $30 each occur once. An interval of $20 will be selected as the standard width.

(c) Open-ended classes must be closed off. It is usual for the width of such classes to be the same as that of the adjoining class. In this example, the class 'up to and including $60' will become > $40 ≤ $60 and the class 'more than $120' will become > $120 ≤ $150.

(d) Each frequency is then multiplied by (standard class width ÷ actual class width) to obtain the height of the bar in the histogram.

(e) The height of bars no longer corresponds to frequency but rather to frequency density and hence the vertical axis should be labelled frequency density.

<table>
<thead>
<tr>
<th>Class interval</th>
<th>Size of interval</th>
<th>Frequency</th>
<th>Adjustment</th>
<th>Height of column (frequency density)</th>
</tr>
</thead>
<tbody>
<tr>
<td>&gt; $40 ≤ $60</td>
<td>20</td>
<td>4</td>
<td>× 20/20</td>
<td>4</td>
</tr>
<tr>
<td>&gt; $60 ≤ $80</td>
<td>20</td>
<td>6</td>
<td>× 20/20</td>
<td>6</td>
</tr>
<tr>
<td>&gt; $80 ≤ $90</td>
<td>10</td>
<td>6</td>
<td>× 20/10</td>
<td>12</td>
</tr>
<tr>
<td>&gt; $90 ≤ $120</td>
<td>30</td>
<td>6</td>
<td>× 20/30</td>
<td>4</td>
</tr>
<tr>
<td>&gt; $120 ≤ $150</td>
<td>30</td>
<td>3</td>
<td>× 20/30</td>
<td>2</td>
</tr>
</tbody>
</table>

(i) The first two bars will be of normal height.

(ii) The third bar will be twice as high as the class frequency (6) would suggest, to compensate for the fact that the class interval, $10, is only half the standard size.

(iii) The fourth and fifth bars will be two thirds as high as the class frequencies (6 and 3) would suggest, to compensate for the fact that the class interval, $30, is 150 per cent of the standard size.
Note that the data is considered to be continuous since the gap between, for example, $79.99 and $80.00 is so small.

**Question 5: Histogram**

The sales force of a company have just completed a successful sales campaign. The performances of individual sales staff have been analysed as follows, into a grouped frequency distribution.

<table>
<thead>
<tr>
<th>Sales</th>
<th>Number of sales staff</th>
</tr>
</thead>
<tbody>
<tr>
<td>Up to $10 000</td>
<td>1</td>
</tr>
<tr>
<td>&gt; $10 000 ≤ $12 000</td>
<td>10</td>
</tr>
<tr>
<td>&gt; $12 000 ≤ $14 000</td>
<td>12</td>
</tr>
<tr>
<td>&gt; $14 000 ≤ $18 000</td>
<td>8</td>
</tr>
<tr>
<td>&gt; $18 000 ≤ $22 000</td>
<td>4</td>
</tr>
<tr>
<td>&gt; $22 000</td>
<td>1</td>
</tr>
</tbody>
</table>

Draw a histogram from this information. Comment on what the histogram shows you.

*(The answer is at the end of the chapter)*

### 7.4 Scatter diagrams

**Definition**

Scatter diagrams (also called scatter plots or scatter graphs) are graphs which are used to exhibit data (rather than equations) in order to compare the way in which two variables vary with each other.

A scatter diagram shows whether there is close correlation between the two variables.

#### 7.4.1 Constructing a scatter diagram

The x axis of a scatter diagram is used to represent the independent variable and the y axis represents the dependent variable.

To construct a scatter diagram or scatter graph, we must have several pairs of data, with each pair showing the value of one variable and the corresponding value of the other variable. Each pair is plotted on a graph.
The resulting graph will show a number of pairs, scattered over the graph. The scattered points might or might not appear to show correlation between the two variables.

- If there is close correlation between the values of the two variables, this means that as the value of one variable increases, the value of the other variable either also increases (if there is positive correlation) or decreases (if there is negative correlation).
- When a scatter diagram represents a time series, with time shown on the x axis, it shows whether there appears to be a rising trend or a declining trend over time.

**Worked Example: Factory output**

The output at a factory each week for the last ten weeks, and the cost of that output, were as follows:

<table>
<thead>
<tr>
<th>Week</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>Output (units)</td>
<td>5</td>
<td>7</td>
<td>5</td>
<td>3</td>
<td>4</td>
<td>6</td>
<td>2</td>
<td>7</td>
<td>4</td>
<td>9</td>
</tr>
<tr>
<td>Cost ($)</td>
<td>17</td>
<td>19</td>
<td>13</td>
<td>9</td>
<td>13</td>
<td>18</td>
<td>5</td>
<td>22</td>
<td>12</td>
<td>25</td>
</tr>
</tbody>
</table>

Required

Plot the relationship between output and costs on a scatter diagram.

**Solution**

The data could be shown on a scatter diagram as follows:

(a) The cost depends on the volume of output: volume is the independent variable and is shown on the x axis.

(b) You will notice from the graph that the points on the graph showing the plotted data, although scattered, lie approximately on a rising trend line, with higher total costs at higher output volumes. This graph therefore indicates that total costs tend to rise as output increases.
Key chapter points

- An attribute is a feature that an object has either got or not got. It cannot be measured. A variable is something which can be measured.
- Variables can be discrete (may take specific values) or continuous (may take any value).
- Data may be primary (collected specifically for the current purpose) or secondary (collected already).
- Examples of secondary data include published statistics and historical records.
- Primary data can be collected using surveys. There are two main types of survey: interviews and questionnaires. Interviews can be face-to-face or performed over the telephone or by email.
- The purpose of statistical analysis is to provide estimates or comparisons on which sound decisions can be made.
- Data is often collected from a sample rather than from a population. A sample can be selected using random sampling (using random number tables or the lottery method), stratified sampling or non-random sampling (cluster sampling). Ensure that you know the characteristics, advantages and disadvantages of each sampling method.
- Once data has been collected it needs to be presented and analysed. It is important to remember that if the data has not been collected properly, no amount of careful presentation or interpretation can remedy the defect.
- Tables are a simple way of presenting information about two variables.
- Charts often convey the meaning or significance of data more clearly than a table would.
- There are three main types of bar chart: simple, component (including percentage component) and multiple (or compound).
- Pie charts are used to show pictorially the relative size of component elements of a total.
- Frequency distributions are used if values of particular variables can occur more than once.
- A frequency distribution can be represented pictorially by means of a histogram. The number of observations in a class is represented by the area covered by the bar, rather than by its height.
- Scatter diagrams (scatter graphs) are graphs which are used to exhibit data (rather than equations) in order to compare the way in which two variables vary with each other.
Quick revision questions

1. Statistics is the science of conducting studies to:
   A. hypothesise, experiment, and form conclusions.
   B. solve a system of equations.
   C. collect, organise, summarise, analyse, and draw conclusions from data.
   D. monitor, study, and report on a subject.

2. If you classified the fruit in a basket as apple, orange, or banana, this would be an example of which level of measurement?
   A. ratio
   B. interval
   C. nominal
   D. ordinal

3. The amount of time needed to run the Dubai marathon is an example of which type of variable?
   A. continuous
   B. discrete
   C. qualitative

4. A survey is being carried out to find out the number of goals scored by the ten most popular men's hockey clubs in Australia in their past ten matches to see if there is much variation between the teams. What sort of data are being collected in such a survey?
   A. quantitative continuous
   B. quantitative discrete
   C. qualitative continuous
   D. qualitative discrete

5. In a histogram, one class is two thirds the width of the other classes. If the score in that class is 20, the correct height to plot on the histogram is
   A. 13.33
   B. 20.00
   C. 30.00
   D. 33.33

6. A histogram uses a set of bars to represent a grouped frequency table. To be correctly presented, the histogram must show the relationship of the rectangles to the frequencies by reference to the
   A. diagonal of each bar
   B. area of each bar
   C. width of each bar
   D. height of each bar

7. In a particular sample survey, an interviewer is required to interview all of the students at a school in Adelaide to find out what they think of the education system in Australia. What is the name used to describe this sampling method?
   A. cluster sampling
   B. random sampling
   C. stratified sampling
1 C Statistical studies entail collecting, organising, summarising, analysing, and drawing conclusions from data. In option A, hypothesis testing can be part of the analysis, and experimenting is a means of collecting data as is solving a system of equations in option B. Monitoring and studying a subject are means of collecting and comparing data over a period of time so option D is also incorrect.

2 C This is an example of nominal level data. There is no ordering scheme and there is no established rule that tells us which fruit should go first or last as there would be if it had been ordinal level (option D). For the level to qualify as ratio – option A, the differences between data values and a zero value would have to be meaningful e.g. a length of 30 metres is three times longer than a length of 10 metres. It is not interval level – option B. If it was, the data would have an ordering scheme and the differences between data values would be meaningful e.g., Columbus set sail for the New World 472 years before The Beatles appeared on The Ed Sullivan Show.

3 A It is an example of a continuous variable, because the time recorded can, in theory, take any value, for example 124 minutes 2.0643 seconds unlike option B – discrete – which can only be a whole number. It is not option C because qualitative data is data that cannot be measured but which reflects some quality of what is being observed.

4 B The number of goals scored is an example of quantitative data as it can be measured. Since the number of goals scored cannot take on any value, they can only score 1, 2, 3 or any whole number of goals (they cannot score 2½ goals) the data is said to be discrete.

You should have been able to eliminate options C and D immediately since qualitative data is data that cannot be measured but which reflects some quality of what is being observed.

5 C Height on histogram = \( \frac{20}{\frac{2}{3}} = 30 \)

If you selected option A, you multiplied 20 by 2/3 instead of dividing it by 2/3.

Option B is incorrect because it represents the score in the class under consideration.

Option D represents the score in the class (20) plus 2/3 of 20 = 20 + 13.33 = 33.33.

6 B A histogram is a chart that looks like a bar chart except that the bars are joined together. On a histogram, frequencies are presented by the area covered by the bars.

7 A Cluster sampling is a non-random sampling method that involves selecting one definable subsection of the population as the sample (the school in Adelaide), that subsection is then taken to be representative of the population in question (all school children in Australia).

Option B is incorrect because with random sampling every item in the population has an equal chance of being selected (which is not the case).

Stratified sampling involves dividing the population into strata and then selecting a random sample from each stratum. Option C is therefore incorrect.
Answers to chapter questions

1  Data types
(a) The number of diagrams in a textbook is a discrete variable, because it can only be counted in whole number steps. You cannot, for example, have 26\(\frac{1}{4}\) diagrams or 47.32 diagrams in a book.
(b) Whether or not a can possesses a sticker is an attribute. It is not something which can be measured. A can either possesses the attribute or it does not.
(c) How long an athlete takes to run a mile is a continuous variable, because the time recorded can, in theory, take any value, for example 4 minutes 2.0643 seconds.
(d) The height of a telegraph pole is a continuous variable.

2  Sampling methods
(a) C Cluster
(b) A Simple random
(c) B Stratified random

3  Bar graph

Workings

\[
\begin{array}{lll}
\text{20X7} & \text{20X8} & \text{20X9} \\
\$ 000 & \$ 000 & \$ 000 \\
3 579 & 2 961 & 2 192 \\
857 & 893 & 917 \\
62 & 59 & 70 \\
4 498 & 3 913 & 3 179 \\
\end{array}
\]

(a) Simple bar graph

This simple bar graph shows that total sales income has been falling each year since 20X7.
This multiple bar graph shows that although commission income and other income have remained at about the same amount in each year, there has been a fall each year in interest income, which is the largest component of total income for the bank.

4 Grouped frequency distribution

We are told what classes to use, so the first step is to identify the lowest and highest values in the data. The lowest value is $25 (in the first row) and the highest value is $73 (in the fourth row). This means that the class intervals must go up to ‘$70 and under $75’.

We can now set out the classes in a column, and then count the number of items in each class using tally marks.

<table>
<thead>
<tr>
<th>Class interval</th>
<th>Tally marks</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>$25 and less than $30</td>
<td>///</td>
<td>3</td>
</tr>
<tr>
<td>$30 and less than $35</td>
<td>///</td>
<td>4</td>
</tr>
<tr>
<td>$35 and less than $40</td>
<td>/// /// ///</td>
<td>10</td>
</tr>
<tr>
<td>$40 and less than $45</td>
<td>/// /// /// ///</td>
<td>15</td>
</tr>
<tr>
<td>$45 and less than $50</td>
<td>/// /// /// /// ///</td>
<td>18</td>
</tr>
<tr>
<td>$50 and less than $55</td>
<td>/// /// /// /// /// ///</td>
<td>20</td>
</tr>
<tr>
<td>$55 and less than $60</td>
<td>/// /// /// /// ///</td>
<td>13</td>
</tr>
<tr>
<td>$60 and less than $65</td>
<td>/// /// ///</td>
<td>8</td>
</tr>
<tr>
<td>$65 and less than $70</td>
<td>/// ///</td>
<td>6</td>
</tr>
<tr>
<td>$70 and less than $75</td>
<td>///</td>
<td>3</td>
</tr>
</tbody>
</table>

Total 100

5 Histogram

This is a grouped frequency distribution for continuous variables.

Before drawing the histogram, we must decide on the following:

(a) A **standard class width**: $2 000 will be chosen.

(b) An **open-ended class width**. In this example, the open-ended class width will therefore be $2 000 for class ‘up to $10 000’ and $4 000 for the class ‘$ > $22 000’.
<table>
<thead>
<tr>
<th>Class interval</th>
<th>Size of interval</th>
<th>Frequency</th>
<th>Adjustment</th>
<th>Height of bar</th>
</tr>
</thead>
<tbody>
<tr>
<td>Up to $10,000</td>
<td>$</td>
<td>2,000</td>
<td>$2/2</td>
<td>1</td>
</tr>
<tr>
<td>&gt; $10,000 ≤ $12,000</td>
<td>2,000</td>
<td>10</td>
<td>$2/2</td>
<td>10</td>
</tr>
<tr>
<td>&gt; $12,000 ≤ $14,000</td>
<td>2,000</td>
<td>12</td>
<td>$2/2</td>
<td>12</td>
</tr>
<tr>
<td>&gt; $14,000 ≤ $18,000</td>
<td>4,000</td>
<td>8</td>
<td>$2/4</td>
<td>4</td>
</tr>
<tr>
<td>&gt; $18,000 ≤ $22,000</td>
<td>4,000</td>
<td>4</td>
<td>$2/4</td>
<td>2</td>
</tr>
<tr>
<td>&gt; $22,000</td>
<td></td>
<td>4,000</td>
<td>$2/4</td>
<td>1/2</td>
</tr>
</tbody>
</table>

This shows that the largest proportion of sales staff made sales in the range $10,000 to $14,000, and relatively few achieved sales in excess of $18,000.
Chapter 13

Descriptive statistics

<table>
<thead>
<tr>
<th>Learning objectives</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Statistical analysis, data, and methods of describing data</td>
<td>LO12</td>
</tr>
<tr>
<td>Construct a bar graph, a pie chart, a histogram, and a scatter diagram from a given set of data</td>
<td>LO12.6</td>
</tr>
<tr>
<td><strong>Descriptive statistics</strong></td>
<td>LO13</td>
</tr>
<tr>
<td>Distinguish between measures of central tendency and measures of variability</td>
<td>LO13.1</td>
</tr>
<tr>
<td>Distinguish between the shapes of a normal distribution, exponential distribution and binomial distribution</td>
<td>LO13.2</td>
</tr>
<tr>
<td>Explain the difference between grouped and ungrouped data</td>
<td>LO13.3</td>
</tr>
<tr>
<td>Calculate and interpret the mean, median, and mode from a given set of data</td>
<td>LO13.4</td>
</tr>
<tr>
<td>Calculate and interpret the range, standard deviation, and variance from a given set of data</td>
<td>LO13.5</td>
</tr>
<tr>
<td>Distinguish between the sample and population standard deviation and the sample and population variance</td>
<td>LO13.6</td>
</tr>
<tr>
<td>Distinguish between kurtosis and skewness</td>
<td>LO13.7</td>
</tr>
<tr>
<td><strong>Frequency distributions and probability</strong></td>
<td>LO14</td>
</tr>
<tr>
<td>Distinguish between class range, class midpoint, relative frequency, and cumulative frequency</td>
<td>LO14.2</td>
</tr>
</tbody>
</table>

**Topic list**

1. Central tendency and variability
2. Averages
3. Measures of variability
4. Distribution shapes
In the previous chapter we saw how data can be summarised and presented in tabular, chart and graphical formats. The graph is a visual image of the data, and is an important first step in data analysis.

In this chapter we go further than the compilation of a frequency distribution and condense the data into two parameters that characterise the distribution. The first is a measure of central tendency, a typical value around which the various items are grouped i.e. an average. The second is a measure of dispersion i.e., some indication of the way in which these items are spread around the average.

An average is a representative figure that is used to give some impression of the mean size of all the items in the population. There are three main types of average: the arithmetic mean, the mode and the median. We will be looking at each of these averages in turn, their calculation, advantages and disadvantages before calculating and interpreting the main measures of variability (dispersion) - the range, standard deviation and variance for both a sample and a population. The interpretation of this variability will allow us to distinguish between the shapes of a normal distribution, exponential distribution and binomial distribution.
Before you begin

If you have studied these topics before, you may wonder whether you need to study this chapter in full. If this is the case, please attempt the questions below, which cover some of the key subjects in the area.

If you answer all these questions successfully, you probably have a reasonably detailed knowledge of the subject matter, but you should still skim through the chapter to ensure that you are familiar with everything covered.

There are references in brackets indicating where in the chapter you can find the information, and you will also find a commentary at the back of the Study Manual.

1. How do we measure central tendency (location) of data?  
   (Sections 1.1, 1.3)

2. **Weekly output**  
   **Cumulative frequency**  
   (Section 2.3.1)

<table>
<thead>
<tr>
<th>Units</th>
<th>Cumulative frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>Less than 150</td>
<td>90</td>
</tr>
<tr>
<td>Less than 200</td>
<td>250</td>
</tr>
<tr>
<td>Less than 300</td>
<td>310</td>
</tr>
<tr>
<td>Less than 400</td>
<td>340</td>
</tr>
<tr>
<td>Less than 600</td>
<td>350</td>
</tr>
</tbody>
</table>

   A. How many employees produced more than 200 units?
   B. How many employees produced between 200 and 300 units?

3. Which measure of central tendency is the most sensitive to extreme scores?  
   (Section 2.3.3)

   A. mean  
   B. median  
   C. mode  
   D. variance

4. Suppose the mean of our first exam is 77 and the median is 82.  
   (Section 2.3.3)
   This would indicate a _____ distribution.

   A. normal  
   B. bimodal  
   C. positively skewed  
   D. negatively skewed

5. Measures of variability give information about the amount of skew or bias in a set of data.  
   (Section 1.3)

   A. true  
   B. false

6. When data is normally distributed, the mean, median, and mode:  
   (Section 4.4)

   A. are all useful for examining the spread of scores.  
   B. all have the same value.  
   C. should never be used together.  
   D. vary randomly.

7. When data is normally distributed, what percentage of scores falls within the one standard deviation interval around the mean (-1 SD to +1 SD)?  
   (Section 4.1)

   A. 68.26 per cent  
   B. 100 per cent  
   C. 34.13 per cent  
   D. 95.00 per cent
8 When a customer calls the ‘Help Line’ at ABC Computer Software Co., the amount of time that a customer must wait ‘on hold’ until somebody answers the line and helps the customer follows an exponential distribution with a mean of 7.5 minutes. (Section 4.2)

What is the probability that a customer waits more than 10 minutes to receive help?
A 0.2636
B 0.75
C 0.7364
D 0

9 Which of the following is not true about the normal distribution? (Section 4.1)
A the mean, median and mode are equal
B the curve is skewed to the right
C the curve never touches the x-axis
D the area under the curve is one

10 Distinguish between positive skewness and negative skewness. (Section 4.4)
1 Central tendency and variability

1.1 Descriptive statistics

Descriptive statistics are ways of describing, organising, summarising, and presenting large sets of quantitative (numerical) data. They are most often used to examine:

(a) Central tendency (location) of data, i.e. where data tend to fall, as measured by the mean, median, and mode.
(b) Dispersion (variability) of data, i.e. how spread out data is, as measured by the variance and its square root, the standard deviation.
(c) Shape i.e. as described by symmetry, skewness, and kurtosis (peakedness) of data.
(d) Unusual features such as gaps (areas of the distribution where there are no observations) and outliers.

Of course, the ‘best way’ of organising data depends on what you want to find out or demonstrate from it. The questions you are interested in answering should guide what statistical presentation you decide to use.

1.2 Grouped and ungrouped data

In statistics, grouped data is an arrangement of raw data with a wide range of values into groups. This process makes the data more manageable. Graphs and frequency diagrams can then be drawn showing the class intervals chosen instead of individual values. To obtain an idea of the distribution, the data is broken down into convenient classes (commonly 6 to 16 classes), which must be mutually exclusive and are usually equal in width to make it easier for histograms to be drawn. The class boundaries should clearly define the range of each class. When dealing with discrete data, suitable intervals would be, for example, 0–2, 3–5, 6–8, and so on. When dealing with continuous data, suitable intervals might be > 0 ≤ 5, > 5 ≤ 10, >10 ≤ 15, >15 ≤ 20 and so on.

The frequency distribution is the foundation of descriptive statistics. It is a prerequisite for the various graphs used to display data and the basic statistics used to describe a data set, such as the mean, median, mode, variance and standard deviation.

In this chapter we will outline the techniques used with both grouped and ungrouped data.

1.3 Central tendency and variability

Researchers are often interested in defining a value that best describes some attribute of the population.

Definition

Measures of central tendency tell us about the values that are in the middle of the data set.

Central tendency is a statistical measure that identifies a single value as representative of an entire distribution of values. The goal of central tendency is to find the single value that is most typical or most representative of the entire distribution. Unfortunately, there is no single, standard procedure for determining central tendency. The problem is that there is no single measure that will always produce a central, representative value in every situation. There are three main measures of central tendency: the arithmetical mean, the median and the mode.

Definition

Measures of variability tell us how spread out the values are in a data set. The terms variability, spread, and dispersion are synonyms, and refer to how spread out a distribution is.
Variability provides a quantitative measure of the degree to which values in a distribution are spread out. The greater the difference between values, the more spread out the distribution is. The more tightly the values group together, the less variability there is in the distribution. Variability is the essence of statistics. The most frequently used methods of measurement of this variance are: range, deviation and variance, inter-quartile range and standard deviation.

2 Averages

2.1 The arithmetic mean

Definition

The arithmetic mean is calculated from the sum of values of items divided by the number of items. The arithmetic mean of a variable \( x \) is denoted by \( \bar{x} \) ("x bar").

\[
\text{Arithmetic mean} = \frac{\text{Sum of values of items}}{\text{Number of items}}
\]

For example, the mean wage of a work force of ten men is the amount each worker would receive if all their earnings were pooled and then shared out equally among them.

Worked Example: The arithmetic mean

The demand for a product on each of 20 days was as follows (in units).

\[
3 \quad 12 \quad 7 \quad 17 \quad 3 \quad 14 \quad 9 \quad 6 \quad 11 \quad 10 \quad 1 \quad 4 \quad 19 \quad 7 \quad 15 \quad 6 \quad 9 \quad 12 \quad 12 \quad 8
\]

Solution

We take the total of the 20 days' demand, which is 185 and divide it by 20.

The arithmetic mean of daily demand is

\[
\text{Sum of demand} = 185 \quad \text{Number of days} = 20
\]

\[
\bar{x} = \frac{185}{20} = 9.25 \text{ units}
\]

In the above example, demand on any one day is never actually 9.25 units. The arithmetic mean is merely an average representation of demand on each of the 20 days.
2.1.1 Finding the arithmetic mean of data in a frequency distribution

When analysing data it is more likely that you will be asked to calculate the arithmetic mean of a frequency distribution. In our previous example, the frequency distribution would be shown as follows:

<table>
<thead>
<tr>
<th>Daily demand</th>
<th>Frequency</th>
<th>Demand × frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>3</td>
<td>6</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td>4</td>
<td>2</td>
<td>12</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>8</td>
</tr>
<tr>
<td>6</td>
<td>2</td>
<td>14</td>
</tr>
<tr>
<td>7</td>
<td>2</td>
<td>18</td>
</tr>
<tr>
<td>8</td>
<td>1</td>
<td>10</td>
</tr>
<tr>
<td>9</td>
<td>1</td>
<td>11</td>
</tr>
<tr>
<td>10</td>
<td>2</td>
<td>20</td>
</tr>
<tr>
<td>11</td>
<td>1</td>
<td>11</td>
</tr>
<tr>
<td>12</td>
<td>3</td>
<td>36</td>
</tr>
<tr>
<td>13</td>
<td>1</td>
<td>14</td>
</tr>
<tr>
<td>14</td>
<td>1</td>
<td>15</td>
</tr>
<tr>
<td>15</td>
<td>1</td>
<td>17</td>
</tr>
<tr>
<td>16</td>
<td>1</td>
<td>19</td>
</tr>
<tr>
<td>17</td>
<td>1</td>
<td>20</td>
</tr>
<tr>
<td>18</td>
<td>1</td>
<td>185</td>
</tr>
<tr>
<td>19</td>
<td>1</td>
<td>185</td>
</tr>
</tbody>
</table>

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Daily demand</td>
<td></td>
</tr>
<tr>
<td>0  ≤  x  &lt;  5</td>
<td>4</td>
</tr>
<tr>
<td>5  ≤  x  &lt;  10</td>
<td>8</td>
</tr>
<tr>
<td>10 ≤ x &lt; 15</td>
<td>6</td>
</tr>
<tr>
<td>15 ≤ x &lt; 20</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>20</td>
</tr>
</tbody>
</table>

\[
\bar{x} = \frac{\Sigma fx}{n} = 9.25
\]

2.1.2 The summation sign, sigma (\(\Sigma\))

The statistical notation for the arithmetic mean of a set of data uses the symbol \(\Sigma\) (sigma). \(\Sigma\) means 'the sum of' and is used as shorthand to mean 'the sum of a set of values'.

Thus, in the previous example:

(a) \(\Sigma f\) would mean the sum of all the frequencies, which is 20;

(b) \(\Sigma fx\) would mean the sum of all the values of 'frequency multiplied by daily demand x', that is, all 14 values of \(fx\) are summed, so \(\Sigma fx = 185\).

2.1.3 The symbolic formula for the arithmetic mean of a frequency distribution

Using the \(\Sigma\) sign, the formula for the arithmetic mean of a frequency distribution is:

\[
\bar{x} = \frac{\Sigma fx}{n}
\]

or

\[
\bar{x} = \frac{\Sigma x}{\Sigma f}
\]

where \(n\) is the number of values recorded, or the number of items measured.

2.1.4 Finding the arithmetic mean of grouped data in class intervals

You might also be asked to calculate (or at least approximate) the arithmetic mean of a frequency distribution, where the frequencies are shown in class intervals.

Using the previous example, the frequency distribution might have been shown as follows:

<table>
<thead>
<tr>
<th>Daily demand</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>0  ≤  x  &lt;  5</td>
<td>4</td>
</tr>
<tr>
<td>5  ≤  x  &lt;  10</td>
<td>8</td>
</tr>
<tr>
<td>10 ≤ x &lt; 15</td>
<td>6</td>
</tr>
<tr>
<td>15 ≤ x &lt; 20</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>20</td>
</tr>
</tbody>
</table>
There is, of course, an extra difficulty with finding the average now; as the data has been collected into classes, a certain amount of detail has been lost and the values of the variables to be used in the calculation of the mean are not clearly specified.

To calculate the arithmetic mean of grouped data we therefore need to decide on a value which best represents all of the values in a particular class interval.

The mid-point of each class interval is conventionally taken, on the assumption that the frequencies occur evenly over the class interval range. In the example above, the variable is discrete, so the first class includes 1, 2, 3, 4 and 5, giving a mid-point of 3. With a continuous variable (such as quantities of fuel consumed in litres), the mid-points would have been 2.5, 7.5 and so on. Once the value of $x$ has been decided, the mean is calculated in exactly the same way as in Paragraph 2.1.

**Worked Example: The arithmetic mean of grouped data**

<table>
<thead>
<tr>
<th>Daily demand</th>
<th>Mid point</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$x$</td>
<td>$f$</td>
</tr>
<tr>
<td>$&gt;0 \leq 5$</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>$&gt;5 \leq 10$</td>
<td>8</td>
<td>8</td>
</tr>
<tr>
<td>$&gt;10 \leq 15$</td>
<td>13</td>
<td>6</td>
</tr>
<tr>
<td>$&gt;15 \leq 20$</td>
<td>18</td>
<td>2</td>
</tr>
</tbody>
</table>

$\sum f = 20$  $\sum fx = 190$

**Solution**

Arithmetic mean $\bar{x} = \frac{\sum fx}{\sum f} = \frac{190}{20} = 9.5$ units.

Because the assumption that frequencies occur evenly within each class interval is not quite correct in this example, our approximate mean of 9.5 does not exactly match the value of 9.25 calculated before when we knew every individual value of daily demand.

**2.1.5 Finding the arithmetic mean of combined data**

Suppose that the mean age of a group of five people is 27 and the mean age of another group of eight people is 32. How would we find the mean age of the whole group of thirteen people?

Remember that the arithmetic mean is calculated as

\[
\text{Sum of values of items} \div \text{Number of items}
\]

The sum of the ages in the first group is $5 \times 27 = 135$

The sum of the ages in the second group is $8 \times 32 = 256$

The sum of all 13 ages is $135 + 256 = 391$

The mean age is therefore $\frac{391}{13} = 30.08$ years.

**Question 1: Mean weight**

The mean weight of 10 units at 5 kg, 10 units at 7 kg and 20 units at $X$ kg is 8 kg. What is the value of $X$?

(The answer is at the end of the chapter)
2.1.6 The advantages and disadvantages of the arithmetic mean

The advantages of the arithmetic mean are as follows:

(a) It is easy to calculate.
(b) It is widely understood.
(c) The value of every item is included in the computation of the mean and so it can be determined with arithmetical precision and is representative of the whole set of data.
(d) It is supported by mathematical theory and is suited to further statistical analysis.

The disadvantages of the arithmetic mean are as follows.

(a) Its value may not correspond to any actual value. For example, the 'average' family might have 2.3 children, but no family can have exactly 2.3 children.
(b) An arithmetic mean might be distorted by extremely high or low values. For example, the mean of 3, 4, 4 and 6 is 4.25, but the mean of 3, 4, 4, 6 and 15 is 6.4. The high value, 15, distorts the average and in some circumstances the mean would be a misleading and inappropriate figure. (Note that extreme values are not uncommon in economic data.)

Question 2: Arithmetic mean

For the week ended 15 November, the overtime earned by the 69 operators employed in the machine shop of Mermaid Ltd was as follows:

<table>
<thead>
<tr>
<th>Overtime</th>
<th>Number of operators</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\leq$ 60</td>
<td>3</td>
</tr>
<tr>
<td>&gt;$60 \leq$ 70</td>
<td>11</td>
</tr>
<tr>
<td>&gt;$70 \leq$ 80</td>
<td>16</td>
</tr>
<tr>
<td>&gt;$80 \leq$ 90</td>
<td>15</td>
</tr>
<tr>
<td>&gt;$90 \leq$ 100</td>
<td>10</td>
</tr>
<tr>
<td>&gt;$100 \leq$ 110</td>
<td>8</td>
</tr>
<tr>
<td>&gt;$110</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td>69</td>
</tr>
</tbody>
</table>

Calculate the arithmetic mean overtime earnings of the machine operators of Mermaid Ltd for the week ended 15 November. 

(The answer is at the end of the chapter)

2.2 The mode

The second measure of central tendency is the mode or modal value.

Definition

The mode is the value which indicates the most frequently occurring value.

Worked Example: Finding the mode

The daily demand for stock in a ten-day period is as follows:

<table>
<thead>
<tr>
<th>Demand Units</th>
<th>Number of days</th>
</tr>
</thead>
<tbody>
<tr>
<td>6</td>
<td>3</td>
</tr>
<tr>
<td>7</td>
<td>6</td>
</tr>
<tr>
<td>8</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>10</td>
</tr>
</tbody>
</table>

Solution

The mode is 7 units, because it is the value which occurs most frequently.
2.2.1 Finding the mode of a grouped frequency distribution

The mode of a grouped frequency distribution can be calculated from a histogram.

Worked Example: Finding the mode from a histogram

Consider the following grouped frequency distribution:

<table>
<thead>
<tr>
<th>Value</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>At least</td>
<td>Less than</td>
</tr>
<tr>
<td>0</td>
<td>10</td>
</tr>
<tr>
<td>10</td>
<td>20</td>
</tr>
<tr>
<td>20</td>
<td>30</td>
</tr>
<tr>
<td>30</td>
<td>40</td>
</tr>
</tbody>
</table>

The modal class (the one with the highest frequency) is 'at least 20, less than 30'. But how can we find a single value to represent the mode?

What we need to do is draw a histogram of the frequency distribution.

Solution

The modal class is always the class with the tallest bar, which may not be the class with the highest frequency if the classes do not all have the same width.

We can estimate the mode graphically as follows.

(a) Join with a straight line the top left-hand corner of the bar for the modal class and the top left-hand corner of the next bar to the right.

(b) Join with a straight line the top right-hand corner of the bar for the modal class and the top right-hand corner of the next bar to the left.

Where these two lines intersect, we find the estimated modal value. In this example it is approximately 27.

A better way of estimating the modal value would be by the formula:

$$\text{Mode} = \frac{b + c - a}{2b} \times \text{width of column} + \text{starting value of column b}$$

where $a =$ the frequency of the lowest value, $b =$ the frequency of the most common value and $c =$ the frequency of the highest value.

$$\text{Mode} = \frac{150 + 100 - 50}{150 \times 2} \times 10 + 20 = 26.7$$
We are assuming that the frequencies occur evenly within each class interval but this may not always be correct. It is unlikely that the 150 values in the modal class occur evenly. Hence the mode in a grouped frequency distribution is only an estimate.

2.2.2 The advantages and disadvantages of the mode

The mode is more appropriate to use than the mean in situations where it is useful to know the most common value. For example, if a manufacturer wishes to start production in a new industry, it might be helpful to him to know what sort of product made by that industry is most in demand with customers.

The advantages of the mode are as follows.

(a) It is easy to find.
(b) It is not influenced by a few extreme values.
(c) It can be used for data which is not even numerical (unlike the mean and median).
(d) It can be the value of an actual item in the distribution.

The mode does have a number of disadvantages.

(a) It may be unrepresentative; it takes no account of a high proportion of the data, only representing the most common value.
(b) It does not take every value into account.
(c) There can be two or more modes within a set of data.
(d) If the modal class is only very slightly bigger than another class, just a few more items in this other class could mean a substantially different result, suggesting some instability in the measure.

2.3 The median

The third measure of central tendency is the median.

Definition

The median is the value of the middle member of a distribution once all of the items have been arranged in order of magnitude.

The median is the middle number or the 50th percentile. The position of a particular percentile \( P \) in \( n \) items arranged in order of magnitude can be calculated by \( P/100 \times (n+1) \). The median of 13 numbers in a sample or population (so \( n = 13 \)) = \( 50/100 \times (13 +1) = 7 \)th number.


**Worked Example: The median**

The median of the following nine values:

| 8 | 6 | 9 | 12 | 15 | 6 | 3 | 20 | 11 |

is found by taking the middle item (the fifth one) in the array:

| 3 | 6 | 6 | 8 | 9 | 11 | 12 | 15 | 20 |

Therefore, the median is 9.

The middle item of an odd number of items is calculated as the \( \frac{(n + 1)\text{th}}{2} \) item.

Consider the following array.

| 1 | 2 | 2 | 2 | 3 | 5 | 6 | 7 | 8 | 11 |

The median is 4 because, with an even number of items, we have to take the arithmetic mean of the two middle ones (in this example, \( \frac{3 + 5}{2} = 4 \)).

**Question 3: The median**

The following times taken to produce a batch of 100 units of Product X have been noted:

| 21 min | 17 min | 24 min | 11 min | 37 min | 27 min |
| 20 min | 15 min | 17 min | 23 min | 29 min | 30 min |
| 24 min | 18 min | 17 min | 21 min | 24 min | 20 min |

What is the median time?

(The answer is at the end of the chapter)

### 2.3.1 Cumulative frequency distributions

**Definition**

A *cumulative distribution* (or cumulative frequency table) can be used to show the total number of times that a value above (or below) a certain amount occurs.

**Worked Example: Cumulative frequency table**

A teacher arranged the marks gained by all year 10 pupils in a mathematics test in a table as shown below:

<table>
<thead>
<tr>
<th>Marks</th>
<th>Frequency of pupils</th>
</tr>
</thead>
<tbody>
<tr>
<td>11-20</td>
<td>2</td>
</tr>
<tr>
<td>21-30</td>
<td>11</td>
</tr>
<tr>
<td>31-40</td>
<td>19</td>
</tr>
<tr>
<td>41-50</td>
<td>36</td>
</tr>
<tr>
<td>51-60</td>
<td>42</td>
</tr>
<tr>
<td>61-70</td>
<td>31</td>
</tr>
<tr>
<td>71-80</td>
<td>13</td>
</tr>
<tr>
<td>81-90</td>
<td>6</td>
</tr>
</tbody>
</table>

This table shows the number of pupils (called the frequency) who gained marks in the various mark bands. For example, the number of pupils who scored between 21 and 30 marks was 11. No pupil scored fewer than 11 marks or more than 90 marks.
To create a cumulative total for the frequency of pupils in each group (called the cumulative frequency) a third column is created as shown below:

<table>
<thead>
<tr>
<th>Marks</th>
<th>Frequency</th>
<th>Cumulative frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>11-20</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>21-30</td>
<td>11</td>
<td>(2+11) = 13</td>
</tr>
<tr>
<td>31-40</td>
<td>19</td>
<td>(13+19) = 32</td>
</tr>
<tr>
<td>41-50</td>
<td>36</td>
<td>(32+36) = 68</td>
</tr>
<tr>
<td>51-60</td>
<td>42</td>
<td>(68+42) = 110</td>
</tr>
<tr>
<td>61-70</td>
<td>31</td>
<td>(110+31) = 141</td>
</tr>
<tr>
<td>71-80</td>
<td>13</td>
<td>(141+13) = 154</td>
</tr>
<tr>
<td>81-90</td>
<td>6</td>
<td>(154+6) = 160</td>
</tr>
</tbody>
</table>

The cumulative frequency column makes it easy to see at a glance that 68 pupils scored 50 marks or fewer, and that 32 pupils scored 40 marks or fewer.

**Cumulative frequency graph for Year 10 Mathematics Results**

A cumulative frequency graph is a way of presenting information visually, which allows other information to be deduced.

For example, from the graph we can obtain the median (or middle) mark. The median is the mark which half of all pupils exceed and half do not reach. As there are 160 pupils, we need to find 80 pupils, at the halfway point on the vertical axis, and then draw a line across until it meets the graph. Drawing a vertical line down from this point and reading the number of marks at that point shows that the median is 53 marks.

**Cumulative frequency graph for Year 10 Mathematics Results**
Note that because we are assuming that the values are spread evenly within each class, the median calculated is only approximate.

### 2.3.2 The advantages and disadvantages of the median

The median is only of interest where there is a range of values and the middle item is of some significance. Perhaps the most suitable application of the median is in comparing changes in incomes or house prices over time.

The median is easy to understand and (like the mode) is unaffected by extremely high or low values. It can be the value of an actual item in the distribution. On the other hand, it fails to reflect the full range of values, it is unsuitable for further statistical analysis and arranging the data in order of size can be tedious.

#### Exam comments

If you are asked to find the median of a set of ungrouped data, remember to arrange the items in order of size first and then to count the number of items in the array. If you have an even number of items, the median may not be the value of one of the items in the data set. The median of an even number of items is found by calculating the arithmetic mean of the two middle items.

### 2.3.3 Relationship between the mean, median and mode

In a symmetrical frequency distribution like the normal distribution, the mean, median, and mode are all the same value, \( M \). Its two halves mirror each other. This can be drawn as follows:

In general the mean is the best measure of central tendency. In its calculation (the sum of all the values divided by the number of values), it does represent all of the values. If any value changes then the mean will change. This is not necessarily true of the median or mode.

This implies that extreme values in either the high or the low direction will have a much greater effect on the mean than they would have on the median or the mode.

#### Worked Example: Asymmetrical frequency distributions

Take the situation where we have a small factory with nine employees and a manager. Five of the employees are paid $20,000 per year, four are paid $27,000, while the manager is paid $92,000. If we were to make a frequency distribution of these salaries it would look somewhat like the following:
This frequency distribution is said to be positively skewed, that is the long slope of the curve is in the positive direction. A positively skewed distribution is caused by a relatively few high scores, or in the present example a single high score (the manager’s salary). The total for salaries is $300 000 so the mean salary would be $30 000. We can see that the mean has been drastically increased relative to the mode and the median. In this case of a skewed distribution, the median is probably a better measure of central tendency than the mean.

Note: The mode is $20 000 and the median – mid-way between the salaries of the 5th and 6th employees – is $(20 000 + 27 000)/2 = $23 500. With this positively skewed distribution, the mode is less than the median, which is less than the mean.

We can also have a skewed distribution because of a relatively few very low scores. For example, you live in a neighbourhood in which there are nine homes valued at $150 000, $140 000, $160 000, $150 000, $160 000, $170 000, $160 000, $150 000, and $160 000. There is one small empty lot in the area and someone builds a garage on it with a valuation of $20 000.

Mean = $142 000 (Arithmetic mean \( \bar{x} = \frac{\text{sum of values}}{n} \))

Median = $155 000 (calculated as the \( \frac{n+1}{2} \)th item)

Mode = $160 000 (because it is the value which occurs most frequently)

If we were to graph the frequency distribution of these 10 values it would be a negatively skewed curve and would look something like:

With this negatively skewed distribution, the mode is greater than the median, which is greater than the mean.

In this case the median might be a more representative measure of central tendency than the mean.

Although in general, the mean is usually the best measure of central tendency, and is the measure we will use to develop further statistics such as the variance and the standard deviation, in the case of a badly skewed distribution (either negatively or positively skewed) the median may be a better measure of central tendency than the mean.

We will be looking at skewness in more detail in a later section of this chapter.
3 Measures of variability

Section overview
- An average is not in itself an adequate summary of a frequency distribution, we also need to know the dispersion of the data, that is, how spread around the mean are the values (e.g. are they all closely clustered around the mean or are they well scattered). Measures of dispersion include the range, mean deviation, variance and standard deviation.

While the measures of central tendency convey information about the middle of the values, the measures of variability quantify the degree to which they vary around the middle.

3.1 The range

Definition
The range is the difference between the highest observation and the lowest observation.

The range is the simplest measure of variation and indicates the 'length' a distribution covers. It is determined by finding the difference between the lowest and highest value in a distribution.

- A large range means there is more variability.
- A small range means there is less variability.

The main properties of the range as a measure of dispersion are as follows:
(a) It is easy to find and to understand.
(b) It is easily affected by one or two extreme values.
(c) It gives no indication of spread between the extremes.
(d) It is not suitable for further statistical analysis.

3.1.1 Ungrouped data
The range for ungrouped data = highest extreme value – lowest extreme value. The range can be expressed as an interval such as 4-10, where 4 is the lowest value and 10 is highest. Sometimes it is expressed as an interval width; that is, the range of 4 – 10 is 6.

Worked Example: Range
Suppose that the marks obtained in an examination were:
24, 27, 36, 48, 52, 52, 53, 53, 59, 60, 85, 90, 95

Solution
Lowest Value = 24 and Highest Value = 95
Therefore, the range = 95 – 24 = 71
The range is 71 marks.

3.1.2 Grouped data
The range for grouped data = the mid-point of the highest group – the mid-point of the lowest group.
Worked Example: The range of grouped data
Consider the following data set of miles recorded by 120 sales people in one week:

<table>
<thead>
<tr>
<th>Mileage</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>400 – 419</td>
<td>12</td>
</tr>
<tr>
<td>420 – 439</td>
<td>27</td>
</tr>
<tr>
<td>440 – 459</td>
<td>34</td>
</tr>
<tr>
<td>460 – 479</td>
<td>24</td>
</tr>
<tr>
<td>480 – 499</td>
<td>15</td>
</tr>
<tr>
<td>500 – 519</td>
<td>8</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Mileage</th>
<th>Mid-range</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>400 – 419</td>
<td>409.5</td>
<td>12</td>
</tr>
<tr>
<td>420 – 439</td>
<td>429.5</td>
<td>27</td>
</tr>
<tr>
<td>440 – 459</td>
<td>449.5</td>
<td>34</td>
</tr>
<tr>
<td>460 – 479</td>
<td>469.5</td>
<td>24</td>
</tr>
<tr>
<td>480 – 499</td>
<td>489.5</td>
<td>15</td>
</tr>
<tr>
<td>500 – 519</td>
<td>509.5</td>
<td>8</td>
</tr>
</tbody>
</table>

The mid-range of the highest group = 509.5
The mid-range of the lowest group = 409.5
Thus the range = 509.5 – 409.5 = 100 miles. Note in this example we are supposing only the grouped data table was available so the exact range of the 120 values could not be calculated.

A great deal of information is ignored when computing the range, since only the largest and smallest data values are considered. The range value of a data set can be greatly influenced by the presence of just one unusually large or small value (outlier).

The disadvantage of using range is that it does not measure the spread of the majority of values in a data set – it only measures the spread between highest and lowest values. As a result, other measures are required in order to give a better picture of the data spread.

Question 4: The mean and the range
Calculate the mean and the range of each of the following sets of data:

(a) \( x_1 = 4 \ 8 \ 7 \ 3 \ 5 \ 16 \ 24 \ 5 \)

(b) \( x_2 = 10 \ 7 \ 9 \ 11 \ 11 \ 8 \ 9 \ 7 \)

What do your calculations show about the dispersion of the data sets?

(The answer is at the end of the chapter)

3.1.3 Quantiles
Quartiles, deciles, and percentiles and any other similar dividing points for analysing a frequency distribution are referred to collectively as quantiles. The purpose of quantiles is to analyse the dispersion of data values.

A measure that expresses position in terms of a percentage is called a percentile for the data set.

Definition
Suppose a data set is arranged in ascending order. The \( p \)th percentile is a number such that \( p \% \) of the observations of the data set fall below it.
We have already defined the median as the 50th percentile. Other statistics can also be obtained from the data set and revolve around the idea of percentiles.

In a similar way, a population could be divided into ten equal groups, and this time the value of each dividing point is referred to as a **decile**.

**Quartiles** are one means of identifying the range within which most of the values in the population occur. The lower quartile is the value below which 25 per cent of the population fall and the upper quartile is the value above which 25 per cent of the population fall. If there were 12 data items the lower quartile would be the fourth item and the upper quartile the ninth item. It follows that 50 per cent of the total population fall between the lower and the upper quartiles.

The three quartiles, or the 25th percentile (Q1), the median (Q2) and 75th percentile (Q3) are often used to describe a data set because they divide the data set into four groups, with each group containing one-fourth (25 per cent) of the observations. They would also divide the relative frequency distribution for a data set into four parts; each contains the same area (0.25), as shown below:

![Graph showing quartiles](image_url)

The first part consists of values in the range from the 1st–25th percentiles, the second part consists of values in the range from the 26th–50th percentiles, the third part consists of values in the range from the 51st–75th percentiles, and the fourth part consists of values in the range from the 76th–100th percentiles.

**Worked Example: Quartiles**

Consider the marks of 19 students obtained in an examination arranged in the ascending order of magnitude.

| 20 | 27 | 29 | 33 | 37 | 40 | 42 | 48 | 50 | 53 | 55 | 62 | 81 | 83 | 88 | 90 | 91 | 95 | 100 |

**Solution**

To find Q1: \( n = 19 \) thus \( Q1 = (n+1)/4 = 20/4 = 5 \)th observation = 37 marks

Q2 is \( (n + 1) /2 = 20/2 = 10 \)th observation (Median) = 53 marks

Q3 is \( 3(n + 1)/4 = 60/4 = 15 \)th observation = 88 marks

Also note that Q1 and Q3 are equidistant from the median. In any perfectly symmetrical distribution (normal distribution) Median – Q1 = Q3 – Median.

**Question 5: Quartiles**

From the data below calculate the median, the upper quartile and the lower quartile:

6, 47, 49, 15, 43, 41, 7, 39, 43, 41, 36

(The answer is at the end of the chapter)

**3.1.4 Quartiles of a grouped distribution**

When the data is grouped, the quartiles cannot normally be determined accurately; they can only be estimated as the median is estimated, either by interpolation or by means of the cumulative frequency graph.
Worked Example: Quartiles

The hours of overtime worked in a particular quarter by the 60 employees of ABC Ltd are shown below. The company has decided to give a bonus of $15 to the 10 per cent of the workforce who worked the most overtime, a bonus of $5 to the 20 per cent of the workforce who worked the least overtime and a $10 bonus to all other employees.

<table>
<thead>
<tr>
<th>Hours</th>
<th>More than</th>
<th>Not more than</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>10</td>
<td>3</td>
</tr>
<tr>
<td>10</td>
<td>10</td>
<td>20</td>
<td>6</td>
</tr>
<tr>
<td>20</td>
<td>20</td>
<td>30</td>
<td>11</td>
</tr>
<tr>
<td>30</td>
<td>30</td>
<td>40</td>
<td>15</td>
</tr>
<tr>
<td>40</td>
<td>40</td>
<td>50</td>
<td>12</td>
</tr>
<tr>
<td>50</td>
<td>50</td>
<td>60</td>
<td>7</td>
</tr>
<tr>
<td>60</td>
<td>60</td>
<td>70</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>60</td>
</tr>
</tbody>
</table>

Required

Calculate the range of overtime hours worked by those employees receiving a $10 bonus.

Solution

<table>
<thead>
<tr>
<th>Hours</th>
<th>Cumulative frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>&gt; 0 ≤ 10</td>
<td>3</td>
</tr>
<tr>
<td>&gt;10 ≤ 20</td>
<td>9</td>
</tr>
<tr>
<td>&gt;20 ≤ 30</td>
<td>20</td>
</tr>
<tr>
<td>&gt;30 ≤ 40</td>
<td>35</td>
</tr>
<tr>
<td>&gt;40 ≤ 50</td>
<td>47</td>
</tr>
<tr>
<td>&gt;50 ≤ 60</td>
<td>54</td>
</tr>
<tr>
<td>&gt;60 ≤ 70</td>
<td>60</td>
</tr>
</tbody>
</table>

The 9th decile is at $60 – (10% of 60) = 54^{th} \text{item}$

The 2nd decile is at 20% of 60 = 12^{th} \text{item}$

From the cumulative frequency graph the range of hours is approximately 23 to 60.
3.1.5 Inter-quartile range

Definition

The **inter-quartile range** is the difference between the values of the upper and lower quartiles and hence shows the range of values of the middle half of the population, i.e., upper quartile (Q3) – lower quartile (Q1).

If we concentrate on two extreme values (as in the case of range), we don’t get any idea about the scatter of the data within the range (i.e. between the two extreme values). For this reason the concept of inter-quartile range is developed. It is the range, which includes the middle 50 per cent of the distribution Q3 – Q1. Here 1/4 (one quarter) of the lower end and 1/4 (one quarter) of the upper end of the observations are excluded.

For example, if the lower and upper quartiles of a frequency distribution were 6 and 11, the inter-quartile range would be 11 – 6 = 5. This shows that the range of values of the middle half of the population is five units.

3.2 The variance and the standard deviation

Definitions

The **variance**, $\sigma^2$, is the average of the squared mean deviation for each value in a population.

The **standard deviation**, $\sigma$, is the square root of the variance.

**Formula to learn**

Variance = $\sigma^2 = \frac{\sum (x - \bar{x})^2}{n}$

$\sigma$ is the lower case Greek letter sigma. The variance is therefore called 'sigma squared'.

Standard deviation = $\sigma = \sqrt{\frac{\sum (x - \bar{x})^2}{n}}$

The standard deviation ($\sigma$) is a measure of how spread out the values are.

The formula is easy: it is the square root of the variance.

3.2.1 Calculating the variance

The variance (which is the square of the standard deviation, $\sigma^2$) is the average of the squared differences from the mean.

It is easy to decipher the step-by-step calculation of variance from the definition above. Variance is the

- average
- squared
- deviation from the
- mean.

These are the four steps needed for calculating variance and you have to start from the end of the definition:

3.2.2 Step 1: Calculating the mean

The mean in general is the central value of a data set. You sum all the numbers up and then divide the sum by the count of numbers used.

<table>
<thead>
<tr>
<th>10</th>
<th>20</th>
<th>30</th>
<th>40</th>
<th>50</th>
<th>Total = 150</th>
</tr>
</thead>
</table>

The sum of the numbers 10, 20, 30, 40, and 50 (which is 150) is divided by the count of numbers (which is 5).

Arithmetic average of 10, 20, 30, 40, and 50 is 30.
3.2.3 Step 2: Calculating deviations from the mean

In the next step we need to calculate the deviations from the mean. For each number in the set, we simply subtract the mean from that number. For our set of numbers 10, 20, 30, 40, and 50 the deviations from the mean (which is 30) are:

\[
\begin{align*}
10 - 30 & \quad 20 - 30 & \quad 30 - 30 & \quad 40 - 30 & \quad 50 - 30 \\
-20 & \quad -10 & \quad 0 & \quad +10 & \quad +20
\end{align*}
\]

3.2.4 Step 3: Squaring the deviations

In step 3 we need to square each deviation. To square a number means to multiply that number by itself. For the numbers in our set, we get:

\[
\begin{align*}
10 - 30 & \quad 20 - 30 & \quad 30 - 30 & \quad 40 - 30 & \quad 50 - 30 \\
-20 \times -20 & \quad -10 \times -10 & \quad 0 \times 0 & \quad +10 \times +10 & \quad +20 \times +20 \\
400 & \quad 100 & \quad 0 & \quad 100 & \quad 400
\end{align*}
\]

Squaring numbers has two effects. Firstly, any negative number squared is a positive number. This way we get rid of the negative signs we had with deviations from the mean for numbers which were smaller than the mean. Secondly, squaring gives much bigger weight to big numbers (or big negative numbers) than to numbers close to zero.

3.2.5 Step 4: Calculating variance as average of squared deviations

We now calculate the arithmetic average of the squared deviations obtained in step 3. It is exactly the same thing as we did in step 1 – the only difference is that in step 1 we were calculating the average of the original numbers (10, 20, 30, 40, and 50), but now in step 4 we are calculating the average of the squared deviations.

\[
\begin{align*}
& x - \bar{x} \quad 10 - 30 \quad 20 - 30 \quad 30 - 30 \quad 40 - 30 \quad 50 - 30 \\
& (x - \bar{x})^2 \quad -20 \times -20 \quad -10 \times -10 \quad 0 \times 0 \quad +10 \times +10 \quad +20 \times +20 \\
& \Sigma (x - \bar{x})^2 \quad 400 \quad +100 \quad +0 \quad +100 \quad +400 \quad \text{Total 1 000}
\end{align*}
\]

We sum them up and get 1 000. Then we divide 1 000 by 5 and get 200. The variance of the set of numbers 10, 20, 30, 40, and 50 is 200. Variance is the average (step 4) squared (step 3) deviation (step 2) from the mean (step 1).

\[
\text{Average of the sum (} = \text{variance} = \sigma^2\text{)} = \frac{\Sigma (x - \bar{x})^2}{n}.
\]

3.2.6 Calculating standard deviation from variance

In finance and in most other disciplines, standard deviation is used more frequently than variance. Both are measures of dispersion or volatility in a data set and they are very closely related. Standard deviation is the square root of variance. Vice versa, variance is standard deviation squared.

To calculate standard deviation from variance, take the square root. In our example, the variance was 200, therefore standard deviation is 14.14.

\[
\text{Square root of the variance} = \text{standard deviation} = \sqrt{\frac{\Sigma (x - \bar{x})^2}{n}}.
\]

Worked Example: Standard deviation

Calculate the standard deviation for the following ten lengths:
Values: 12, 9, 3, 10, 12, 22, 7, 11, 15 and 19 cm
Solution

Mean = \frac{(12 + 9 + 3 + 10 + 12 + 22 + 7 + 11 + 15 + 19)}{10} = 120 \div 10 = 12\ cm

<table>
<thead>
<tr>
<th>Value</th>
<th>(x - \bar{x})</th>
<th>((x - \bar{x})^2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>12</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>9</td>
<td>-3</td>
<td>9</td>
</tr>
<tr>
<td>3</td>
<td>-9</td>
<td>81</td>
</tr>
<tr>
<td>10</td>
<td>-2</td>
<td>4</td>
</tr>
<tr>
<td>12</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>22</td>
<td>10</td>
<td>100</td>
</tr>
<tr>
<td>7</td>
<td>-5</td>
<td>25</td>
</tr>
<tr>
<td>11</td>
<td>-1</td>
<td>1</td>
</tr>
<tr>
<td>15</td>
<td>3</td>
<td>9</td>
</tr>
<tr>
<td>19</td>
<td>7</td>
<td>49</td>
</tr>
</tbody>
</table>

\[
\text{Variance} = \frac{278}{10} = 27.8\ \text{cm}\text{ and standard deviation} = \sqrt{27.8} = 5.27\ \text{cm}
\]

3.2.7 The variance and the standard deviation of several items together

You may need to calculate the variance and standard deviation for \(n\) items together, given the variance and standard deviation for one item alone.

**Worked Example: Several items together**

The daily demand for an item of inventory has a mean of 6 units, with a variance of 4 and a standard deviation of 2 units. Demand on any one day is unaffected by demand on previous days or subsequent days.

**Required**

Calculate the arithmetic mean, the variance and the standard deviation of demand for a five day week.

**Solution**

If we let

- Arithmetic mean = \(\bar{x} = 6\)
- Variance = \(\sigma^2 = 4\)
- Standard deviation = \(\sigma = 2\)
- Number of days in week = \(n = 5\)

The following rules apply to \(\bar{x}\), \(\sigma^2\) and \(\sigma\) when we have several items together.

- **Arithmetic mean** = \(n \bar{x} = 5 \times 6 = 30\) units per week.
- **Variance** = \(n\sigma^2 = 5 \times 4 = 20\) units per week.
- **Standard deviation** = \(\sqrt{n\sigma^2} = \sqrt{20} = 4.47\) units per week.

3.2.8 The main properties of the standard deviation

Sigma (\(\sigma\)) is the mathematical symbol for standard deviation. As we have shown, it provides a measure of the dispersion, or variation, within the (population) data set. The meaning of the standard deviation is most easily seen when the underlying population is a normal distribution, which graphically looks like a bell-curve. It is self-evident that in any normal distribution 50 per cent of the population will be more than average; 50 per cent will be lower. The standard deviation tells us more: it is known that 68.26 per cent of the population (from which the sample is derived) lies within ± 1\(\sigma\) of the mean; 95.44 per cent within ± 2\(\sigma\) and 99.73 per cent within ± 3\(\sigma\) from the mean.
The standard deviation's main properties are as follows:

(a) It is based on all the values in the distribution and so is more comprehensive than dispersion measures based on quantiles.

(b) It is suitable for further statistical analysis.

(c) It is more difficult to understand than some other measures of dispersion.

3.3 Population v sample

In general statistics performs two main tasks. Its goal is either to describe something that has already happened or already exists (descriptive statistics), or to estimate something that has not happened yet or is not fully known (inferential statistics).

Descriptive statistics deals with the problem how to effectively look at data we already have. Inferential statistics (the estimating and forecasting part of statistics) deals with the problem of not having all the data.

The primary task of inferential statistics is making an opinion about something by using only an incomplete sample of data.

In statistics it is very important to distinguish between population and sample. A population is defined as all members (e.g. occurrences, prices, annual returns) of a specified group. Population is the whole group.

A sample is a part of a population that is used to describe the characteristics (e.g. mean or standard deviation) of the whole population. The size of a sample can be less than 1 per cent, or 15 per cent, or 50 per cent of the population, but it is never the whole population.

3.3.1 Population v sample for variance and standard deviation

\[ \sigma^2 \text{ is the variance for a population and } \sigma = \sqrt{\sigma^2} \text{ is the population standard deviation} \]

\[ s^2 \text{ is the variance for a sample and } s = \sqrt{s^2} \text{ is the sample standard deviation} \]

When calculating variance and standard deviation, it is important to know whether we are calculating them for the whole population using all the data, or we are calculating them using only a sample of data. In the first case we call them population variance and population standard deviation. In the second case we call them sample variance and sample standard deviation.

When calculating the sample variance, the sum of the squared differences from the mean is divided by n-1 instead of n.

Worked Example: population variance and standard deviation

Population variance: I want to find the standard deviation of last year’s returns of the 12 equity funds have invested in. There is no estimating or forecasting in this task. My population is only these 12 funds have all the data available, as it is very easy to find these 12 funds’ performance data.

When I calculate population variance, I divide the sum of squared deviations from the mean by the number of items in the population (in this example I divide by 12).
Sample variance: I now want to find the standard deviation of last year’s returns of all equity funds in the world. To do this, I take a sample, consisting of the returns of 100 equity funds. When I calculate sample variance, I divide the sum of squared deviations from the mean by the number of items in the sample less one. In our example, I divide by 99 (100 minus 1).

Formula to learn

Sample standard deviation = \( s = \sqrt{\frac{\sum(x - \bar{x})^2}{n-1}} \)

4 Distribution shapes

The shape of a distribution is described by the following characteristics:

1 Symmetry. When it is graphed, a symmetric distribution can be divided at the centre so that each half is a mirror image of the other.

2 Number of peaks. Distributions can have few or many peaks. Distributions with one clear peak are called unimodal, and distributions with two clear peaks are called bimodal. When a symmetric distribution has a single peak at the centre, it is referred to as bell-shaped.

3 Skewness. When they are displayed graphically, some distributions may be asymmetric i.e. have many more observations on one side of the graph than the other. Distributions with most of their observations on the left (toward lower values) are said to be skewed left or negatively skewed; and distributions with most of their observations on the right (toward higher values) are said to be skewed right or positively skewed.

4 Peakedness or kurtosis – the relative peakedness or flatness of a distribution compared with the normal distribution.

5 Uniformity. When the observations in a set of data are equally spread across the range of the distribution, the distribution is called a uniform distribution. A uniform distribution has no clear peaks.

4.1 Normal distribution

We looked at the shape of the normal distribution when we discussed the standard deviation earlier in the chapter. A normal distribution is a continuous, symmetric, bell-shaped distribution of a variable.

The properties of a normal distribution are as follows:

(a) The mean, median, and mode are equal and are located at the centre of the distribution. 50 per cent of the scores are above the mean and 50 per cent are below the mean.

(b) The frequency curve is bell shaped.

(i) the bell shape has perfect bilateral symmetry – the left balances exactly with the right.

(ii) the score at -2 is balanced by a score at +2 and the frequencies from the centre to +2 and from the centre to -2 are equal.
(iii) the area under the curve from the centre to +2 is exactly the same as the area under the curve from the centre to -2.

c) The curve is continuous, that is, there are no gaps or holes. For each value of X, there is a corresponding value of Y.

d) The curve never touches the x axis. Theoretically, no matter how far in either direction the curve extends, it never meets the x axis — but it gets increasingly closer.

e) The total area under a normal distribution curve is equal to 1.00, or 100 per cent. This holds for any distribution function.

(f) The area under the part of a normal curve that lies within one standard deviation of the mean is approximately 0.68, or 68 per cent; within two standard deviations, about 0.95, or 95 per cent; and within three standard deviations, about 0.997, or 99.7 per cent.

4.2 Exponential distribution

**Formula to learn**

The probability density function (PDF) of an exponential distribution is:

$$f(x; \mu) = \frac{1}{\mu} e^{-\frac{x}{\mu}}$$

where: $$\mu > 0$$, and $$x > 0$$.

where:

- $$1/\mu = \text{constant event rate, in events per unit of measurement, e.g. events per hour, per cycle, etc.}$$
- $$\mu = \text{mean time between events, or to an event.}$$
- $$x = \text{operating time, life or age, in hours, cycles, miles, actuations, etc.}$$

An exponential distribution is frequently described as the waiting time distribution since it can be used to model the time until some event happens or something stops working. For example, the distances travelled by a bus between major mechanical breakdown or the lengths of time between earthquakes with a magnitude greater than seven in California.

It is often helpful also to consider the **cumulative density function** of the distribution. This is the probability that the event will occur in the time period between 0 and x.

For example, a help-line operator might take an average of 3 minutes to deal with each query. Thus $$\mu = 3$$.

The probability density function is then given by

$$f(x; 3) = \frac{1}{3} e^{-\frac{x}{3}}$$

and would be shown on a graph as follows.

![Graph of exponential distribution](image-url)
Let us say that we are interested in the probability that the operator takes less than 2 minutes to deal with the next call. We are interested in the shaded area in the graph below.

![Graph showing probability distribution](image)

It is the cumulative density function that allows us to make such calculations. It is the probability that the call will be dealt with in a time between 0 and 2 minutes.

**Formula to learn**

The cumulative density function (CDF) of an exponential distribution is:

\[ F(x; \mu) = 1 - \frac{x}{\mu} \]

The shaded area therefore represents a probability of

\[ F(2; 3) = 1 - e^{-\frac{2}{3}} \]

\[ = 1 - 0.51 \]

\[ = 0.49 \]

The probability that the operator takes less than 2 minutes in their next call is 0.49, or 49 per cent.

**Worked Example: Exponential distribution**

Suppose that the time to machine a kitchen cabinet door follows an exponential distribution with a mean of 4 minutes.

(a) What is the probability that a door will be completed in more than 4 minutes?

(b) What is the probability that a door will be completed in more than 4 minutes and less than 8 minutes?

**Solution**

(a) \[ P(< 4 \text{ minutes}) = 1 - e^{-\frac{4}{4}} \]

\[ = 1 - 0.37 \]

\[ = 0.63 \]

\[ \therefore P(> 4 \text{ minutes}) = 1 - 0.63 = 0.37 \]
Features of the PDF include:

- The probabilities associated with intervals of the random variable are determined by calculating the area under the density function curve over that interval.
- The exponential PDF only has one shape.
- The distribution starts at 1/μ and decreases thereafter exponentially and monotonically as x increases.
- The exponential PDF is always convex and is stretched to the right.
- μ = mean time between events, or to an event.
- The median is given by 0.693 μ
- The mode = 0
- The standard deviation = μ

For skewed distributions, the mean and median are not the same. The mean will be pulled in the direction of the skewness, which in the case of the exponential distribution is always to the right.

For skewed distributions, the median is something more useful than the mean.

4.3 The binomial distribution

The binomial distribution measures the probabilities of the number of successes over a given number of trials with a specific chance of success in each try. In the simplest scenario of a coin toss (with a fair coin), where the probability of getting a head with each toss is 0.50 and there are six trials, the binomial distribution will measure the likelihood of getting anywhere from no heads in six tosses (very unlikely) to 3 heads (more probable) to 6 heads (also very unlikely). The binomial distribution in this case will be symmetric, reflecting the even odds. If the probabilities shift from even odds, the distribution will become skewed.

4.3.1 Characteristics of the binomial distribution

Binomial probabilities can be calculated using the binomial formula.

\[ P(x) = \frac{n!}{x!(n-x)!} \pi^x (1-\pi)^{n-x} \]

Note: The ! symbol denotes ‘factorial’. The factorial of a number n = 1 × 2 × 3 × 4 × 5 ....× n

For example 4! = 1 × 2 × 3 × 4 = 24.

When using the binomial formula to solve problems, all that is necessary is that we are able to identify three things: the number of trials (n), the probability of a success on any one trial (π), and the number of successes desired (x). The formulas used to compute the mean and the standard deviation of a binomial distribution are as follows:

- mean = nπ
- standard deviation = \sqrt{n \pi(1-\pi)}

Two binomial distributions are shown below. Notice that for \( \pi = 0.5 \) (even odds), the distribution is symmetric whereas for \( \pi = 0.3 \) (an unfair coin), the distribution has a positive skewness.
Note that the binomial distribution is a discrete distribution. That is, it only makes sense for integer values of $x$. You can’t ask: what is the probability of observing 4.3 heads in ten coin tosses? Also note that the binomial distribution is determined by two parameters: $n$ and $\pi$.

We have chosen to represent the probability density function (PDF) with a bar plot, and not a line graph as this emphasises the discrete nature of the probability density function.

Different values of $n$ and $\pi$ lead to different distributions with different shapes.

For instance, with $n = 10$ and $\pi = 0.5$, the distribution is symmetric:
4.3.2 Graphing the binomial distribution

The graph of a binomial distribution can be constructed by using all the possible X values of a distribution and their associated probabilities. The X values are graphed along the X axis, and the probabilities are graphed along the Y axis. Note that the graph of the binomial distribution has three shapes: if \( \pi < 0.5 \), the graph is positively skewed, if \( \pi > 0.5 \), the graph is negatively skewed, and if \( \pi = 0.5 \), the graph is symmetrical. The skewness lessens as \( n \) gets large. In other words, if \( n \) remains constant but \( \pi \) becomes larger and larger up to 0.50, the shape of the binomial probability distribution becomes more symmetrical. If \( \pi \) remains the same but \( n \) becomes larger and larger, the shape of the binomial probability distribution becomes more symmetrical.

4.4 Skewed distributions

In everyday language, the terms ‘skewed’ and ‘askew’ are used to refer to something that is out of line or distorted on one side. When referring to the shape of frequency or probability distributions, ‘skewness’ refers to asymmetry of the distribution. A distribution with an asymmetric tail extending out to the right is referred to as ‘positively skewed’ or ‘skewed to the right’, because most of the variance from the mean is in the positive direction. Income in most societies has positive skewness, because the gap between the wealthy and the middle class has the room to be much greater than the gap between the poor and the middle class. If the opposite is the case, the distribution is negatively skewed, or skewed left. Last, if the distribution is symmetrical (like a normal distribution) it has zero skewness. Skewness can range from minus infinity to positive infinity.

As well as being able to calculate the average and spread of a frequency distribution, you should be aware of the skewness of a distribution.

We briefly looked at skewed distributions in Section 2.3.3 when discussing the relation between the mean, median and mode.
**Definition**

**Skewness** is the asymmetry of a frequency distribution curve.

A frequency distribution must be either symmetrical or skewed (asymmetrical).

It may happen that two distributions have the same standard deviation. For example, see the following diagram.

Although the two distributions have roughly the same means and standard deviations they are not identical. They differ in symmetry. The left-hand side distribution is a symmetrical one whereas the distribution on the right-hand side is asymmetrical or skewed.

Three examples of distributions, each with a different skewness, are shown below:

**Positive skewness**

A *positively skewed* distribution's graph will lean towards the left hand side, with a tail stretching out to the right. The mean, the median and the mode will have different values. The mode will have a lower value than the median and the mean will have a higher value than the median.

Mode < Median < Mean
No skewness

**Distribution with skewness = 0; normal distribution.**

The arithmetic mean, the median and the mode will all have the same value.

\[
\text{Mode} = \text{Median} = \text{Mean}
\]

![Histogram showing a normal distribution](image)

**Negative skewness**

A **negatively skewed** distribution's graph will lean towards the right-hand side, with a tail stretching out to the left. The mean, the median and the mode will have different values. The mode will have a higher value than the median and the mean will have a lower value than the median.

\[
\text{Mean} < \text{Median} < \text{Mode}
\]

![Histogram showing a negatively skewed distribution](image)

### 4.5 Kurtosis

**Definition**

Kurtosis characterises the relative peakedness or flatness of a distribution compared with the normal distribution. Positive kurtosis indicates a relatively peaked distribution. Negative kurtosis indicates a relatively flat distribution.

Kurtosis is a measure of a distribution's vertical stretch. It quantifies whether the shape of the data distribution matches the normal distribution. If the increase to the mode is sharper (more peaked) than that of a normal distribution, it has positive kurtosis. This would happen if the distribution is tightly packed for...
the most part, with a few extreme data points contributing to the bulk of the variance. If the opposite is the case (a flatter distribution), the distribution has negative kurtosis.

Two examples of distributions with differing values of kurtosis are shown below:

Distribution with positive kurtosis $> 0$

![Graph of a distribution with positive kurtosis]

Distribution with negative kurtosis $< 0$

![Graph of a distribution with negative kurtosis]

Three degrees of kurtosis are noted:

1. **Mesokurtic** distributions are, like the normal bell curve, neither peaked nor flat. Its kurtosis $= 0$
2. **Platykurtic** distributions are flatter than the normal bell curve, and so have negative kurtosis.
3. **Leptokurtic** distributions are more peaked than the normal bell curve and so have positive kurtosis.
Key chapter points

- Measures of central tendency tell us about the values that are in the middle of the data set.
- The mean, median, and mode are the most widely used measures of central tendency.
- The arithmetic mean is the best known type of average and is widely understood. It is used for further statistical analysis.
- The arithmetic mean of ungrouped data = sum of items ÷ number of items.
- The arithmetic mean of grouped data, \( \bar{x} = \frac{\sum fx}{n} \) or \( \bar{x} = \frac{\sum fx}{\sum f} \) where n is the number of values recorded, or the number of items measured.
- The mode or modal value is a value which means 'the most frequently occurring value'.
- The mode of a grouped frequency distribution can be calculated from a histogram.
- The median is the value of the middle member of an array. The middle item of an odd number of items is calculated as the \( \frac{(n+1)^{th}}{2} \) item.
- The median of a grouped frequency distribution can be established from a cumulative frequency table.
- Measures of variability tell us how spread out, or dispersed, the values are in a data set.
- The range is the difference between the highest and lowest observations.
- The quartiles and the median divide the population into four groups of equal size.
- The inter-quartile range is the difference between the upper and lower quartiles (Q3 – Q1) and hence shows the range of values of the middle half of the population.
- The variance, \( \sigma^2 \), is the average of the squared mean deviation for each value in a distribution.
- The standard deviation, which is the square root of the variance, is the most important measure of spread used in statistics. Make sure you understand how to calculate the standard deviation of a set of data.
- \( \sigma^2 \) is the variance for a population and \( \sigma = \sqrt{\sigma^2} \) is the population standard deviation.
- \( s^2 \) is the variance for a sample and \( s = \sqrt{s^2} \) is the sample standard deviation.
- A normal distribution is a continuous, symmetric, bell-shaped distribution.
- The exponential distribution can be used to model the time until an event occurs, or the time between two events. Examples of quantities that can be modelled using the exponential distribution are the time between phone calls in a call centre or the time until a machine fails.
- The binomial distribution measures the probabilities of the number of successes over a given number of trials with a specific chance of success in each trial.
- Skewness is the asymmetry of a frequency distribution curve.
- Kurtosis is a measure of a distribution’s vertical stretch. It quantifies whether the shape of the data distribution matches the normal distribution.
Quick revision questions

1. Insert the formulae in the box below into the correct position.
   (a) The arithmetic mean of ungrouped data =
   (b) The arithmetic mean of grouped data = or
   \[ \frac{\sum x}{n} \]
   \[ \frac{\sum fx}{n} \]
   \[ \frac{\sum fx}{\sum f} \]

2. What is the name given to the average which means ‘the most frequently occurring value’?
   A. Arithmetic mean
   B. Median
   C. Mode

3. The mean weight of a group of components has been calculated as 133.5. The individual weights of the components were 143, 96, x, 153.5, 92.5, y, 47.
   If y = 4x, what is the value of x?

4. Calculate the mid-points for both discrete and continuous variables in the table below.

<table>
<thead>
<tr>
<th>Class interval</th>
<th>Mid-point (Discrete data)</th>
<th>Mid-point (Continuous data)</th>
</tr>
</thead>
<tbody>
<tr>
<td>25 &lt; 30</td>
<td></td>
<td></td>
</tr>
<tr>
<td>30 &lt; 35</td>
<td></td>
<td></td>
</tr>
<tr>
<td>35 &lt; 40</td>
<td></td>
<td></td>
</tr>
<tr>
<td>40 &lt; 45</td>
<td></td>
<td></td>
</tr>
<tr>
<td>45 &lt; 50</td>
<td></td>
<td></td>
</tr>
<tr>
<td>50 &lt; 55</td>
<td></td>
<td></td>
</tr>
<tr>
<td>55 &lt; 60</td>
<td></td>
<td></td>
</tr>
<tr>
<td>60 &lt; 65</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

5. A group of children have the following ages in years: 10, 8, 6, 9, 13, 12, 7, 11.
   What is the median age?

6. Measures of variability give information about the amount of skew or bias in a set of data.
   A. true
   B. false

7. Fill in the blanks in the statements below using the words in the box.
   (a) .......... quartile = Q₁ = value .......... which 25% of the population fall.
   (b) .......... quartile = Q₃ = value .......... which 25% of the population fall.
   | Upper | Above | Below | Lower |
8 In a negatively skewed distribution
   A the mean is the same as the median.
   B the mean is smaller than the median.
   C the mean lies between the median and the mode.
   D the mean is larger than the median.

9 The standard deviation of a sample of data is 36. What is the value of the variance?
Answers to quick revision questions

1 (a) \( \frac{\sum x}{n} \)

(b) \( \frac{\sum fx}{n} \) or \( \frac{\sum fx}{\sum f} \)

2 Mode

3 \( 80.5 \)

\[ \text{Mean} = \frac{\text{Total}}{7} \]

So Total = 7 \times 133.5 = 934.5

934.5 = 143 + 96 + x + 153.5 + 92.5 + y + 47

934.5 = 532 + x + y

y = 4x

So 934.5 = 532 + x + 4x

5x = 934.5 - 532

5x = 402.5

x = 80.5

4

<table>
<thead>
<tr>
<th>Class interval</th>
<th>Mid-point (Discrete data)</th>
<th>Mid-point (Continuous data)</th>
</tr>
</thead>
<tbody>
<tr>
<td>25 &lt; 30</td>
<td>27</td>
<td>27.5</td>
</tr>
<tr>
<td>30 &lt; 35</td>
<td>32</td>
<td>32.5</td>
</tr>
<tr>
<td>35 &lt; 40</td>
<td>37</td>
<td>37.5</td>
</tr>
<tr>
<td>40 &lt; 45</td>
<td>42</td>
<td>42.5</td>
</tr>
<tr>
<td>45 &lt; 50</td>
<td>47</td>
<td>47.5</td>
</tr>
<tr>
<td>50 &lt; 55</td>
<td>52</td>
<td>52.5</td>
</tr>
<tr>
<td>55 &lt; 60</td>
<td>57</td>
<td>57.5</td>
</tr>
<tr>
<td>60 &lt; 65</td>
<td>62</td>
<td>62.5</td>
</tr>
</tbody>
</table>

5 6, 7, 8, 9, 10, 11, 12, 13

Median is 9½

6 B false

Measures of variability tell us how spread out, or dispersed, scores are in a data set.

7 (a) Lower quartile = \( Q_1 \) = value below which 25 per cent of the population fall

(b) Upper quartile = \( Q_3 \) = value above which 25 per cent of the population fall

8 B The mean is smaller than the median.

9 The variance is the square of the standard deviation.

\[ 36^2 = 1\,296 \]
1 Mean weight
The value of X is 10

Workings

\[ \text{Mean} = \frac{\text{Sum of values of items}}{\text{Number of items}} \]

Sum of first 10 units = 5 \times 10 = 50 kgs
Sum of second 10 units = 7 \times 10 = 70 kgs
Sum of next 20 units = 20 \times X = 20X
Sum of all 40 units = 50 + 70 + 20X = 120 + 20X

\[ \therefore \text{Arithmetic mean} = 8 = \frac{120 + 20X}{40} \]

\[ 8 \times 40 = 120 + 20X \]
\[ 320 = 120 + 20X \text{ (subtract 120 from both sides)} \]
\[ 320 - 120 = 20X \]
\[ 200 = 20X \]
\[ 10 = X \text{ (divide both sides by 20)} \]

2 Arithmetic mean

The mid point of the range 'under $60' is assumed to be $55 and that of the range over $110 to be $115, since all other class intervals are $10. This is obviously an approximation which might result in a loss of accuracy, but there is no better alternative assumption to use. Because overtime can vary in steps of 1c, it is virtually a continuous variable and hence the mid-points of the classes are halfway between their end points.

<table>
<thead>
<tr>
<th>Mid-point of class</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>$55</td>
<td>3</td>
</tr>
<tr>
<td>$65</td>
<td>11</td>
</tr>
<tr>
<td>$75</td>
<td>16</td>
</tr>
<tr>
<td>$85</td>
<td>15</td>
</tr>
<tr>
<td>$95</td>
<td>10</td>
</tr>
<tr>
<td>$105</td>
<td>8</td>
</tr>
<tr>
<td>$115</td>
<td>6</td>
</tr>
</tbody>
</table>

\[ \text{Arithmetic mean} = \frac{\sum fx}{\sum f} = \frac{5835}{69} = $84.57 \]

3 The median

The times can be arranged as follows.

11 15 17 17 17 18 20 20 21
21 23 24 24 24 27 29 30 37

The median = 21 minutes. (We could have used the average of the 9th and 10th items if we had wanted to.)
4 Mean and the range

(a) \( \bar{x}_1 = \frac{72}{8} = 9 \)

The figures have a mean of 9 and a range of 24 – 3 = 21.

(b) \( \bar{x}_2 = \frac{72}{8} = 9 \)

The figures have a mean of 9 and a range of 11 – 7 = 4.

The set of data \( x_1 \) is more widely dispersed than the set of data \( x_2 \).

5 Quartiles

<table>
<thead>
<tr>
<th>Data</th>
<th>6 47 49 15 43 41 7 39 43 41 36 (11 numbers)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ordered data</td>
<td>6 7 15 36 39 41 41 43 43 47 49</td>
</tr>
<tr>
<td>Median = (11 + 1)/2 = 6\text{th}</td>
<td>= 41</td>
</tr>
<tr>
<td>Lower quartile (11 + 1)/4 = 3\text{rd}</td>
<td>= 15</td>
</tr>
<tr>
<td>Upper quartile 3 (11 + 1)/4 = 9\text{th}</td>
<td>= 43</td>
</tr>
</tbody>
</table>
Chapter 14

Frequency distributions and probability

<table>
<thead>
<tr>
<th>Learning objectives</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Frequency distributions and probability</td>
<td>LO14</td>
</tr>
<tr>
<td>Develop a frequency distribution from a given set of data</td>
<td>LO14.1</td>
</tr>
<tr>
<td>Distinguish between class range, class midpoint, relative frequency and cumulative frequency</td>
<td>LO14.2</td>
</tr>
<tr>
<td>Define the concept of probability</td>
<td>LO14.3</td>
</tr>
<tr>
<td>Explain the different ways of assigning probability</td>
<td>LO14.4</td>
</tr>
<tr>
<td>Explain and apply marginal, union, joint, and conditional probabilities</td>
<td>LO14.5</td>
</tr>
<tr>
<td>Explain the use of probability matrices to solve probability problems</td>
<td>LO14.6</td>
</tr>
</tbody>
</table>

Topic list

1. Quantifying probability
2. The concept of probability
3. Different ways of assigning probability
4. Probability rules and laws
5. Probability matrices
Probability is of fundamental importance in the theory of statistics. Key principles of probability are most easily explained by using examples of coin tossing, dice throwing and games of chance. One great advantage they have is that the chance of, for example, heads and tails when tossing a coin can be assumed to be known, whereas in most practical situations the probabilities can only be estimated. For example, a man’s expectation of life can be calculated only from current mortality rates, which may not apply to future generations.
Before you begin

If you have studied these topics before, you may wonder whether you need to study this chapter in full. If this is the case, please attempt the questions below, which cover some of the key subjects in the area.

If you answer all these questions successfully, you probably have a reasonably detailed knowledge of the subject matter, but you should still skim through the chapter to ensure that you are familiar with everything covered.

There are references in brackets indicating where in the chapter you can find the information, and you will also find a commentary at the back of the study manual.

1. What is the difference between a frequency histogram and a relative frequency histogram? (Section 1.3.2)
2. Give an example of a symmetrical outcome. (Section 2.2.1)
3. Explain the Law of Large Numbers. (Section 3.1)
4. A white counter was taken from a bag of different coloured counters, and then replaced, and the experiment repeated many times. The relative frequency of getting a white counter was found to be 0.3. If the bag contained 20 counters, estimate the number of white counters. (Section 4.2)
5. A survey showed that 60 per cent of all adults in your city take public transportation to work. If three people are chosen at random, what is the probability that they will all take public transportation to work? (Section 4.2)
6. If you throw a die four times, what is the probability that one or more of your throws will come up with a four? (Section 4.2)
7. What is the probability that you draw two cards from a deck and both of them are spades? (Section 4.2)
8. Multiple union or disjunction: $A \cup B \cup C \cup D$ means $A$ or $B$ or $C$ or $D$ (Section 4.2.1)
   A. true
   B. false
9. You have a bag of marbles. There are 3 red marbles, 2 green marbles, 7 yellow marbles and 3 blue marbles. What is the probability of drawing a yellow or red marble? (Section 4.2)
10. Explain the term 'replaced' when we are talking about events being independent. (Section 4.3)
11. In a survey at a California ski resort, 100 skiers and snowboarders of different age groups were surveyed. The information is shown in the table below. Assume one person is randomly selected. Find the probability that the person was a skier OR was aged 11 - 20 from the following contingency table.

<table>
<thead>
<tr>
<th>Age group</th>
<th>0 - 10</th>
<th>11 - 20</th>
<th>21 - 40</th>
<th>40+</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ski</td>
<td>10</td>
<td>12</td>
<td>30</td>
<td>8</td>
<td>60</td>
</tr>
<tr>
<td>Snowboard</td>
<td>6</td>
<td>17</td>
<td>12</td>
<td>5</td>
<td>40</td>
</tr>
<tr>
<td>Total</td>
<td>16</td>
<td>29</td>
<td>42</td>
<td>13</td>
<td>100</td>
</tr>
</tbody>
</table>

A. 77/100
B. 60/100
C. 12/100
D. 89/100
1 Quantifying probability

1.1 Introducing probability

**Definition**

Probability is a measure of likelihood and can be stated as a percentage, a ratio, or more usually as a number from 0 to 1.

Consider the following.

- Probability = 0 = impossibility
- Probability = 1 = certainty
- Probability = 1/2 = a 50% chance of something happening
- Probability = 1/4 = a 1 in 4 chance of something happening

1.2 Expressing probabilities

In statistics, **probabilities** are more commonly expressed as **proportions** than as **percentages**. Consider the following possible outcomes:

<table>
<thead>
<tr>
<th>Possible outcome</th>
<th>Probability as a percentage</th>
<th>Probability as a proportion</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>15.0</td>
<td>0.150</td>
</tr>
<tr>
<td>B</td>
<td>20.0</td>
<td>0.200</td>
</tr>
<tr>
<td>C</td>
<td>32.5</td>
<td>0.325</td>
</tr>
<tr>
<td>D</td>
<td>7.5</td>
<td>0.075</td>
</tr>
<tr>
<td>E</td>
<td>12.5</td>
<td>0.125</td>
</tr>
<tr>
<td>F</td>
<td>12.5</td>
<td>0.125</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>100.0</strong></td>
<td><strong>1.000</strong></td>
</tr>
</tbody>
</table>

It is useful to consider how probability can be quantified. A businessman might estimate that if the selling price of a product is raised by 20 cents, there would be a 90 per cent probability that demand would fall by 30 per cent, but how would he have reached his estimate of 90 per cent probability?

1.3 Developing a frequency distribution

One of the more commonly used pictorials in statistics is the frequency histogram, which in some ways is similar to a bar chart and tells how many items are in each numerical category. Frequency distributions were introduced in Chapter 12. For example, suppose that after a day’s trading on eBay, you want to determine which items were the most popular: the high-priced items, the low-priced items, and so forth. Let’s say you sold a total of 32 items for the following prices: $1, $2, $2, $2, $4, $4, $4, $4, $7, $8, $9, $9, $9, $9, $9, $11, $15, $15, $15, $19, $20, $21, $21, $25, $25, $29, $29, $29, $30, $30, $30, $35, $35.

As you can see, the items sold ranged in price from $1 to $35.

1.3.1 Class intervals

First, divide this range of $1 to $35 into a number of categories, called **class intervals**. Typically, no fewer than 5 or more than 20 class intervals work best for a frequency histogram.

Choose the first class interval to include your lowest (smallest value) data and make sure that no overlap exists so that one piece of data does not fall into two class intervals. For example, you would not have your first class interval be $1–$5 and your second class interval $5–$10 because any items that sold for $5 would belong in both the first and the second intervals. Instead, use $1–$5 for the first interval and $6–$10 for the second. Class intervals are mutually exclusive.
Distribution of items sold on eBay

<table>
<thead>
<tr>
<th>Class</th>
<th>Interval</th>
<th>Midpoint</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>$1-$5</td>
<td>$3.00</td>
<td>8</td>
</tr>
<tr>
<td>2</td>
<td>$6-$10</td>
<td>$8.00</td>
<td>6</td>
</tr>
<tr>
<td>3</td>
<td>$11-$15</td>
<td>$13.00</td>
<td>4</td>
</tr>
<tr>
<td>4</td>
<td>$16-$20</td>
<td>$18.00</td>
<td>2</td>
</tr>
<tr>
<td>5</td>
<td>$21-$25</td>
<td>$23.00</td>
<td>4</td>
</tr>
<tr>
<td>6</td>
<td>$26-$30</td>
<td>$28.00</td>
<td>6</td>
</tr>
<tr>
<td>7</td>
<td>$31-$35</td>
<td>$33.00</td>
<td>2</td>
</tr>
</tbody>
</table>

Note that each class interval has the same width. That is, $1–$5 has a width of five dollars, inclusive; $6–$10 has a width of five dollars, inclusive; $11–$15 has a width of five dollars, inclusive, and so forth.

After calculating the interval range, take the midpoint for each interval. This makes it easier to estimate the value of sales in a particular range.

The value of items sold for $10 or less is approximately $(3.00 \times 8) + (8 \times 6) = 72$.

1.3.2 Frequency histogram

From the data, a frequency histogram would look like that shown below:

Frequency histogram of items sold on eBay

Unlike in a bar chart, the class intervals are drawn immediately adjacent to each other.

The next step is to construct a relative frequency histogram using the same table.

A relative frequency histogram compares each class interval to the total number of items. For example, the first interval ($1–$5) contains 8 out of the total of 32 items, so the relative frequency of the first class interval is $8/32 = 0.25$. 
### Distribution of items sold on eBay including relative frequencies

<table>
<thead>
<tr>
<th>Class</th>
<th>Interval</th>
<th>Midpoint</th>
<th>Frequency</th>
<th>Relative frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>$1-$5</td>
<td>$3.00</td>
<td>8</td>
<td>0.25</td>
</tr>
<tr>
<td>2</td>
<td>$6-$10</td>
<td>$8.00</td>
<td>6</td>
<td>0.1875</td>
</tr>
<tr>
<td>3</td>
<td>$11-$15</td>
<td>$13.00</td>
<td>4</td>
<td>0.125</td>
</tr>
<tr>
<td>4</td>
<td>$16-$20</td>
<td>$18.00</td>
<td>2</td>
<td>0.0625</td>
</tr>
<tr>
<td>5</td>
<td>$21-$25</td>
<td>$23.00</td>
<td>4</td>
<td>0.125</td>
</tr>
<tr>
<td>6</td>
<td>$26-$30</td>
<td>$28.00</td>
<td>6</td>
<td>0.1875</td>
</tr>
<tr>
<td>7</td>
<td>$31-$35</td>
<td>$33.00</td>
<td>2</td>
<td>0.0625</td>
</tr>
</tbody>
</table>

The only difference between a frequency histogram and a relative frequency histogram is that the vertical axis uses relative or proportional frequency instead of simple frequency.

---

#### 1.4 Frequency distributions – a summary

Students often find frequency distributions tricky. The following summary might help to clarify the different types of frequency distribution we have covered in this section.

(a) **Frequency distribution.** Individual data items are arranged in a table showing the frequency each individual data item occurs.

(b) **Grouped frequency distribution – discrete variables.** Data items which are discrete variables, (eg the number of marks obtained in an examination) are divided into classes of say 10 marks. The numbers of students (frequencies) scoring marks within each band are then grouped into a single frequency.

(c) **Grouped frequency distribution – continuous variables.** These are similar to the grouped frequency distributions for discrete variables (above.) However, as they are concerned with continuous variables note the following points.

   (i) There is an open-ended class at the end of the range.

   (ii) Class intervals must be carefully considered so that they capture all of the data once (and only once!).

(d) **Cumulative frequency distribution.** These distributions are used to show the number of times that a value above or below a certain amount occurs. Cumulative frequencies are obtained by adding the individual frequencies together.

---

Economics and Markets
2 The concept of probability

2.1 So 'what is probability'?

The concept of probability could be stated simply as

**Definition**

*Probability* is 'a measure of the degree of belief that an event will occur'.

An alternative definition is:

**Definition**

*Probability* is a way of summarising the uncertainty of statements or events. It gives a numerical measure for the degree of certainty (or degree of uncertainty) of the occurrence of an event.

2.2 Interpretation of the probability concept

Inferential statistics (the branch of statistics concerned with drawing conclusions about a population from a sample) is built on the foundation of probability theory, and has been remarkably successful in guiding opinion about the conclusions to be drawn from data. Yet (paradoxically) the very idea of probability has been plagued by controversy from the beginning of the subject to the present day. In this section we provide a glimpse of the debate about the interpretation of the probability concept.

2.2.1 Classical approach – symmetrical outcomes

One conception of probability is drawn from the idea of symmetrical outcomes. For example, the two possible outcomes of tossing a fair coin seem not to be distinguishable in any way that affects which side will land up or down. Therefore, the probability of heads is taken to be ½, as is the probability of tails. In general, if there are n symmetrical outcomes, the probability of any given one of them occurring is taken to be 1/n. Thus, if a six-sided die is rolled, the probability of any one of the six sides coming up is 1/6.

Classical probability suffers from a serious limitation. The definition of probability implicitly defines all outcomes to be equally probable. While this might be useful for drawing cards, rolling dice, or pulling balls from urns, it offers no method for dealing with outcomes with unequal probabilities. This limitation can even lead to mistaken statements about probabilities. An often given example goes like this:

- I could be hit by a meteorite tomorrow. There are two possible outcomes: I will be hit, or I will not be hit. Therefore, the probability I will be hit by a meteorite tomorrow is ½ = 50% probability.

Of course, the problem here is not with the classical theory, merely the attempted application of the theory to a situation to which it is not well adapted.

2.2.2 Relative frequencies approach

Probabilities can also be thought of in terms of relative frequencies. If we tossed a coin millions of times, we would expect the proportion of tosses that came up heads to be pretty close to ½. As the number of tosses increases, the proportion of heads approaches ½. Therefore, we can say that the probability of a head is ½. However, sometimes a situation may be too complex to understand the physical nature of it well enough to calculate probabilities and, by running a large number of trials and observing the outcomes, we can estimate the probability. This is the empirical probability based on long-run relative frequencies and is defined as the ratio of the number of observed outcomes favourable to the event divided by the number of observed outcomes. The larger the number of trials, the more accurate the estimate of probability. If the system can be modelled by computer, then simulations can be performed in place of physical trials.
The following example illustrates the frequencies attitude about probabilities. Suppose you wish to know what the weather will be like next Saturday because you are planning a picnic. You turn on your radio, and the weather person says, 'There is a 10 per cent chance of rain.' You decide to have the picnic outdoors and, lo and behold, it rains. You are furious with the weather person. But was she wrong? No, she did not say it would not rain, only that rain was unlikely. She would have been flatly wrong only if she said that the probability is 0 and it subsequently rained. However, if you kept track of her weather predictions over a long period of time and found that it rained on 50 per cent of the days that the weather person said the probability was 0.10, you could say her probability assessments are wrong.

So when is it accurate to say that the probability of rain is 0.10? According to our frequency interpretation, it means that it will rain 10 per cent of the days on which rain is forecast with this probability.

### 2.2.3 Subjective approach

For some purposes, probability is best thought of as subjective. A manager frequently faces situations in which neither classical nor empirical probabilities are useful. For example, in a one-off situation, such as the launch of a new, unique product, the probability of success can neither be calculated nor estimated from repeated trials. However, the manager may make an educated guess of the probability. This subjective probability can be thought of as a person’s degree of confidence that the event will occur. In the absence of better information upon which to rely, subjective probability may be used to make logically consistent decisions, but the quality of those decisions depends on the accuracy of the subjective estimate.

### 3 Different ways of assigning probability

As we noted in the last section, there are three distinctly different ways of assigning probability:

1. The first one is the classical probability approach. Here the probability of success is based on prior knowledge of the process involved.
2. The second approach to probability is the relative frequency or empirical method approach where the outcome is based on observed data, rather than upon prior knowledge of a process. The assumption of this approach is that the random process can be replicated many times under identical conditions.
3. The third approach to probability is the subjective approach.

#### 3.1 Classical approach

Classical probability is founded on the assumption that the outcomes of the most elementary experiments are equally likely to happen and those outcomes are mutually exclusive and collectively exhaustive. It involves experimenting and using rules and laws, which we will look at in detail in the next section.
3.1.1 Mutually exclusive and collectively exhaustive events

Those events that cannot happen together are called mutually exclusive events. For example, in the toss of a single coin, the events of heads and tails are mutually exclusive. The probability of two mutually exclusive events occurring at the same time is zero.

A list of collectively exhaustive events contains all possible elementary events for an experiment. For example, for a die-tossing experiment, the set of events consists of 1, 2, 3, 4, 5, and 6. The set is collectively exhaustive because it includes all possible outcomes.

The following equation is used to assign classical probability:

\[ P(X) = \frac{\text{Number of favourable outcomes}}{\text{Total number of possible outcomes}} \]

Here, ‘favourable’ refers not to any subjective value given to the outcomes, but is rather the classical terminology used to indicate that an outcome belongs to a given event of interest.

The classical theory assumes that all possible outcomes are equally likely, and that we know the total number of possible outcomes \( n \). For example, we know the number of cards in a deck of regular cards \( n = 52 \), for the first draw, and we assume that the deck is fair, i.e. that every card has an equal chance of being drawn. What is the probability that an ace is drawn, if there are four aces in the deck?

\[ P(A) = \frac{4}{52}, \text{ or } \frac{1}{13}, \text{ or } .0769 \text{ or } 7.69 \text{ per cent} \]

If a card is drawn and is not an ace and it is not returned to the pack, the probability that the next draw yields an ace goes up, since \( n \) has gone down.

\[ P(A) = \frac{4}{51}, \text{ or } .0784 \text{ or } 7.84 \text{ per cent} \]

As \( n \) gets smaller and no aces are drawn, the probability of an ace on the next draw increases.

If all the possible outcomes are known and equally likely, then classical probability yields a definite number every time. There is no generalisation; its answers are quite specific. Classical probability is not concerned with sample size or representativeness.

The results of classical probability obey the Law of Large Numbers. That is, the results will be closer and closer to the percentages predicted the more often we carry out the draw. If we draw the first card from a fair deck one million times, we’ll find that we will get an ace in very close to 7.7 per cent of the cases. This is why casinos consistently make a profit on the games they offer; they know the odds.

Rolling dice or tossing a coin are activities associated with a classical approach to probability. In these cases, you can list all the possible outcomes of an experiment and determine the actual probabilities of each simple event.

3.1.2 Terminology

Before we go too far, we should define some terms.

Definitions

Experiment – is an activity that is either observed or measured, such as tossing a coin, or drawing a card. It is any process that can be repeated in which the results are uncertain.

Outcome – is a list of possibilities.

Sample space, \( S \) – the set of all possible outcomes in an experiment. The sample space for the roll of a single die is 1, 2, 3, 4, 5, and 6. The sample space of the experiment of tossing a coin three times is:

<table>
<thead>
<tr>
<th></th>
<th>T</th>
<th>T</th>
<th>T</th>
<th>T</th>
<th>H</th>
<th>H</th>
<th>H</th>
</tr>
</thead>
<tbody>
<tr>
<td>First toss</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Second toss</td>
<td>T</td>
<td>T</td>
<td>H</td>
<td>H</td>
<td>T</td>
<td>T</td>
<td>H</td>
</tr>
<tr>
<td>Third toss</td>
<td>T</td>
<td>H</td>
<td>T</td>
<td>H</td>
<td>T</td>
<td>H</td>
<td>T</td>
</tr>
</tbody>
</table>
Sample space can aid in finding probabilities. However, using the sample space to express probabilities is hard when the sample space is large. Hence, we usually use other approaches to determine probability.

**Event** – is any collection of outcomes from a probability experiment. An event may consist of one or more simple events. Events are denoted using capital letters such as $E$.

The **probability of an event**, denoted $P(E)$, is the likelihood of that event occurring.

### 3.2 Relative frequency approach

This approach is sometimes called the **Empirical method**.

Relative probability is based on cumulated historical data. The following equation is used to assign this type of probability:

**Formula to learn**

$$P(X) = \frac{\text{Number of times an event occurred in the past}}{\text{Total number of opportunities for the event to occur}}$$

Note that relative probability is not based on rules or laws but on what has happened in the past. For example, your company wants to decide on the probability that its inspectors are going to reject the next batch of raw materials from a supplier. Data collected from your company record books show that the supplier had sent your company 80 batches in the past, and inspectors had rejected 15 of them. By the method of relative probability, the probability of the inspectors rejecting the next batch is $15/80$, or $0.19$. If the next batch is rejected, the relative probability for the subsequent shipment would change to $16/81 = 0.20$.

We need to use relative frequency probability when the number of possible outcomes ($n$) is so large we can’t observe all of them. For example, what is the probability that a 60-year-old woman will die of a heart attack within ten years? We can’t observe all 60-year-old women.

But we can predict the probability for random typical women who resemble one another, if we observe enough of them. Here is where we need to watch for the fallacy of hasty generalisation, because size and representativeness of the sample really matter. We need to make sure we observe enough instances ($n$ should be large enough), and the ones we observe need to be typical.

$$P(a) = \frac{\text{observed } f}{\text{observed } n}$$

The probability of 60-year-old women dying of a heart attack within ten years (event $a$) equals the number of 60-year-olds who do die of a heart attack within ten years ($f$), divided by the number of 60-year-old women in the sample ($n$). So if we observed 1 000 60 year old women over ten years, and observed that 35 of them died of heart attacks, we could say the probability of a random 60 year old woman dying of a heart attack is around 3.5 per cent ($35/1 000$).

$$P(a) = 35 / 1 000$$

There’s always the chance that our numbers are unreliable because our sample is unrepresentative or too small.

Naturally, we can’t predict with certainty for any particular woman. If we could, we wouldn’t need probability theory.

Relative frequencies are always between 0 per cent (the event never happens or is extremely rare) and 100 per cent (the event always happens, or very nearly so).
**Worked Example: Relative frequencies to approximate probabilities**

The following data represent the number of homes with various types of home heating fuels based on a survey of 1,000 homes.

(a) Approximate the probability that a randomly selected home uses electricity as its home heating fuel.
(b) Would it be unusual to select a home that uses coal or coke as its home heating fuel?

<table>
<thead>
<tr>
<th>House heating fuel</th>
<th>Frequency</th>
<th>Probability of home using this fuel</th>
</tr>
</thead>
<tbody>
<tr>
<td>Utility gas</td>
<td>504</td>
<td>0.504</td>
</tr>
<tr>
<td>Bottled, tank or LP gas</td>
<td>64</td>
<td>0.064</td>
</tr>
<tr>
<td>Electricity</td>
<td>307</td>
<td>0.307</td>
</tr>
<tr>
<td>Fuel oil, kerosene, etc.</td>
<td>94</td>
<td>0.094</td>
</tr>
<tr>
<td>Coal or coke</td>
<td>2</td>
<td>0.002</td>
</tr>
<tr>
<td>Wood</td>
<td>17</td>
<td>0.017</td>
</tr>
<tr>
<td>Solar energy</td>
<td>1</td>
<td>0.001</td>
</tr>
<tr>
<td>Other fuel</td>
<td>4</td>
<td>0.004</td>
</tr>
<tr>
<td>No fuel used</td>
<td>7</td>
<td>0.007</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>1,000</strong></td>
<td><strong>1.000</strong></td>
</tr>
</tbody>
</table>

**Solution**

(a) 0.307

(b) probability of 0.002 – yes, it would be unusual.

### 3.3 Subjective approach

Subjective probabilities are probabilities obtained based upon an educated guess, personal judgment, accumulation of knowledge, and experience. For example, medical doctors sometimes assign subjective probabilities to the length of life expectancy for people having cancer.

With the subjective approach, probability measures the speaker’s ‘degree of belief’ that the event will occur, on a scale of 0 per cent (complete disbelief that the event will happen) to 100 per cent (certainty that the event will happen). According to the subjective theory, what it means for me to say that ‘the probability that A occurs is 2/3’ is that I believe that A will happen twice as strongly as I believe that A will not happen. The subjective theory is particularly useful in assigning meaning to the probability of events that in principle can occur only once. For example, how might one assign meaning to a statement like ‘there is a 25 per cent chance of an earthquake on the San Andreas fault with magnitude 8 or larger before 2050?’ It is very hard to use either the classical approach with its equally likely outcomes or the frequency approach to make sense of the assertion.

### 4 Probability rules and laws

**Section overview**

- There are many ways which can be used in solving probability problems. These methods include the laws of probability, sample space, insight, contingency tables and probability matrices. Because of the individuality and variety of probability problems, some approaches apply more readily in certain cases than in others. There is no best method for solving all probability problems.
- Three laws of probability are discussed in this section: the addition law, the multiplication law, and the conditional law.
4.1 Venn diagrams

A Venn diagram is a pictorial method of showing probability. We can show all the possible outcomes and the outcome we are interested in (A).

The sample space (S) = the set of all possible outcomes. For example, the sample space for a six-sided die consists of the integer values 1, 2, 3, 4, 5, and 6.

\[ P(S) = 1. \]

![Figure 2 A Venn diagram](image)

4.1.1 Complementary outcomes

The complement of an event such as A, written as \( \bar{A} \), contains all events not included in A. For example, if in rolling a die, event A is getting an odd number, the complement of A is getting an even number. Thus, the complement of event A contains whatever portion of the sample space that event A does not contain. A and \( \bar{A} \) together cover every possible eventuality.

![Figure 3 Venn diagram showing the complement of A](image)

**Worked Example: Complementary outcome**

You throw two dice. What is the probability that the sum of the two dice will be six? To solve this problem, list all the possible outcomes.

**Solution**

There are 36 of them since each die can come up one of six ways. The 36 possibilities are shown below.

<table>
<thead>
<tr>
<th>Die 1</th>
<th>Die 2</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>1</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>1</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>1</td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>1</td>
<td>6</td>
<td>7</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>2</td>
<td>3</td>
<td>5</td>
</tr>
<tr>
<td>2</td>
<td>4</td>
<td>6</td>
</tr>
<tr>
<td>2</td>
<td>5</td>
<td>7</td>
</tr>
<tr>
<td>2</td>
<td>6</td>
<td>8</td>
</tr>
</tbody>
</table>

You can see that five of the 36 possibilities total six. Therefore, the probability is 5/36.
If you know the probability of an event occurring, it is easy to compute the probability that the event does not occur – the complementary outcome. If $P(A)$ is the probability of Event A, then $1 - P(A)$ is the probability that the event does not occur. For the last example, the probability that the total is 6 is $5/36$. Therefore, the probability that the total is not 6 is $1 - 5/36 = 31/36$.

### 4.1.2 Mutually exclusive outcomes

**Definition**

Mutually exclusive outcomes are outcomes where the occurrence of one of the outcomes excludes the possibility of any of the others happening.

In the first Venn diagram (left), events A and B are mutually exclusive. In the second Venn diagram (right), events A and B are not mutually exclusive.

Events A and B are **mutually exclusive** if they have no events in common. In other words, if A occurs B cannot occur and vice-versa. On a Venn Diagram, this would mean that the circles representing events A and B would not overlap.

If, for example, we are asked to pick a card from a pack of 52, the probability that the card is red is $\frac{1}{2}$. The probability that the card is a club is $\frac{1}{4}$. However, if the card is red it can't be a club. These events are therefore mutually exclusive.

### 4.2 Addition law

If A and B are events, the probability of obtaining either of them is:

$$P(A \text{ or } B) = P(A) + P(B) - P(A \text{ and } B)$$

For example, what is the probability that a card chosen at random from a deck of cards will either be a king or a heart?

$$P(\text{king or heart}) = P(A \text{ or } B) = \frac{4}{52} + \frac{13}{52} - \frac{1}{52} = 30.77\%$$

If the events A and B are mutually exclusive (that is, if both events cannot occur simultaneously), the last term $[P(A \text{ and } B)]$ will be 0. Thus the addition rule with mutually exclusive events becomes:

$$P(A \text{ or } B) = P(A) + P(B)$$

For example, suppose a high school consists of 25 per cent juniors, 15 per cent seniors, and the remaining 60 per cent is students of other grades. The relative frequency of students who are either juniors and seniors is 40 per cent. We can add the relative frequencies of juniors and seniors because no student can be both junior and senior.

$$P(J \text{ or } S) = 0.25 + 0.15$$

which equals 0.40

The above formula can be expanded to consider more than two exclusive events:

$$P(A \text{ or } B \text{ or } C \text{ or } D \ldots \text{ or } Z) = P(A) + P(B) + P(C) + \ldots + P(Z)$$

### 4.2.1 Unions and Intersections

In statistics, the sign $\cap$ means intersection and $\cup$ means union.

If two events are mutually exclusive, $P(A \cap B) = 0$, so
\[ P(A) + P(B) = P(A \cup B) \]

An element qualifies for the **union** of A and B if it is in either A or B or in both. For example, if \( A = (2, 8, 14, 18) \) and \( B = (4, 6, 8, 10, 12) \), then the union of A and B = \( (2, 4, 6, 8, 10, 12, 14, 18) \). The key word indicating the union of two or more events is 'or'.

An element qualifies for the **intersection** of A and B if it is in both A and B. For example, if \( A = (2, 8, 14, 18) \) and \( B = (4, 6, 8, 10, 12) \), then the intersection of A and B = \( (8) \). The key word indicating the intersection of two or more events is 'and'.

See the following figures showing events 'A and B' and 'A or B' when A and B are NOT mutually exclusive.

**Figure 6 Venn diagram showing the intersection of two events**

**Figure 7 Venn diagram showing the union of two events**

Shaded area is the intersection (\( \cap \)) of A and B: \( A \cap B \)

Shaded area is the union (\( \cup \)) of A and B: \( A \cup B \)

From the Venn diagram (right panel) we can see that the probability of A or B is the sum of the individual probabilities, minus the intersection, which otherwise gets added twice:

\[ P(A \cup B) = P(A) + P(B) - P(A \cap B). \]

If A and B are mutually exclusive, then \( P(A \cap B) \) is empty and so \( P(A \cup B) = P(A) + P(B) \); i.e. the sum of the probabilities.

### 4.2.2 Venn diagram: General rule of addition

We can show how to calculate \( P(A \cup B) \) from three diagrams.

The shaded area is the probability of A and not B = \( P(A) - P(A \cap B) \)

The shaded area is the probability of A and B = \( P(A \cap B) \)
The shaded area is the probability of B and not A = P(B) – P(A ∩ B)

If we add these three sections together we get the formula for the probability of A or B = P(A) + P(B) – P(A ∩ B)

**Question 1: Addition law**

Suppose we have a machine that inserts a mixture of beans, broccoli, and other types of vegetables into a plastic bag. Most of the bags contain the correct weight, but because of slight variation in the size of the beans and other vegetables, a package might be slightly underweight or overweight. A check of many packages in the past indicates that:

<table>
<thead>
<tr>
<th>Weight</th>
<th>Event</th>
<th>No. of packages</th>
<th>Probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>Underweight</td>
<td>X</td>
<td>100</td>
<td>0.025</td>
</tr>
<tr>
<td>Correct weight</td>
<td>Y</td>
<td>3600</td>
<td>0.9</td>
</tr>
<tr>
<td>Overweight</td>
<td>Z</td>
<td>300</td>
<td>0.075</td>
</tr>
<tr>
<td></td>
<td></td>
<td>4000</td>
<td>1.000</td>
</tr>
</tbody>
</table>

What is the probability of selecting a package at random and having the package be under weight or overweight?

(The answer is at the end of the chapter)

**Worked Example: Addition rule**

A hamburger chain found that 75 per cent of all customers use mustard, 80 per cent use ketchup, 65 per cent use both.

What is the probability that a particular customer will use at least one of these?

**Solution**

A = customer uses mustard
B = customer uses ketchup

A ∪ B = a particular customer will use at least one of these

Given P(A) = .75, P(B) = .80 and P(A ∩ B) = .65

P(A ∪ B) = P(A) + P(B) - P(A ∩ B) = 0.75 + 0.80 - 0.65 = 0.90.

**4.3 Multiplication law**

**Definition**

Independent events are events where the outcome of one event in no way affects the outcome of the other events.

The simple multiplication law for two independent events, A and B, is as follows:

P(A and B) = P(A ∩ B) = P(A) x P(B)

Note that P(A and B) = 0 when A and B have mutually exclusive outcomes.
Two events are independent if the first one does not influence the second. For example, if a bag contains two blue balls and two red balls and two balls are selected randomly, the events are:

- independent if the first ball is replaced after being selected.
- not independent if the first ball is removed without being replaced. In this instance, there are only three balls remaining in the bag so the probabilities of selecting the various colours have changed.

Two events are independent if (and only if):

\[ P(A \cap B) = P(A) \times P(B) \]

**Worked Example: Independent events**

A die is thrown and a coin is tossed simultaneously. What is the probability of throwing a five and getting heads on the coin?

**Solution**

The probability of throwing a 5 on a die is \( \frac{1}{6} \)

The probability of a tossed coin coming up heads is \( \frac{1}{2} \)

The probability of throwing a 5 and getting heads on a coin is \( \frac{1}{2} \times \frac{1}{6} = \frac{1}{12} \)

**Definition**

Dependent or conditional events are events where the outcome of one event depends on the outcome of the other.

The general rule of multiplication for two dependent events, A and B is as follows.

\[ P(A \text{ and } B) = P(A) \times P(B|A) \]

We can interpret this as the probability of A and B is equal to the probability of A, multiplied by the probability of B given that outcome A has happened.

For example, suppose there are 10 marbles in a bag, and three are defective. Two marbles are to be selected, one after the other without replacement. What is the probability of selecting a defective marble followed by another defective marble?

Probability that the first marble selected is defective: \( P(A) = \frac{3}{10} \)

If the first marble is defective, the probability that the second marble selected is defective: \( P(B) = \frac{2}{9} \)

\[ P(A \text{ and } B) = \left( \frac{3}{10} \right) \times \left( \frac{2}{9} \right) = 6.67\% \]

When A and B are independent events, then \( P(B|A) = P(B) \) since, by definition, the occurrence of B (and therefore \( P(B) \)) does not depend upon the occurrence of A. Similarly \( P(A|B) = P(A) \).

**Question 2: Multiplication rule**

The independent probabilities that the three sections of an accounting department will encounter one computer error in a week are respectively 0.1, 0.2 and 0.3. There is never more than one computer error encountered by any one section in a week.

Calculate the probability that there will be the following number of errors encountered by the accounting department next week.

(a) At least one computer error.

(b) One and only one computer error.

(The answer is at the end of the chapter)
4.4 Conditional law

Conditional probabilities are based on knowledge of one of the variables. The conditional probability of an event, such as A, occurring given that another event, such as B, has occurred is expressed as:

\[ P(A|B) = \frac{P(A \text{ and } B)}{P(B)} \]

Note that when using the conditional law of probability, you always divide the joint probability by the probability of the event after the word given. Thus, to get \( P(A \text{ given } B) \), you divide the joint probability of A and B by the unconditional probability of B. In other words, the above equation is used to find the conditional probability for any two dependent events. When two events, such as A and B, are independent their conditional probability is calculated as follows:

\[ P(A|B) = P(A) \quad \text{and} \quad P(B|A) = P(B) \]

Worked Example: Given event

If a single card is selected at random from a deck of cards, what is the probability that the card is a king given that it is a club?

Solution

\[ P(\text{king given club}) = P(A|B) = \frac{P(B \cap A)}{P(B)} \]
\[ P(B) = \frac{13}{52}, \text{ and } P(\text{king and club}) = \frac{1}{52}, \text{ thus} \]
\[ P(\text{king given club}) = \frac{P(A|B) = (1/52)}{(13/52)} = \frac{1}{13} \]

Recall the previous hamburger chain example (4.2.2), what is the probability that a ketchup user uses mustard?

\[ P(A|B) = P(A \cap B) / P(B) \]
\[ = 0.65 / 0.80 = 0.8125 \]

Question 3: Conditional probability

The board of directors of Shuttem Co has warned that there is a 60 per cent probability that a factory will be closed down unless its workforce improves its productivity. The factory’s manager has estimated that the probability of success in agreeing a productivity deal with the workforce is only 30 per cent.

Required

Assuming a deal would improve productivity, which would keep the factory open, determine the likelihood that the factory will be closed.

(The answer is at the end of the chapter)

5 Probability matrices

Section overview

- We are going to use a contingency table to explain and apply marginal, union, joint and conditional probabilities. A contingency table is a table with rows and columns with totals for each. The entries in the table describe data gathered from a survey or from research.
5.1 Probability matrix

A contingency table provides a different way of calculating probabilities. The table helps in determining conditional probabilities quite easily. The table displays sample values in relation to two different variables that may be dependent or contingent on one another. The entries should be mutually exclusive and together cover the sample space.

Worked Example: Flashback survey

Flashback recently conducted a survey of 200 selected purchasers of their newly introduced gadget to obtain a gender and age profile of its new customers. The findings have been summarised in the contingency table below, which allows you to see how they interact.

<table>
<thead>
<tr>
<th>Age (Years)</th>
<th>Sex (L)</th>
<th>25 - 40</th>
<th>&gt;40</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male (M)</td>
<td>60</td>
<td>20</td>
<td>40</td>
<td>120</td>
</tr>
<tr>
<td>Female (F)</td>
<td>40</td>
<td>30</td>
<td>10</td>
<td>80</td>
</tr>
<tr>
<td>Total</td>
<td>100</td>
<td>50</td>
<td>50</td>
<td>200</td>
</tr>
</tbody>
</table>

We can define the following events:

M = a male is selected
F = a female is selected
L = the person selected is under 25
B = the person selected is between 25 and 40
G = the person selected is over 40

The contingency table, also called a raw values matrix, shows the frequency counts for each category in the survey and the subtotals and totals containing a breakdown of these by age and gender.

One person from the total group of 200 is to be selected at random to receive a prize.

5.1.1 Union probability

Definition

Union probability is denoted by \( P(A \cup B) \) or \( P(A \cup B) \), where \( A \) and \( B \) are two events. \( P(A \cup B) \) is the probability that \( A \) will occur or that \( B \) will occur or that both \( A \) and \( B \) will occur.

What is the probability that the person selected from the Flashback survey is female or over 40?

Solution

Let \( F \) denote the event of female and \( G \) denote the event of over 40 years of age. The question is \( P(F \cup G) \).

By the general law of addition \( P(F \cup G) = P(F) + P(G) - P(F \cap G) \)

Of the 200 in the group, 80 are female. Therefore, \( P(F) = 80/200 = .4 \). The group includes 50 whose age is greater than 40, therefore \( P(G) = 50/200 = .25 \). Because 10 from the group are both female and over 40, \( P(F \cap G) = 10/200 = .05 \). The union probability is solved as \( P(F \cup G) = .4 + .25 - .05 = .6 \)

To solve this probability using a matrix, you can either use the contingency table (raw values matrix) as we have just shown or convert the table to a probability matrix by dividing every value in the matrix by the value of \( n \), 200. The contingency table can be used in a manner similar to the Addition Law. To compute the
union probability of selecting a person who is either female or over 40 years of age from the contingency table, add the number of people in the Female row (80) to the number of people in the >40 column (50), then subtract the number of people in the intersection cell of Female and >40 (10). This step yields the value 80 + 50 - 10 = 120. Dividing this value by the value of n (200) produces the union probability. \( P(F \cup G) = \frac{120}{200} = 0.6 \)

A second way to produce the answer from the contingency table is to add all the cells once that are in the Female row (80) or the >40 column (a further 40) = 120 and then divide by the total number in the group = 200, which gives \( P(F \cup G) = \frac{120}{200} = 0.6 \)

5.1.2 Marginal probability

Definition

Marginal probability or unconditional probability is the probability of an event, regardless of the results of any other events. It is denoted \( P(A) \), where A is the event.

What is the probability that the person selected from the Flashback survey is female?

Solution

The probability of selecting a female \( P(F) = \frac{80}{200} = 0.4 \)

From the contingency table we can also calculate the probability of the following:

- The probability of selecting a male \( P(M) = \frac{120}{200} = 0.6 \)
- The probability of selecting a person under 25 \( P(L) = \frac{100}{200} = 0.5 \)
- The probability of selecting a person aged between 25 and 40 \( P(B) = \frac{50}{200} = 0.25 \)
- The probability of selecting a person who is older than 40 \( P(G) = \frac{50}{200} = 0.25 \)

5.1.3 Joint probability

Definition

Joint probability measures the likelihood that two or more events will happen concurrently. It is denoted by \( P(A \text{ and } B) \) or \( P(A \cap B) \). To become eligible for the joint probability, both events A and B must occur.

What is the probability that the person selected from the Flashback survey is female and under 25?

Solution

The probability of selecting a female and under 25 = \( P(F \text{ and } L) = \frac{40}{200} = 0.2 \)

We could also calculate the probability of selecting a male who is younger than 40 = \( P(M \text{ and } <25) + P(M \text{ and } 25-40) = \frac{80}{200} + \frac{40}{200} = 0.4 \)

5.1.4 Conditional probability

Definition

A conditional probability is denoted by \( P(A|B) \). This phrase is read: the probability that A will occur given that B is known to have occurred.
Suppose that information has been leaked about the winner of the prize. The selected individual is under 25 years of age.

What is the probability that the winner from the Flashback survey is male?

**Solution**

Given that event L occurred, we have narrowed the number of outcomes to the 100 under 25-year olds. Each is equally likely to be chosen and 60 of them are male so the answer is \( P = P(M|L) = 0.6 \)

**Question 4: Contingency tables**

A cosmetics company has developed a new anti-dandruff shampoo which is being tested on volunteers. Seventy percent of the volunteers have used the shampoo whereas others have used a normal shampoo, believing it to be the new anti-dandruff shampoo. Two sevenths of those using the new shampoo showed no improvement whereas one third of those using the normal shampoo had less dandruff.

**Required**

A volunteer shows no improvement. What is the probability that he used the normal shampoo?

(The answer is at the end of the chapter)

### 5.2 Probability matrices

Probability matrices are another useful tool for solving probability problems. A probability matrix displays the marginal probabilities and the intersection probabilities of a given problem. Union probabilities and conditional probabilities can be computed from the matrix. Generally a probability matrix is constructed as a two dimensional table.

Shown here are the raw values matrix and corresponding probability matrix for the results of a national survey of 200 executives who were asked to identify the geographic locale of their company and their company’s industry type. The executives were only allowed to select one locale and one industry type.

#### Raw values matrix

<table>
<thead>
<tr>
<th>Geographic location</th>
<th>North</th>
<th>South</th>
<th>East</th>
<th>West</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Finance</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A</td>
<td>24</td>
<td>10</td>
<td>8</td>
<td>14</td>
<td>56</td>
</tr>
<tr>
<td>Manufacturing</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>B</td>
<td>30</td>
<td>6</td>
<td>22</td>
<td>12</td>
<td>70</td>
</tr>
<tr>
<td>Communications</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C</td>
<td>28</td>
<td>18</td>
<td>12</td>
<td>16</td>
<td>74</td>
</tr>
<tr>
<td>Total</td>
<td>82</td>
<td>34</td>
<td>42</td>
<td>42</td>
<td>200</td>
</tr>
</tbody>
</table>

By dividing every value of the raw values matrix by the total (200), the corresponding probability matrix (shown below) can be constructed.
Probability matrix

Geographic location

<table>
<thead>
<tr>
<th></th>
<th>North</th>
<th>South</th>
<th>East</th>
<th>West</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>D</strong></td>
<td>.12</td>
<td>.05</td>
<td>.04</td>
<td>.07</td>
<td>.28</td>
</tr>
<tr>
<td><strong>E</strong></td>
<td>.15</td>
<td>.03</td>
<td>.11</td>
<td>.06</td>
<td>.35</td>
</tr>
<tr>
<td><strong>F</strong></td>
<td>.14</td>
<td>.09</td>
<td>.06</td>
<td>.08</td>
<td>.37</td>
</tr>
<tr>
<td><strong>G</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>.41</td>
<td>.17</td>
<td>.21</td>
<td>.21</td>
<td>1.00</td>
</tr>
</tbody>
</table>

Suppose a respondent is selected randomly from these data.

A  What is the probability that the respondent is from the East (F)?

B  What is the probability that the respondent is from the communications industry (C) or from the North (D)?

C  What is the probability that the respondent is from the South (E) or from the finance industry (A)?

**Solution**

A  \[ P(\text{East}) = P(F) = .21 \text{ (the total of column F)} \]

B  \[ P(C \cup D) = P(C) + P(D) - P(C \cap D) = .37 + .41 - .14 = .64 \text{ (Row C total + column D total - cell CD)} \]

C  \[ P(E \cup A) = P(E) + P(A) - P(E \cap A) = .17 + .28 - .05 = .40 \text{ (Total of column E + total of row A - cell AE)} \]

Note in the previous probability matrix the row totals and the column totals are the marginal probabilities.

**Question 5: Probability matrix**

Convert the Flashback survey matrix from its raw value state to a probability matrix then design and solve four problems; one for each of the types of probability – union, joint, marginal and conditional.

(The answer is at the end of the chapter)
Key chapter points

- Probability is a measure of likelihood and can be stated as a percentage, a ratio, or more usually as a number from 0 to 1.
- The difference between a frequency histogram and a relative frequency histogram is that the vertical axis uses relative or proportional frequency instead of simple frequency.
- One conception of probability is drawn from the idea of symmetrical outcomes. For example, the two possible outcomes of tossing a fair coin are not distinguishable in any way that affects which side will land up or down.
- The empirical probability approach is based on long-run relative frequencies and is defined as the ratio of the number of observed outcomes favourable to the event divided by the number of observed outcomes. The larger the number of trials, the more accurate the estimate of probability.
- In the absence of better information upon which to rely, subjective probability may be used to make logically consistent decisions.
- There are three distinct ways of assigning probability: the classical, the relative frequency or empirical method, and the subjective approach.
- The simple addition law for two mutually exclusive events, A and B, is as follows:
  \[ P(A \text{ or } B) = P(A) + P(B) \]
  Mutually exclusive outcomes are outcomes where the occurrence of one of the outcomes excludes the possibility of any of the others happening.
- The simple multiplication law for two independent events A and B, is as follows:
  \[ P(A \text{ and } B) = P(A) \times P(B) \]
  Independent events are events where the outcome of one event in no way affects the outcome of the other events.
- The general rule of addition for two events, A and B, which are not mutually exclusive, is as follows:
  \[ P(A \text{ or } B) = P(A) + P(B) - P(A \text{ and } B) \]
- The general rule of multiplication for two dependent events, A and B, is:
  \[ P(A \text{ and } B) = P(A) \times P(B|A) = P(B) \times P(A|B) \]
  Dependent or conditional events are events where the outcome of one event depends on the outcome of the others.
- The conditional law is \( P(A|B) = P(A \text{ and } B)/P(B) \)
- Contingency tables are useful for dealing with conditional probability.
- Probability matrices are used to solve probability problems.
Quick revision questions

1. Which of the three approaches to probability subscribes to the view that the probability may be found by counting the number of successes in the sample and dividing by the number of outcomes in the sample space?
   A. classical
   B. subjective
   C. empirical

2. 100 people were asked whether they were left-handed. Four people answered 'yes'. What is the relative frequency of 'left-handed'?

3. What is an event?

4. Explain what is meant by independent events.

5. If you roll a six-sided die, what is the probability that either a one or a six will come up?

6. If you throw a die three times, what is the probability that one or more of your throws will come up with a one?

7. Two events and conditional probability
   Let event A be that the team makes the playoffs. Let event B be that the coach gets rehired for next season. List the four possibilities.

8. An analysis of 480 working days in a factory shows that on 360 days there were no machine breakdowns. Assuming that this pattern will continue, what is the probability that there will be a machine breakdown on a particular day?
   A. 0 per cent
   B. 25 per cent
   C. 33 per cent
   D. 75 per cent

9. A production director is responsible for overseeing the operations of three factories – North, South and West. He visits one factory per week. He visits the West factory as often as he visits the North factory, but he visits the South factory twice as often as he visits the West factory.
   What is the probability that in any one week he will visit the North factory?
   A. 0.17
   B. 0.20
   C. 0.25
   D. 0.33

10. A project may result in profits of $15 000 or $20 000, or in a loss of $5 000. The probabilities of each profit are 0.2, 0.5 and 0.3 respectively.
    What is the expected profit?
1 A The classical approach.

2 The relative frequency is $4/100 = 1/25 = 0.04$

3 An event is any collection of outcomes from a probability experiment.

4 Independent events are events where the outcome of any one event in no way affects the outcome of the other events.

5 The two outcomes about which we are concerned (a one or a six coming up) are called favourable outcomes. Given that all outcomes are equally likely, we can compute the probability of a one or a six using the simple addition (or) law for two mutually exclusive events, A and B, as follows:

$$P(A \text{ or } B) = P(A \cup B) = P(A) + P(B)$$

In this case there are two favourable outcomes and six possible outcomes. So the probability of throwing either a one or six is $1/3$.

6 It makes more sense to re-write the question to be – what is the probability of getting a one on the first throw OR a one on the second throw OR a one on the third throw? The easiest way to approach this problem is to compute the probability of

NOT getting a 1 on the first throw
AND not getting a 1 on the second throw
AND not getting a 1 on the third throw.

The answer will be 1 minus this probability. The probability of not getting a 1 on any of the three throws is $5/6 \times 5/6 \times 5/6 = 125/216$. Therefore, the probability of getting a 1 on at least one of the throws is $1 - 125/216 = 91/216$.

7 The four possibilities are:

1 A and B (team makes the playoffs, and the coach gets rehired)
2 A and not B (team makes the playoffs, but the coach does not get rehired)
3 B and not A (team does not make the playoffs, but the coach does get rehired)
4 not B and not A (team does not make the playoffs, the coach does not get rehired)

8 B The data tells us that there was a machine breakdown on 120 days ($480 - 360$) out of a total of 480.

$$P(\text{machine breakdown}) = \frac{120}{480} \times 100\% = 25\%$$

You should have been able to eliminate option A immediately since a probability of 0 per cent is impossibility.

If you selected option C, you calculated the probability of a machine breakdown as 120 out of a possible 365 days instead of 480 days.

If you selected option D, you incorrectly calculated the probability that there was not a machine breakdown on any particular day.
Factory | Ratio of visits
--- | ---
North | 1
South | 2
West | 1

$P(\text{visiting North factory}) = \frac{1}{4} = 0.25$

If you didn't select the correct option, make sure that you are clear about how the correct answer has been arrived at. Remember to look at the ratio of visits since no actual numbers of visits are given.

10 Expected profit = ($15,000 \times 0.2) + ($20,000 \times 0.5) + (– $5,000 \times 0.3)$

= $3,000 + $10,000 – $1,500$

= $11,500$
Answers to chapter questions

1. What is the probability of selecting a package at random and having the package be under weight or over weight? The events are mutually exclusive, since a package cannot be underweight and overweight at the same time. The answer is: \( P(X \text{ or } Z) = P(X) + P(Z) = 0.025 + 0.075 = 0.1 \)

2. (a) The probability of at least one computer error is 1 minus the probability of no error. The probability of no error is \( 0.9 \times 0.8 \times 0.7 = 0.504 \).

(Since the probability of an error is 0.1, 0.2 and 0.3 in each section, the probability of no error in each section must be 0.9, 0.8 and 0.7 respectively.)

The probability of at least one error is \( 1 - 0.504 = 0.496 \).

(b) \( Y = \text{yes}, \ N = \text{no} \)

<table>
<thead>
<tr>
<th>Section 1</th>
<th>Section 2</th>
<th>Section 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Error?</td>
<td>Y</td>
<td>N</td>
</tr>
<tr>
<td>Error?</td>
<td>N</td>
<td>Y</td>
</tr>
<tr>
<td>Error?</td>
<td>N</td>
<td>N</td>
</tr>
</tbody>
</table>

Probabilities

<table>
<thead>
<tr>
<th>(i) ( 0.1 \times 0.8 \times 0.7 = 0.056 )</th>
<th>(ii) ( 0.9 \times 0.2 \times 0.7 = 0.126 )</th>
<th>(iii) ( 0.9 \times 0.8 \times 0.3 = 0.216 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total 0.398</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The probability of one and only one error is 0.398.

3. If outcome A is the shutdown of the factory and outcome B is the failure to improve productivity:

\[
P(A \text{ and } B) = P(B) \times P(A|B) = 0.7 \times 0.6 = 0.42
\]

4. The problem is solved by drawing a contingency table, showing 'improvement' and 'no improvement', volunteers using normal shampoo and volunteers using the new shampoo.

Let us suppose that there were 1 000 volunteers (we could use any number) then 700 used the new shampoo and 300 did not. \( \frac{2}{7} \) of the 700 (= 200) and \( \frac{2}{3} \) of the 300 (= 200) showed no improvement. We can depict the results of the test on the 1 000 volunteers as follows:

<table>
<thead>
<tr>
<th></th>
<th>New shampoo</th>
<th>Normal shampoo</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Improvement</td>
<td>500</td>
<td>100</td>
<td>600</td>
</tr>
<tr>
<td>No improvement</td>
<td>200</td>
<td>200</td>
<td>400</td>
</tr>
<tr>
<td>Total</td>
<td>700</td>
<td>300</td>
<td>1000</td>
</tr>
</tbody>
</table>

We can now calculate \( P(\text{shows no improvement}) \)

\[
P(\text{shows no improvement}) = \frac{400}{1000} = \frac{1}{2}
\]

\( P(\text{used normal shampoo} | \text{shows no improvement}) = \frac{200}{400} = \frac{1}{2} \)

Other probabilities are just as easy to calculate.

\( P(\text{shows improvement} | \text{used new shampoo}) = \frac{500}{700} = \frac{5}{7} \)

\( P(\text{used new shampoo} | \text{shows improvement}) = \frac{500}{600} = \frac{5}{6} \)
Because students will all design different questions, there is no model answer to this part of the question. Some questions that apply to the matrix follow for students who wish to be more challenged.

Q1 The prize giving is to be televised. The channel is watched by 20 per cent of men and 40 per cent of women. What proportion of Flashback’s customers are likely to be watching?

Q2 Four of the women have had birthdays since the survey started and are now in the 25-40 year old bracket. Calculate the new row on the probability matrix and the new probability of selecting a person aged between 25 and 40 or male.

Q3 Because there has been no denial about the leaked information on the winner, what is the new probability of a female under 25 winning? How much better are their chances of winning with the reduced numbers?

Q4 What is the probability of a 30-year old purchaser being a male?

Solutions

Q1 Assuming the sample is representative of all Flashback’s customers, if W is the event of a customer watching the prize giving:

\[ P(M \cap W) = P(M) \times P(W|M) = .6 \times .2 = .12 \]
\[ P(F \cap W) = P(F) \times P(W|F) = .4 \times .4 = .16 \]

Total = .12 + .16 = .28

Q2 The female row in the contingency table was 40, 30, 10 and is now 36, 34, 10 and the new probability matrix is:

<table>
<thead>
<tr>
<th>Age (Years)</th>
<th>&lt;25</th>
<th>25 - 40</th>
<th>&gt;40</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sex</td>
<td>L</td>
<td>B</td>
<td>G</td>
<td></td>
</tr>
<tr>
<td>Male (M)</td>
<td>.3</td>
<td>.1</td>
<td>.2</td>
<td>.6</td>
</tr>
<tr>
<td>Female (F)</td>
<td>.18</td>
<td>.17</td>
<td>.05</td>
<td>.4</td>
</tr>
<tr>
<td>Total</td>
<td>.48</td>
<td>.27</td>
<td>.25</td>
<td>1.0</td>
</tr>
</tbody>
</table>

\[ P(B \cup M) = .6 + .17 = .77 \]

Q3 Still keeping the assumption of the leak, \[ P(F|L) = 0.18 / 0.48 = 0.375 \] and was previously 0.2 / 0.5 = 0.4

i.e. there is less chance of a female under 25 winning, but if you are a female under 25, then your personal chance (after the leaked information) was 1/(60 + 40) = .01 but has now increased to 1/(60 + 36) = .0104

Q4 \[ P(M|B) = P(M \cap B) / P(B) = 0.1 / 0.27 = 0.37 \]
Chapter 15

Hypothesis testing

<table>
<thead>
<tr>
<th>Learning objectives</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hypothesis testing</td>
<td>LO15</td>
</tr>
<tr>
<td>Explain the concept of hypothesis testing</td>
<td>LO15.1</td>
</tr>
<tr>
<td>Construct null and alternative hypotheses</td>
<td>LO15.2</td>
</tr>
<tr>
<td>Distinguish between Type I and Type II errors</td>
<td>LO15.3</td>
</tr>
<tr>
<td>Test population mean using one-tail and two-tail tests</td>
<td>LO15.4</td>
</tr>
<tr>
<td>Test population proportion</td>
<td>LO15.5</td>
</tr>
<tr>
<td>Calculate and interpret the probability value (p-value) in hypothesis testing</td>
<td>LO15.6</td>
</tr>
</tbody>
</table>

Topic list

1. Concept of hypothesis testing
2. Formulate an analysis plan
3. Analyse sample data
4. Test methods
Introduction

When interpreting an experimental finding, a natural question arises as to whether the finding could have occurred by chance. Hypothesis testing is a statistical procedure for testing whether chance is a plausible explanation of an experimental finding.
Before you begin

If you have studied these topics before, you may wonder whether you need to study this chapter in full. If this is the case, please attempt the questions below, which cover some of the key subjects in the area.

If you answer all these questions successfully, you probably have a reasonably detailed knowledge of the subject matter, but you should still skim through the chapter to ensure that you are familiar with everything covered.

There are references in brackets indicating where in the chapter you can find the information, and you will also find a commentary at the back of the Study Manual.

1. A researcher believes that second graders will score higher than first graders on a particular test. Which of the following is the null hypothesis? (Section 1.3)
   A. Mean of the 1st graders < mean of the second graders
   B. Mean of the 1st graders > mean of the second graders
   C. Mean of the 1st graders = mean of the second graders

2. A recent report indicated that waiters and waitresses at casual dining restaurants make an average of $100 per night in tips with a standard deviation of $15. Maureen works in a casual dining restaurant and doesn’t think this is correct. She feels she makes much less than this in an average night. Over the next five work nights, she computes her tips and the average is $93. Does Maureen make significantly less than what the report stated at the .05 level of significance? Use a z score statistic for your test. At the .05 level of significance, the critical z score statistic is -1.64. (Section 2.2)

3. Consider the statement, 'New cars are expected to last an average of three years before needing a major service.' With a p-value of 0.0079, we conclude that (Section 3.1)
   A. no conclusion can be made.
   B. new cars last an average of less than three years before needing a major service.
   C. new cars last an average of at least three years before needing a major service.

4. Given the set of hypotheses: $H_0: p = 0.4$ $H_a: p < 0.4$. This test is: (Section 2)
   A. two-tailed
   B. right-tailed
   C. left-tailed
   D. no-tailed

5. Consider the statement, 'New cars are expected to last an average of three years before needing a major service.' With a p-value of 0.2456, which is the correct decision? (Section 3.1)
   A. no decision can be made.
   B. reject the null hypothesis.
   C. do not reject the null hypothesis.
Consider the statement, 'New cars are expected to last an average of at least three years before needing a major service.' Which of the following is the null hypothesis?

A the population mean <= 3.
B the population mean < 3.
C the population mean > 3.
D the population mean is >= 3.

A study is done to see if the average age a 'child' moves permanently out of his parents' home in the United States is at most 23. 43 US Adults, all age 40, were surveyed. The sample average age was 24.2 with a standard deviation of 3.7. Which is the Type I error?

A Conclude that the average age is greater than 23, when it is 24.2.
B Conclude that the average age is at most 24.2, when it is at most 24.2.
C Conclude that the average age is at most 23, when it is greater than 23.
D Conclude that the average age is greater than 23, when it is at most 23.

A study is done to see if the average age a 'child' moves permanently out of his parents' home in the United States is at most 23. 43 US Adults, all age 40, were surveyed. The sample average age was 24.2 with a standard deviation of 3.7. Which is the Type II error?

A Conclude that the average age is greater than 23, when it is at most 23.
B Conclude that the average age is greater than 23, when it is 24.2.
C Conclude that the average age is at most 24.2, when it is at most 24.2.
D Conclude that the average age is at most 23, when it is greater than 23.
1 Concept of hypothesis testing

1.1 Steps in hypothesis testing

When interpreting an experimental finding, a natural question arises as to whether the finding could have occurred by chance. Hypothesis testing is a statistical procedure for testing whether chance is a plausible explanation of an experimental finding.

Hypothesis testing is based on testing a null hypothesis. This is the hypothesis or assumption that there is no significant difference between the results we have obtained and the results that we might expect. We then test this hypothesis by comparing our results with what we might expect, using a test statistic.

Depending on this comparison, we either ‘accept’ the null hypothesis, or we reject the null hypothesis and accept instead an alternative hypothesis. (Strictly, we do not accept the null hypothesis. We do not reject it, which is not quite the same thing.)

All hypothesis tests are conducted the same way. The researcher states a hypothesis is to be tested, formulates an analysis plan, analyses sample data according to the plan, and accepts or rejects the null hypothesis, based on results of the analysis.

1 State the hypotheses. This involves stating a null hypothesis and an alternative hypothesis. The hypotheses are stated in such a way that they are mutually exclusive. That is, if one is true, the other must be false. With hypothesis testing, we are often testing whether the mean of a sample indicates that the sample comes from a known population. The null hypothesis is that there is no significant difference between the sample mean and the known or assumed mean for the population as a whole. (Instead of testing the mean of a sample, we may be testing the proportion of an item in the sample, and comparing this with a known or assumed proportion in the population as a whole.)

2 Formulate an analysis plan. The analysis plan describes how to use sample data to accept or reject the null hypothesis. The do not reject/reject decision often focuses around a single test statistic. The test statistic makes an assumption about the expected probability distribution of sample results.

3 Analyse sample data. The mean (or proportion) of the sample is most unlikely to be exactly the same as the actual or assumed population mean (or proportion). There will be a difference, and we need to decide whether this difference seems to be significant. Because we assume the probability distribution of sample results, we can assess whether the result from the sample is within a statistically acceptable difference. This is done by measuring a test statistic, which may be a z score or a t test statistic. We compare this test statistic with a critical value for the z score or t test score.

4 Interpret results. Apply the decision rule described in the analysis plan. If the test statistic supports the null hypothesis, do not reject the null hypothesis. Otherwise, reject the null hypothesis and accept the alternative hypothesis.

Definitions

A **hypothesis** is a statement or claim regarding a characteristic of one or more populations.

**Hypothesis testing** is a procedure, based on sample evidence and probability, used to test claims regarding a characteristic of one or more populations.

The characteristic is usually population mean or population proportion.
1.2 The concept of hypothesis testing

Suppose that the average time for employees to complete a standard task is 14 minutes with a standard deviation of 1.5 minutes (normally distributed around the mean of 14 minutes). A change is made in the way that the task is performed and the average completion time for a sample of 20 performances of the task is 11 minutes on average. Can we conclude that the new procedure for the task has reduced the average completion time, given that we have sampled only 20 performances of the new procedure?

The difference between the expected time and the time from the sample is 2 minutes, but could this simply be a matter of chance, and within a ‘realistic’ limit?

With hypothesis testing, we establish the null hypothesis. This is that there is no significant difference between the sample average time of 11 minutes and the expected or normal average time of 14 minutes. The alternative hypothesis is that there is a significant difference (and the performance time has been reduced from 14 minutes).

The null hypothesis is tested. A z-score is used in statistics to model any normal distribution as a standard normal distribution. It is the number of standard deviations that the value \( x \) is above (to the right of) or below (to the left of) the mean, \( \mu \).

Formula to learn

A z-score is the number of standard deviations that a given \( x \) value is above or below the mean. If \( z \) represents the z-score for a given \( x \) value then

\[
  z = \frac{x - \mu}{\sigma}
\]

In the example above, the difference between the expected time and the sample time is 3 minutes, which is 2 standard deviations. To complete the hypothesis test, we need to decide on a level of confidence that we want to have in deciding whether to accept the null hypothesis or not. The test statistic is the number of standard deviations from the mean (above the mean, below the mean or either above or below the mean) within which the sample value must fall if we are not to reject the null hypothesis and accept that the difference between actual and expected results was simply a matter of chance.

Why is the z score important in hypothesis testing?

If we took a large number of samples and calculated the mean value of each sample, the sample means would be distributed around the true population mean. The sample means would in fact be normally distributed around the true population mean. If we know the standard deviation of this probability distribution of sample means, we can assess the probability that a sample mean lies within this probability distribution. The z score is important in hypothesis testing because it measures the number of standard deviations between the sample mean and the actual or assumed population mean, and we can analyse the normal distribution to decide whether the difference between the sample mean and the population mean is significant, and suggests that the sample does not come from the population.

Calculating the z score

The z score is simply a measure of the number of standard deviations between the sample mean and the population mean.

Using the example above, we can calculate the z score by \( z = \frac{x - \mu}{\sigma} \) where \( x \) is observed performance time of 11 minutes and \( \mu \) is the expected performance time of 14 minutes. The standard deviation is 1.5 minutes, so:

\[
  z = \frac{(11 - 14)}{1.5} = -2.
\]
By looking at the area under the normal curve we can see that the area to the left of –2 is about 2.3% or .023. This is a low probability, and therefore the performance times in the sample would have to be unusually quick if they the average time remains at 14 minutes. The null hypothesis that there has been no reduction in the average performance time must have considerable doubt cast on it. There is strong evidence that average performance times have been reduced by the new procedure.

### 1.3 Null and alternative hypotheses

The best way to determine whether a statistical hypothesis is true would be to examine the entire population. Since that is often impractical, researchers typically examine a random sample from the population. If sample data is consistent with the statistical hypothesis, the hypothesis is accepted; if not, it is rejected.

#### Definitions

**Null hypothesis.** The null hypothesis, denoted by $H_0$, is usually the hypothesis that sample observations result purely from chance. In other words, the null hypothesis is usually that there is no significant difference between the sample statistic and the known or assumed population statistic. Typically, the statistic we are looking at is the sample mean and the population mean, or a proportion in the sample and the proportion in the population.

**Alternative hypothesis.** The alternative hypothesis, denoted by $H_1$ or $H_a$, is the hypothesis that sample observations are influenced by some non-random cause. The alternative hypothesis will be accepted only if the null hypothesis is rejected.

#### 1.3.1 Introducing the null and alternative hypotheses

The first step for every hypothesis test requires the analyst to state a null hypothesis and an alternative hypothesis. The null hypothesis is typically a hypothesis of no difference e.g. no difference between what is observed and what could happen by chance. That is why the word 'null' is used.

The hypotheses are stated in such a way that they are mutually exclusive. That is, if one is true, the other must be false; and vice versa.

For example, suppose we wanted to determine whether a coin was fair and balanced. A null hypothesis might be that half the flips would result in heads and the other half, in tails. The alternative hypothesis might
be that the number of heads and tails would be very different. Symbolically, these hypotheses would be expressed as:

\[ \text{H}_0: P = 0.5 \]
\[ \text{H}_a: P \neq 0.5 \]

Suppose we flipped the coin 50 times, resulting in 40 heads and 10 tails. Given this result, we would be inclined to reject the null hypothesis. We would conclude, based on the evidence, that the coin was probably not fair and balanced.

The main goal of hypothesis testing is to tell us whether we have enough evidence to reject the null hypothesis. In our case we want to know whether the coin is biased or not, so our null hypothesis should be 'the coin is fair.' If we get enough evidence that contradicts this hypothesis, for example, by flipping it 100 times and having it come up heads only once, then we can safely reject it.

All of this is perfectly quantifiable, of course but what constitutes 'enough' and 'safely' are a matter of statistics.

1.3.2 Similarity to the justice system

A statistical test procedure is comparable to a trial. A defendant stands trial and is considered innocent as long as his or her guilt is not proven. The prosecutor tries to prove the guilt of the defendant. Only when there is enough evidence is the defendant condemned.

In the start of the procedure there are two hypotheses \( \text{H}_0 \): 'the defendant is innocent', and \( \text{H}_a \): 'the defendant is guilty'. The first one is called the null hypothesis, and is for the time being accepted. The second one is called the alternative (hypothesis). It is the hypothesis one tries to prove.

In good law practice one does not want to condemn an innocent defendant. That is why the hypothesis of innocence is only rejected when an error is very unlikely. Such an error is called error of the first kind (i.e. the condemnation of an innocent person), and the occurrence of this error is controlled to be rare. As a consequence of this asymmetric behaviour, the error of the second kind (setting free a guilty person), is often rather large.

<table>
<thead>
<tr>
<th>Purpose</th>
<th>Hypothesis testing</th>
<th>Justice procedures</th>
</tr>
</thead>
<tbody>
<tr>
<td>Step 1</td>
<td>Null versus alternative hypothesis</td>
<td>Innocence versus guilty</td>
</tr>
<tr>
<td>Step 2</td>
<td>Evaluate the evidence from data</td>
<td>Evaluate the evidence</td>
</tr>
<tr>
<td>Step 3</td>
<td>Make a decision</td>
<td>Jury hand out the decision</td>
</tr>
</tbody>
</table>

1.4 Defining the null and alternative hypotheses

Any hypothesis test will yield one of two outcomes:

- We will reject the hypothesis based on some evidence at hand
- We will not reject the hypothesis based on that same evidence.

Related to these two possible outcomes are the concepts of the null and alternative hypotheses. In statistical terms, the null hypothesis is often a statement that outlines the possible outcome of the test that we do not want to observe. The alternative hypothesis would represent the outcome of the hypothesis test that we wish to find supporting evidence for. The null hypothesis (commonly denoted as \( \text{H}_0 \)) also represents the starting point of hypothesis testing.

To demonstrate, consider an example in which company X will introduce a new product to market if the individuals in the intended target population are likely to spend, on average, more than $10 per week on the product. From the company's perspective, it would be hoped that those acting in the market will spend more than the $10 per week necessary for product launch. The null hypothesis representing the contrary position is that those in the market will not spend the necessary amount required for the product launch to take place.

Our discussion has not made any reference yet to any sample. At present, we have confined ourselves to discussion of a population. In establishing the null hypothesis of a test, we discuss population parameters.
and not the sample parameters. For the example, that means that the null, and alternative, hypothesis is framed in terms of the population mean \( \mu \), not the sample mean \( \bar{x} \). In this instance, the null hypothesis may be framed as:

\[
H_0 : \mu \leq 10
\]

The alternative hypothesis is therefore given as:

\[
H_1 : \mu > 10
\]

In establishing our example, we have framed the problem in terms of what is known as a 'one-tailed' test. If we drew a single sample of respondents from the intended population target market and asked them how much they would be willing to spend on the product per week (provided it was introduced to market), a response on average less than $10 would provide evidence supporting the null hypothesis.

Alternatively, a response on average of greater than or equal to $10 would provide evidence rejecting the null hypothesis, thus supporting the introduction to market of the new product. We show this in the diagram above.

Such tests are known as one tailed test because the range of values resulting in the support of the null hypothesis are confined to one tail of the population distribution from which the sample is drawn. Hypothesis tests may also be created such that evidence supporting the null hypothesis may come from values observed in both tails of the distribution.

The most common test in statistics is testing whether a parameter is statistically different from zero. Such a test is not direction specific.

### 1.4.1 Can we accept the null hypothesis?

Some researchers say that a hypothesis test can have one of two outcomes: you accept the null hypothesis or you reject the null hypothesis. Many statisticians, however, take issue with the notion of 'accepting the null hypothesis'. Instead, they say: you reject the null hypothesis or you fail to reject the null hypothesis.

Why the distinction between 'acceptance' and 'failure to reject'? Acceptance implies that the null hypothesis is true. Failure to reject implies that the data is not sufficiently persuasive for us to prefer the alternative hypothesis over the null hypothesis.

### 2 Formulate an analysis plan

#### Section overview

- The analysis plan describes how to use sample data to accept or reject the null hypothesis. The accept/reject decision focuses around a single test statistic.

#### 2.1 Selecting the test statistic

Once the null and alternative hypotheses have been decided, the next stage in hypothesis testing is to determine the appropriate test statistic to use. Previously, we defined the concepts of population parameter and sample statistics. These respectively define some aspect of the population and sample distributions. A
**Test statistic** is defined as the results of a statistical test that relates some sample statistic to that of a population parameter. Once calculated, the test statistic acts as a basis of comparison between what we observe in the sample data and what we would expect to observe in the population given certain properties of the assumed population distribution.

Hypothesis testing, as we have discussed, is about testing the probability of whether a sample observation is likely to have come from (or represent) the true population value. There is therefore some level of probability that any statistical test we perform will provide an erroneous result. One benefit of classical statistics is that while the analyst will never know for certain whether the hypothesis test conducted has given the wrong result or not, the probability that an erroneous result will occur can be calculated.

Two statistical tests are commonly associated with hypothesis tests involving normally distributed random variables. These are the **z-score and t-test statistic**. Under the hypothesis:

\[ H_0: \mu = \mu_x \]
\[ H_1: \mu \neq \mu_x \]

where \( \mu_x \) is the assumed population mean, the z statistic is:

\[ z = \frac{\bar{X} - \mu_x}{\sigma / \sqrt{n}} \]

where \( z \) is the test statistic, \( \bar{X} \) is the sample mean, \( \mu_x \) is the population mean defined in the hypothesis, \( \sigma \) is the population standard deviation, \( n \) is the sample size drawn for the test.

We write \( z \sim \mathcal{N}(0,1) \) where \( \sim \) (tilde) means 'is distributed as' and \( \mathcal{N}(0, 1) \) is normally distributed with a mean of zero and a standard deviation of one.

The z test statistic assumes knowledge of the population standard deviation, \( \sigma \), which means that unless this value is known or can be determined accurately from the sample, because \( n \) is large, the test cannot be performed. The z score statistic is therefore used when the population standard deviation is known or when the sample size is large.

The t-test refers to a statistical distribution known as the t distribution, and in hypothesis testing it is assumed that the test statistic follows a t distribution, rather than the assumption of a normal distribution when a z statistic is used. The t-test statistic is given as:

\[ t = \frac{\bar{X} - \mu_x}{S / \sqrt{n}} \]

where \( t \) is the test statistic, \( \bar{X}, \mu \) and \( n \) are as previously defined and \( S \) is the sample standard deviation calculated as:

\[ S = \sqrt{\frac{\sum (X_i - \bar{X})^2}{n-1}} \]

The t-test statistic requires no prior knowledge of the population standard deviation as the test uses the sample standard deviation instead. The population standard deviation is unlikely to be known in advance. Therefore, the analyst uses the sample standard deviation as a proxy. The t-test statistic is used when the sample size is small.

(Note: Other test statistics are possible in hypothesis testing. Two common distributions for test statistics are the Chi-square distribution and the F-distribution, but we will not be covering these in this Study Manual.)

### 2.2 Significance of the test and alpha

In hypothesis testing, we test for significance. In other words, we test whether the difference between the sample value and the population value appears to be significant. The upper level of ‘acceptable error’ must be chosen by the analyst. Typically, we test at the 95% level of significance or the 99% level of significance, but other levels of significance may be used. The 95% level of significance may also be expressed as a 5% level of significance, and the 99% level of significance may be expressed as a 1% level of significance.
Alpha

The level of significance is the upper level of acceptable error. This is called alpha (represented as $\alpha$). In most research, a value of 0.05 is generally accepted as the upper level of acceptable error, although other values can be valid. The $\alpha$ level of the test determines what is known as the critical value of the test.

Hypothesis testing works by comparing the critical value of the test as determined by the value of $\alpha$ selected by the analyst (and the number of observations over the sample) to that of the test statistic calculated for the test conducted on the sample. Needless to say, the lower the significance level selected by the analyst, the more stringent the test. That is, the smaller we define $\alpha$, the lower the probability of rejecting the null hypothesis when it is true.

Over time, we have come to understand quite well the properties of various sampling distributions. The accumulated knowledge of this understanding is often reflected in the tables at the back of most statistical texts. For given values of degrees of freedom (d.f. are the number of independent measurements) statisticians have established for different distributions, the critical values at which we may reject or not reject null hypotheses.

- If we use a z score statistic, we compare the z score from the sample with a critical z score value. This critical value depends on the level of significance chosen. If the z score value from the sample is larger than the critical x score, we reject the null hypothesis at the chosen level of significance.

- Similarly if we use a t-test statistic, we compare t-test score from the sample with a critical t-test value. If the t-test value from the sample is larger than the critical t-test score, we reject the null hypothesis at the chosen level of significance.

p-value

Instead of using z scores or t-test statistics, we can use p-values for hypothesis testing.

The p-value is an estimated measure of the degree to which the test result obtained from the sample reflects the truth about the population, i.e. how reliable the conclusion drawn from the sample is in inferring what is occurring within the population.

A p value is the probability of obtaining a sample result as extreme as the one that has been obtained (observed), assuming that the null hypothesis is true. For example, a p-value of .2 indicates that there is a 20 per cent probability that we would find a relationship within the sample that does not exist within the population.

The null hypothesis is rejected (and the alternative hypothesis is accepted) if the p-value is less than a statistically significant probability level, known as alpha ($\alpha$).

It is usual to use an alpha value of .05 or .01, which means that we are testing at the 95 per cent or 99 per cent level of confidence.

The upper level of acceptable error is the alpha. We can compare the p-value with the alpha. If the p-value is less than the alpha, the probability that there is no significant difference between the sample and the population is less than the selected level of significance. We can conclude therefore that the null hypothesis should be rejected at the chosen level of significance.

For example if we obtain a p-value of .032 and we are testing at the 95% level of confidence and an alpha of .05, we would reject the null hypothesis, because it is less than 0.05.

Needless to say, the lower the significance level selected by the analyst (the lower the value of $\alpha$) the more stringent the test. The lower we define $\alpha$, the lower the probability that we will reject the null hypothesis when it is true. To understand this concept, it is necessary to consider Type I and Type II errors in testing. We shall begin by defining Type I and Type II errors.
2.3 Type I and Type II errors

Section overview
- There are two types of error that may be made in significance testing.
  1. a null hypothesis that is true may be rejected
  2. a null hypothesis that is incorrect may not be rejected.

The first type of error is called a Type I error and the second type of error is called a Type II error.

Definitions

We define a **Type I error** as occurring when we reject the null hypothesis when it is true.

A **Type II error** is said to occur when we fail to reject the null hypothesis when it is false.

When we talk about Type I and Type II errors we are talking about making a mistake in terms of the inferences drawn from our statistical tests.

With Type I and Type II errors defined we may now go on to define the level of significance, alpha ($\alpha$) and beta ($\beta$).

Definitions

**Alpha** ($\alpha$) is the probability of making a Type I error.

**Beta** ($\beta$) is the probability of making a Type II error.

These two types of errors are defined in the table:

<table>
<thead>
<tr>
<th>True state of the null hypothesis</th>
<th>H₀ True</th>
<th>H₀ False</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reject H₀</td>
<td>Type I error</td>
<td>Correct</td>
</tr>
<tr>
<td>Do not reject H₀</td>
<td>Correct</td>
<td>Type II error</td>
</tr>
</tbody>
</table>

The probability of a Type I error is designated by the Greek letter alpha ($\alpha$) and is called the Type I error rate; the probability of a Type II error (the Type II error rate) is designated by the Greek letter beta ($\beta$). A Type II error is an error in the sense that an opportunity to reject the null hypothesis correctly was lost.

Failing to reject the null when the null is true is a correct decision. However, we are usually interested in trying to find true differences, and therefore look to reject null hypotheses. Rejecting the null when it is really not true is a correct decision as well. More specifically, the probability a test has to do this is referred to as **power**. Power may be defined as the probability of correctly rejecting a false null hypothesis. Some people also refer to power as precision or sensitivity.

Power is actually a function of Type II error, or, beta ($\beta$). More specifically, Power is equal to $1 - \beta$. How these values are calculated is shown in examples below.

When you calculate power, you need to identify two distributions: the distribution under the null, and the distribution under the alternative hypothesis. Remember, we are talking about power, which means we are talking about correctly rejecting the null. If we are rejecting the null, we are therefore assuming our sample comes from a population different from the one specified under the null.
Let us assume that we have a null hypothesis that is really true, but we reject it. In this case we have made a Type I error. Incorporating two distributions, we would draw this as follows. Notice the area shaded – it is the area beyond our criterion, or alpha. The area beyond alpha is the region (probability) for making a Type I error (as shaded below):

![Type I Error Diagram](image)

Let us now assume that we have a null hypothesis that is really false, but we fail to reject. Now we have made a Type II error. In this case the alternative is true, so that is the distribution we’re working with. Now, the area to the left of the original alpha under the alternative is our probability of error – Type II (light blue shading below). If we rejected the null as we should have, we would have made a correct decision. This would be defined as the probability to the right of the criterion under the alternative. More specifically, this probability is power (dark blue shading below).

![Type II Error Diagram](image)

There are four outcomes from hypothesis testing.

- We reject the null when in fact the alternative is true. This decision would be correct.
- We fail to reject the null when in fact the null is true. This decision would be correct.
- We reject the null when in fact the null is true. This decision would be incorrect. This type of error is called a Type I error.
- We fail to reject the null when in fact the alternative is true. This decision would be incorrect. This type of error is called a Type II error.

The level of significance, $\alpha$, is the probability of making a Type I error.

We refer to the probability of making a Type II error as $\beta$. One minus $\beta$ is known as the power of the test.

In establishing the level of significance for the test, $\alpha$, the analyst effectively determines some critical value for the hypothesis test which is compared to the test-statistic. That is, you the analyst are setting the probability that you may make an erroneous conclusion. As stated previously, usually we set $\alpha$ to equal 5% and in doing so we are resigning ourselves to the fact that there is a 5% chance that we will draw the wrong conclusion from the test as a result of bad luck. The temptation is therefore to use ever-smaller values of $\alpha$ so as to ensure that we will not make a Type I error.

While it is typical in academic settings to set $\alpha$ to equal 0.05 i.e. a 5% level of confidence, why not use $\alpha = 0.01$ instead and therefore have a 1% chance of making such a Type I error rather than a 5% chance? Or why not set $\alpha$ to 0.001 for an even more stringent test? While this is a perfectly legitimate response we need to realise that if we do this we greatly increase the probability of making a Type II error. That is, the more stringent the criteria you set to reduce the possibility of making a Type I error, the more likely we will make a Type II error.
This is shown in the diagram below.

The diagram shows that as we decrease the level of \( \alpha \), \( \beta \) increases. Given that the probability of an outcome is given as the area under the curve, decreasing the probability of a Type I error occurring must, for any specific alternative hypothesis, necessarily increase the probability of a Type II error.

The relationships between the null and alternative hypotheses and \( \alpha \) and \( \beta \) can be shown as:

<table>
<thead>
<tr>
<th>Decision</th>
<th>( H_0 ) true</th>
<th>( H_0 ) false</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reject ( H_0 )</td>
<td>Type I error</td>
<td>Correct decision made</td>
</tr>
<tr>
<td></td>
<td>Probability = ( \alpha )</td>
<td>Probability = 1 - ( \beta )</td>
</tr>
<tr>
<td>Do not reject ( H_0 )</td>
<td>Correct decision made</td>
<td>Type II error</td>
</tr>
<tr>
<td></td>
<td>Probability = 1 - ( \alpha )</td>
<td>Probability = ( \beta )</td>
</tr>
</tbody>
</table>

Once the significance level has been determined by the analyst the critical value for the test may be determined using Microsoft Excel. The formulas to calculate the critical values for normal and student t-test distribution are:

\[
\text{NORMINV}(\text{probability 1-}\alpha/2, \text{<mean>, <standard deviation>})
\]
(for two-tailed tests)
\[
\text{TINV}(\text{probability } \alpha, \text{<degrees of freedom>})
\]

For example:

\[
\text{NORMINV}(0.975,0,1) = 1.96
\]
\[
\text{TINV}(0.05,99) = 1.98
\]

### 2.3.1 Rejection region

The critical value of the test obtained is then used to create what is known as the rejection region(s) for the test. Two-tailed tests will have two rejection regions, one in each tail; one-tailed tests will have a single rejection region located in only one tail of the distribution. Test statistics that fall into the rejection or critical region(s) of the test support the rejection of the null hypothesis.

For one-tailed tests, the rejection region of the test is dependent upon the sign of the test statistic and the null hypothesis stated. Generally, the rejection region for positive critical values encompass values greater than the critical value. Negative critical values generally are associated with rejection regions less than the critical value (i.e. more negative in value – see diagram below).

Two-tailed tests relate to both tails of the distribution, hence values in the rejection region encompass the area of the distribution greater than the critical value and less than the negative of the critical value. The
The statistical hypothesis-testing method outlined in the previous example requires that the analyst choose the significance level $\alpha$ independent of the collection of any data and the computing of the test-statistic. The selection of $\alpha$ in effect fixes the rejection region of the test. The decision rule is fixed: reject the null hypothesis if the test statistic falls into the region of the distribution known as the rejection region, as dictated by the critical value for the test.

**Comparing the probability value to alpha**

<table>
<thead>
<tr>
<th>Test</th>
<th>Action taken by analyst</th>
</tr>
</thead>
<tbody>
<tr>
<td>The p-value is less than alpha</td>
<td>Reject the null hypothesis of the test</td>
</tr>
<tr>
<td>The p-value is greater than alpha</td>
<td>Do not reject the null hypothesis of the test</td>
</tr>
</tbody>
</table>

**2.3.2 Using p-values**

The fixed nature of the rejection region means that a hypothesis is rejected if the value of the test statistic relative to that of the critical value places the test statistic in the rejection region of the distribution for the test. Such decision rules allow for no margin for interpretation as to how significant a test statistic is.

For this reason, a growing number of researchers prefer to examine the probabilities i.e. the p-values, of the outcome of tests rather than use the cut-and-dried results dictated by the use of critical values. The use of p-values allows the analyst to determine the relative significance of the test. If the p-value is less than the level of $\alpha$ then the analyst rejects the null hypothesis of the test. If on the other hand the p-value exceeds the level of alpha then the analyst cannot reject the null hypothesis.

p-values for the more common distributions may be calculated in Microsoft Excel.

\[
\text{NORMDIST(<test statistic>, <mean>, <standard deviation>,1)}
\]

\[
\text{TDIST(<test statistic>,<degrees of freedom>, <number of tails>)}
\]

For example:

\[
\text{NORMDIST(-1.64,0,1,1) = 0.050503}
\]

\[
\text{TDIST(1.99,265,2) = 0.047618}
\]

In explaining the above examples, we will use a value of $\alpha$ equal to 0.05. The first example shows the p-value from a normal distribution test (z-score) in which the test-statistic is -1.64. The test is calculated for a standard normal (mean = 0, standard deviation = 1). The p-value calculated is 0.050503, which strictly speaking would not result in the rejection of the null hypothesis if the level of $\alpha$ was 0.05. The next example shows a two-tailed student-t-distribution test with a test-statistic of 1.99 and 265 degrees of freedom. The p-value of 0.047618 is less than an $\alpha$ of 0.05 (we do not need to divide $\alpha$ by 2: Excel automatically adjusts the p-value for this) suggesting that the test statistic is statistically significant. The above shows the arbitrary nature in selecting a value for $\alpha$. At $\alpha$ equal to 0.05 only the first example would result in the rejection of the null hypothesis. However, had we selected $\alpha$ equal to 0.06 both examples would result in the rejection of the null hypothesis.
Question 1: p-value

In hypothesis testing, which of the following statements are always true?

I. the p-value is greater than the significance level.
II. the p-value is computed from the significance level.
III. the p-value is the parameter in the null hypothesis.
IV. the p-value is a test statistic.
V. the p-value is a probability.

A. I only
B. II only
C. III only
D. IV only
E. V only

(The answer is at the end of the chapter)

2.4 Performing the test

Now that the stages of hypothesis testing have been outlined, all that remains is to perform the test and draw the appropriate conclusions. The example we will use to demonstrate is the t-test. This compares the mean of a sample with the known value of the population mean, or with a given hypothetical mean, and determines how likely it is that this difference is due to chance.

Values in the sample should be independent and normally distributed. The 'test statistic' t is calculated by:

\[ t = \frac{\overline{X} - \mu}{S/\sqrt{n}} \]

where \( \overline{X} \), \( \mu \) and \( n \) are as previously defined and \( S \) is the sample standard deviation calculated as:

\[ S = \sqrt{\frac{\sum (X - \overline{X})^2}{n-1}} \]

Example

A new product will be introduced to the market if it receives a mean of over six on a ten-point scale. A sample of 40 respondents was asked to rate a prototype of the product. The results indicated a mean rating of 6.9 with a standard deviation of 1.4. The researcher chose an alpha level of 0.05. Should the new product be introduced?

The analysis would proceed thus:

H0: \( \mu \leq 6 \)

H1: \( \mu > 6 \)

\[ t = \frac{\overline{X} - \mu}{S/\sqrt{n}} = \frac{6.9 - 6}{1.4/\sqrt{40}} = 4.065786 \]

The test is a one-tailed test because we are looking at whether the mean is more than six, not whether the mean differs above or below 6.

The critical value at which we reject the null hypothesis is calculated using Excel as shown below.

\[ TINV(0.1,39) \]

We use 0.1 instead of 0.05 as Excel calculates the test as a two-tailed test. As this is a one-tailed test, using 0.1 instead of 0.05 forces Excel to calculate the critical value at correct probability for the one-tailed test.

The critical value for this test is 1.6848, and hence \( t \) observed is greater than \( t \) critical, suggesting that the null hypothesis should be rejected and that the new product should be introduced to the market. To confirm this finding we can also calculate the p-value for the test using Excel. We show this below.

\[ TDIST(4.065786,39,1) \]
Excel calculates the p-value for the test as 0.00011242. This is much smaller than our $\alpha$ value of 0.05, and hence using the p-value for the test, we also would reject the null hypothesis for the test.

**Question 2: Probability**

Tommy claims that he blindly guessed on a 20 question true/false test, but then he got an 80 per cent. Using the binomial calculator, you find out that the probability of getting 16 or more correct out of 20 when $p = .5$ is .0059. This probability of .0059 is the probability that

A he would get an 80 per cent if he took the test again.
B he would get this score or better if he were just guessing.
C he was guessing blindly on the test.

(The answer is at the end of the chapter)

### 3 Analyse sample data

**Section overview**

- The analysis plan describes how to use sample data to accept or reject the null hypothesis. It should specify the significance level and the test method.

#### 3.1 Significance level

Often, researchers choose significance levels equal to 0.01, 0.05, or 0.10; but any value between 0 and 1 can be used.

Many researchers are excited when they have discovered a 'significant' finding, without really understanding what it means. When a statistic is significant, it simply means that you are very sure that the statistic is reliable. It does not mean the finding is important.

For example, suppose we give 1 000 people an IQ test, and we ask if there is a significant difference between male and female scores. The mean score for males is 98 and the mean score for females is 100. We use an independent groups t-test and find that the difference is significant at the .001 level. The big question is, 'so what?'. The difference between 98 and 100 on an IQ test is a very small difference. So small, in fact, that it is not even important.

Then why did the t-statistic come out significant? Because there was a large sample size. When you have a large sample size, very small differences will be detected as significant. This means that you are very sure that the difference is real – it didn’t happen by fluke. If we had only given the IQ test to 25 people instead of 1 000, the two-point difference between males and females would not have been significant.

Significance is a statistical term that tells how sure you are that a difference or relationship exists. To say that a significant difference or relationship exists only tells half the story. We might be very sure that a relationship exists, but is it a strong, moderate, or weak relationship? After finding a significant relationship, it is important to evaluate its strength. Significant relationships can be strong or weak. Significant differences can be large or small. It just depends on your sample size.

Much as we would like to think we make correct decisions when we either reject or fail to reject the null, there is no way of knowing for sure. However, it would only make sense that occasionally we will be wrong. Therefore, we can only focus on the probability of making such errors occurring - we will never know for sure if we have made an error or not. There are two forms of errors we can estimate in terms of their probabilities, both of which are readily calculable.

#### 3.1.1 Procedure used to test for significance

Whenever we perform a significance test, it involves comparing a test value that we have calculated to some critical value for the statistic. The critical value(s) for a hypothesis test is a threshold to which the value of the test statistic in a sample is compared to determine whether or not the null hypothesis is rejected.
The critical value for any hypothesis test depends on the significance level at which the test is carried out, and whether the test is one-sided or two-sided.

1. Decide on the critical alpha level you will use (i.e., the error rate you are willing to accept).
2. Conduct the research/Obtain the sample.
3. Calculate the statistic.
4. Compare the statistic to a critical value obtained from a table.

If your statistic is higher than the critical value from the table:
- Your finding is significant.
- You reject the null hypothesis.
- The probability is small that the difference or relationship happened by chance, and p-value is less than the critical alpha level (p-value < α).

If your statistic is lower than the critical value from the table:
- Your finding is not significant.
- You fail to reject the null hypothesis.
- The probability is high that the difference or relationship happened by chance, and p-value is greater than the critical alpha level (p-value > α).

3.2 One-tailed and two-tailed significance tests

One important concept in significance testing is whether you use a one-tailed or two-tailed test of significance. The answer is that it depends on your hypothesis. When your research hypothesis states the direction of the difference or relationship (more than, less than), then you use a one-tailed probability. For example, a one-tailed test would be used to test these null hypotheses:
- Females will not score significantly higher than males on an IQ test.
- Blue collar workers will not buy significantly more products than white collar workers.
- Superman is not significantly stronger than the average person.

In each case, the null hypothesis (indirectly) predicts the direction of the difference.

A two-tailed test would be used to test these null hypotheses:
- There will be no significant difference in IQ scores between males and females.
- There will be no significant difference in the amount of product purchased between blue collar and white collar workers.
- There is no significant difference in strength between Superman and the average person.

A two-tailed test is concerned with the probability of 'either more than or less than'.

The one-tailed probability is exactly half the value of the two-tailed probability.

3.3 The classical method of testing a hypothesis

If a claim is made regarding the population mean with the standard deviation (σ) known, we use the following steps to test the claim provided:

Step 1  The sample is obtained using simple random sampling.
Step 2  The population from which the sample is drawn is normally distributed or the sample size, n, is large. (A normal distribution may be assumed when the sample size is 30 or more as a general guide. Whether a normal distribution is appropriate depends not just on sample size, in fact, but also on the distribution and range of the sample data.)

Use the following steps to test the hypothesis:

Step 1  Hypothesis
Step 2  Critical value
Step 3  Test statistic
Step 4  Compare
Step 5  Conclusion
• A claim is made regarding the population mean. The claim is used to determine the null and alternative hypotheses.

• Select a level of significance $\alpha$ based upon the seriousness of making a Type I error. The level of significance is used to determine the critical value. The critical value represents the maximum number of standard deviations the sample mean can be from $\mu_0$ before the null hypothesis is rejected.

• Provided the population from which the sample is drawn is normally distributed or the sample size is large and the population standard deviation, $\sigma$, is known, the distribution of the sample mean $\bar{X}$ is normal with mean $\mu_0$ and standard deviation $\frac{\sigma}{\sqrt{n}}$.

This standard deviation is known as the standard error.

Therefore, $z = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}}$ represents the number of standard deviations the sample mean is from the assumed mean $\mu_0$. $z$ is called the test statistic.

• Compare the value of the test statistic to that of the critical value to make a decision regarding the null hypothesis.

<table>
<thead>
<tr>
<th>Two-tailed</th>
<th>Left-tailed</th>
<th>Right-tailed</th>
</tr>
</thead>
<tbody>
<tr>
<td>If $z &lt; -z_{\alpha/2}$ or $z &gt; z_{\alpha/2}$ reject the null hypothesis</td>
<td>if $z &lt; -z_{\alpha}$ reject the null hypothesis</td>
<td>if $z &gt; z_{\alpha}$ reject the null hypothesis</td>
</tr>
</tbody>
</table>

$z_{\alpha} = \text{NORMINV}(\alpha, \mu, \sigma) - \mu \over \sigma/\sqrt{n}$

### 3.4 Interpret test statistics with p-values

There is another way to interpret the test statistic. In hypothesis testing, we make a yes or no decision without discussing borderline cases. For example with $\alpha = .05$, a two tailed test will indicate rejection of $H_0$ for a test statistic of $z = 2$ or for $z = 6$, but $z = 6$ is much stronger evidence than $z = 2$. To show this difference we write the p-value which is the lowest significance level such that we will still reject $H_0$. For a two-tailed test, we use twice the table value to find $p$, and for a one tailed test, we use the table value.

**Worked Example: p-value**

We want to test the hypothesis with a significance level of .05 that the climate has changed since industrialisation. Suppose that the mean temperature throughout history is 9.5°C. During the last 40 years, the mean temperature has been 10.5 degrees and suppose the population standard deviation is two degrees. What can we conclude?

**Solution**

We are testing whether there has been a significant increase or decrease in the mean temperature, and so a two-tailed test will be used.

We have

$H_0: \mu = 10^\circ$

$H_1: \mu \neq 10^\circ$

We compute the $z$ score:

$z = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}}$

$z = \frac{10.5 - 9.5}{2/\sqrt{40}}$

$z = 1.316 = 3.16$

$= 3.16$
Normal distribution tables give us .992 when z = 3.16, so that
\[ p = \frac{(1 - .992)}{2} = .004 \]
We divide by 2 because we have a two-tailed test.
Since .004 < 0.05 we reject the null hypothesis and conclude that there has been a change in temperature.

3.5 Decision rules

The analysis plan includes decision rules for accepting or rejecting the null hypothesis. In practice, statisticians describe these decision rules in two ways – with reference to a p-value or with reference to a region of acceptance.

1. **p-value.** The strength of evidence in support of a null hypothesis is measured by the p-value. Suppose the test statistic is equal to T. The p-value is the probability of observing a test statistic as extreme as T, assuming the null hypothesis is true. If the p-value is less than the significance level, we reject the null hypothesis.

2. **Region of acceptance.** The region of acceptance is a range of values. If the test statistic falls within the region of acceptance, the null hypothesis is accepted. The region of acceptance is defined so that the chance of making a Type I error is equal to the significance level.

The set of values outside the region of acceptance is called the region of rejection. If the test statistic falls within the region of rejection, the null hypothesis is rejected. In such cases, we say that the hypothesis has been rejected at the \( \alpha \) level of significance.

These approaches are equivalent. Some statistics texts use the p-value approach; others use the region of acceptance approach.

**Question 3: Hypothesis testing for population mean**

50 smokers were questioned about the number of hours they sleep each night. We want to test the hypothesis that the smokers need less sleep than the general public which needs an average of 7.7 hours of sleep.

Compute a rejection region for a significance level of .05.

If the sample mean is 7.5 and the population standard deviation is 0.5, what can you conclude?

Note: At the .05 level of significance, the critical z score value is 1.645 for a one-tailed test and 1.96 for a two-tailed test.

(The answer is at the end of the chapter)

4 Test methods

**Section overview**

- Typically, the test method involves a test statistic and a sampling distribution. Computed from sample data, the test statistic might be a mean score, proportion, difference between means, difference between proportions, z-score, t-score, chi-square, and so on. Given a test statistic and its sampling distribution, a researcher can assess probabilities associated with the test statistic. If the test statistic probability is less than the significance level, the null hypothesis is rejected.

**4.1 Which test?**

Why another statistical test?

As a general guide, the normal distribution and the z test are most suitable when:

- the sample size is over 30 (although the distribution and range of values in the sample should also affect whether a normal distribution may be assumed), or
- the sample size is under 30 and the population standard deviation is known beforehand.
Under other conditions, such as when the sample size is under 30 and the population standard deviation is not known beforehand, another technique must be used.

**Statistical test (t-test for mean)**

The t-test is a statistical test for the mean of a population. It should be used whenever all of the following criteria apply:

- the population standard deviation $\sigma$ is unknown
- the population is known to be normally distributed.

Regardless of whether you use $z$ or $t$, you will need to have the following information for the hypothesis test:

- the null hypothesis
- the alternative hypothesis (this will tell you if you are using a one-sided or two-sided test)
- the mean of the sample data
- the standard deviation of the population ($z$-test) or the sample ($t$-test)
- the number in your sample ($n$)

**Calculation of a z-score or t-score**

$$Z\text{-score or t-score} = \frac{\text{Range}}{\text{Standard deviation}}$$

If testing new data: $\text{Range} = y - \mu$

$$\text{Standard deviation} = \sigma = \sqrt{\frac{\sum(x - \bar{x})^2}{N}}$$

If testing proportions: $\text{Range} = P_1 - P_2$

$$\text{Standard deviation} = \frac{\sqrt{NP_1(1-P_1)}}{N}$$

**Testing one sample (section 4.5)**

$$Z\text{-score or t-score} = \frac{\text{Range}}{\text{Standard deviation}}$$

If $n$ is large (say, $>30$), use $z$-score. $\text{Range} = \bar{x} - \mu$

- If population standard deviation $\sigma$ is known, standard deviation of distribution $= \frac{\sigma}{\sqrt{n}}$
- If population standard deviation $\sigma$ is not known, standard deviation of distribution $= \frac{s}{\sqrt{n}}$, where $s = \frac{\sum(x - \bar{x})^2}{n-1}$

If $n$ is small (say, $\leq 30$), use $t$-score.

**Testing two samples (section 4.6)**

$$Z\text{-score or t-score} = \frac{\text{Range}}{\text{Standard deviation}}$$

$\text{Range} = \bar{x}_1 - \bar{x}_2$

$$\text{Standard deviation} = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}} \times \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$
4.2 The z-score

We can conduct a hypothesis test of a proportion, when the following conditions are met:

- The sampling method is simple random sampling.
- Each sample point can result in just two possible outcomes. We call one of these outcomes a success and the other, a failure.
- The sample includes at least five successes and five failures

The easiest and most frequent thing we do is find probabilities of events less extreme or more extreme than an event. We do this by using the z-score. Using an example of IQ tests that have means of 100 and standard deviations of 15, let us say you take an IQ test and get a score of 125. Are you really smart or just mediocre? To find your relative standing we convert your score to a z-score using the formula below.

**Formula to learn**

The z-score is the number of standard deviations you are from the mean of 0 (recall that by subtracting the mean and dividing the result by the standard deviation you convert your data to the standard normal distribution which has a mean of 0 and standard deviation of 1).

\[
 z = \frac{x - \mu}{\sigma} 
\]

\[
 z = \frac{125 - 100}{15} = 1.6667 
\]

So you are 1.6667 standard deviations above the mean. We look up 1.6667 in a table of normal values, use the Excel function =NORMSDIST(1.6667) or use the online calculator and we get .952213. That means you are in the 95th percentile or you have a higher IQ score than about 95 per cent of the population.

4.3 Z-test for the mean

Statistics is all about understanding the role of chance in our measurements and we often want to know what the chances are of obtaining a sample mean given the population mean is a certain value.

The z-score indicates the number of standard deviation units of the sample from the population mean.

Note that the z-test is not the same as the z-score, although they are closely related.

The standard error (SE) of the mean tells us how much the sample mean varies from sample to sample (it is the standard deviation of the sample mean given a particular sample size (n)).

\[
 SE = \frac{\sigma}{\sqrt{n}} \quad \text{where } \sigma \text{ is the population standard deviation.} 
\]

The empirical rule tells us that 95 per cent of the time the sample mean will fall within two standard errors of the population mean. We can extend the principle of the empirical rule and use the normal curve to find the probabilities for a given sample mean using a statistical test called the 1-sample z-test.
**Formula to learn**

The z measure is calculated as: \( z = \frac{(x - \mu)}{SE} \)

\[
Z = \frac{(\bar{x} - \mu)}{\sigma} \div \sqrt{n}
\]

The Z-test compares sample and population means to determine if there is a significant difference. It requires a simple random sample from a population with a normal distribution and where the mean is known.

The z measure is calculated as: \( z = \frac{(x - \mu)}{SE} \)

where \( x \) is the mean sample to be standardised, \( \mu \) (mu) is the population mean and SE is the standard error of the mean.

The z measure is then calculated or looked up in a table to find the probability \( x \) being that far from the mean by chance. A negative z value means it is below the population mean (the sign is ignored in the lookup table).

**Worked Example: The test statistic**

Suppose we want to show that American children have an average higher cholesterol level than the national average. It is known that the mean cholesterol level for all Americans is 190. Construct the relevant hypothesis test.

**Solution**

\( H_0: \mu = 190 \)

\( H_1: \mu > 190 \)

We test 100 children and find that \( \bar{x} = 198 \) and suppose we know the population standard deviation \( \sigma = 15 \).

Do we have evidence to suggest that children have an average higher cholesterol level than the national average? We have

\[
z = \frac{198 - 190}{15 / \sqrt{100}} = 5.33
\]

Since \( z \) is so high, the probability that \( H_0 \) is true is so small that we decide to reject \( H_0 \) and accept \( H_1 \).

Therefore, we can conclude that children have a higher average cholesterol level than the national average.

Suppose that \( \alpha = .05 \). We can draw the appropriate picture and find the z score for -.025 and .025. We call the outside regions the rejection regions.

We call the shaded areas the rejection region since if the value of z falls in these regions, we can say that the null hypothesis is very unlikely so we can reject the null hypothesis.
4.4 Z-test for population mean

Consider the following exercise: the policy of a particular bank branch is that its ATMs must be stocked with enough cash to satisfy customers making withdrawals over an entire weekend. At this branch the expected (i.e. population) average amount of money withdrawn from ATM machines per customer transaction over the weekend is $160 with an expected (i.e. population) standard deviation of $30. Suppose that a random sample of 36 customer transactions is examined and it is observed that the sample mean withdrawal is $172. Note that the standard deviation that we are to use for this problem did not come from the sample. Therefore, this will be a z-test, not a t-test. For this problem, we have

\[ \sigma = 30, \quad n = 36, \quad \bar{x} = 172. \]

(a) State the null and alternative hypothesis.

This is not very clear, but apparently we are to check if the average exceeds $160, which would mean the ATMs are not ‘stocked with enough cash.’ This is what we will try to show, so it should go in the alternative hypothesis. Therefore the hypotheses would be:

\[ H_0: \mu \leq 160 \]
\[ H_1: \mu > 160 \]

(b) At the .05 level of significance, using the critical value approach to hypothesis testing, is there enough evidence to believe that the true average withdrawal is greater than $160?

Let us get the test statistic:

\[ z = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}} = \frac{172 - 160}{30 / \sqrt{36}} = 2.4 \]

Set up the rejection region by drawing a z-curve and shade the last 5% of the right tail (Use the function NORMINV(.95,1,0). We need the z critical value associated with this area, which is \[ z_{0.05} = 1.645 \].

The test statistic falls into the rejection region, i.e., 2.4 > 1.645, therefore we reject \( H_0 \). Yes, there is enough evidence that the average is more than $160.

(c) At the .05 level of significance, using the p-value approach to hypothesis testing, is there enough evidence to believe that the true average withdrawal is greater than $160?

Now you should draw another z-curve, this time shading the area to the right of the test statistic, 2.4. Alternatively, you could use the same curve you drew in (b) and shade the new area with a different coloured pen. This would make it clear that the area we wish to obtain will be less than \( \alpha = 0.05 \). The shaded area is of course the p-value, and we can get it with Excel expression 1–NORMSDIST(2.4) = .0082.

Now the probability we need is p-value  = P(z > 2.4) = 0.0081975

This p-value is smaller than \( \alpha \), thus we reject \( H_0 \).
(d) Interpret the meaning of the p-value in this problem.

The probability of obtaining a sample whose mean is $172 or more when $H_0$ is true is .0082.

(e) Compare your conclusions in (b) and (c).

The conclusions are the same, of course. A 'large' test statistic, one that is larger than the critical value, is associated with a 'small' p-value, one that is smaller than alpha.

And the decision rule is:

Reject $H_0$ if the test statistic falls in the rejection region [part (b)], or

Reject $H_0$ if the p-value is less than alpha [part (c)].

4.5 One population hypothesis tests

4.5.1 One population hypothesis test for a mean

For one sample of 25 library visitors the time spent in Melbourne Central Library had mean of 41.64 and standard deviation of 40.13 minutes.

National survey finds large-city library users spend mean of 38 minutes.

Is population mean for Melbourne Central Library users different from national mean?

1 Assumptions

- Random sample
- Population normally distributed

2 Hypotheses

- Null hypothesis $H_0$: $\mu = 38$
- Alternative hypothesis $H_1$: $\mu \neq 38$

3 Level of significance

- Probability of error in making decision to reject null hypothesis
- Choose $\alpha = 0.05$

4 Test statistic

$$t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}} = \frac{41.64 - 38}{40.13/\sqrt{25}} = 0.454$$

5 p-value

- Probability of test statistic this extreme or greater
- Obtained from table of $t$ distribution for $n - 1 = 24$ degrees of freedom
- $t = 0.454 < t = 1.318$ for 24 df, one-tail probability of 0.100
- Double probability for two-tailed test
- p-value > 0.200

6 Conclusion

- p-value > 0.200 implies that if null hypothesis true, it is quite likely that sample value could have been obtained by chance
- Since p-value is greater than significance level $\alpha = 0.05$, cannot reject null hypothesis; cannot conclude that population mean is different from national mean of 38 minutes

4.5.2 One-tailed test to illustrate calculation of p-value

- Same sample of 25 for time spent in library as before.
- Now research hypothesis is that population mean is greater than 27 minutes.
- Null hypothesis is that population mean is less than or equal to 27 minutes.
- One-tailed test.
1 **Hypotheses**

Null hypothesis  \( H_0: \mu \leq 27 \)
Alternative hypothesis  \( H_1: \mu > 27 \)

2 **Test statistic**

\[
\frac{\overline{x} - \mu_0}{s/\sqrt{n}} = \frac{41.64 - 27}{40.13/\sqrt{25}} = 1.824
\]

3 **p-value**

Test statistic  \( t = 1.824 \)

\( n = 25 \) so 24 df

\( t = 1.824 \) falls between 1.7110 and 2.0640

So p-value is between 0.025 and 0.050

4 **Conclusion**

p-value between 0.025 and 0.050 is small, so probability of getting sample mean of 41.64 or more by chance if null hypothesis is true is small.

Since p-value is less than significance level  \( \alpha = 0.05 \), reject null hypothesis; conclude that population mean is greater than 27 minutes.

4.5.3 **One sample test for proportion**

- National poll interviewing 1 100 adults is conducted
- 42 per cent of respondents say they intend to vote for John Smith
- Consensus poll results previous month put figure at 39 per cent

Can we conclude that level of support has increased, that population proportion is greater than 39 per cent?

1 **Assumptions**

- Random sample
- Large sample
- Sample size of at least 20
- At least five in each category would be preferable

2 **Hypotheses**

- Null hypothesis  \( H_0: \pi = \pi_0 = 0.39 \)
- Alternative hypothesis  \( H_1: \pi > \pi_0 = 0.39 \)

3 **Level of significance**

- Probability of error in making decision to reject null hypothesis
- For this test choose  \( \alpha = 0.05 \)

4 **Test statistic**

\[
z = \frac{\pi - \pi_0}{\sqrt{\pi_0(1-\pi_0)/n}} = \frac{0.42-0.39}{\sqrt{0.39(1-0.39)/1100}} = 2.040
\]

1 Need p-value
2 Probability of test statistic this extreme or greater
3 For  \( z = 2.040 \), obtain probability in tail of 0.0207
4 Since this is one-tailed test, p-value is therefore 0.0207
Conclusion

- p-value = 0.0207 implies that if null hypothesis true, it is very unlikely that sample value this
great or greater could have been obtained by chance.
- Since p-value is less than significance level $\alpha = 0.05$, reject null hypothesis; conclude that
population proportion is greater than last month's proportion of 0.39.

4.5.4 One-sample test if population s.d. is known

A one-sample z-test can be used to assess whether a sample drawn at random from a population tends to
have the same characteristics as the population from which it is taken.

Here is how to use the test.

Define hypotheses. The table below shows three sets of null and alternative hypotheses. Each makes a
statement about how the true population mean $\mu$ is related to some hypothesised value $M$. (In the table, the
symbol $\neq$ means 'is not equal to'.)

<table>
<thead>
<tr>
<th>Set</th>
<th>Null hypothesis</th>
<th>Alternative hypothesis</th>
<th>Number of tails</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>$= M$</td>
<td>$\neq M$</td>
<td>2</td>
</tr>
<tr>
<td>2</td>
<td>$\geq M$</td>
<td>$&lt; M$</td>
<td>1</td>
</tr>
<tr>
<td>3</td>
<td>$\leq M$</td>
<td>$&gt; M$</td>
<td>1</td>
</tr>
</tbody>
</table>

1 Specify significance level. Often, researchers choose significance levels equal to 0.01, 0.05, or 0.10;
but any value between 0 and 1 can be used.

2 Compute test statistic. The test statistic is a z-score ($z$) defined by the following equation:

$$z = \frac{(x - M)}{\sigma}$$

where $x$ is the observed sample mean, $M$ is the hypothesised population mean (from the null
hypothesis), and $\sigma$ is the standard deviation of the population.

3 Compute p-value. The p-value is the probability of observing a sample statistic as extreme as the test
statistic. Since the test statistic is a z-score, use the normal distribution calculator to assess the
probability associated with the z-score.

4 Evaluate null hypothesis. The evaluation involves comparing the p-value to the significance level, and
rejecting the null hypothesis when the p-value is less than the significance level.

The one-sample z-test can be used when the population is normally distributed, and the population
variance is known.

4.6 Two population hypothesis tests

There is a lot of discussion on whether or not it is ever appropriate to use a one-tailed test. The rationale
is that if you already know the direction of the difference, why bother doing any statistical tests. While it is
generally safest to use a two-tailed test, there are situations where a one-tailed test seems more
appropriate. The bottom line is that it is the choice of the researcher whether to use one-tailed or two-
tailed research questions.

4.6.1 Two-tail testing of two sample means from independent populations

Populations are independent when a sample selected from one is not related to a sample selected from the
other. Examples of independent populations include production time using two different assembly
procedures and industrial accidents at two plants.

These tests assume the populations are approximately normal with equal variances.
These equal variances make a weighted (pooled) point estimate the best estimate of the population σ²

\[ s^2_w = \frac{(n_1 - 1)s^2_1 + (n_2 - 1)s^2_2}{n_1 + n_2 - 2} \]

\( s^2_1 \) is the variance of sample 1 and \( s^2_2 \) is the variance of sample 2.

**Worked Example: Two tail-testing**

Freddy Murphy wants to compare the time sales people spend with customers at two of his stores. A sample of 6 sales people from one store had a mean of 4.5 minutes and a variance of 3. A sample of 5 from a second store had a mean of 5.1 and a variance of 3.1. Freddy will conduct a .05 level test to determine whether the means are the same for these normally distributed populations.

**Solution**

\[ \bar{x}_1 = 4.5, \quad s^2_1 = 3.0, \quad \bar{x}_2 = 5.1, \quad s^2_2 = 3.1 \]

\( n_1 = 6 \) and \( n_2 = 5 \)

\( s^2_w \) is the weighted or pooled estimate of the population variance.

\( df = n_1 + n_2 - 2 = 6 + 5 - 2 = 9 \)

**The five-step approach to hypothesis testing:**

**Step 1** The null and alternative hypotheses:

\[ H_0: \mu_1 = \mu_2 \]

\[ H_{1, \neq} \]

**Step 2** The level of significance is 0.05 for this two-tail test.

**Step 3** The relevant test statistic is.

\[ t = \frac{\bar{x}_1 - \bar{x}_2}{s^2_w \left( \frac{1}{n_1} + \frac{1}{n_2} \right)} \]

**Step 4** Reject the null hypothesis if the test statistic is beyond the critical value.

**Step 5** Apply the decision rule.

\[ s^2_w = \frac{(6 - 1)3.0 + (5 - 1)3.1}{6 + 5 - 2} = 3.0 \]

\[ t = \frac{4.5 - 5.1}{\sqrt{3.0 \left( \frac{1}{6} + \frac{1}{5} \right)}} = -0.57 \]

\( df = n_1 + n_2 - 2 = 6 + 5 - 2 = 9 \)

\( t = \pm 2.262 \)

\( H_0 \) is not rejected because -0.57 is not beyond -2.262. Average customer waiting time at these stores is the same.

**4.6.2 Two sample test of proportions**

To compare proportions (or probabilities, \( p_1 \) and \( p_2 \)) found in samples (of size \( n_1 \) and \( n_2 \)) from two independent populations,

\[ z = \frac{p_1 - p_2}{\sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}} \]

Set \( p_1 \) to the larger probability. The test is two tailed.
Worked Example: Two sample proportions

In town A in a poll of 200 people, 55 per cent said they would vote for the United Separatists, but in a poll of 100 people in town B this dropped to 50 per cent.

Is there any difference between the two towns at the 5 per cent level?

\[ z = \frac{0.55 - 0.50}{\sqrt{\frac{0.55 \times 0.45}{200} + \frac{0.50 \times 0.50}{100}}} = 0.818 \]

giving a probability of 0.206 that this difference or more occurred by chance.

i.e. There is no difference at the 5 per cent level.

The sample size is large enough for this result to be compared with the Two Sample Means test.

The mean (for each person) is just the percentage, and the variance is \( p(p-1) \), so

\[ t = \frac{0.55 - 0.50}{\sqrt{\frac{(200 - 1) \times 0.55 \times 0.45 + (100 - 1) \times 0.50 \times 0.50}{200 + 100 - 2} \times \left( \frac{1}{200} + \frac{1}{100} \right)}} = 0.819 \]

And the number of degrees of freedom is large enough for the \( t \) and \( Z \) distributions to be the same.
Key chapter points

- The steps in hypothesis testing include formulating null and alternative hypotheses, analysing sample data, constructing a test statistic, and interpreting results.
- A statistical hypothesis ALWAYS includes a parameter (e.g., \( H_0: \mu=500 \)), and never a statistic (\( \bar{x} \)).
- The null hypothesis ALWAYS includes an equals sign (=).
- In a hypothesis test, we ALWAYS begin the test assuming the null hypothesis is true.
- The claim we usually seek evidence for is the alternative hypothesis.
- The alternative hypothesis includes <, >, or \( \neq \). Look for key phrases in the claim. For example, 'more than' means >; 'different from' means \( \neq \); and 'less than' means <.
- The alternative hypothesis can be one-tailed or two-tailed.
- If you are unsure whether to use a one-tailed or two-tailed hypothesis test, ALWAYS use a two-tailed test.
- \( \alpha \) = significance level = probability of a Type I error (reject a true \( H_0 \)).
- \( \beta \) = probability of a Type II error (do not reject a false \( H_0 \)).
- When testing a statistical hypothesis, there is always a possibility that your conclusion will be wrong (Type I error or Type II error). And, to make matters worse, you won’t know whether you are wrong or not.
- You can never say the null hypothesis is TRUE unless you have access to all the population data (and that never occurs). Rather, we say we do not reject the null hypothesis.
- A significant result occurs when you reject the null hypothesis.
- p-value is the probability that the test statistic takes a value equal to or more extreme than the value actually observed (in both directions for a two-tail test), assuming \( H_0 \) is true.
- A 'large' p-value is evidence for \( H_0 \) and a 'small' p-value is evidence against \( H_0 \).
- An extremely small p-value (<0.01) means that \( H_0 \) is strongly rejected or the result is highly statistically significant.
Quick revision questions

1. The diameter of a spindle in a small motor is supposed to be 5 mm. If the spindle is either too small or too large, the motor will not work properly. The manufacturer measures the diameter in a sample of motors to determine whether the mean diameter has moved away from the target. What are the null and alternative hypotheses?
   A. $H_0: \mu = 5$ and $H_1: \mu > 5$
   B. $H_0: \mu = 5$ and $H_1: \mu < 5$
   C. $H_0: \mu = 5$ and $H_1: \mu \neq 5$
   D. $H_0: \mu < 5$ and $H_1: \mu > 5$

2. A study claims that the proportion of adults in the Philippines with rudimentary literary skills is 21 per cent. A researcher believes the true percentage differs from this one that is published. Which of the following demonstrates a Type I error?
   A. Not rejecting $H_0: \pi = 0.21$ when actually $\pi = 0.21$.
   B. Not rejecting $H_0: \pi = 0.21$ when actually $\pi \neq 0.21$.
   C. Rejecting $H_0: \pi = 0.21$ when actually $\pi = 0.21$.
   D. Rejecting $H_0: \pi = 0.21$ when actually $\pi \neq 0.21$.

3. The examinations in a large accounting class are scaled after grading so that the mean score is 50. The professor thinks that one teaching assistant is a poor teacher and suspects that his students have a lower mean score than the class as a whole. The TA’s students this semester can be considered a sample from the population of all students in the course, so the professor compares their mean score with 50. Which of the following demonstrates a Type II error?
   A. not rejecting $H_0: \mu = 50$ when actually $\mu < 50$.
   B. not rejecting $H_0: \mu = 50$ when actually $\mu > 50$.
   C. rejecting $H_0: \mu = 50$ when actually $\mu < 50$.
   D. rejecting $H_0: \mu = 50$ when actually $\mu > 50$.

4. The level of significance represents which of the following?
   A. the probability of rejecting $H_0$
   B. the probability of rejecting $H_1$
   C. the probability of making a Type I error
   D. the probability of making a Type II error

5. The mean weight of a sample of Australian Postal Service shipments is 0.40 kg. It is believed that the true mean is greater. Suppose that the sample evidence indicated that the null hypothesis is rejected. Which of the following would be a correct conclusion?
   A. There is sufficient evidence to support the claim that the mean weight of all Australian Postal Service shipments is greater than 0.40 kg.
   B. There is sufficient evidence to support the claim that the mean weight of all Australian Postal Service shipments is less than 0.40 kg.
   C. There is not sufficient evidence to support the claim that the mean weight of all Australian Postal Service shipments is greater than 0.40 kg.
   D. There is sufficient evidence to support the claim that the mean weight of all Australian Postal Service shipments is 0.40 kg.
6 In order to test hypotheses regarding the population mean, which of the following requirements must be satisfied?
A the population from which the sample is drawn is uniformly distributed.
B a simple random sample is obtained.
C the population standard deviation is estimated.
D the sample size is greater than 20.

7 Briefly explain what is meant by the region of acceptance.
Answers to quick revision questions

1. C  $H_0: \mu = 5$ and $H_1: \mu \neq 5$

2. C  Rejecting $H_0: \pi = 0.21$ when actually $\pi = 0.21$

3. A  Not rejecting $H_0: \mu = 50$ when actually $\mu < 50$.

4. C  The probability of making a Type I error.

5. A  There is sufficient evidence to support the claim that the mean weight of all Australian Postal Service shipments is greater than 0.40 kg.

6. B  A simple random sample is obtained.

7. The region of acceptance is a range of values. If the test statistic falls within the region of acceptance, the null hypothesis is not rejected. The region of acceptance is defined so that the chance of making a Type I error is equal to the significance level.

   The set of values outside the region of acceptance is called the region of rejection. If the test statistic falls within the region of rejection, the null hypothesis is rejected. In such cases, we say that the hypothesis has been rejected at the stated level of significance.
Answers to chapter questions

1 The correct answer is E. The p-value is the probability of observing a sample statistic as extreme as the test statistic. It can be greater than the significance level, but it can also be smaller than the significance level. It is not computed from the significance level, it is not the parameter in the null hypothesis, and it is not a test statistic.

2 B If Tommy were guessing blindly, the probability that he would have answered 16 out of the 20 questions correctly is .0059. This is not the probability that he was guessing blindly. Remember, the probability value is the probability of an outcome given the hypothesis. It is not the probability of the hypothesis given the outcome.

3 First, we write down the null and alternative hypotheses

\[ H_0: \mu = 7.7 \]
\[ H_1: \mu < 7.7 \]

This is a left-tailed test. The z-score that corresponds to 0.05 is -1.645. The critical region is the area that lies to the left of -1.645. If the z-score is less than -1.645 then we will reject the null hypothesis and accept the alternative hypothesis. If it is greater than -1.645, we will fail to reject the null hypothesis and say that the test was not statistically significant.

We have

\[ z = \frac{7.5 - 7.7}{0.5 / \sqrt{50}} = -2.83 \]

Since -2.83 is to the left of -1.645, it is in the critical region. Therefore, we reject the null hypothesis and accept the alternative hypothesis. We can conclude that smokers need less sleep.
Chapter 16

Linear regression and correlation

<table>
<thead>
<tr>
<th>Learning objectives</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Simple linear regression and correlation</td>
<td>LO16</td>
</tr>
<tr>
<td>Calculate the equation of a simple regression line from a sample of data</td>
<td>LO16.1</td>
</tr>
<tr>
<td>Explain and interpret the slope and intercept of the equation</td>
<td>LO16.2</td>
</tr>
<tr>
<td>Calculate and interpret estimated values of y using the regression line</td>
<td>LO16.3</td>
</tr>
<tr>
<td>Calculate and interpret the coefficient of correlation</td>
<td>LO16.4</td>
</tr>
<tr>
<td>Calculate and interpret the coefficient of determination</td>
<td>LO16.5</td>
</tr>
</tbody>
</table>

Topic list

1 Regression and correlation
2 Correlation
3 Regression
4 Finding the best-fitting line
The overall basic aim of correlation and regression is to ascertain the extent to which one variable is related to another. Scatter diagrams were introduced in Chapter 12. We are now going to look at how the inter-relationship shown between variables in a scatter diagram can be described and calculated. The first two sections deal with correlation, which is concerned with assessing the strength of the relationship between two variables.

We will then see how, if we assume that there is a linear relationship between two variables (such as selling costs and sales volume) we can determine the equation of a straight line to represent the relationship between the variables and use that equation to make forecasts or predictions.

**Correlation**
- Scattergraphs show the correlation between two related variables
- Degrees of correlation include:
  - perfectly correlated
  - partly correlated
  - uncorrelated
- Correlation coefficient measures degree of correlation between -1 and +1
- Coefficient of determination is the square of the correlation coefficient

**Lines of best fit**
- On the graph of the linear equation:
  - the intercept is the point at which the straight line crosses the y axis
  - the slope is the gradient of the straight line
- The equation of a straight line is $y = a + bx$
- The least squares method of linear regression analysis refers to the square of the differences between actual and predicted values
Before you begin

If you have studied these topics before, you may wonder whether you need to study this chapter in full. If this is the case, please attempt the questions below, which cover some of the key subjects in the area.

If you answer all these questions successfully, you probably have a reasonably detailed knowledge of the subject matter, but you should still skim through the chapter to ensure that you are familiar with everything covered.

There are references in brackets indicating where in the chapter you can find the information, and you will also find a commentary at the back of the Study Manual.

1. Give some examples of variables which might be correlated. (Section 1.1)

2. The value of the correlation coefficient between \( x \) and \( y \) is 0.9. Which of the following is correct? (Section 2.3)
   - A. there is a weak relationship between \( x \) and \( y \).
   - B. \( x \) is 90 per cent of \( y \).
   - C. if the values of \( x \) and \( y \) were plotted on a graph, the line relating them would have a slope of 0.9.
   - D. there is a very strong relationship between \( x \) and \( y \).

3. If the correlation coefficient of a set of data is 0.95, what is the coefficient of determination? (Section 2.4.1)

4. Explain the equation of a straight line. (Section 3.1)

5. In calculating the regression equation linking two variables, the standard formulae for the regression coefficients are given in terms of \( X \) and \( Y \). Which of the following is true? (Section 4.3)
   - A. \( X \) must be the variable which will be forecast.
   - B. It does not matter which variable is which.
   - C. \( Y \) must be the dependent variable.
   - D. \( Y \) must be the variable shown on the vertical axis of a scattergraph.
1 Regression and correlation

1.1 Introduction

In this chapter we are concerned with relationships between variables. For example:

(i) How do the sales of a product depend on the price charged?
(ii) How does the strength of a material depend on temperature?
(iii) To what extent is metal pitting related to pollution?
(iv) How strong is the link between inflation and employment rates?
(v) How can we use the amount of fertiliser used to predict crop yields?

These are essentially two types of problem:

(a) Correlation problems which involve measuring the strength of a relationship.
(b) Regression problems which are concerned with the form or nature of a relationship.

Definitions

Correlation measures the degree of relatedness of two variables. Correlation can be positive, negative or zero. When the value of one variable is related to the value of another, they are said to be correlated.

Simple regression is used to examine the relationship between one dependent and one independent variable.

After performing an analysis, the regression statistics can be used to predict the dependent variable when the independent variable is known. Regression goes beyond correlation by adding prediction capabilities.

People use regression on an intuitive level every day. In business, a well-dressed man is thought to be financially successful. A mother knows that more sugar in her children’s diet results in higher energy levels. The ease of waking up in the morning often depends on how late you went to bed the night before. Quantitative regression adds precision by developing a mathematical formula that can be used for predictive purposes.

For example, a medical researcher might want to use body weight (independent variable) to predict the most appropriate dose for a new drug (dependent variable). The purpose of running the regression is to find a formula that fits the relationship between the two variables. Then you can use that formula to predict values for the dependent variable when only the independent variable is known. A doctor could prescribe the proper dose based on a person’s body weight.

Question 1: Correlation or regression?

Are the five examples at the beginning of this Introduction correlation or regression problems?

(The answer is at the end of the chapter)

1.2 Scattergraphs

We will be concerned with relationships between just two variables X and Y which is referred to as simple correlation and simple regression.

One way of showing the correlation between two related variables is on a scatter diagram or scattergraph, plotting a number of pairs of data on the graph. For example, a scatter diagram showing monthly selling costs against the volume of sales for a 12-month period might be as follows:
The independent variable (the cause) is plotted on the horizontal (x) axis and the dependent variable (the effect) is plotted on the vertical (y) axis.

This scattergraph suggests that there is some correlation between selling costs and sales volume, so that as sales volume rises, selling costs tend to rise as well.

2 Correlation

2.1 Degrees of correlation

Two variables can be one of the following:

- Perfectly correlated.
- Partly correlated.
- Uncorrelated.

These differing degrees of correlation can be illustrated by scatter diagrams.

**Perfect correlation**

All the pairs of values lie on a straight line. An exact linear relationship exists between the two variables.

**Partial correlation**

In (c), although there is no exact relationship, low values of X tend to be associated with low values of Y, and high values of X with high values of Y.

In (d) again, there is no exact relationship, but low values of X tend to be associated with high values of Y and vice versa.
2.2 **Positive and negative correlation**

Correlation, whether perfect or partial, can be **positive** or **negative**.

**Definitions**

- **Positive correlation** means that low values of one variable are associated with low values of the other, and high values of one variable are associated with high values of the other.
- **Negative correlation** means that low values of one variable are associated with high values of the other, and high values of one variable with low values of the other.

**Question 2: Negative correlation**

Which of the four diagrams in 2.1 showing perfect and partial correlation demonstrate negative correlation?  
(The answer is at the end of the chapter)

**Exam comments**

An assessment question could ask you to select which one of a collection of scatter diagrams shows a strong negative linear correlation or a weak positive linear correlation between \( X \) and \( Y \).

2.3 **The correlation coefficient**

**Definition**

The **degree of correlation** between two variables is measured by a **correlation coefficient**, \( r \). The nearer \( r \) is to +1 or –1, the stronger the relationship.

**Pearson’s correlation coefficient**, \( r \) (also known as the **sample correlation coefficient**) is used to measure how strong the connection is between two variables, known as the degree of correlation.

It is calculated using a formula which will be given to you in the assessment. It looks complicated but with a systematic approach and plenty of practice, you will be able to answer correlation questions in the assessment.
Formula to learn

**Correlation coefficient,** \( r = \frac{n\Sigma XY - \Sigma X \Sigma Y}{\sqrt{[n\Sigma X^2 - (\Sigma X)^2][n\Sigma Y^2 - (\Sigma Y)^2]}} \)

where: \( X \) and \( Y \) represent pairs of data for two variables \( X \) and \( Y \).
\( n = \) the number of pairs of data used in the analysis.

### 2.3.1 The correlation coefficient range

The correlation coefficient, \( r \) must always fall between –1 and +1. If you get a value outside this range you have made a mistake.

- \( r = +1 \) means that the variables are **perfectly positively correlated**.
- \( r = -1 \) means that the variables are **perfectly negatively correlated**.
- \( r = 0 \) means that the variables are **uncorrelated**.

**Worked Example: The correlation coefficient**

The cost of output at a factory is thought to depend on the number of units produced. Data have been collected for the number of units produced each month in the last six months, and the associated costs, as follows:

<table>
<thead>
<tr>
<th>Month</th>
<th>Output 000s of units</th>
<th>Cost $ 000</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( X )</td>
<td>( Y )</td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>9</td>
</tr>
<tr>
<td>2</td>
<td>3</td>
<td>11</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>7</td>
</tr>
<tr>
<td>4</td>
<td>4</td>
<td>13</td>
</tr>
<tr>
<td>5</td>
<td>3</td>
<td>11</td>
</tr>
<tr>
<td>6</td>
<td>5</td>
<td>15</td>
</tr>
</tbody>
</table>

**Required**

Assess whether there is any correlation between output and cost.

**Solution**

\[ r = \frac{n\Sigma XY - \Sigma X \Sigma Y}{\sqrt{[n\Sigma X^2 - (\Sigma X)^2][n\Sigma Y^2 - (\Sigma Y)^2]}} \]

We need to find the values for the following:

(a) \( \Sigma XY \) Multiply each value of \( X \) by its corresponding \( Y \) value, so that there are six values for \( XY \). Add up the six values to get the total.

(b) \( \Sigma X \) Add up the six values of \( X \) to get a total. \((\Sigma X)^2\) will be the square of this total.

(c) \( \Sigma Y \) Add up the six values of \( Y \) to get a total. \((\Sigma Y)^2\) will be the square of this total.

(d) \( \Sigma X^2 \) Find the square of each value of \( X \), so that there are six values for \( X^2 \). Add up these values to get a total.

(e) \( \Sigma Y^2 \) Find the square of each value of \( Y \), so that there are six values for \( Y^2 \). Add up these values to get a total.

Set out your workings in a table.
Question 3: Correlation

A company wants to know if the money they spend on advertising is effective in creating sales. The following data have been collected.

<table>
<thead>
<tr>
<th>Monthly advertising expenditure</th>
<th>Sales in following month</th>
</tr>
</thead>
<tbody>
<tr>
<td>$ 000</td>
<td>$ 000</td>
</tr>
<tr>
<td>1.2</td>
<td>132.5</td>
</tr>
<tr>
<td>0.9</td>
<td>98.5</td>
</tr>
<tr>
<td>1.6</td>
<td>154.3</td>
</tr>
<tr>
<td>2.1</td>
<td>201.4</td>
</tr>
<tr>
<td>1.6</td>
<td>161.0</td>
</tr>
</tbody>
</table>

Required

Calculate the correlation coefficient for the data and explain the result.

(The answer is at the end of the chapter)

2.3.2 Correlation in a time series

Correlation exists in a time series if there is a relationship between the period of time and the recorded value for that period of time. The correlation coefficient is calculated with time as the x variable although it is convenient to use simplified values for x instead of year numbers.

For example, instead of having a series of years 20X2 to 20X8, we could have values for x from 0 (20X2) to 6 (20X8).

Note that whatever starting value you use for x (be it 0, 1, 2, ..., 721, ..., 953 or anything else), the value of r will always be the same.
Question 4: Correlation in a time series
Sales of product A between 20X5 and 20X9 were as follows:

<table>
<thead>
<tr>
<th>Year</th>
<th>Units sold (000s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>20X5</td>
<td>20</td>
</tr>
<tr>
<td>20X6</td>
<td>18</td>
</tr>
<tr>
<td>20X7</td>
<td>15</td>
</tr>
<tr>
<td>20X8</td>
<td>14</td>
</tr>
<tr>
<td>20X9</td>
<td>11</td>
</tr>
</tbody>
</table>

Required
Determine whether there is a trend in sales. In other words, decide whether there is any correlation between the year and the number of units sold.  
(The answer is at the end of the chapter)

2.4 The coefficient of determination, $r^2$
Symbolised as $r^2$, the coefficient of determination is the square of the correlation coefficient.

Definition
The coefficient of determination $r^2$ measures the proportion of the total variation in the value of the dependent variable that can be explained by variation in the value of the independent variable.

Finding the coefficient of determination takes only three steps:

Step 1. Find the correlation coefficient, $r$
Example, $r = 0.543$.

Step 2. Square the correlation coefficient.
$0.543^2 = 0.295$

Step 3. Convert the square of the correlation coefficient to a percentage.
$0.295 = 29.5$ per cent

The coefficient of determination can have only positive values ranging from $r^2 = +1.0$ for a perfect correlation (positive or negative) down to $r^2 = 0.0$ for a complete absence of correlation. The advantage of the correlation coefficient, $r$, is that it provides you with the positive or negative direction of the correlation. The advantage of the coefficient of determination, $r^2$, is that it provides a measure of the strength of the correlation. The correlation coefficient, $r$, gives you the true direction of the correlation (+ or -) but only the square root of the strength of the correlation; while the coefficient of determination, $r^2$, gives you the true strength of the correlation but without an indication of its direction. Both of them together give you the whole picture.

If the coefficient of determination between $X$ and $Y$ is 0.55 (55 per cent), the correlation coefficient is either +0.74 or -0.74, we can say that 55 per cent of the variability in $Y$ is explained by a variability in $X$.

2.4.1 Interpreting $r^2$
In the question above, $r = -0.992$, therefore $r^2 = 0.984$. This means that over 98 per cent of variations in sales can be explained by the passage of time, leaving 0.016 (less than 2 per cent) of variations to be explained by other factors.

Similarly, if the correlation coefficient between a company’s output volume and maintenance costs was 0.9, $r^2$ would be 0.81, meaning that 81 per cent of variation in maintenance costs could be explained by variation in output volume, leaving only 19 per cent of variation to be explained by other factors (such as the age of the equipment).
Note, however, that if \( r^2 = 0.81 \), we would say that 81 per cent of the variation in \( y \) can be explained by variation in \( x \). We do not necessarily conclude that 81 per cent of variation in \( y \) are caused by the variation in \( x \). We must beware of reading too much significance into our statistical analysis.

### 2.4.2 Non-linear relationships

The formulae used above for \( r \) and \( r^2 \) only work for linear or near linear relationships. All the points on a scatter diagram might lie on a smooth curve as follows:

![Graph of non-linear relationship](image)

If the formula for \( r \) were used in a situation such as this, a low value of \( r \) would be obtained, suggesting that very little correlation exists, whereas in fact the two sets of variables are perfectly correlated by a non-linear relationship. There are methods of testing correlations of this type, but they are outside the scope of this Study Manual. We are only concerned with linear correlation.

### 2.5 Correlation and causation

If two variables are well correlated, either positively or negatively, this may be due to pure chance or there may be a reason for it. The larger the number of pairs of data collected, the less likely it is that the correlation is due to chance, though that possibility should never be ignored entirely.

If there is a reason, it may not be causal. For example, monthly net income is well correlated with monthly credit to a person’s bank account, for the logical (rather than causal) reason that for most people the one equals the other.

Even if there is a causal explanation for a correlation, it does not follow that variations in the value of one variable cause variations in the value of the other. For example, sales of ice cream and of sunglasses are well correlated, not because of a direct causal link but because the weather influences both variables. Having said this, it is of course possible that where two variables are correlated, there is a direct causal link to be found.

### 3 Regression

#### 3.1 The dependent variable and the explanatory variable

Here we are concerned with the form of the relationship between the variables. This can be summarised by an equation that enables us to predict or estimate the values of one variable \( (y – the \ dependent \ variable) \) given values of the other variable \( (x – the \ independent \ or \ explanatory \ variable) \). The first thing to do then is to decide which variable is which.

Regression analysis is concerned with how the values of \( y \) depend on the corresponding values of \( x \). The variable that is eventually to be predicted or estimated should be labelled \( y \).
Formula to learn

The equation of a straight line has the form \( y = a + bx \)

where: \( x \) and \( y \) are related variables
- \( x \) is the independent variable
- \( y \) is the dependent variable
- \( a \) is the intercept of the line on the vertical axis
- \( b \) is the gradient of the line.

Worked Example: Drawing graphs

Plot the graph for \( y = 4x + 5 \).

Consider the range of values from \( x = 0 \) to \( x = 10 \).

Solution

The first step is to draw up a table for the equation. Although the problem mentions \( x = 0 \) to \( x = 10 \), it is not necessary to calculate values of \( y \) for \( x = 1, 2, 3 \) and so on. A graph of a linear equation can actually be drawn from just two \((x, y)\) values but it is always best to calculate a number of values in case you make an arithmetical error. We have calculated five values, but three would be enough in your assessment.

<table>
<thead>
<tr>
<th>( x )</th>
<th>( y )</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>5</td>
</tr>
<tr>
<td>2</td>
<td>13</td>
</tr>
<tr>
<td>4</td>
<td>21</td>
</tr>
<tr>
<td>6</td>
<td>29</td>
</tr>
<tr>
<td>8</td>
<td>37</td>
</tr>
<tr>
<td>10</td>
<td>45</td>
</tr>
</tbody>
</table>

Graph of \( y = 4x + 5 \)

3.2 The intercept and the gradient (slope)

The graph of a linear equation is determined by two things:
- The gradient (or slope) of the straight line.
- The point at which the straight line crosses the y axis.

Definitions

The intercept is the point at which a straight line crosses the \( y \)-axis.

The gradient of the graph of a linear equation is \[ \frac{\text{change in } y}{\text{change in } x} = \frac{(y_2 - y_1) / (x_2 - x_1)}{\text{where } (x_1, y_1) \text{ and } (x_2, y_2) \text{ are two points on the straight line.}} \]
3.2.1 The intercept

The intercept of \( y = 4x + 5 \) is where \( y = 5 \). It is no coincidence that the intercept is the same as the constant represented by \( a \) in the general form of the equation \( y = a + bx \). \( a \) is the value \( y \) takes when \( x = 0 \).

3.2.2 The gradient

If we take two points on the line (see graph in 3.1):

1. \( x = 2, \ y = 13 \)
2. \( x = 4, \ y = 21 \)

The gradient of \( y = 4x + 5 \) = \( \frac{\text{change in } y}{\text{change in } x} = \frac{(21-13)}{(4-2)} = \frac{8}{2} = 4 \)

Notice that the gradient is also given by the number multiplied by \( x \) in the equation (\( b \) in the general form of the equation).

Question 5: The gradient
If \( y = 10 - x \), the gradient = ________

(The answer is at the end of the chapter)

3.2.3 Positive and negative gradients

Note that the gradient of \( y = 4x + 5 \) is positive whereas the gradient of \( y = 10 - x \) is negative.

- A positive gradient slopes upwards from left to right.
- A negative gradient slopes downwards from left to right.
- The greater the value of the gradient, the steeper the slope.

Question 6: Intercept and gradient

What is the intercept and gradient of the graph of \( 4y = 16x - 12 \)?

<table>
<thead>
<tr>
<th>Intercept</th>
<th>Gradient</th>
</tr>
</thead>
<tbody>
<tr>
<td>A -3</td>
<td>+4</td>
</tr>
<tr>
<td>B -4</td>
<td>+3</td>
</tr>
<tr>
<td>C +3</td>
<td>-4</td>
</tr>
<tr>
<td>D +4</td>
<td>-3</td>
</tr>
</tbody>
</table>

(The answer is at the end of the chapter)

4 Finding the best-fitting line

4.1 Equation of a line of best fit

If we assume that there is a linear relationship between the two variables and we determine the equation of a straight line \( (y = a + bx) \), which is a good fit for the available data plotted on a scattergraph, we can use the equation for forecasting. We do this by substituting values for \( x \) into the equation and deriving values for \( y \).

There are a number of techniques for estimating the equation of a line of best fit. We will be looking at the scattergraph method and simple linear regression analysis. Both provide a technique for estimating values for \( a \) and \( b \) in the equation, \( y = a + bx \).
Definition
The **scattergraph method** involves plotting pairs of data for two related variables on a graph, to produce a scattergraph, and then using judgment to draw what seems to be a line of best fit through plotted data.

**Worked Example: The scattergraph method**

Suppose we have the following pairs of data about output and costs:

<table>
<thead>
<tr>
<th>Month</th>
<th>Output 000 units</th>
<th>Costs $ 000</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>20</td>
<td>82</td>
</tr>
<tr>
<td>2</td>
<td>16</td>
<td>70</td>
</tr>
<tr>
<td>3</td>
<td>24</td>
<td>90</td>
</tr>
<tr>
<td>4</td>
<td>22</td>
<td>85</td>
</tr>
<tr>
<td>5</td>
<td>18</td>
<td>73</td>
</tr>
</tbody>
</table>

These pairs of data can be plotted on a scattergraph (the horizontal axis representing the independent variable and the vertical axis the dependent) and a line of best fit might be judged as the one shown below. It is drawn to pass through the middle of the data points, thereby having as many data points below the line as above it.

**Worked Example: Scattergraph**

A **formula for the line** can be found. In our example, suppose that we read the following data from the graph.

(i) When $X = 0$, $Y = 22 000$. This must be the value of $a$ in the formula $Y = a + bX$.

(ii) When $X = 25 000$, $Y = \sim 95 000$. Since $Y = a + bX$, and $a = 22 000$

\[
95 000 = 22 000 + (b \times 25 000) \\
b \times 25 000 = 73 000 \\
b = 2.92
\]

In this example the estimated equation from the scattergraph is $Y = 22 000 + 2.92X$. 
4.2 Forecasting and scattergraphs

If the company to which the data in Paragraph 4.1 relates wanted to predict costs at a certain level of output (say 13 000 units), the value of 13 000 could be substituted into the equation

\[ Y = 22 000 + 2.92X \]

and an estimate of costs made.

If \( X = 13 000 \), \( Y = 22 000 + (2.92 \times 13 000) \)

\[ \therefore Y = 60 000 \] (rounded)

Predictions can be made directly from the scattergraph, but this will usually be less accurate.

![Scattergraph](image)

The prediction of the cost of producing 13 000 units from the scattergraph is $60 000.

4.3 The least squares method of linear regression analysis

Definition

The least squares method of linear regression analysis provides a technique for estimating the equation of a line of best fit.

The equation of a straight line has the form \( Y = a + bX \)

Where: \( X \) and \( Y \) are related variables
- \( X \) is the independent variable
- \( Y \) is the dependent variable
- \( a \) is the intercept of the line on the vertical axis
- \( b \) is the gradient of the line.

Formula to learn

The least squares method of linear regression analysis involves using the following formulae for \( a \) and \( b \) in \( Y = a + bX \).

\[
 b = \frac{n\Sigma XY - \Sigma X \Sigma Y}{n\Sigma X^2 - (\Sigma X)^2}
\]

\[
 a = \bar{Y} - b \bar{X}
\]

where:
- \( n \) is the number of pairs of data
- \( \bar{X} \) is the mean \( X \) value of all the pairs of data
- \( \bar{Y} \) is the mean \( Y \) value of all the pairs of data
There are some points to note about these formulae:

(a) The line of best fit that is derived represents the regression of $Y$ upon $X$.

A different line of best fit could be obtained by interchanging $X$ and $Y$ in the formulae. This would then represent the regression of $X$ upon $Y$ ($x = a + by$) and it would have a slightly different slope. For examination purposes, always use the regression of $Y$ upon $X$, where $X$ is the independent variable, and $Y$ is the dependent variable whose value we wish to forecast for given values of $X$. In a time series, $X$ will represent time.

(b) Since $a = \frac{\Sigma Y \cdot b \Sigma X}{n}$, it follows that the line of best fit must always pass through the point $\left(\frac{\Sigma X}{n}, \frac{\Sigma Y}{n}\right)$.

(c) If you look at the formula for $b$ and compare it with the formula we gave for the correlation coefficient (Paragraph 2.3) you should see some similarities between the two formulae.

(d) The sign of $b$ in $Y = a + bX$ is the same as the sign of $r$. In other words they are both positive or both are negative.

### 4.3.1 The meaning of 'least squares'

The term 'squares' in 'least squares regression analysis' refers to the squares of the differences between actual values of the dependent variable ($Y$) and predicted values given by the regression line of best fit. These differences are referred to as residuals or residual errors. 'Least squares' means that the line of best fit that is calculated is the one that minimises the sum of the squares of all the residuals. The differences are measured vertically on a graph, not at an angle to take the shortest route to the regression line.

![Graph showing least squares regression](image)

### 4.3.2 Some helpful hints

(a) The value of $b$ must be calculated first as it is needed to calculate $a$.

(b) $\bar{X}$ is the mean of the $X$ values $= \frac{\Sigma X}{n}$

$\bar{Y}$ is the mean of the $Y$ values $= \frac{\Sigma Y}{n}$

(c) Remember that $X$ is the independent variable and $Y$ is the dependent variable.

(d) Set your workings out in a table to find the figures to put into the formulae.

(e) These values can also be calculated using Excel.

(f) Excel also calculates $r$ and $r^2$.

### Worked Example: Linear regression analysis

(a) Given that there is a fairly high degree of correlation between the output and the costs detailed in Paragraph 4.1 (so that a linear relationship can be assumed), calculate an equation to determine the expected level of costs, for any given volume of output, using the least squares method.

(b) Prepare a budget for total costs if output is 22,000 units.

(c) Confirm that the degree of correlation between output and costs is high by calculating the correlation coefficient.
Solution

(a) Workings

<table>
<thead>
<tr>
<th>X</th>
<th>Y</th>
<th>XY</th>
<th>X^2</th>
<th>Y^2</th>
</tr>
</thead>
<tbody>
<tr>
<td>20</td>
<td>82</td>
<td>1 640</td>
<td>400</td>
<td>6 724</td>
</tr>
<tr>
<td>16</td>
<td>70</td>
<td>1 120</td>
<td>256</td>
<td>4 900</td>
</tr>
<tr>
<td>24</td>
<td>90</td>
<td>2 160</td>
<td>576</td>
<td>8 100</td>
</tr>
<tr>
<td>22</td>
<td>85</td>
<td>1 870</td>
<td>484</td>
<td>7 225</td>
</tr>
<tr>
<td>18</td>
<td>73</td>
<td>1 314</td>
<td>324</td>
<td>5 329</td>
</tr>
</tbody>
</table>

\[ \sum X = 100 \quad \sum Y = 400 \quad \sum XY = 8104 \quad \sum X^2 = 2040 \quad \sum Y^2 = 32278 \]

\[ n = 5 \] (There are five pairs of data for X and Y values)

\[ b = \frac{n \sum XY - \sum X \sum Y}{n \sum X^2 - (\sum X)^2} = \frac{(5 \times 8104) - (100 \times 400)}{(5 \times 2040) - 100^2} \]

\[ = \frac{40520 - 40000}{10200 - 10000} = \frac{520}{200} = 2.6 \]

\[ a = \bar{Y} - b \bar{X} = \frac{400}{5} - 2.6 \times \left( \frac{100}{5} \right) = 28 \]

\[ Y = 28 + 2.6X \]

where: \[ Y = \text{total cost, in thousands of dollars} \quad X = \text{output, in thousands of units} \]

Compare this equation to that determined in Paragraph 4.1.

Note that the fixed costs are $28,000 (when X = 0 costs are $28,000) and the variable cost per unit is $2.60.

(b) If the output is 22,000 units, we would expect costs to be

\[ 28 + (2.6 \times 22) = 85.2 = $85,200. \]

(c) \[ r = \frac{520}{\sqrt{200 \times (5 \times 32278 - 400^2)}} = \frac{520}{\sqrt{200 \times 1390}} = \frac{520}{527.3} = +0.986 \]

---

Exam comments

In an assessment, you might be required to use the linear regression analysis technique to calculate the values of a and b in the equation \( y = a + bx \). This type of task is ideally tested by objective test questions.

---

Question 7: Linear regression analysis

If \( \sum X = 79, \sum Y = 1466, \sum X^2 = 1083, \sum Y^2 = 363076, \sum XY = 19736 \) and \( n = 6 \), then the value of \( b \), the gradient, to two decimal places, is:

A  10.12
B  111.03
C  13.62
D  -8.53

(The answer is at the end of the chapter)
Question 8: Forecasting

In a forecasting model based on \( y = a + bx \), the intercept is $262. If the value of \( y \) is $503 when \( x \) is 23, then the value of the gradient, to two decimal places, is:

- A –20.96
- B –10.48
- C 10.48
- D 20.96

(The answer is at the end of the chapter)

Question 9: Expense claims and recorded car mileages

The expense claims and recorded car mileages on company business, relating to a particular day, of a sample of salesmen of a company are as follows:

<table>
<thead>
<tr>
<th>Salesman</th>
<th>Mileage</th>
<th>Expenses</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>100</td>
<td>60</td>
</tr>
<tr>
<td>B</td>
<td>80</td>
<td>48</td>
</tr>
<tr>
<td>C</td>
<td>20</td>
<td>20</td>
</tr>
<tr>
<td>D</td>
<td>120</td>
<td>55</td>
</tr>
<tr>
<td>E</td>
<td>70</td>
<td>38</td>
</tr>
<tr>
<td>F</td>
<td>50</td>
<td>38</td>
</tr>
<tr>
<td>G</td>
<td>80</td>
<td>44</td>
</tr>
<tr>
<td>H</td>
<td>40</td>
<td>30</td>
</tr>
<tr>
<td>I</td>
<td>50</td>
<td>40</td>
</tr>
<tr>
<td>J</td>
<td>60</td>
<td>50</td>
</tr>
</tbody>
</table>

Required

(a) Plot the data on a scatter diagram.
(b) Determine, by the method of least squares, a linear model to predict expenses, mileage having been given.
(c) Plot the model on your scatter diagram.
(d) Three further salesmen submit expense claims with mileages as follows:

<table>
<thead>
<tr>
<th>Salesman</th>
<th>Mileage</th>
<th>Expenses</th>
</tr>
</thead>
<tbody>
<tr>
<td>K</td>
<td>110</td>
<td>64</td>
</tr>
<tr>
<td>L</td>
<td>30</td>
<td>48</td>
</tr>
<tr>
<td>M</td>
<td>160</td>
<td>80</td>
</tr>
</tbody>
</table>

Discuss in each case whether or not the claim is reasonable.

(The answer is at the end of the chapter)

4.4 Interpretation of the regression coefficients

We said earlier that the least squares method of linear regression analysis provides estimates for \( a \) and \( b \) and, given values for \( x \) and \( y \), you should now be able to determine \( a \) and \( b \) yourself.

\( a \) and \( b \) are known as the coefficients of the regression line \( Y = a + bX \) but what do they actually mean?

We explained in Paragraph 3.2 that \( a \) is the intercept of the line of best fit on the vertical axis. The scatter diagram from the previous question – the expense claims and recorded car mileages – has been reproduced below but the regression line has been extended so that it meets the \( y \) axis. The intercept of this line on the \( y \) axis would appear to be approximately 18 or 19; the value of \( a \) was calculated as 18.46 using linear
regression analysis. This means that when mileage is zero, expenses are $18. In other words, there are other costs, unrelated to the distance travelled, of $18 per day. These might include subsistence costs.

In the scatter diagram below the gradient = \( \frac{\text{change in } y}{\text{increase in } x} = \frac{61 - 18}{120} = 0.36 \)

This means that an increase of one mile travelled produces an increase in costs of $0.36. The coefficient 0.36 therefore indicates that there is a variable motoring cost of 36c per mile on average.

\( b \) is the gradient or slope of the line of best fit. The slope of a line is the change in \( y \) for a unit increase in \( x \). In diagram (a) below, an increase in the value of \( x \) from 5 to 6 produces an increase in \( y \) of 10 from 10 to 20. In diagram (b), a similar increase in \( x \) produces an increase of 5 in \( y \).

The gradient of the line in (a) = \( \frac{\text{change in } y}{\text{increase in } x} = \frac{20 - 10}{6 - 5} = \frac{10}{1} = 10 \)

The gradient of the line in (b) = \( \frac{\text{change in } y}{\text{increase in } x} = \frac{15 - 10}{6 - 5} = 5 \)

The gradient of the line in (c) = \( \frac{\text{change in } y}{\text{increase in } x} = \frac{10 - 15}{6 - 5} = -5 \)

In (a) and (b), \( y \) increases as \( x \) increases and the gradient is positive. In (c), however, \( y \) decreases as \( x \) increases and hence the gradient is negative.

In numerical terms the gradient gives the rate of change in \( y \) for a unit increase in \( x \).
4.5 Reliability of estimates

The coefficient of determination ($r^2$) which is computed by squaring the correlation coefficient provides an estimate of the variation in $y$ which is due to the regression (i.e. changes in the other variable).

If there is a perfect linear relationship between $x$ and $y$ ($r = \pm 1$) then we can predict $y$ from any given value of $x$ with great confidence.

If correlation is high (for example, $r = 0.9$) the actual values will all lie quite close to the regression line and so predictions should not be far out. If correlation is below about 0.7, predictions will only give a very rough guide as to the likely value of $y$.

As with any analytical process, the amount of data available is very important. Even if correlation is high, if we have fewer than about 10 pairs of values, we must regard any estimate as being somewhat unreliable.

When calculating a line of best fit, there will be a range of values for $x$. In the least squares method example following Paragraph 4.3 the line $y = 28 + 2.6x$ was predicted from data with output values ranging from $x = 16$ to $x = 24$. Depending on the degree of correlation between $x$ and $y$, we might safely use the estimated line of best fit to predict values for $y$, provided that the value of $x$ remains within the range 16 to 24. We would be on less safe ground if we used the formula to predict a value for $y$ when $x = 10$, or 30, or any other value outside the range $16 – 24$, because we would have to assume that the fitted line applies outside the range of $x$ values used to establish the line in the first place.

(a) **Interpolation** means using a line of best fit to predict a value within the two extreme points of the observed range.

(b) **Extrapolation** means using a line of best fit to predict a value outside the two extreme points.

4.6 The advantages and disadvantages of regression analysis

The advantages of the least squares method of regression analysis are as follows:

(a) It can be used to estimate a line of best fit using all the data available. It is likely to provide a more reliable estimate than any other technique of producing a straight line of best fit (for example, estimating by eye).

(b) The reliability of the estimated line can be evaluated by calculating the correlation coefficient $r$.

The disadvantages of the method are as follows:

(a) It assumes a linear relationship between the two variables, whereas a non-linear relationship may exist.

(b) It assumes that what has happened in the past will provide a reliable guide to the future. For example, if a line is calculated for total costs of production, based on historical data, the estimate could be used to budget for future costs. However, if there has been cost inflation, a productivity agreement with the workforce, a move to new premises, the dismissal of large numbers of office staff and the introduction of new equipment, future costs of production might bear no relation to costs in the past.

(c) The technique assumes that the value of one variable, $y$, can be predicted or estimated from the value of one other variable, $x$. In reality, the value of $y$ might depend on several other variables, not just on $x$. 
Key chapter points

- When the value of one variable is related to the value of another, they are said to be correlated.
- Two variables might be perfectly correlated, partly correlated or uncorrelated. Correlation can be positive or negative.
- The degree of correlation between two variables is measured by the sample correlation coefficient, \( r \). The nearer \( r \) is to +1 or \(-1\), the stronger the relationship.
- The coefficient of determination, \( r^2 \), measures the proportion of the total variation in the value of the dependent variable that can be explained by the variation in the value of the independent variable.
- The scattergraph method involves the use of judgment to draw what seems to be a line of best fit through plotted data.
- Linear regression analysis (the least squares method) is one technique for estimating a line of best fit. Ensure that you know how to use the formulae to calculate \( a \) and \( b \) in \( y = a + bx \).
- Correlation and regression analysis do not indicate cause and effect. Even if \( r = 1 \), the correlation could still be spurious, both variables being influenced by a third.
- The coefficient of determination (\( r \)-squared) is the square of the correlation coefficient. Its value may vary from zero to one. It has the advantage over the correlation coefficient in that it may be interpreted directly as the proportion of variance in the dependent variable that can be accounted for by the regression equation.
1. The regression equation $Y = 5 + 4X$ has been calculated from 6 pairs of values, with $X$ ranging from 1 to 10. The correlation coefficient is 0.9. It is estimated that $Y = 85$ when $X = 20$. Which of the following are true?

I. The estimate is not reliable because the sample is small.
II. The estimate is reliable.
III. The estimate is not reliable because the correlation is low.
IV. The estimate is not reliable because $X$ is outside the range of the data.

A. I and II only  
B. I and III only  
C. I and IV only  
D. II and IV only

2. When the value of one variable is related to the value of another, they are said to be correlated. Correlation therefore means an inter-relationship or correspondence.

If the following points were plotted on a graph, what sort of correlation would they display? (Tick as appropriate)

(a) $(2, 3)$
   $(3, 4.5)$
   $(4, 6)$

   Perfect positive correlation  
   Perfect negative correlation  
   Uncorrelated

(b) $(2, 3)$
   $(3, 1.5)$
   $(4, 0)$

   Perfect positive correlation  
   Perfect negative correlation  
   Uncorrelated

(c) $(2, 3)$
   $(4, 0)$
   $(4, 6)$

   Perfect positive correlation  
   Perfect negative correlation  
   Uncorrelated

3. The relationship between expenditure in $ on advertising ($X$) in one time period and sales (in $) ($Y$) in the next period has been found to be $Y = 50 + 7X$. Which of the following interprets the value ‘7’ correctly?

A. For every $1 spent on advertising, sales increase by $7 on average.  
B. For every $1 increase in sales, $7 must on average be spent on advertising.  
C. For every $1 spent on advertising, sales increase by $8.  
D. When advertising is zero, sales are $50/7.

4. If $\sum X = 21$  
   $\sum Y = 184$  
   $\sum X^2 = 91$  
   $\sum XY = 587$  
   $n = 7$

Which of the following values for $a$ and $b$ are correct in the formula $Y = a + bX$?

$a$  
$b$

A. $-22.5$  
B. $-22.5$  
C. $+22.5$  
D. $+22.5$
5. The following scatter diagrams can be associated with which correlation coefficients?

![Scatter Diagrams]

I. 0  
II. –1  
III. +0.8

A = 
B = 
C =

6. The value of the correlation coefficient between x and y is 0.9. Which of the following is correct?

A. There is a weak relationship between x and y.  
B. x is 90 per cent of y.  
C. If the values of x and y were plotted on a graph, the line relating them would have a slope of 0.9.  
D. There is a very strong relationship between x and y.

7. The value of \( a \) in the regression equation

A. cannot be negative.  
B. measures the value of the independent variable.  
C. is the value of the dependent variable when the value of the independent variable is zero.  
D. equals one.

8. If the correlation coefficient for two variables is –0.8, the coefficient of determination for the same two variables is
Answers to quick revision questions

1. C  The sample of only six pairs of values is very small and is therefore likely to reduce the reliability of the estimate. Statement I is therefore true.

With such a small sample and the extrapolation required, the estimate is unlikely to be reliable. Statement II is therefore not true.

Since a correlation coefficient of 0.9 would be regarded as strong (it is a high value) the estimate would be reliable. Statement III is therefore not true.

When \( X = 20 \), we don't know anything about the relationship between \( X \) and \( Y \) since the sample data only goes up to \( X = 10 \). Statement IV is therefore true.

2.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th>Perfect positive correlation</th>
<th>Perfect negative correlation</th>
<th>Uncorrelated</th>
</tr>
</thead>
<tbody>
<tr>
<td>(a)</td>
<td>(2, 3)</td>
<td>?</td>
<td>Perfect negative correlation</td>
<td>Uncorrelated</td>
</tr>
<tr>
<td></td>
<td>(3, 4.5)</td>
<td>?</td>
<td>Perfect negative correlation</td>
<td>Uncorrelated</td>
</tr>
<tr>
<td></td>
<td>(4, 6)</td>
<td>?</td>
<td>Perfect negative correlation</td>
<td>Uncorrelated</td>
</tr>
<tr>
<td>(b)</td>
<td>(2, 3)</td>
<td>?</td>
<td>Perfect negative correlation</td>
<td>Uncorrelated</td>
</tr>
<tr>
<td></td>
<td>(3, 1.5)</td>
<td>?</td>
<td>Perfect negative correlation</td>
<td>Uncorrelated</td>
</tr>
<tr>
<td></td>
<td>(4, 0)</td>
<td>?</td>
<td>Perfect negative correlation</td>
<td>Uncorrelated</td>
</tr>
<tr>
<td>(c)</td>
<td>(2, 3)</td>
<td>?</td>
<td>Perfect negative correlation</td>
<td>Uncorrelated</td>
</tr>
<tr>
<td></td>
<td>(4, 0)</td>
<td>?</td>
<td>Perfect negative correlation</td>
<td>Uncorrelated</td>
</tr>
<tr>
<td></td>
<td>(4, 6)</td>
<td>?</td>
<td>Perfect negative correlation</td>
<td>Uncorrelated</td>
</tr>
</tbody>
</table>

**Workings**

(a) If these points were plotted on a scatter diagram, all the pairs of values would lie on an upward-sloping straight line with a gradient of \( +1.5 \). \[ \text{Gradient} = \frac{6-3}{4-2} = 1.5 \]  This would be indicative of PERFECT POSITIVE CORRELATION.

(b) If these points were plotted on a scatter diagram, all the pairs of values would lie on a downward-sloping straight line with a gradient of \( -1.5 \). \[ \text{Gradient} = \frac{0-3}{4-2} = -1.5 \]  This would be indicative of PERFECT NEGATIVE CORRELATION.

(c) If these points were plotted on a scatter diagram there would not be evidence of any correlation.

3. A  \( Y = 50 + 7X \)

If we increase expenditure on advertising by $1, then we increase \( Y \) by 7.

Therefore, sales increase by $7.

If advertising costs are increased by $7

\( Y \) increases by \( 7 \times 7 = 49 \)

Sales are increased by $49.

Option B is therefore incorrect.

If \( Y = 50 + 7X \)

when advertising is zero, \( X = 0 \)

\[ \therefore Y = 50 + (7 \times 0) \]

\[ Y = 50 + 0 \]

\[ Y = 50 \]

Therefore option D is incorrect.
Where \( y = a + bx \)

\[
\begin{align*}
\beta &= \frac{\sum XY - \sum X \sum Y}{\sum X^2 - (\sum X)^2} \\
&= \frac{(7 \times 587) - (21 \times 184)}{(7 \times 91) - (21^2)} \\
&= \frac{245}{196} \\
&= 1.25 \\
\alpha &= \bar{Y} - \beta \bar{X} \\
&= \frac{184}{7} - \frac{(1.25 \times 21)}{7} \\
&= 22.5
\end{align*}
\]

The correct answer is therefore D.

5  A  II

B  III

C  I

6  D  The correlation coefficient of 0.9 is very close to 1 and so there is a very strong relationship between \( x \) and \( y \).

7  C  \( a \) is the intercept on the \( y \) axis which is the value of the dependent variable when the value of the independent variable is zero.

8  The coefficient of determination for the same two variables is \((-0.8)^2 = 0.64\)
Answers to chapter questions

1 (i) regression
(ii) regression
(iii) correlation
(iv) correlation
(v) regression

2 Diagram (b) showing perfect correlation and diagram (d) showing partial correlation both demonstrate negative correlation.

3

<table>
<thead>
<tr>
<th>Monthly advertising expenditure</th>
<th>Sales</th>
</tr>
</thead>
<tbody>
<tr>
<td>$X$</td>
<td>$Y$</td>
</tr>
<tr>
<td>1.2</td>
<td>132.5</td>
</tr>
<tr>
<td>0.9</td>
<td>98.5</td>
</tr>
<tr>
<td>1.6</td>
<td>154.3</td>
</tr>
<tr>
<td>2.1</td>
<td>201.4</td>
</tr>
<tr>
<td>1.6</td>
<td>161.0</td>
</tr>
<tr>
<td>7.4</td>
<td>747.7</td>
</tr>
</tbody>
</table>

$(\Sigma X)^2 = 7.4^2 = 54.76$
$(\Sigma Y)^2 = 747.7^2 = 559 055.29$

$$r = \frac{(5 \times 1 175.07) - (7.4 \times 747.7)}{\sqrt{[(5 \times 11.78) - 54.76] \times [(5 \times 117 549.95) - 559 055.29]}}$$

$$= \frac{5 875.35 - 5 532.98}{\sqrt{4.14 \times 28 694.46}}$$

$$= \frac{342.37}{\sqrt{118 795.06}}$$

$$= \frac{342.37}{344.67} = 0.993$$

0.993 is very close to 1, therefore there is a strong positive correlation and sales are dependent on advertising expenditure.

4 Let 20X5 to 20X9 be years 0 to 4.

<table>
<thead>
<tr>
<th>$x$</th>
<th>$y$</th>
<th>$xy$</th>
<th>$x^2$</th>
<th>$y^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>20</td>
<td>0</td>
<td>0</td>
<td>400</td>
</tr>
<tr>
<td>1</td>
<td>18</td>
<td>18</td>
<td>1</td>
<td>324</td>
</tr>
<tr>
<td>2</td>
<td>15</td>
<td>30</td>
<td>4</td>
<td>225</td>
</tr>
<tr>
<td>3</td>
<td>14</td>
<td>42</td>
<td>9</td>
<td>196</td>
</tr>
<tr>
<td>4</td>
<td>11</td>
<td>44</td>
<td>16</td>
<td>121</td>
</tr>
</tbody>
</table>

$\Sigma x = 10 \quad \Sigma y = 78 \quad \Sigma xy = 134 \quad \Sigma x^2 = 30 \quad \Sigma y^2 = 1 266$

$(\Sigma x)^2 = 100 \quad (\Sigma y)^2 = 6 084$
\[ n = 5 \]
\[ r = \frac{(5 \times 134) - (10 \times 78)}{\sqrt{(5 \times 30 - 100) \times (5 \times 1266 - 6084)}} \]
\[ = \frac{670 - 780}{\sqrt{(150 - 100) \times (6330 - 6084)}} = \frac{-110}{\sqrt{50 \times 246}} \]
\[ = -110 \times \frac{1}{\sqrt{12300}} = -110 \times \frac{1}{110.90537} = -0.992 \]

There is partial negative correlation between the year of sale and units sold. The value of \( r \) is close to \(-1\), therefore a high degree of correlation exists, although it is not quite perfect correlation. This means that there is a clear downward trend in sales.

5 The gradient = \( -1 \)

If \( y = 10 - x \), then \( a = 10 \) and \( b = -1 \) \((-1 \times x = -x)\).

Therefore, gradient = \(-1\)

6 \( 4y = 16x - 12 \)

Equation must be in the form \( y = a + bx \)

\( y = 4x - 3 \) (divide both sides by 4)
\( y = -3 + 4x \) (rearrange the RHS)

Intercept = \( a = -3 \)
Gradient = \( b = 4 \)

Therefore, the correct answer is A.

If you selected option D, you have obviously confused the intercept and the gradient. Remember that with an equation in the form \( y = a + bx \), \( a \) = intercept (i.e. where the line of the graph crosses the \( y \) axis) and \( b \) = the slope or gradient of the line.

7 A

\[ b = \frac{(6 \times 19736) - (79 \times 1466)}{(6 \times 1083) - 79^2} \]
\[ = \frac{118416 - 115814}{6,498 - 6241} = \frac{2602}{257} = 10.12 \]

8 C \( y = a + bx \)

\( 503 = 262 + (b \times 23) \)
\( 241 = b \times 23 \)
\( b = 10.48 \)
(a) The independent variable, x, is the mileage. The dependent variable, y, is the expenses in dollars.

\[
\begin{array}{cccc}
 x & y & x^2 & xy \\
 100 & 60 & 10000 & 6000 \\
 80 & 48 & 6400 & 3840 \\
 20 & 20 & 400 & 400 \\
 120 & 55 & 14400 & 6600 \\
 70 & 38 & 4900 & 2660 \\
 50 & 38 & 2500 & 1900 \\
 80 & 44 & 6400 & 3520 \\
 40 & 30 & 1600 & 1200 \\
 50 & 40 & 2500 & 2000 \\
 60 & 50 & 3600 & 3000 \\
 670 & 423 & 52700 & 31120 \\
\end{array}
\]

\[
b = \frac{(10 \times 31 \times 120) - (670 \times 423)}{10 \times 52700 - 670^2}
\]

\[
= \frac{311200 - 283410}{527000 - 448900}
\]

\[
= \frac{27790}{78100}
\]

\[
= 0.3558
\]

\[
a = \frac{423}{10} - 0.3558 \times \frac{670}{10} = 18.46
\]

The linear model is \( y = 18.46 + 0.3558x \).

(b) The independent variable, x, is the mileage. The dependent variable, y, is the expenses in dollars.

(c) The line may be plotted using two points as follows:

\[ 
\]

\[ x = 20, \ y = 25.58 \]

\[ x = 120, \ y = 61.16 \]

(d) For mileage of 110 miles, the model would predict expenses of \( 18.46 + (0.3558 \times 110) = 57.60 \), so K’s claim of $64 is not unreasonably high.

For mileage of 30 miles, the model would predict expenses of \( 18.46 + (0.3558 \times 30) = 29.13 \), so L’s claim of $48 is very high and should be investigated.

The model is based on data for mileages from 20 to 120 miles. It should not be used to extrapolate to 160 miles, but if it were to be so used it would predict expenses of \( 18.46 + (0.3558 \times 160) = 75.39 \). On this basis, M’s claim for $80 is not unreasonable.
Revision questions
Chapter 1

1 Fill in the gap.

In a . . . . . . . . . . . . . . . . . . . . economic decisions are made partly by free market forces of supply and demand, and partly by government decisions.

A command economy
B mixed economy
C free market economy
D centrally planned economy

2 The basic economic problem facing all economies is:

A maximising economic growth
B unemployment
C inflation
D allocating scarce resources

3 There are four main factors of production, each of which has an economic reward. Which one of the following statements about the factors of production is not correct?

A capital is rewarded with interest
B enterprise is rewarded with profit
C labour is rewarded with wages
D land is rewarded with property

4 Which of the following would not be regarded by economists as a factor of production?

A labour
B enterprise
C management
D capital

5 Which one of the following would not shift a country's production possibility frontier outwards (further away from the origin)?

A an increase in exports
B technical progress reducing production costs
C an increase in the working population
D an improvement in the literacy rate

6 Opportunity cost is:

A the cost of producing one extra unit of the commodity
B the lowest average cost of the commodity
C the total cost of the commodity
D the loss of the next best alternative

7 Which of the following best describes the opportunity cost of a program of immunisation?

A the actuarial valuation of the lives of those who are protected against the disease
B the cost of the vaccine
C the cost of providing the medical staff
D the work the medical staff cannot undertake as a result of the program
8 Assume that two small countries, X and Y, produce two commodities P and Q, and that there are no transport costs. One unit of resource in country X produces 4 units of P or 8 units of Q. One unit of resource in country Y produces 1 unit of P or 3 units of Q.

Which one of the following statements is true?

A country X has an absolute advantage over country Y in producing P and Q, and so will not trade.
B country X does not have an absolute advantage over country Y in producing P and Q.
C country Y has a comparative advantage over country X in producing Q.
D country X has a comparative advantage over country Y in producing both P and Q.

9 A good or service has a price if:
A it is both useful and scarce
B it is scarce only
C it is useful only
D it is useful, scarce and efficiently produced

10 Mr Smith has a limited income which restricts the number of different goods he can buy. Which one of the following best describes the position at which Mr Smith’s utility from purchasing different goods is maximised?
A total utility from each good is equal
B marginal utility from each good is equal
C marginal utility from each good is 0
D ratio of marginal utility to price is equal for each good
Chapter 2

1 Which one of the following would not lead directly to a shift in the demand curve for overseas holidays?
   A an advertising campaign by holiday tour operators
   B a fall in the disposable incomes of consumers
   C a rise in the price of domestic holidays
   D a rise in the price of overseas holidays

2 Which of the following is not a substitute for carpet?
   A ceramic floor tiles
   B wooden floorboard
   C vinyl flooring
   D carpet underlay

3 The demand for fashion goods is not influenced by:
   A price
   B allocative inefficiency among producers
   C the distribution of income among households
   D expectation of future price changes

4 Which one of the following would normally cause a rightward shift in the demand curve for a product?
   A a fall in the price of a substitute product
   B a reduction in direct taxation on incomes
   C a reduction in price of the product
   D an increase in the price of a complementary product

5 Consider the price and demand for flower vases. The price of cut flowers goes up sharply. Which of the following would you expect to happen?
   A The demand curve for flower vases will shift to the left and their price will rise.
   B The demand curve for flower vases will shift to the right and their price will rise.
   C There will be a movement along the demand curve for flower vases and their price will go down.
   D The demand curve for flower vases will shift to the left and their price will go down.

6 The demand curve for a resource may shift because of
   A a change in the demand for a good whose production is dependent on the resource.
   B concerns about potential harmful pollution from the resource.
   C a change in the price of a substitute resource.
   D all of the above.

7 All of the following are likely to lead an outward shift in the supply curve for a good, except
   A the introduction of cost-reducing technology.
   B an increase in the price of the good.
   C a decrease in the price of a resource used to make the good.
   D a decrease in taxes on producers.
8 The supply curve for sofas has moved to the right. Which of the following could have caused this shift?
A a decrease in the price of sofas
B a decrease in the price of futons (a substitute)
C a decrease in the cost of horsehair (a raw material used in making sofas)
D a decrease in the wage rate in the futon industry

9 Consumer surplus is
A the excess between what consumers are prepared to pay for a good or service, and the prevailing market price.
B the indirect tax producers pay on a good or service.
C the marginal utility gained by consuming one more unit of a good or service.
D the indirect tax consumers pay on a good or service.

10 When a price floor is imposed above the market equilibrium level the result is
A a shortage
B a surplus
C scarcity
D the good is no longer scarce
Chapter 3

1. When price is increased and there is zero change in demand, demand is:
   A. perfectly elastic
   B. unitary
   C. inelastic
   D. perfectly inelastic

2. The demand for a good rises from 20,000 to 25,000 following a reduction in price from $20 to $18. Using the point elasticity of demand method what is the price elasticity of demand?
   A. -2.1
   B. -2.5
   C. +2.1
   D. +2.5

3. Elasticity of demand for labour is influenced by all of the factors below except which one?
   A. the elasticity of supply of alternative factors of production
   B. the elasticity of supply of the final product
   C. the proportion of labour costs to total costs
   D. the ease of substituting other factors of production

4. The demand for a product will tend to be inelastic when
   A. it has very few close substitutes.
   B. it is very quickly consumed.
   C. it tends to be purchased by people on subsistence incomes.
   D. it has a wide range of different uses.

5. A company sells two products, widgets and fidgets. Widgets have a high price elasticity of demand. Fidgets are relatively price inelastic. The company decides to spend $2 million on an advertising campaign for each product, in order to increase demand. Which of the following statements would be true?
   A. The supply curve of both products would shift to the right, and for widgets by a greater proportion than for fidgets.
   B. The advertising campaign would be more successful for fidgets than for widgets.
   C. The cost of the advertising campaign for fidgets could be covered by raising the price of the product.
   D. The supply curve of both products would shift to the left, and for widgets by a greater proportionate amount than for fidgets.

6. The supply of skilled basketweavers is inelastic but not perfectly inelastic. There is an improvement in the productivity of basketweavers. Which of the following would you now expect to happen?
   A. The number of basketweavers in employment will go down, but their wages will go up.
   B. The number in employment will be unchanged, and their wages will go up.
   C. The number in employment will go up, and their wages will go up.
   D. The number in employment will go up, but their wages will go down.
7 The current price of lawnmowers is \( P \). The supply of lawnmowers is inelastic in the short run, but more elastic in the longer run. The demand for lawnmowers falls. What would you expect to happen?

A In the short term the price will fall quite sharply, but in the longer term, the price will rise a little as supply is reduced although it will remain less than \( P \).

B In the short term the price will be unchanged, but in the longer term, the new equilibrium will be at a lower output quantity and a price less than \( P \).

C In the short term the price will fall quite sharply, but in the longer term supply will be reduced so that the price rises above \( P \).

D In the short term, the price will be unchanged, but in the longer term the new equilibrium will be at a lower output quantity and a price higher than \( P \).

8 A demand curve, relating price to output, which is a straight line sloping downwards

A has a constant elasticity along its whole length.

B has a falling elasticity as you move down the line.

C has a rising elasticity as you move down the line.

D has an elasticity at each point on the line which has no predictable pattern.

9 Over the last two years, the average income in country A has increased from \$25 000 to \$27 000 per year. Over the same period, the quantity demanded of product P has increased from 8 000 to 8 320 per year.

Which one of the following is correct?

A demand for the good is price elastic

B the product is normal

C the income elasticity of the good is -0.5

D demand for the good is price inelastic

10 Sales of good X have increased from 15 000 to 17 000 following a rise in household incomes from \$40 000 to \$42 000.

What is the income elasticity of demand for good X?

A – 2.67

B – 1

C + 1

D + 2.67

11 A business currently sells 20 000 units of its product per month, at a price of \$10 per unit. The product has an income elasticity of – 0.8. If incomes rise by 5%:

A total revenue from the product will fall by \$40 000.

B total revenue from the product will fall by \$8 000.

C total revenue from the product will increase by \$8 000.

D total revenue from the product will increase by \$40 000.
Chapter 4

1. According to the traditional theory of the firm, the equilibrium position for all firms will be where:
   A. profits are maximised
   B. output is maximised
   C. revenue is maximised
   D. costs are minimised

2. Mark the following items on the graph below:
   I. Average total cost
   II. Marginal cost
   III. Average fixed cost
   IV. Average variable cost

3. The difference between total revenues and total costs (including both implicit and explicit costs) is the firm's
   A. accounting profit
   B. opportunity cost
   C. opportunity profit
   D. economic profit

4. Economic costs are
   A. the same as accounting costs.
   B. the opportunity costs of the factors of production employed by the firm.
   C. future costs incurred as a consequence of a decision.
   D. explicit costs.

5. Marginal cost is
   A. the addition to total cost of producing one more unit of output.
   B. the total cost divided by the quantity produced.
   C. the total fixed costs and the total variable cost.
   D. the difference between fixed and variable cost.
6 Which of the following always rise when a manufacturing business increases output?
I marginal costs
II fixed costs
III total costs
IV total variable costs
A I and IV only
B III and IV only
C I, III and IV only
D I, II, III and IV

7 Which of the following would not cause cost curves to shift?
A technological advances
B changes in labour efficiency
C changes in demand
D increased cost of factors of production

8 Which of the following best describes the law of diminishing returns?
As more labour is added to a fixed amount of capital
A total output will fall.
B increases in total output will become smaller for each additional unit of labour employed.
C the marginal revenue from each additional unit of output produced will decline.
D production costs will rise because higher wages will have to be paid to attract more labour.

9 The long-run average cost curve for a business will eventually rise because of
A the law of diminishing returns.
B increasing competition in the industry.
C limits to the size of the market for the good.
D diseconomies of scale.

10 ATC = average total cost
AVC = average variable cost
MC = marginal cost
Which of the following statements is correct?
A MC will equal ATC when ATC is at its minimum amount, but will not equal AVC when AVC is at its minimum.
B MC will equal AVC when AVC is at its minimum amount, but will not equal ATC when ATC is at its minimum.
C MC will equal ATC when ATC is at its minimum amount and AVC when AVC is at its minimum, which is at the same output level.
D MC will equal ATC when ATC is at its minimum amount and AVC when AVC is at its minimum, but this will occur at different output levels.
11 Which one of the following is not a source of economies of scale?
A the introduction of specialist capital equipment
B bulk buying
C the employment of specialist managers
D cost savings resulting from new production techniques

12 The law of diminishing returns can apply to a business only when
A all factors of production can be varied.
B at least one factor of production is fixed.
C all factors of production are fixed.
D capital used in production is fixed.

13 Diseconomies of scale occur in a business when:
A minimum efficient scale is reached
B short-run variable costs begin to rise
C diminishing returns to a limited factor of production begin to occur
D long-run average costs begin to rise

14 Which of the following propositions are false?
I It is possible for the average total cost curve to be falling while the average variable cost curve is rising.
II It is possible for the average total cost curve to be rising while the average variable cost curve is falling.
III Marginal fixed costs per unit will fall as output increases.
IV Marginal costs will be equal to marginal variable costs.
A I and III only
B I and IV only
C II and III only
D II and IV only
Chapter 5

1. The diagram shows a firm operating in a perfectly competitive market. The shaded area ABCD represents
   A. supernormal profit
   B. normal profit
   C. accounting profit
   D. short term loss

2. Which of the following are characteristics of perfect competition?
   I. large numbers of producers
   II. differentiated goods
   III. the absence of long-run excess profits
   IV. freedom of entry to and exit from the industry
   A. I, II and III only
   B. I, III and IV only
   C. II, III and IV only
   D. I, II, III and IV

3. What is a monopsony?
   A. a market with a single buyer
   B. a market with a single supplier
   C. a market dominated by a small number of large buyers
   D. a market dominated by a small number of large suppliers

4. Oligopoly markets typically do not display price competition because
   A. barriers to entry exist.
   B. products are clearly differentiated.
   C. producers’ decisions are interdependent.
   D. there is always a price leader.
5. A monopolist’s average revenue curve always slopes downwards because
   A. economies of scale exist in distribution.
   B. there are allocative inefficiencies.
   C. market demand increases as price falls.
   D. marginal revenue can be negative.

6. In a perfectly competitive market, all producers charge the same price because
   A. they are all profit maximisers.
   B. they have the same costs.
   C. the product is homogeneous.
   D. none of the above.

7. The purpose of a cartel is to
   A. rationalise production.
   B. reduce consumer uncertainty.
   C. standardise product quality.
   D. ensure that a dominant group of producers all charge the same price.

8. Monopolies maintain supernormal profits in the long run because
   A. they are more efficient than other firms.
   B. unlike other firms, they benefit from economies of scale.
   C. they can advertise.
   D. there are barriers to the entry of competitors.

9. The diagram above shows a profit-maximising monopoly producer of high-definition television sets
   that was originally producing at Q₁ with price P₁ and marginal cost curve MC and marginal revenue
   curve MR. Now, as a result of changed conditions (not shown), it sets a price P₂.

   This change could have occurred because of
   A. an increase in consumer incomes.
   B. a reduction in labour costs.
   C. an increase in research costs.
   D. a fall in the price of normal television sets.

10. Price discrimination in monopoly markets can be used to
    A. increase output and allocative inefficiency.
    B. increase output and reduce allocative inefficiency.
    C. reduce output and allocative inefficiency.
    D. reduce output and increase allocative inefficiency.
11 CM Co operates in a perfectly competitive market. The best way for the firm to increase its profits is to
A raise prices.
B reduce costs.
C set barriers to entry.
D increase production.

12 In an industry with no barriers of entry and with ease of both exit and entry,
A oligopolies will prevail.
B monopoly will prevail.
C higher than normal economic profits will appear in the long run.
D higher than normal economic profits will be eliminated in the long run.

13 Which of the following are reasons why monopoly is inefficient?
I Under monopoly, price will be higher than marginal cost.
II Monopolists do not produce at the level of output where average costs are at their lowest level.
III Profits can be used to invest in research and development.
A II only
B I and II only
C I and III only
D II and III only
Chapter 6

1. The diagram below illustrates the effect of an indirect tax on the supplier and the consumer of a good. The total amount of the tax is the distance AC. Which part of this is paid by the consumer of the good?

   ![Diagram](image)

   A. AB  
   B. AC  
   C. CB  
   D. none

2. A reduction in government regulation of industry is unlikely to produce which of the following undesirable effects?

   A. an increase in market imperfections  
   B. lower quality of service  
   C. cyclical unemployment  
   D. reduced provision of public goods

3. Which of the following are merit goods?

   I. national defence  
   II. health services  
   III. education  
   IV. street lighting  

   A. II and III only  
   B. II and IV only  
   C. I, II, and III only  
   D. I, II, III and IV

4. Arguments for allocating resources through the market mechanism rather than through government direction include three of the following. Which one is the exception?

   A. It provides a more efficient means of communicating consumer wants to producers.  
   B. It ensures a fairer distribution of income.  
   C. It gives more incentive to producers to reduce costs.  
   D. It encourages companies to respond to consumer demand.
5. The primary burden of a tax will fall most heavily on buyers when
   A. demand and supply are both elastic.
   B. demand is elastic and supply is inelastic.
   C. demand is inelastic and supply is elastic.
   D. demand and supply are both inelastic.

6. The excess burden of tax is minimised when
   A. demand and supply are both elastic.
   B. demand is elastic and supply is inelastic.
   C. demand is inelastic and supply is elastic.
   D. demand and supply are both inelastic.

7. Which of the following solutions to market failure is least likely to distort economic decision?
   A. establishing property rights
   B. imposition of a price floor
   C. imposition of a quota
   D. subsidy to encourage production

8. Which one of the following is the best example of a social good?
   A. national defence system
   B. health service
   C. public transport
   D. education service

9. Which one of the following statements most accurately describes merit goods.
   A. They will not be provided by the free market economy.
   B. They are non-exclusive.
   C. They are under-provided in the free market economy.
   D. They are indivisible.

10. Which one of the following is the best example of an external social cost?
    A. restrictions on the supply of oil leading to an increase in fuel prices
    B. bad weather leading to a poor harvest and an increase in cereal prices
    C. possible illnesses which people may suffer from drinking too much
    D. pollution emitted from a factory which has caused local residents to become ill
Chapter 7

1. GNP (Gross National Product) at factor cost may be best defined as
   A the total of goods and services produced within an economy over a given period of time.
   B the total expenditure of consumers on domestically produced goods and services.
   C all incomes received by residents in a country in return for factor services provided domestically and abroad.
   D the value of total output produced domestically plus net property income from abroad, minus capital consumption.

2. Which of the following is a measure of economic development?
   A output
   B income
   C expenditure
   D poverty

3. Net National Product at factor cost
   + Capital consumption
   + Indirect taxes on expenditure
   – Subsidies
   equals
   A Gross National Product at market prices
   B Gross National Product at factor cost
   C Gross Domestic Product at market prices
   D Gross Domestic Product at factor cost

4. Which of the following cannot be termed a 'transfer payment' for the purpose of national income accounting?
   A interest paid to holders of government stock
   B salaries paid to Members of Parliament
   C payments of State pensions
   D social security payments

5. Which of the following investments creates an injection into the circular flow of income?
   A An increase by a firm in its inventories of finished goods, prior to a marketing campaign
   B The purchase by a superannuation fund of shares in a newly privatised company
   C The purchase of a second-hand piece of farming machinery with their savings by a farming co-operative group
   D The takeover of one company by another company
6 The following data relate to national Income statistics in Muvovia, which are compiled in the same way as in Australia:

<table>
<thead>
<tr>
<th></th>
<th>20X8 Actual prices</th>
<th>20X9 Actual prices</th>
</tr>
</thead>
<tbody>
<tr>
<td>Consumers' expenditure</td>
<td>$200 000</td>
<td>$225 000</td>
</tr>
<tr>
<td>General government final consumption</td>
<td>$70 000</td>
<td>$74 000</td>
</tr>
<tr>
<td>Gross domestic fixed capital formation</td>
<td>$54 000</td>
<td>$60 000</td>
</tr>
<tr>
<td>Imports</td>
<td>$92 000</td>
<td>$99 000</td>
</tr>
<tr>
<td>Exports</td>
<td>$93 000</td>
<td>$94 000</td>
</tr>
<tr>
<td>Taxes on expenditure</td>
<td>$52 000</td>
<td>$50 000</td>
</tr>
<tr>
<td>Subsidies</td>
<td>$8 000</td>
<td>$10 000</td>
</tr>
</tbody>
</table>

The general rate of inflation in Muvovia between 20X8 and 20X9 was 10%. The real change in Gross Domestic Product at market prices between 20X8 and 20X9, in percentage terms, was:

A a fall of about 1%
B a rise of about 1%
C a rise of about 5%
D a rise of about 9%

7 Which one of the following statements is correct?

A Two countries with the same total national income will have roughly the same living standards.
B Services provided free to the public, such as police work and state education, are valued at opportunity cost in the national income statistics.
C Official statistics might over-estimate the national income for a country with a strong black economy.
D Gross National Product figures are often used in preference to Net National Product figures because of the difficulty in calculating capital consumption.

8 A country has a GDP at factor cost of $150m. Government expenditure was $25m, subsidies were $10m and taxes were $17m. What is its GDP at market prices?

A $143m
B $157m
C $168m
D $182m

9 In recent years, country W has enjoyed a sustained period of economic growth. However, this rate of growth has now started to slow. Which one of the following is country W most likely to have observed when the rate of growth began to slow?

A the level of imports increased
B inventory levels increased
C inflation rates increased
D investment in new equipment increased
Chapter 8

1. Identify the full employment level of real national income on the diagram below:

![Diagram of AS and AD curves with Y1 and Y2 as the axes and P1 and P2 as the price levels.]

A. \( P_1 \)
B. \( Y_1 \)
C. \( Y_2 \)
D. \( P_2 \)

2. The marginal propensity to consume measures

A. the relationship between changes in consumption and changes in consumer utility.
B. the proportion of household incomes spent on consumer goods.
C. the proportion of total national income spent on consumer goods.
D. the relationship between changes in income and changes in consumption.

3. The accelerator principle states

A. how an initial increase in a component of national income leads to much greater eventual rise in national income.
B. that a small change in investment will lead to a much greater change in the output of consumer goods.
C. that a small change in the output of consumer goods will lead to a much greater change in the production of capital goods.
D. that changes in investment level are the cause of trade cycles.

4. Other things remaining the same, according to Keynes, an increase in the money supply will tend to reduce

A. interest rates.
B. liquidity preference.
C. the volume of bank overdrafts.
D. prices and incomes.
5 The government has no involvement in the economy in Ostland, and the country has no overseas trade (no imports or exports). If the marginal propensity to consume is 0.6 in Ostland, what is its marginal propensity to save?

A 0.4  
B 1.67  
C 2.5  
D it will depend on the level of national income

6 In Gondwana, an open economy, for every additional $1 of income, 20 per cent is taken as taxes, 10 per cent is saved and 50 per cent is spent on domestically produced goods. What is the value of the multiplier in Gondwana?

A 0.8  
B 1.25  
C 2  
D 3.33

7 The government has decided to introduce a deflationary policy to manage the rate of economic growth in Kelgravia. This policy could include

A reducing taxation rates.  
B increasing the levels of investment in the economy.  
C increasing levels of government spending.  
D increasing interest rates.

8 Which one of the following is most likely to increase aggregate demand in an economy?

A increased saving  
B increased spending on imports  
C increased taxation  
D increased investment

9 Interest rates in Isopia have recently been reduced. Which one of the following is most likely to result from this cut in interest rates?

A an increase in savings  
B an increase in spending  
C a decrease in borrowing  
D a decrease in consumption

10 An increase in the marginal propensity to consume will

A shift aggregate supply to the right.  
B lead to a fall in the level of national income.  
C increase the size of the multiplier.  
D reduce injections into the economy.

11 Which one of the following will not cause a fall in the level of economic activity in a country?

A a reduction in the government budget deficit  
B an increase in interest rates  
C an increase in the level of cyclical unemployment  
D an increase in tax rates
12. The following diagram shows aggregate demand for a closed economy with no government sector:

What is the equilibrium level of spending in this economy?

A. OP  
B. OQ  
C. OR  
D. OS
Chapter 9

1. A government may seek to reduce the rate of demand-pull inflation by any of the following means except:
   A. reducing interest rates
   B. increasing sales tax
   C. applying more stringent controls over bank lending
   D. reducing public expenditure

2. Structural unemployment is best defined as unemployment caused by:
   A. defects in the industrial and commercial structure
   B. a long-term decline in a particular industry
   C. a mismatch between available jobs and the unemployed
   D. a switch from labour-intensive to capital-intensive production methods

3. Unemployment that rises and falls in a regular pattern not associated with the overall economic cycle is called ........................................ unemployment.
   A. cyclical
   B. frictional
   C. seasonal
   D. residual

4. Which of the following would not lead to cost push inflation?
   A. rising import prices
   B. increase in wages
   C. increases in indirect taxation
   D. high consumer expenditure such that aggregate demand exceeds aggregate supply

5. Unemployment that occurs in specific industries as a result of long-term changes in the patterns of demand and supply is:
   A. seasonal
   B. structural
   C. cyclical
   D. frictional

6. The ........................................ says that the expansion of aggregate demand to reduce unemployment below its natural rate will only produce inflation.
   A. short-run Phillips curve
   B. natural rate hypothesis
   C. equilibrium rate hypothesis
   D. wage control hypothesis

7. Which of the following is a homogenous property of money?
   A. money needs to be divisible
   B. money must be easy to use
   C. money should be scarce
   D. money should not deteriorate over time
8 The Reserve Bank of Australia's measure of money $M_3$ is made up of
A  bank currency and all deposits of the private non-banking sector.
B  bank currency and all deposits and borrowings from the private sector.
C  bank currency and notes and coins in circulation.
D  bank currency and bank deposits placed with the RBA.

9 Which one of the following measures has not been recommended by Friedman and the monetarist economists as a means of reducing the natural rate of unemployment to a lower level?
A  measures to stimulate consumer demand for more goods
B  schemes to retrain workers in new job skills
C  measures to cut trade union power
D  restructuring the income tax system

10 If expected inflation is negative, the nominal interest rate is........................................than the real interest rate, and........................................
A  greater, positive
B  greater, negative
C  less, positive
D  less, negative

11 If the real rate of interest is 3 per cent per annum and the expected rate of inflation is 6 per cent per annum, the nominal interest rate will be approximately:
A  ½ per cent
B  2 per cent
C  3 per cent
D  9 per cent

12 If the nominal rate of interest in Erewhon is 5 per cent per annum and the annual rate of inflation is 2 per cent, what is the real rate of interest (to the nearest whole number)?
A  2 per cent
B  3 per cent
C  5 per cent
D  7 per cent

13 According to monetarist economists, in the long-run the Phillips curve is:
A  horizontal
B  vertical
C  downward-sloping
D  upward-sloping

14 According to Keynesian economists:
A  the equilibrium level of national income will be reached at the point of full employment.
B  the equilibrium level of national income may occur at any level of unemployment.
C  the equilibrium level of national income will be reached at the point where inflation is zero.
D  the equilibrium level of national income will always occur at a high level of unemployment.
Chapter 10

1. The budget deficit is defined as
   A. the gap between government expenditure and government receipts.
   B. the gap between government expenditure and tax receipts.
   C. the gap between government purchases and tax revenues.
   D. the net interest paid on outstanding debt.

2. Which of the following are effects of reduced interest rates?
   I. Consumer spending will increase.
   II. Business investment will be encouraged.
   III. Saving will increase.
   A. I only
   B. I and II only
   C. II and III only
   D. I, II and III

3. If a reduction in the taxes on alcoholic drinks resulted in a less even distribution of wealth in society, with a greater proportion of wealth in the hands of the rich sections of society, we could conclude that, on average
   A. people with low incomes spend more on alcoholic drinks than people with high incomes.
   B. people with low incomes spend less on alcoholic drinks than people with high incomes.
   C. people with low incomes spend a bigger proportion of their income on alcoholic drinks than people with high incomes.
   D. people with low incomes spend a lower proportion of their income on alcoholic drinks than people with high incomes.

4. Which one of the following is an aspect of fiscal policy measures by the government?
   A. to raise short-term interest rates in the money markets
   B. to support the exchange rate for the country’s currency
   C. to control growth in the money supply
   D. to alter rates of taxation

5. Which of the following are true?
   I. If interest rates are low but expected to rise, liquidity preference will be high.
   II. If interest rates are high but expected to fall, liquidity preference will be high.
   III. The demand for money is high when interest rates are low.
   IV. The demand for money is high when interest rates are high.
   A. I only
   B. II only
   C. I and III only
   D. II and IV only
6 If a government wants to exercise an expansionary fiscal policy it should
A reduce taxes, increase government expenditure.
B increase taxes, reduce government expenditure.
C increase money supply, reduce government expenditure.
D reduce money supply, increase government expenditure.

7 Which one of the following is likely to be a government objective rather than part of its macroeconomic policy?
A lower interest rates
B lower inflation
C lower taxation rates
D lower government spending

8 Recent legislation in country E has made the labour market more flexible and this has led to a shift in the aggregate supply curve. This is likely to
A reduce price levels in the economy but increase national income.
B reduce price levels in the economy and reduce national income.
C increase price levels in the economy but reduce national income.
D increase price levels in the economy and increase national income.

9 The government budget deficit in country AB has significantly increased during the last fiscal year. Which of the following is most likely to result from the increase in the budget deficit?
A Country AB’s equilibrium level of national income will fall.
B Employment levels in country AB will fall.
C Aggregate demand in country AB’s economy will rise.
D Government spending in country AB in the short term will be reduced.

10 The government of Newland has decided to implement a contractionary fiscal policy in order to control economic growth in the country. What will be the most likely impact of this policy, if all other factors remain the same?
A National output will fall but the rate of inflation will rise.
B National output will fall and the rate of inflation will fall.
C National output will fall but the rate of interest will rise.
D National output will rise and the rate of interest will rise.

11 The central bank in country K is trying to reduce the growth of money supply and has decided to control bank lending. The central bank has recently imposed a restriction on the amount commercial banks can lend to property companies.
This restriction is best described as being an example of
A quantitative controls
B a special directive (a direct control)
C open market operations
D managing reserve requirements
12 Which one of the following is not a function of a central bank?
A the conduct of fiscal policy
B management of the national debt
C holder of the foreign exchange reserves
D lender of last resort

13 Which of the following is a central bank least likely to be responsible for?
A fixing the general level of interest rates
B regulating the banking industry
C determining social security payments
D maintaining national reserves of foreign currency

14 The central bank influences interest rates by
A controlling the amount of notes and coins in circulation.
B changing the reserve requirements.
C setting the rate at which it lends funds to commercial banks.
D imposing credit limits.
Chapter 11

1 How many of the following are arguments in favour of privatisation?
   I Privatised companies may be more efficient than nationalised companies.
   II Denationalisation will create wider share ownership, which will encourage staff to develop a better understanding of their business and profit drivers.
   III Privatised companies may have a more profit-oriented management culture than nationalised ones.
   IV Industries can benefit from significant economies of scale arising from natural monopolies.
   A 1  
   B 2  
   C 3  
   D 4

2 In Australia, who would be responsible for investigating suspected price fixing or collusion in the retail sector?
   A The Treasurer  
   B The Retail Industry Association  
   C ACCC  
   D NSW Office of Fair Trading

3 Which of the following is not an argument in favour of privatisation?
   A Privatisation provides a source of money for the government.
   B Privatisation reduces bureaucracy and political meddling in the industries concerned.
   C Privatised industries are more likely to respond to the public interest.
   D Privatisation encourages a more profit-oriented management culture.

4 A number of industries in Freeland have recently been privatised. As a result of the privatisation
   A assets have been transferred from the public sector, and payments for those assets went to the private sector.
   B there is likely to be a more even distribution of wealth and income in Freeland.
   C assets have been transferred from the public sector and payments for those assets went to the government.
   D there has been a reduction in competition.

5 In Australia, the Competition and Consumer Act (CCA) 2010 prohibits
   A monopolies from making supernormal profits.
   B monopolies from maintaining barriers to entry.
   C mergers in all cases.
   D anti-competitive arrangements.
6 Which one of the following statements is incorrect?
A Privatisation can be a means of widening share ownership.
B Privatising a public sector industry will permit economies of scale.
C Denationalisation provides a source of money for the government.
D Privatising a public sector industry is likely to make it more responsive to the profit motive.

7 In Australia, the 2007-08 Gini coefficient shows that
A there has been an increase in income inequality since 1994/95.
B there has been a decrease in income inequality since 1994/95.
C Australia is a highly unequal society in terms of income distribution.
D Australia almost has a line of perfect equality.

8 Which form of taxation is the most effective in addressing income inequality?
A regressive taxation
B proportional taxation
C progressive taxation
D indirect taxation

9 What is the Gini coefficient of perfect income equality?
A 1
B 0.75
C 0.25
D 0
Chapter 12

1. A company has recorded the following data on days lost through employee sickness in a year.

<table>
<thead>
<tr>
<th>Days lost per employee</th>
<th>Number of employees</th>
</tr>
</thead>
<tbody>
<tr>
<td>At least</td>
<td>Less than</td>
</tr>
<tr>
<td>2</td>
<td>6</td>
</tr>
<tr>
<td>6</td>
<td>8</td>
</tr>
<tr>
<td>8</td>
<td>10</td>
</tr>
<tr>
<td>10</td>
<td>14</td>
</tr>
</tbody>
</table>

Which of the following histograms represents the data most accurately?

A: Graph 1

B: Graph 2

C: Graph 3

D: Graph 4

2. In a histogram in which one class interval is one and a half times as wide as the remaining classes, the height to be plotted in relation to the frequency for that class is

A. \( \times 1.5 \)
B. \( \times 1.00 \)
C. \( \times 0.75 \)
D. \( \times 0.67 \)

3. A pie chart is used to display the following data:

<table>
<thead>
<tr>
<th>Sales</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Region A</td>
<td>26</td>
</tr>
<tr>
<td>Region B</td>
<td>41</td>
</tr>
<tr>
<td>Region C</td>
<td>33</td>
</tr>
</tbody>
</table>

Economics and Markets
What angle (in degrees) on the pie chart will be used to represent Region C's share of sales?

A 33
B 237.6
C 118.8
D 59.4

4 In a histogram, one class is two thirds of the width of the remaining classes. If the score in that class is 25, the correct height to plot on the histogram is

A 16.67
B 25.00
C 37.50
D 41.67

5 In an experiment to determine if antibiotics increase the final dressed weight of cattle, the following were measured on each animal in the study:

- sex
- initial weight
- weight gain
- grade of meat, where grade is recorded as (A, B, or C)

The scale of measurement of these variables is:

A nominal, ratio, interval, nominal
B nominal, ratio, ratio, nominal
C nominal, ratio, ratio, ordinal
D ordinal, ratio, ratio, ordinal

6 The following table shows the typical salary of part qualified accountants in five different regions of Australia:

<table>
<thead>
<tr>
<th>Area</th>
<th>Typical salary</th>
</tr>
</thead>
<tbody>
<tr>
<td>New South Wales</td>
<td>$41 500</td>
</tr>
<tr>
<td>Western Australia</td>
<td>$40 800</td>
</tr>
<tr>
<td>Queensland</td>
<td>$38 200</td>
</tr>
<tr>
<td>Northern Territory</td>
<td>$37 500</td>
</tr>
<tr>
<td>Tasmania</td>
<td>$36 700</td>
</tr>
</tbody>
</table>

The best diagram to draw to highlight the differences between areas is

A a pie chart
B a multiple bar chart
C a percentage component bar chart
D a simple bar chart

7 A random sample of 500 households was selected and several variables are recorded for each household. Which of the following is not correct?

A Household total income is a ratio scaled variable.
B Household income (which averages about $35 000) and was rounded to the nearest $100 can be treated as a continuous variable even though it is 'discrete'.
C Socioeconomic status was coded as 1 = low income, 2 = middle income, 3 = high income and is an interval scaled variable.
D The primary language used at home is a nominal variable.
Which of the following statements are not true?

I. If a sample is selected using random sampling, it will be free from bias.
II. A sampling frame is a numbered list of all items in a sample.
III. In cluster sampling there is very little potential for bias.
IV. Stratified sampling involves dividing the population into strata and then selecting a random sample from each stratum.

A. II only
B. II and III only
C. I, II and III only
D. I, II, III and IV
Chapter 13

1. A population consists of four observations: \{1, 3, 5, 7\}. What is the variance?
   A 2
   B 4
   C 5
   D 6

2. A sample consists of four observations: \{1, 3, 5, 7\}. What is the standard deviation?
   A 2
   B 2.58
   C 6
   D 6.67

3. Which of these options is true where data is highly positively skewed?
   A The least representative average is the median.
   B The mode will overestimate the average.
   C The mean will tend to overestimate the average.
   D The mean, mode and median will produce equally representative results.

4. Four friends take an IQ test. Their scores are 96, 100, 106, 114. Which of the following statements is true?
   A The mean is 103.
   B The mean is 104.
   C The median is 100.
   D The median is 106.

5. You took a test and got 75 per cent. Three possibilities of how the rest of the class performed on this test appear below. In which of the three possibilities did you score well above the centre of the distribution?

<table>
<thead>
<tr>
<th>Outcome A</th>
<th>Outcome B</th>
<th>Outcome C</th>
</tr>
</thead>
<tbody>
<tr>
<td>67</td>
<td>54</td>
<td>74</td>
</tr>
<tr>
<td>69</td>
<td>64</td>
<td>75</td>
</tr>
<tr>
<td>71</td>
<td>67</td>
<td>77</td>
</tr>
<tr>
<td>75</td>
<td>71</td>
<td>79</td>
</tr>
<tr>
<td>78</td>
<td>71</td>
<td>86</td>
</tr>
<tr>
<td>81</td>
<td>75</td>
<td>93</td>
</tr>
<tr>
<td>84</td>
<td>77</td>
<td>97</td>
</tr>
</tbody>
</table>

   A Outcome A
   B Outcome B
   C Outcome C
   D It is impossible to say

6. When measurement data is not normally distributed, which of the following measures of central tendency and variability are recommended?
   A median and variance
   B mean and standard deviation
   C median and range
   D mean and range
When data is normally distributed, what percentage of scores falls within the one standard deviation interval around the mean (-1 SD to +1 SD)?

A 68.26%
B 100%
C 34.13%
D 95.00%

The graph showing the age of getting a driver's license in California starts and peaks at age 16, and decreases from there. This shape most closely resembles what type of distribution?

A normal
B binomial
C uniform
D exponential

In a binomial distribution we

A count the number of successes until an event is obtained.
B count the number of trials until a success is obtained.
C count the number of successes in a finite number of trials.
D count the number of trials until the number of successes equals the number of events.

In an exponential distribution, the mean is larger than the median.

A true always
B true if the data is positively skewed
C true if the data is negatively skewed
D false always

Would the variance of 10, 12, 17, 20, 25, 27, 42 and 45 be larger if these numbers represented a sample or a population?

A sample
B population
C it depends on the size of the sample
D it depends on the size of the population
Chapter 14

1. Probability can be thought of as
   A. symmetrical outcomes
   B. relative frequencies
   C. subjective
   Which of the statement(s) are correct?
   1. A only
   2. B only
   3. C only
   4. A and B only
   5. A and C only
   6. B and C only
   7. none of them
   8. All of them

2. A golfer has 12 golf shirts in his closet. Suppose nine of these shirts are white and the others blue. He gets dressed in the dark with a fresh shirt each day, so he just grabs a shirt and puts it on. He plays golf two days in a row and does not do laundry.
   What is the likelihood both shirts selected are white?

3. If we randomly draw a card using a standard deck of cards (without jokers), what is the probability of drawing a king OR a club?

4. You draw a card from a deck of cards, put it back, and then draw another card. What is the probability that the first card is a heart and the second card is black?
5 Two Events in a Venn Diagram
Suppose that the probability that the team makes the playoffs is .6, the probability that the coach gets rehired is .8, and the probability that the team makes the playoffs AND the coach gets rehired is .5. Call making the playoffs event A and call getting rehired event B. Identify the correct Venn diagram:

A

B

C

D

6 Select all that apply. Which of the following pairs are independent events?
A two coin flips
B a student’s mid-term and final grade in a class
C draw an ace, leave it out, then draw an ace again
D draw an ace, put it back, then draw an ace again

7 You win $1 every time you flip a coin and get heads. You win $3 every time you roll a dice and get a five. You roll a dice and flip the coin one time each. What is the probability that you win money from at least one game?

8 A survey showed that 60 per cent of all adults in your city take public transportation to work. If three people are chosen at random, what is the probability that they will all take public transportation to work?

9 If you throw a dice four times, what is the probability that one or more of your throws will come up with a four?
Chapter 15

1 Suppose a researcher conducts an experiment to test a hypothesis. If she doubles her sample size, which of the following will increase?

I. the power of the hypothesis test.
II. the effect size of the hypothesis test.
III. the probability of making a Type II error.

A. I only
B. II only
C. III only
D. I, II and III

2 Other things being equal, which of the following actions will reduce the power of a hypothesis test?

I. increasing sample size.
II. increasing significance level.
III. increasing beta, the probability of a Type II error.

A. I only
B. II only
C. III only
D. I, II and III

3 Sulphur compounds cause 'off-odours' in wine, so winemakers want to know the odour threshold, the lowest concentration of a compound that the human nose can detect. The odour threshold for dimethyl sulphide (DMS) in trained wine tasters is about 25 micrograms per litre of wine. The untrained noses of consumers may be less sensitive, however. Here are the DMS odour thresholds for 10 untrained students.

31 31 43 36 26 34 32 30 20 24

Assume that the standard deviation of the odour threshold for untrained noses is known to be $\sigma = 7$ micrograms per litre of wine. We are trying to determine if the mean odour threshold for students is higher than the published threshold.

What is the value of the test statistic (z score statistic) to two decimal places?

4 Here are the IQ test scores of 31 seventh-grade female students in a Midwest school district.

114 100 104 89 102 91 114 114 103 105
108 130 120 132 111 128 118 119 86 72
111 103 74 112 107 103 98 96 112 93

Treat the 31 girls as a simple random sample of all seventh-grade girls in the school district. Suppose that the standard deviation of IQ scores in this population is known to be $\sigma = 15$. We are trying to determine if the mean IQ score for the population differs from 100.

What is the value of the test statistic (z score statistic) to two decimal places?

5 A pharmaceutical manufacturer does a chemical analysis to check the potency of products. The standard release potency for cephalothin crystals is 910 with a standard deviation of 8.5. An assay of 16 lots gives the following potency data.

897 914 913 906 916 918 905 921
918 906 895 893 908 906 907 901

We are trying to determine if there is significant evidence that the mean potency is greater than the standard release potency.

What is the test statistic (z score statistic) rounded to two decimal places?
Chapter 16

1 Monthly sales of overcoats (Y in $000s) are found to be related to the average monthly daytime temperature (X in degrees C) by the regression equation Y = 32 – 1.6X. Which of the following is true?

A When temperature increases by 1 degree, sales fall by $3 200.
B When temperature increases by 1 degree, sales fall by $1 600.
C When temperature increases by 1 degree, sales fall by $1.60.
D When temperature increases by 1 degree, sales increase by $1 600.

2 20 pairs of values of (X, Y) with X ranging from 15 to 45 were used to obtain the regression equation Y = 480 – 5X. The correlation coefficient is –0.95. It has been estimated that when X = 10, Y = 430. Which of the following reduces the reliability of the estimate?

A the sample size
B the magnitude of the correlation coefficient
C X = 10 being outside the range of the sample data
D the correlation being negative

3 The value of b in the regression equation

A cannot be negative.
B equals one when there is perfect correlation.
C measures the scatter of points around the regression line.
D measures the increase in the y variable for unit increase in x.

4 On the basis of the scatter diagram above, which of the following equations would best represent the regression line of Y on X?

A Y= –X + 8
B Y= X + 8
C Y= X – 8
D Y= –X – 8

5 A correlation coefficient with a value of 0.9

A indicates that x causes y.
B indicates that y causes x.
C indicates that the relationship between x and y is probably too strong to have been caused by chance.
D shows that x and y have something in common.
6. The regression equation linking $x$ and $y$ is $y = 5x - 24$. Which of the following is correct?

I. The slope of the equation if plotted on a graph is 5.
II. The line cuts the y-axis at $-24$ if plotted on a graph.
III. The slope of the equation if plotted on a graph is $-24$.
IV. The line cuts the y-axis at 4.8 if plotted on a graph.

A. I only
B. I and II only
C. I, II and III only
D. I, II, III and IV

7. The value of the product moment correlation coefficient, $r$, for the data, is $r = -0.9260$.

The value of the coefficient of determination is

A. 0.9260
B. $-0.8575$
C. 0.8575
D. $-0.9623$

8. What is the equation of the least-squares regression line of $x$ on $y$ for the following five data pairs?

<table>
<thead>
<tr>
<th>$x$</th>
<th>$y$</th>
</tr>
</thead>
<tbody>
<tr>
<td>9</td>
<td>2</td>
</tr>
<tr>
<td>10</td>
<td>3</td>
</tr>
<tr>
<td>9</td>
<td>1</td>
</tr>
<tr>
<td>8</td>
<td>1</td>
</tr>
<tr>
<td>9</td>
<td>2</td>
</tr>
</tbody>
</table>

$x = \underline{\text{ }}$

State each value to two decimal places.
Answers to revision questions
Chapter 1

1 B In a mixed economy, economic decisions are made partly by free market forces of supply and demand, and partly by government decisions. In a 'command' or 'centrally planned' economy, decisions are made entirely by the government. In a 'free market' economy, decisions are made solely on the basis of the market forces of supply and demand.

2 D The basic economic problem is one of allocating scarce resources and economics is the study of how those scarce resources are or should be used.

3 D The correct answer should be: Land is rewarded with rent.

4 C Management is a specialised type of labour. A, B and D are factors of production; the missing factor is land.

5 A This is merely a use of production; the other three options will increase the production capacity in the country.

6 D This is a definition of opportunity cost.

7 D The opportunity costs represent the alternative work which is forgone to provide the vaccination program.

8 C Country X has an absolute advantage over country Y in making P and Q, because 1 unit of resource in country X will make more of either P or Q than one unit of resource in country Y. However, international trade should still take place because of comparative advantage in producing P and Q.

The opportunity costs of producing a unit of P is \( \frac{4}{8} = \frac{1}{2} \) unit of Q in country X and only \( \frac{1}{3} \) unit of Q in country Y.

Similarly, the opportunity cost of producing a unit of Q is 2 units of P in country X and 3 units of P in country Y.

Country X has a comparative advantage in producing P and country Y has a comparative advantage in the production of Q. International trade should be beneficial for both countries, with country X exporting P and country Y exporting Q.

9 A A good or service has a price if it is useful as well as scarce.

10 D Imagine you can buy a second car for $10 000 or buy a bike for $100 and they both give you the same extra utility. You wouldn't choose to buy the car as you're paying much more to achieve the same utility as you could get from buying the bike. If you get 10 times more utility for one thing compared to another you would be prepared to pay 10 times more for it.
Chapter 2

1. D A rise in the price of overseas holidays will lead to a movement along the demand curve rather than a shift in the demand curve.

2. D Carpet underlay is a complement to carpet.

3. B This is a supply-side factor.

4. B A reduction in income tax will increase real household income, and so demand for normal products will shift to the right – quantity demanded will be greater at any given price. Items A and D will cause a leftward shift in the demand curve. Item C would cause a movement to the right along the demand curve.

5. D It is assumed that cut flowers and flower vases are complementary goods. The rise in price of cut flowers will have an adverse effect on demand for flower vases, and the demand curve for flower vases will shift to the left. Given no change in supply conditions for vases, the new equilibrium price for vases will be lower.

6. D Variations in the conditions of demand create a shift in the demand curve. A variation in the conditions of demand for a resource will result from a change in the demand for a good whose production is dependent on the resource (A), concerns about the potential harmful pollution from the resource (B) and a change in the price of a substitute resource (C).

7. B A change in the price of a good will lead to a movement along the supply curve, not a shift in the curve itself.

8. C As costs incurred in a producing sofas have fallen, producers will be prepared to produce more at any given price. A change in the price of sofas will lead to a movement along the supply curve. A decrease in the price of a substitute will lead to a decrease in the demand for sofas.

9. A Consumer surplus is the excess between what consumers are prepared to pay for a good or service and the prevailing market price they have to pay to purchase it.

10. B Although there is a surplus, this does not mean that the good is no longer scarce. At the current price, however, buyers desire less of a good than sellers want to bring to the market so there is a surplus. Scarcity indicates that a good is less freely available than consumers would like.
Chapter 3

1. D  Demand is perfectly inelastic if a change in price has no impact on the quantity demanded.

2. B  Price elasticity of demand = –2.5

\[ \text{Price elasticity of demand} = \frac{% \text{ change in quantity demanded}}{% \text{ change in price}} = \frac{5000/20,000}{-2/20} = \frac{25}{-10} = -2.5 \]

3. B  The elasticity of supply of the final product will not be an influencing factor.

4. A  Statement A is correct: demand will tend to be elastic when the product has a large number of close substitutes. The rate of consumption (statement B) could be irrelevant to elasticity. However, a high rate of demand/consumption might suggest consumer goods, which tend to have elastic demand. Statement C is incorrect. If a product is bought by people on subsistence incomes, a rise in its price is unlikely to result in higher total spending on the product (i.e. demand will not be inelastic) and if demand switches to cheaper substitutes, which is likely, demand for the product will be price elastic. Luxury goods (statement D) tend to be price elastic.

5. B  Oddly enough, advertising (a form of non-price competition) is more likely to be successful for products with a low price elasticity of demand – i.e. for products whose demand is influenced by factors other than price. Statements A and D are incorrect because the supply curves of the product will be unaffected. Statement C is not necessarily correct because the higher total profits (and revenue) from their higher price will not necessarily cover the costs of the advertising.

6. C  The diagram shows that a shift in the MRP of labour due to productivity improvements, from MRP$_1$ to MRP$_2$, will result in a relatively small increase in employment numbers from $L_1$ to $L_2$ and a relatively large increase in wages from $W_1$ to $W_2$. 

The diagram shows that a shift in the MRP of labour due to productivity improvements, from MRP$_1$ to MRP$_2$, will result in a relatively small increase in employment numbers from $L_1$ to $L_2$ and a relatively large increase in wages from $W_1$ to $W_2$. 

---

486  Economics and Markets
7 A This is best illustrated by a diagram.

The original equilibrium is price $P$ and output $Q$. When demand falls, supply will fall in the short run to $Q_1$ and the price will fall to $P_1$. In the longer term, supply capacity is reduced, and the supply curve becomes more elastic. The output quantity falls further to $Q_2$, and price recovers to $P_2$, which is less than the original equilibrium price $P$.

8 B A demand curve has falling elasticity as you move downwards to the right (i.e. as quantity increases).

9 B Price elasticity measures the responsiveness in quantity demanded to a change in price, but the question doesn’t mention price so you should have ruled out options A and D immediately. Income elasticity measures the responsiveness in quantity demanded to a change in income. In this case, income has increased 8 per cent, but the quantity demanded has only increased 4 per cent. This means that income elasticity is 0.5, but note the income elasticity is positive (+0.5, not –0.5) because both income and quantity demanded have increased.

Therefore, the only thing we can know for certain is that product P is a normal good: demand for it increases as household income increases.

10 D The income elasticity of demand is:

\[
\text{Income elasticity} = \frac{\frac{\% \text{ change in quantity demanded}}{\% \text{ change in income}}}{\frac{2000}{15000}} = \frac{13.33\%}{5\%} = +2.67
\]

11 B The negative income elasticity means that the quantity demanded of the good will fall following the increase in incomes.

Income elasticity is –0.8 so when incomes rise by 5 per cent, the quantity demanded of the product will fall by 4 per cent (5 × –0.8).

So the new quantity demanded will be 19 200 (96 per cent of the original 20 000), and revenue will fall from $200 000 to $192 000.
Chapter 4

1. A In economics, profit maximisation is assumed to be the basic aim for all firms.

Note: The average fixed cost line AFC is shown in the graph. Average fixed cost is also the difference between average total cost (AC) and average variable cost (AVC).

2. D Economic profit. Accounting profit only includes explicit costs. Economic profit also includes implicit costs (opportunity costs).

3. B Economic costs are the opportunity costs of the factors of production employed by the firm.

4. A Marginal cost is the addition to total cost of producing one more unit of output.

5. B III and IV. Fixed costs are, by definition, fixed. Whether marginal costs rise or fall depends on the level of output.

6. C Cost curve shifts are caused by changes to the cost of factors of production not changes in demand.

7. D This is a classic example of the operation of diminishing returns.

8. D Do not confuse the long and short run effects. The long run suffers diseconomies of scale, but diminishing returns are a short run phenomenon.

9. D
11. **D** Technical improvements could apply at any scale of operations.

12. **B** This defines the short run, and the law of diminishing returns is a short run phenomenon.

13. **D** Diseconomies of scale are a long run effect. The long run cost curve does not necessarily rise once MES has been reached. Diminishing returns are a short run phenomenon.

14. **C** Proposition III is false: it is **average** fixed costs per unit (AFC) that fall as output increases. Marginal fixed costs = 0. Since AFC falls, any fall in average variable cost (AVC) must mean a falling average total cost (ATC) since AVC + AFC = ATC. Proposition II must therefore also be false.
Chapter 5

1. A. The area represents supernormal profit, because average revenue (AR) is greater than average total cost. Efficient firms in a perfectly competitive market are able to earn supernormal profits in the short run.

2. B. Perfect competition requires a homogeneous product. Product differentiation is an important aspect of monopolistic competition.

3. A. A monopsony is a market with a single buyer. This is a definition.

4. C. Any unilateral price change brings a disadvantage to the supplier concerned. This is illustrated by a kinked demand curve.

5. C. The monopolist’s demand curve is also the market demand since there are no other producers.

6. C. Homogeneity means the producers are in competition with one another.

7. D. This enables them to control the market and thereby sustain a super-normal profit.

8. D. There are barriers to the entry of competitors. Note that economies of scale (option B) are not confined to any particular market structure.

9. B. This will cause a downward shift in the marginal cost curve. A could have resulted in output reaching $Q_2$ by increasing the demand (shifting AR and therefore MR to the right), but this would have produced a higher price than $P_2$.

10. B. If price discrimination is possible, the monopolist can segment the market. This is likely to increase output, and with lower prices being offered to some segments, allocative inefficiency will be reduced.

11. B. CM Co will not be able to raise prices (option A) as being in a perfectly competitive market it is a price taker. Option C (creating barriers to entry) is not possible either in a perfectly competitive market. Option D increasing production will not work if it is currently making normal profits, because increasing production will mean that $MC > MR$.

12. D. Both A and B are incorrect as with no barriers to entry no single firm or group of firms will be able to acquire the market share for a monopoly or oligopoly.

13. B. Allocative efficiency occurs where $P = MC$, as it does under perfect competition. However, under monopoly, price is greater than marginal cost. Technical efficiency is achieved if a firm produces at the level of output where average total costs are minimised. It is argued that monopolies restrict output compared to perfect competition, and monopolists’ average costs could be lowered by increasing output. The fact that monopolists can sustain supernormal profits, and use them for investment in research and development, is a potential benefit of a monopoly rather than a disadvantage.
Chapter 6

1. B The consumer bears increase AB. The producers incur the cost BC.
2. C This is related to the trade cycle.
3. A Health and education are merit goods. Defence and street lighting are public goods.
4. B The price mechanism has nothing to do with income.
5. C When demand is inelastic, then consumers will not react significantly to a rise in price. Consequently, producers are able to pass much of the price rise on to consumers. When supply is inelastic, the burden of tax will fall primarily on the producer. Therefore, when demand is inelastic and supply elastic, the primary burden of a tax will fall most heavily on buyers.
6. D The excess burden of tax or deadweight loss is lower when the imposition of tax does not significantly reduce the number of units traded (i.e. there is no significant decrease in trade). When demand is inelastic, the deadweight loss is lower as consumers will not react significantly to a rise in prices. When supply is inelastic, the deadweight loss is again lower as producers will not change output levels significantly in response to a price change. As a result, the excess burden of tax is minimised when demand and supply are both inelastic.
7. A By establishing property rights (which may be tradable), market forces ought to be able to resolve a market failure in the least distortive manner. Floor prices, quotas and subsidies all distort the workings of the price mechanism.
8. A A social good is one which is indivisible, and so will not be provided by a private company because that company would have no way of ensuring that only those people who have paid for the good or service benefit from it. Of the four options, this is only true for a national defence system. It would be possible to restrict the provision of health, transport and education services to those people who have paid for them.
9. C The characteristics of non-exclusivity and indivisibility relate to public goods, which are reasons why public goods will not be provided at all by the free market. There will be some supply and demand for merit goods under a free market system, but this will be below the optimal level.
10. D An external social cost arises when an economic activity produces negative externalities, meaning the total cost of the activity to society is greater than the private cost to the people or firms involved in it. When the factory analyses its costs it will only consider its own private costs of production. However, its production processes have a wider impact on society through the pollution they emit. Option D is a better example of an external social cost than Option C, because in option C the people who suffer the illness are the same people who have been drinking too much.
Chapter 7

1 C All income received. This is the definition.
2 D All others are measures of economic activity.
3 A Net National Product at factor cost plus capital consumption equals Gross National Product at factor cost. By adding back taxes on expenditure and subtracting subsidies, we then get from GNP at factor cost to GNP at market prices.
4 B Transfer payments are payments where the recipient does not make any contribution to national output in return. They involve the transfer of wealth rather than a reward for creating economic wealth, and a redistributing of income from taxpayers to others. Salaries of Members of Parliament are a part of general government expenditure and so are included in the national income figures.
5 A Increasing inventory levels is an investment, because it involves incurring expenditures now for some benefit in the future time. Although the purchase of shares (item B), second-hand machinery (item C) or an already-existing company (item D) are all investments for the individuals or organisations concerned, they are merely the transfer of ownership of already-existing assets, and there is no creation of new non-current asset capital investment or inventories. From the point of view of the national economy as a whole, these do not count as investment and do not provide an injection into the circular flow.
6 A

<table>
<thead>
<tr>
<th></th>
<th>20X8</th>
<th>20X9</th>
</tr>
</thead>
<tbody>
<tr>
<td>Consumers' expenditure</td>
<td>$200 000</td>
<td>$225 000</td>
</tr>
<tr>
<td>Government expenditure</td>
<td>$70 000</td>
<td>$74 000</td>
</tr>
<tr>
<td>Fixed capital formation</td>
<td>$54 000</td>
<td>$60 000</td>
</tr>
<tr>
<td>Exports</td>
<td>$93 000</td>
<td>$94 000</td>
</tr>
<tr>
<td>Imports</td>
<td>($92 000)</td>
<td>($99 000)</td>
</tr>
<tr>
<td>GDP at market prices</td>
<td>$325 000</td>
<td>$354 000</td>
</tr>
</tbody>
</table>

Increase (354 – 325) = $29 000 million

% Increase in money terms = \( \frac{29 000}{325 000} \times 100\% = 8.9\% \)

% change in real terms, with 10% inflation, is a fall of about 1 per cent

Note. No adjustment is needed for taxes and subsidies because the question asks about GDP at market prices.

7 D Capital consumption represents an estimated cost based on current prices for the gradual using up of the nation’s productive non-current assets. It is difficult to estimate accurately. Statement A is incorrect, largely because inter-country comparisons of living standards would be based on national income per head rather than total national income. Statement B is incorrect because services provided free such as policing are included in the statistics at actual cost. Statement C is wrong because when there is a strong black economy, with economic activity not reported to the government to avoid taxation, official statistics will underestimate national income.

8 B

<table>
<thead>
<tr>
<th></th>
<th>$m</th>
</tr>
</thead>
<tbody>
<tr>
<td>GDP at factor cost</td>
<td>$150</td>
</tr>
<tr>
<td>Add taxes</td>
<td>$17</td>
</tr>
<tr>
<td>Less subsidies</td>
<td>$-10</td>
</tr>
<tr>
<td>GDP at market prices</td>
<td>$157</td>
</tr>
</tbody>
</table>

Note. Government expenditure is already included in GDP at factor cost.
Initially, firms will carry on producing the same level of output as they did when the economy was growing quickly. However, the slowdown in the rate of growth means that demand will reduce, and so goods will be held as inventory rather than being sold.

The slowdown in the rate of growth is likely to mean that both the overall demand for goods and the demand for imported goods are likely to fall rather than rise.
Chapter 8

1. C  Y, represents full employment level. The vertical aggregate supply curve shows that the economy cannot produce any additional output because its resources are already fully employed.

2. D  The relationship between changes in income and changes in consumption.

3. C  Items A and B describe the multiplier effect. Item D is not correct, because although Keynes believed that a combination of the multiplier and the accelerator helped to cause trade cycles, this is not the accelerator principle as such. Item C correctly states that if firms have a fixed capital:output ratio, an increase in output will create a bigger proportional increase in investment in new capital equipment, so that more capital goods will be produced.

4. A  Lower interest rates should be a consequence of an increase in the money supply, with a movement along the liquidity preference curve rather than a shift in the liquidity preference curve (item B).

5. A  In a closed economy (no imports or exports) with no government involvement (no tax), the marginal propensity to consume plus the marginal propensity to save must equal 1. The only withdrawal from the economy in this case is saving.

6. C  The multiplier in an open economy = \( \frac{1}{s + m + t} \)

   If taxes are 20 per cent, savings are 10 per cent and domestic consumption is 50 per cent, the balance (imports) must be 20 per cent.

   So the multiplier is \( \frac{1}{0.1 + 0.2 + 0.2} = \frac{1}{0.5} = 2 \)

7. D  A deflationary policy will aim to reduce the level of aggregate demand in the economy. Increasing interest rates is a monetarist policy designed to reduce aggregate demand. Increased levels of government spending and investment both constitute injections into the economy which should increase the rate of growth. Reducing tax rates is also likely to boost aggregate demand.

8. D  Investment is an injection into the circular flow of income in an economy. Saving, imports and taxation are all withdrawals.

9. B  Lower interest rates are likely to lead to an increase in spending. The cost of borrowing will decrease, so people can borrow more, and use their borrowings to spend more. Conversely, people will save less, because they will earn less interest on their savings.

10. C  The multiplier is calculated as \( \frac{1}{1 - MPC} \). Therefore an increase in the marginal propensity to consume will lead an increase in the size of the multiplier.

11. C  Cyclical (or demand-deficient) unemployment will occur as a result of a fall in the level of economic activity in an economy; it will not, in itself cause the fall in the level of economic activity.

12. C  The economy is in equilibrium at the point where income = expenditure. Expenditure in this economy is consumption (C) + investment (I); so the equilibrium point occurs where 'C+I' crosses the 45 degree line.
Chapter 9

1. A. Items B, C and D will all be measures which reduce the demand for goods and services. Public expenditure (item D) represents the government’s own demand. Bank lending (item C) is largely used for spending on goods and services by the people who borrow the money. Higher sales tax (item B) could increase total spending on goods and services inclusive of the tax, but spending net of tax will fall, and this should result in a reduction in demand-pull inflation. Item A, lower interest rates, is likely to result in higher consumer borrowing and even stronger demand-pull inflation.

2. C. Structural unemployment is caused by a mismatch between available jobs and the unemployed. This could be caused by a geographical mismatch or by a mismatch of skills. Items B and D could be causes of structural unemployment, but don’t fully describe it.

3. C. Seasonal unemployment fluctuates in seasonal patterns throughout the year.

4. D. This is an example of demand pull inflation.

5. B. Structural unemployment occurs as a result of long-term changes in the conditions of an industry.

6. B. This is a definition of the natural rate hypothesis or the non-accelerating inflation rate of unemployment (NAIRU).

7. A. A homogenous quality of money is that it must be consistent, divisible and of measurable value. Money being easy to use is a requirement of convenience, money being scarce is a requirement of acceptability and money not deteriorating is an example of durability.

8. A. M₃ is a measure of the broad money supply, consisting of bank currency and all deposits of the private non-banking sector.

9. A. Friedman argued that stimulating demand will only have a temporary effect on unemployment, and that demand-led expansion of the economy would soon become inflationary (with no increase in the real output). He argued in favour of supply side measures to reduce the natural rate of unemployment. Retraining schemes (item B) should reduce structural unemployment. Cutting trade union power (item C) was seen as a way of reducing unemployment. Lower income taxes (item D) and lower benefits for the unemployed would make individuals more willing to work and less willing to remain unemployed.

10. C. If expected inflation is negative, the nominal interest rate is less than the real interest rate, and positive.

11. D. Nominal rate of interest = Real rate of interest + Rate of inflation (approx) = 3% + 6% = 9%.

12. B. Real rate of interest = \( \frac{1 + \text{nominal rate}}{1 + \text{inflation rate}} = \frac{1.05}{1.02} = 1.029 \)

So real rate of interest = 2.9%

As an alternative approach, you could have applied the Fisher equation:

Nominal interest rate = real interest rate + inflation rate

If this is re-arranged: Real interest rate = nominal interest rate – inflation rate

3% = 5% – 2%

13. B. Monetarists argue that the downward-sloping Phillips curve (and the negative relationship between inflation and unemployment) only holds in the short term. They argue that in the long term, there is an expectations-augmented Phillips curve which is vertical, and identifies the natural rate of unemployment in an economy.

14. B. For Keynesian economists, the equilibrium level of national income is determined by the intersection between aggregate demand and aggregate supply. This does not in itself identify what the level of unemployment will be at the point of equilibrium.
# Chapter 10

1. A The gap between government expenditure and government receipts.

2. B Low interest rates will encourage spending rather than saving. Investment will be encouraged because the opportunity cost of investing is reduced.

3. D A reduction in taxes on alcoholic drinks will leave all consumers of alcohol with more income. A less even distribution of wealth in society means that richer people will now be relatively better off than before, which means that they have obtained a bigger benefit from the tax cuts. The conclusion points to either answer B or answer D. The benefit has to be relative, since the distribution of wealth refers to relative (proportionate) wealth, and so answer D must be correct.

4. D Fiscal policy is concerned with the government’s tax income, expenditure and borrowing (to make up the difference between income and expenditure).

5. C If interest rates are low, but expected to rise this implies that bond prices are likely to fall. People will hold funds so that they can invest in bonds later, so liquidity preference is high (option I). If interest rates are low, the speculative demand for money will be high (option III).

6. A Options C and D relate to monetary policy. Fiscal policy looks at levels of government expenditure and how they can be funded through tax revenues.

7. B Lower inflation is a government objective. Changing interest rates, tax rates or government spending are all possible ways of changing the level of economic activity in a country.

8. A The labour market becoming more flexible will lead to an outward (rightward) shift in the aggregate supply curve. The new intersection between aggregate supply and aggregate demand will mean that prices are lower than before, but the level of national income has increased. This is the basic idea behind supply side policy.

9. C A budget deficit indicates that government spending exceeds government income from tax revenues. Governments often run a budget deficit to try to stimulate aggregate demand during a period of recession, and a significant increase in the budget deficit is likely to boost aggregate demand. In the longer term, government spending may be reduced to try to reduce the deficit, but option D is ruled out here because it refers specifically to the short term.

10. B A contractionary fiscal policy will cause an inward (leftward) shift in the aggregate demand curve. This shift will mean that the new equilibrium level of national income (the intersection between the aggregate demand and aggregate supply curves) will be at lower price levels and a lower level of national output than the initial equilibrium.

11. B The central bank is trying to control the growth of the money supply by imposing direct controls on bank lending. The central bank hasn’t imposed quantitative controls (restricting the overall amount banks can lend) but has imposed qualitative control – restricting the amount of lending to a particular industry. A special directive is a type of qualitative control.

12. A Fiscal policy is about government finance, not bank finance.

13. C Determining social security payments is a government role.

14. C Setting the rate at which it lends to commercial banks.
Chapter 11

1. C  The first three options are arguments in favour of privatisation. The fourth (management of natural monopolies) is an argument in favour of nationalisation.

2. C  The ACCC is responsible for all anti-competitive conduct by corporations.

3. C  State-owned industries are more likely to respond to the public interest, ahead of the profit motive which underlies privatised industries. This is a reason why key industries (for example, health) may be retained under central control.

4. C  As a result of the privatisation, assets will be transferred from public ownership to private ownership, and the government will receive the proceeds from the sale. Privatisation is likely to lead to a less even distribution of wealth and income, although it should also lead to the creation of competition between firms (rather than a reduction in competition).

5. D  The presumption of the Act is that anti-competitive arrangements are necessarily against the public interest and so should be illegal.

6. B  Privatisation often leads to smaller firms as public sector monopolies are broken up into smaller companies. Although privatisation may be seen as a way of increasing competitiveness, it does not do so through promoting economies of scale.

7. A  Australia’s Gini co-efficient was 0.319 in 2007-08, up 5.6 per cent from the 1994-95 measure of 0.302. It would be inaccurate to say Australia is highly unequal as the coefficient is still under 0.5.

8. C  Regressive taxation is a flat rate of tax applied on goods and services so actually punishes the poor. This is generally the same with indirect taxation. Proportional taxation is based on taking the same proportion of income across the population, so it impacts all levels of wealth accordingly. The correct answer is therefore progressive taxation.

9. D  The Gini coefficient of perfect income equality is zero: that is to say, the Lorenz curve would not deviate from the 45 degree line.
Chapter 12

1  B  Since the class intervals are different, adjusted frequencies need to be calculated. The adjusted frequencies are calculated as follows. (Standard class width is taken as 2.)

<table>
<thead>
<tr>
<th>Class interval</th>
<th>Size of interval</th>
<th>Frequency</th>
<th>Adjustment</th>
<th>Height of bar</th>
</tr>
</thead>
<tbody>
<tr>
<td>2-6</td>
<td>4</td>
<td>12</td>
<td>× 2/4</td>
<td>6</td>
</tr>
<tr>
<td>6-8</td>
<td>2</td>
<td>30</td>
<td>× 2/2</td>
<td>30</td>
</tr>
<tr>
<td>8-10</td>
<td>2</td>
<td>18</td>
<td>× 2/2</td>
<td>18</td>
</tr>
<tr>
<td>10-14</td>
<td>4</td>
<td>12</td>
<td>× 2/4</td>
<td>6</td>
</tr>
</tbody>
</table>

The histogram which represents the above bar heights correctly is graph 2. The correct answer is therefore B.

Options A, C and D are incorrect because the class intervals need to be adjusted to take account of unequal class widths (as shown above).

2  D  If a distribution has unequal class intervals, the heights of the bars have to be adjusted for the fact that the bars do not have the same width. If the width of one bar is one and a half times the standard width, we must divide the frequency by one and a half, i.e. multiply by 0.67 (1/1.5 = 2/3 = 0.67).

The correct answer is D.

3  C  Region C’s angle is given by 33% × 360 = 118.8 degrees

4  C  If a distribution has unequal class intervals, remember that the heights of the bars must be adjusted for the fact that the bars do not have the same width. Therefore, if one class is two thirds the standard width, you need to divide the frequency by 2/3, i.e. 25 ÷ 2/3 = 37.5.

If you selected option A then you multiplied the frequency by 2/3 instead of dividing it by 2/3.

Option B represents the unadjusted frequency of the class under consideration, which is incorrect; an adjustment needs to be made.

If you selected option D you incorrectly took 2/3 of 25 (= 16.67) and added this to the unadjusted class frequency of 25. 25 + 16.67 = 41.67.

5  C  The scale for the variables is nominal, ratio, ratio, ordinal. We can eliminate option D immediately because the first variable must be nominal – sex can only be male or female.

Although ordinal variables provide information concerning the relative position of participants or observations in a research study, ordinal variables do not tell us anything about the absolute magnitude of the difference between quality grades A, B and C. This eliminates options A and B.

6  D  The best diagram to draw to highlight the differences between areas is a simple bar chart.

A simple bar chart is a chart consisting of one or more bars, in which the length of each bar indicates the magnitude of the corresponding data items. This is the best diagram to draw to highlight the differences of typical salaries in different areas.

We are not interested in showing the breakdown of the total salary, therefore a pie chart and a percentage component bar chart are not really appropriate.

A multiple bar chart is a bar chart in which two or more separate bars are used to present sub-divisions of data. The data available relating to salaries is not subdivided and this type of chart is therefore not appropriate in this situation.

7  C  Because socioeconomic status was coded as 1=low income, 2=middle income, 3=high income and claimed to be an interval scaled variable, this is the only option that was incorrect. It is ordinal level data. Interval level data is similar to ordinal level data in that it has a definite ordering scheme, but the differences between data are meaningful and can be measured.

8  B  A sampling frame is a numbered list of all items in a population (not a sample).

Cluster sampling involves selecting one definable subsection of the population which therefore makes the potential for bias considerable.

Statements II and III are therefore not true. The correct answer is therefore B.
Chapter 13

1. The correct answer is C. First, we need to compute the population mean.
   \[ \mu = \frac{(1 + 3 + 5 + 7)}{4} = 4 \]
   Then we plug all of the known values into the formula for the variance of a population, as shown below:
   \[ \sigma^2 = \frac{\sum (X_i - \mu)^2}{N} \]
   \[ \sigma^2 = \frac{[(1 - 4)^2 + (3 - 4)^2 + (5 - 4)^2 + (7 - 4)^2]}{4} \]
   \[ \sigma^2 = \frac{[(-3)^2 + (-1)^2 + (1)^2 + (3)^2]}{4} = \frac{20}{4} = 5 \]

2. The correct answer is B. First, we need to compute the sample mean.
   \[ \bar{x} = \frac{(1 + 3 + 5 + 7)}{4} = 4 \]
   Then we plug all of the known values into the formula for the standard deviation of a sample, as shown below:
   \[ s = \sqrt{\frac{\sum (x_i - \bar{x})^2}{(n - 1)}} \]
   \[ s = \sqrt{\frac{[(1 - 4)^2 + (3 - 4)^2 + (5 - 4)^2 + (7 - 4)^2]}{(4 - 1)}} \]
   \[ s = \sqrt{\frac{[(-3)^2 + (-1)^2 + (1)^2 + (3)^2]}{3}} = \sqrt{\frac{20}{3}} = \sqrt{6.67} = 2.58 \]

3. C The mean will tend to overstate the average, and the mode will underestimate it.

4. The correct answer is B. The mean score is computed from the equation:
   \[ \text{Mean score} = \frac{\sum x}{n} = \frac{(96 + 100 + 106 + 114)}{4} = 104 \]
   Since there are an even number of scores (four scores), the median is the average of the two middle scores. Therefore, the median is \( \frac{(100 + 106)}{2} = 103. \)

5. B
   In Outcome B, your score is higher than all but two of the rest of the scores. Just by looking at the distribution, you can tell that you did very well compared to the rest of the class. Therefore, you scored well above the centre of the distribution.
   In Outcome A, you did about average, with your score falling in the middle of the group. This puts you about at the centre of the distribution.
   In Outcome C, you did worse than almost everyone else. This puts you below the centre of the distribution.

6. C The other measurements are preferred when data are normally distributed.
7 A  The Empirical Rule: If a variable is normally distributed, then:
• within one standard deviation of the mean there will be approximately 68.2 per cent of the data.
• within two standard deviations of the mean there will be approximately 95.4 per cent of the data.
• within three standard deviations of the mean there will be approximately 99.7 per cent of the data.

8 D  The exponential distribution is the only distribution that starts at $1/\mu$ and decreases thereafter exponentially and monotonically as $x$ increases.

9 C  In a binomial distribution we count the number of successes in a finite number of trials.

10 A  In an exponential distribution, the mean is always larger than the median because the median is given by $0.693 \mu$ and $\mu =$ mean time between failures, or to a failure.

11 A  The variance would be larger if these numbers represented a sample because you would divide by $N - 1$ instead of just $N$. 
Chapter 14

1 All of them. There are three distinctly different ways of assigning probability:

   1 The first one is the classical probability approach. Here the probability of success is based on prior knowledge of the process involved, using symmetrical outcomes.

   2 The second approach to probability is the frequency (or empirical method) approach where the outcome is based on observed data, rather than upon prior knowledge of a process. The assumption of this approach is that the random process can be replicated many times under identical conditions, and relative frequencies recorded.

   3 The third approach to probability is the subjective approach

2 The event that the first shirt selected is white is W1. The probability is P(W1) = 9/12
The event that the second shirt selected is also white is identified as W2. The conditional probability that the second shirt selected is white, given that the first shirt selected is also white, is P(W2 | W1) = 8/11.

To determine the probability of 2 white shirts being selected we use formula: P(A and B) = P(A) × P(B|A)
P(W1 and W2) = P(W1)P(W2 |W1) = (9/12)(8/11) = 0.55

3 This example is slightly more complicated. We cannot simply add together the number of outcomes for each event separately (4 + 13 = 17) as this inadvertently counts one of the outcomes twice (the king of clubs). The correct answer is 16/52, which we get from 13/52 + 4/52 - 1/52 where this is essentially p(club) + p(king) - p(king of clubs).

Let’s take another example. If you flip a coin and roll a six-sided die, what is the probability that the coin comes up heads and the die comes up 1? Since the two events are independent, the probability is simply the probability of a head (which is 1/2) times the probability of the die coming up 1 (which is 1/6). Therefore, the probability of both events occurring is 1/2 x 1/6 = 1/12.

4 The card is replaced, so the two events are independent and the probability is P(heart) x P(black) = ¼ x ½ = 0.125

5 D Correct. The overlapping area is 0.5

   A Wrong: total probability = 0.6 + 0.8, is > 1.
   'Coach rehired' area should extend into the 'Both' area.

   B Wrong: total probability = 0.6 + 0.5 + 0.8, is > 1.
   'Makes playoffs' and 'Coach rehired' areas should extend into the 'Both' area.

   C Wrong: total area cannot be greater than any individual area.
   Total area is in fact 0.6 + 0.8 - 0.5 = 0.9, and the remaining 0.1 is the probability that the team does not make the playoffs and the coach is not rehired.

6 A and D

   Two events are independent if the occurrence of one has no effect on the probability of the occurrence of the other. A student’s test grades are not independent because a student who does well on the midterm understands the material and is therefore more likely to do well on the final.
   Two draws of an ace without replacement are not independent because what you get on your first draw affects the probability of getting an ace on the second draw. This is a conditional probability.

7 The probability that you win $1 OR $3 (or both games) is: (1/2) + (1/6) - (1/2)(1/6) = 7/12 = 0.58

8 The probability that person 1 AND person 2 AND person 3 will take public transportation is: .6 x .6 x.6 = 0.216

9 The probability of NOT getting a 4 on any roll is: (5/6)(5/6)(5/6)(5/6) = .48, so the probability of rolling at least one 4 is: 1 - .48 = 0.52.
Chapter 15

1. The correct answer is A. Increasing sample size makes the hypothesis test more sensitive – more likely to reject the null hypothesis when it is, in fact, false. Therefore, it increases the power of the test. The effect size is not affected by sample size. And the probability of making a Type II error gets smaller, not bigger, as sample size increases.

2. The correct answer is C. Increasing sample size makes the hypothesis test more sensitive – more likely to reject the null hypothesis when it is, in fact, false. Increasing the significance level reduces the region of acceptance, which makes the hypothesis test more likely to reject the null hypothesis, thus increasing the power of the test. Since, by definition, power is equal to one minus beta, the power of a test will get smaller as beta gets bigger.

3. Mean = 307 / 10 = 30.7
   
   \[ z = \frac{30.7 - 25}{7/\sqrt{10}} = 2.57 \]

4. Mean = 3281 / 31 = 105.84
   
   \[ z = \frac{105.84 - 100}{15/\sqrt{31}} = 2.17 \]

5. Mean = 14524 / 16 = 907.75
   
   \[ z = \frac{907.75 - 910}{8.5/\sqrt{16}} = -1.06 \]
Chapter 16

1. B  If temperatures increase by one degree, Y decreases by 1.6, i.e. sales have now fallen by $1,600.

   If you selected option C, you have forgotten that sales are given in $000s.

   If you selected option D, you have forgotten that a negative sign in the regression equation means that as X increases, Y decreases.

2. C  The value X = 10 does not lie between 15 and 45 so we do not know whether or not the relationship between the variables still holds for this value of X. We therefore cannot rely upon the estimate.

   The sample size is quite large for this type of analysis and will provide reliable estimates. Option A will not therefore reduce the reliability of the estimate.

   The correlation coefficient is very close to –1, indicating a very strong relationship between the variables which will provide reliable estimates. Option B will not therefore reduce the reliability of the estimate.

   The fact that the correlation coefficient is negative tells us that Y decreases as X increases but this is not a problem in assessing the strength of the relationship between the variables nor the reliability of the estimates. Option D is therefore an incorrect answer.

3. D  b is the slope of the line which measures the increase in the y variable for unit increase in x.

4. A  The scatter diagram displays near perfect negative correlation. The regression line has a gradient of –1 and cuts the y axis at y= 8 (i.e. when x= 0, y= 8).

5. C  A correlation coefficient of 0.9 indicates that the relationship between x and y is probably too strong to have been caused by chance.

6. B  I  When an equation has the form y = a + bx, b gives the slope which is 5 in the equation y = 5x – 24.

   II  When an equation has the form y = a + bx, a gives the intersection on the y axis which is –24 in the equation y = 5x – 24.

   III  The slope of the equation is 5. The intercept on the y axis is -24.

   IV  The line cuts the y axis at –24. The line cuts the x axis at 4.8.

   I and II are correct and the answer is therefore B.

7. C  The value of the coefficient of determination $r^2$ is 0.8575. This is calculated by finding the square of the correlation coefficient, r, for the data. If you had chosen –0.8575, (option B) it was because of an apparent inability to correctly square a negative number using a calculator.
Since the question asks for the regression line of \( x \) on \( y \); \( x \) and \( y \) must therefore be interchanged in the formula.

\[
\begin{array}{cccc}
 x & y & y^2 & xy \\
 9 & 2 & 4 & 18 \\
 10 & 3 & 9 & 30 \\
 9 & 1 & 1 & 9 \\
 8 & 1 & 1 & 8 \\
 9 & 2 & 4 & 18 \\
 \hline
 45 & 9 & 19 & 83
\end{array}
\]

\[
n = 5
\]

\[
b = \frac{n \Sigma xy - \Sigma x \Sigma y}{n \Sigma y^2 - (\Sigma y)^2} = \frac{10}{14} = 0.71
\]

\[
a = \bar{x} - b \bar{y} = \frac{45}{5} - \left(0.71 \times \frac{9}{5}\right) = 7.722 = 7.72 \text{ (to 2 decimal places)}
\]

Therefore, the equation is \( x = 7.72 + 0.71y \)
Before you begin:
answers and commentary
Chapter 1

1 Economics studies the ways in which society decides what to produce, how to produce it, who to produce it for and how to apportion it.

2 Students should list their own examples. As an example one human need is basic shelter: a want is an MP3 player.

3 The production possibility frontier shows the allocation of resources. To take a simple example, suppose that an imaginary society can use its available resources to produce two products, A and B. The society’s resources are limited. Therefore, there are restrictions on the amounts of A and B that can be made. The possible combinations of A and B can be shown by a production possibility frontier.

4 The theory of comparative advantage is based on the idea of opportunity cost and the production possibility frontier. Within a country, the opportunity cost for any category of product may be established in terms of the next most advantageous use of national resources. If two countries produce different goods most efficiently and can exchange them at an advantageous rate in terms of the comparative opportunity costs of importing and home production, then it will be beneficial for them to specialise and trade. Total production of each good will be higher than if they each produce both goods.

5 Total utility is the total satisfaction that people derive from spending their income and consuming goods. Marginal utility is the satisfaction gained from consuming one additional unit of a good or the satisfaction forgone by consuming one unit less.
Chapter 2

1. Demand for a good or service is the quantity of that good or service that potential purchasers would be willing and able to buy, or attempt to buy, at any possible price.

2. A normal market demand curve is as follows:

3. A change in the price of one good will not necessarily change the demand for another good. Substitution takes place when the price of one good rises relative to a substitute good. Examples of substitute goods and services:
   - Rival brands of the same commodity, like Coca-Cola and Pepsi-Cola.
   - Tea and coffee.
   - Some different forms of entertainment.

4. Supply refers to the quantity of a good that existing suppliers or would-be suppliers would want to produce for the market at a given price.

5. The price mechanism brings demand and supply into equilibrium, and the equilibrium price for a good is the price at which the volume demanded by consumers and the volume that firms would be willing to supply is the same.

6. A price ceiling is where the government attempts to prevent prices of goods rising by establishing a price ceiling below the equilibrium price.

7. A price floor aims to ensure that suppliers earn at least the minimum price (or floor price) for each unit of output they sell.

8. Yes, it is one of the 90 per cent of countries in the world that has some form of minimum wage legislation in place.
Chapter 3

1. The coefficient of PED is measured as:

\[
\frac{\text{Percentage change in quantity demanded}}{\text{Percentage change in price}}
\]

2. When \( \eta = 0 \), there is no change in quantity demanded, regardless of the change in price. In this case, the demand curve is a vertical straight line.

3. It is upward sloping. The price elasticity of demand in such a case would be positive whereas, mathematically, the price elasticity of demand for normal goods is negative.

4. The demand curve for a good will also slope upwards (as with a Giffen good).

5. Elasticity of supply = \[
\frac{\text{\% change in quantity supplied}}{\text{\% change in price}}
\]

6. Choose any three of the following:
   (a) Existence of inventories of finished goods.
   (b) Availability of labour.
   (c) Spare capacity.
   (d) Availability of raw materials and components.
   (e) Barriers to entry.
   (f) Time scale.
Chapter 4

1. Total revenue: is the total income obtained from selling a given quantity of output.

   Average revenue is the total revenue divided by the number of units sold.

   Marginal revenue is the addition to total revenue earned from the sale of one extra unit of output.

2. Total revenue minus total cost.

3. Breakeven occurs where total revenue equals total cost, and therefore by extension, average revenue equals average cost.

4. The short run is a time period in which the amount of at least one factor of production (land, labour, capital or enterprise) is fixed. The long run is a period sufficiently long to allow full flexibility in all the factors of production used.

5. Economic profit consists of sales revenue minus both the explicit costs and the implicit costs of the business. Implicit costs are benefits forgone by not using the factors of production in their next most profitable way (opportunity costs).


7. The law of diminishing returns says that if one or more factors of production are fixed, but the input of another is increased, the extra output generated by each extra unit of input will eventually begin to fall. In our factory, as we add staff, we start to see queues forming at machines; it becomes more difficult to co-ordinate work; machinery starts to break down through over-use, and there simply is not enough space to work efficiently.

8. The factors which cause average cost to decline in the long run as output increases.

9. The main reasons for possible diseconomies of scale are the human and behavioural problems of managing a large firm.
Chapter 5

1 Imperfect competition.

2 In a **monopoly** there is only one firm, the sole producer of a good which has no closely competing substitutes.

3 Price discrimination occurs when a firm sells the same product at different prices in different markets. Four basic conditions are necessary for price discrimination to be effective and profitable.
   
   (a) The seller must be able to control the supply of the product and keep out any competitors.
   
   (b) There must be at least two distinct markets with no cross-over between them.
   
   (c) The seller must be able to prevent the resale of the good by one buyer to another.
   
   (d) There must be significant differences in the willingness to pay among the different classes of buyers.

4 Choose two of the following advantages:
   
   - Economies of scale.
   - Natural monopolies exist because of a very high ratio of fixed costs to variable costs. Monopolies can afford to spend more on research and development.
   - Monopolies may find it easier than small firms to raise new capital.
   - Temporary monopolies can stimulate competition, and are in the longer-term interests of consumers.

Choose two of the following disadvantages:

   - The profit-maximising output of a monopoly is at a point where total market output is lower and prices are higher than they would be under perfect competition.
   - Monopolies do not achieve allocative efficiency since the prices they charge are greater than marginal cost.
   - Monopolies do not use resources in the most efficient way possible (technical efficiency).
   - Monopolists can carry out restrictive practices, such as price discrimination.
   - Product differentiation can be beneficial for producers, but at the expense of consumers.
   - Monopolies might become slack about cost control and adopt a complacent attitude to innovation.
   - Monopolies might stifle competition.
   - If a monopoly controls a vital resource, it might make decisions which are damaging to the public interest.
   - There might be diseconomies of scale in a large monopoly firm.

5 Monopolistic competition is a market structure in which firms’ products are comparable rather than homogeneous. A firm operating in conditions of monopolistic competition has a downward sloping demand curve like a monopoly: the quantity of output customers demand responds to the price at which the firm is prepared to sell. The downward sloping demand curve is possible because of product differentiation created by the firm. However, unlike a monopoly firm, a firm operating under monopolistic competition is unable to utilise barriers to entry against other firms. Indeed, the firm already competes with rivals, which can take retaliatory competitive action if the firm makes big profits.

6 An oligopoly is a market structure where a few large suppliers dominate the market.

7 Examples are: Coles and Woolworths (Australian supermarkets) and the ‘Big Four’ of ASDA, Tesco, Sainsbury’s and Morrisons (UK supermarkets).

8 An **oligopsony** is a market structure where there are a small number of buyers and a large number of sellers.
Chapter 6

1 Market failure occurs when the free market mechanism fails to produce the most efficient allocation of scarce resources. It is caused by a number of factors:

- Imperfections in a market.
- Divergence between private costs and social costs.
- The need to provide social goods.
- The need to consider non-market goals, such as the consumption of merit goods.

2 **Private good**: A private good must be both exclusive and rivalrous. Exclusivity means that it is reasonable to prevent a class of consumers (or firms) from consuming the good. Rivalrous means that consumption by one party prevents simultaneous consumption by another. A private good satisfies an individual want and is almost always produced for profit.

**Social cost** measures the cost to society as a whole of the resources that a firm uses.

**Private cost** measures the cost to the firm of the resources it uses to produce a good.

**Social benefit** measures the total benefit to society from a transaction.

**Private benefit** measures the benefit obtained directly by a supplier or by a consumer.

3 Externalities are spill-over effects of a transaction which extend beyond the parties to the transaction and affect society as a whole. In other words, externalities are the differences between the private and the social costs, or benefits, arising from an activity.

4 **Social good**: Some goods, by their very nature, involve so much 'spill-over' of externalities that they are difficult to provide except as social goods whose production is organised by the government.

**Merit goods** are considered to be worth providing to everyone irrespective of whether everyone can afford to pay for them, because their consumption is in the long-term public interest. Education is one of the chief examples of a merit good.

5 If an indirect tax is imposed on a product, the tax will shift the supply curve upwards (leftwards) by the amount the tax adds to the price of each item. This is because although the price to consumers includes the tax, the revenue the suppliers receive is only the net-of-tax price.

6 Choose two of the three reasons listed below:

(a) To encourage more production of the good, by offering a further incentive to suppliers.

(b) To keep prices lower for socially desirable goods whose production the government wishes to encourage.

(c) To protect a vital industry when demand in the short term is low and threatening to cause an excessive contraction of the industry.
Chapter 7

1 Economic development assesses the non-financial indicators of an economy and the overall impact of economic growth on human development, social progress and technological advancement. Improvement in key indicators such as poverty, literacy, health, life expectancy, environment, political freedom and social justice are all examples of economic development.

2 (a) National income.
(b) Gross national product (GNP).
(c) Gross domestic product (GDP).

3 GDP is Gross Domestic Product and it measures the value of the goods and services produced by an economy in a given period.

4 GNP is Gross National Product and it measures GDP plus income accruing to domestic residents from investments abroad less income accruing to foreign residents from investments in the domestic economy.

5 (a) Expenditure approach
This approach measures the economic wealth created in a period by calculating the amount of expenditure on the goods and services that are produced by the nation’s economy.

(b) Income approach
This approach measures the income of individuals from employment and from self-employment, the profits of firms and public corporations and rent on property. (Interest earnings will be included within the profits of companies or the income of individuals.)

(c) Value-added approach
This approach is to measure the value added by all activities which produce goods and services, that is their net output in the period. However, we must take the incremental value added at each stage of production to avoid double counting. (One firm’s output might become another firm’s input, so if gross totals were used the gross value would be overstated.)

6 Select three of the following:
- Non-market production.
- The underground economy.
- Leisure and human costs.
- Quality variation and new goods.
- Economic ‘bads’ – harmful side effects, such as pollution, are ignored.
- Economic welfare, but rather measures the value of goods and services produced in the year only.
Chapter 8

1. In the national income equation:
   \[ Y = \text{national income} \]
   \[ C = \text{consumption} \]
   \[ G = \text{total government spending} \]
   \[ I = \text{investment} \]
   \[ M = \text{total imports} \]
   \[ X = \text{total exports} \]

2. When a household receives an increase in income, some will be spent and some will be saved. The proportion which is spent is the marginal propensity to consume, while the proportion which is saved is the marginal propensity to save.

3. The multiplier explains how the increase in total national income will be much greater than an initial injection into an economy, due to the injection being recycled through the economy.

4. The accelerator principle assumes that if there is a small change in the output of consumer goods, there will be a much greater change in the output of capital equipment required to make those consumer goods. This change in production of capital equipment (investment spending) speeds up the rate of economic growth, or slump.

5. If one aim of a country’s economic policy is to achieve full employment, then the ideal equilibrium level of national income will be where \( AD \) and \( AS \) are in balance at the full employment level of national income, without any inflationary gap – in other words, where aggregate demand at current price levels is exactly sufficient to encourage firms to produce at an output capacity where the country’s resources are fully employed.
Chapter 9

1. The rate of unemployment in an economy can be calculated as:

\[
\frac{\text{Number of unemployed}}{\text{Total workforce}} \times 100\%
\]

2. Choose three of the following: frictional, seasonal, technological, cyclical, real wage unemployment and voluntary unemployment.

3. Full employment means that the country's economic resources are fully employed. However, as far as labour is concerned, full employment does not mean that everyone has a job all the time. There will always be some normal or transitional unemployment as people lose their job or give up one job for another, and so full employment might mean, for example, that 3 to 5 per cent of the total working population is unemployed at any time.

4. The CPI (in Australia) and the RPI (in the UK) are adjusted to exclude mortgage costs and often other elements as well. The effects of interest rate changes on mortgage costs (as well as other changes due to season, volatility or policy) help to make the CPI fluctuate more widely than the underlying rate of inflation.

5. Financial assets must have a high degree of liquidity to be regarded as narrow money and it is normally money for transaction purposes. Broad money extends the range of assets that are regarded as money to include money held in the form of savings.

6. List three of the following: need for a real return, uncertainty over future rates of inflation, changes in government borrowing, higher individual demand for borrowing, monetary policy, overseas interest rates.

7. The classical theory of money is based on the view that money is used only as a medium of exchange and people require it only in order to settle transactions in goods and services.

8. Keynes identified three reasons why people hold wealth as money rather than as interest-bearing securities. These are the transactions motive, precautionary motive and the speculative motive. This is the crux of the Keynesian theory of money.
Chapter 10

1 A government’s fiscal policy is concerned with taxation, borrowing and spending; and their effects upon the economy. Monetary policy is concerned with money, the money supply, interest rates, inflation and the exchange rate.

2 Regressive, proportional and progressive.

3 Direct taxes are levied on income while indirect taxes are levied on expenditure. Indirect taxes are regressive. Direct taxes can be progressive.

4 Quantitative controls might be imposed on either bank lending (assets), for example a ‘lending ceiling’ limiting annual lending growth, or bank deposits (liabilities). The purpose of quantitative controls might be seen as a means of keeping bank lending in check without having to resort to higher interest rates. Qualitative controls might be used to alter the type of lending by banks. For example, the government (via the bank) can ask the banks to limit their lending to the personal sector, and lend more to industry, or to lend less to a particular type of firm (such as, for example, property companies) and more to manufacturing businesses.

5 Inflationary expectations reflect the rates of inflation that are expected in the future. The inflationary expectations of the work force will be reflected in the level of wage rises that is demanded in the annual round of pay negotiations between employers and workers.

6 Choose four of the following:
   • Setting interest rates.
   • Banker to the government.
   • Central issuer of banknotes.
   • Manager of the national debt.
   • Manager of the nation’s foreign currency reserves.
   • Banker to the clearing banks.
   • Lender to the clearing banks (lender of last resort).
   • Supervision of the banking system

7 (a) Australia – the Reserve Bank of Australia.
    (b) United Kingdom – the Bank of England.
    (c) European Union – the European Central Bank.
Chapter 11

1. Deregulation can be defined as the removal or weakening of any form of statutory (or voluntary) regulation of free market activity. Deregulation allows free market forces more scope to determine the outcome.

2. List two of the following:
   
   (a) Privatised companies may be **more efficient** than State monopolies and private sector managers are likely to try to reduce costs and strip out unproductive labour. Private companies may also provide **better quality** because they will have to compete to survive. The threat of competition may also lead to innovation.

   (b) Denationalisation provides an immediate **source of money** for the government, through the sale of assets or businesses.

   (c) Privatisation **reduces bureaucratic and political meddling** in the industries concerned.

   (d) Privatised companies may have a **more flexible** and **profit-oriented management culture**.

   (e) There is a view that **wider share ownership** should be encouraged. Denationalisation is one method of creating wider share ownership, as the sale of Telstra and NRMA shares have demonstrated in Australia. If workers own shares in their company, they are more likely to want it to be successful.

3. List two of the following:

   (a) State-owned industries are more likely to respond to the **public interest**, ahead of the profit motive. For example, State-owned industries are more likely to cross-subsidise unprofitable operations from profitable ones. For example, Australia Post will continue to deliver letters to the remote outback stations even though the service might be very unprofitable. But privatisation may lead to fewer deliveries and higher prices.

   (b) Encouraging private competition to State-run industries might be inadvisable where **significant economies of scale** can be achieved by monopoly operations.

   (c) Government can **provide capital more cheaply** than the market to industries whose earning potential is low, but which are deemed to be of strategic importance, such as aircraft manufacture. Opponents of privatisation suggest that the very idea of privatising a strategic industry is spurious.

   (d) State-owned industries can be run in a way that **protects employment**, as is the case in China. The problem with this is that the taxpayer is effectively **subsidising technical inefficiency**.

   (e) Surpluses from State-run industries can be used for **public welfare** rather than private wealth. However, the problem here is that points (a) and (d) above tend to preclude the creation of surpluses.


5. The ACCC administers the Federal Government’s Trade Practices Act 1974. It is responsible for monitoring any anti-competitive practices in business and industries as well as monitoring regulated industries (such as electricity services, water authorities and telecommunications).

6. The **Gini coefficient** measures the deviation of the Lorenz curve from the 45 degree line. It is the ratio of the area between the curve and the 45 degree line to the whole area below the 45 degree line. Any Gini coefficient will be between zero and one: the higher it is, the more unequal the distribution of income.

7. 0.319
Chapter 12

1. There are basically two types of decision making. The first is objective or quantitative (using numbers). The second is subjective or qualitative (not using numbers). The goal is to make any decision making as objective as possible.

For example, it is easy to say that a piece of equipment or a return on investment of some kind will yield X because A + B = C. However, this is only half the picture. Often, a much larger and more important aspect of decision making is evaluating the subjective criteria, which sometimes leads to contradictory conclusions. For example, if you are selling your business and there are two different offers, the higher one from a less reputable firm, the other lower one from a more reputable and trustworthy firm. Which do you choose? Your decision will depend on a variety of factors solely dependent on the particular situation.

2. The number of maths students with red hair is an example of quantitative data, because it can be measured. Because the number of red headed maths students can only be counted in whole number steps, the resulting data is discrete. You cannot for example have 12½ students, but you can have 12 or 13.

Whether or not a tin of peas is damaged is not something that can be measured. This is therefore an example of an attribute, as a tin is either damaged or not damaged.

The time taken to swim 50 metres may be measured and is therefore quantitative data. Because the time recorded can take on any value, in theory, the data is said to be continuous.

The number of pensioners over 75 is an example of quantitative data as it can be measured. However, the results can only take on whole number values so the data is therefore discrete.

3. Listed below are a few of the arguments in favour of using a sample:

(a) Reduced cost: It is obviously less costly to obtain data for a selected subset of a population, rather than the entire population. Furthermore, data collected through a carefully selected sample are highly accurate measures of the larger population.

(b) Speed: Observations are easier to collect and summarise with a sample than with a complete count. This consideration may be vital if the speed of the analysis is important, such as through exit polls in elections.

(c) Greater scope: Sometimes highly trained personnel or specialised equipment which is limited in availability must be used to obtain the data. A complete census (enumeration) is not practical or possible. Therefore, surveys that rely on sampling have greater flexibility regarding the type of information that can be obtained.

4. Stratified sampling involves dividing the population into strata and then selecting a random sample from each stratum.

5. Retail prices are very important to a wide variety of users.

(a) For the government, the Consumer Price Index (CPI) indicates the degree of success there has been in fighting inflation.

(b) For employees, the CPI may give an indication of how much wages need to rise to keep pace with inflation.

(c) For consumers, the CPI indicates the increases to be expected in the prices of goods in shops.

(d) For businesses, the CPI may give a broad indication of how much costs should have been expected to rise over recent years and months.

(e) For pensioners and social security recipients, the movement in the CPI is used to update benefit levels.

6. The correct answer is A. ‘Tiny’ < ‘little’ < ‘medium-sized’ <‘huge’ < ‘infinite’ is an example of an ordinal scale level.

7. The correct answer is B. A pie chart would be most suitable to illustrate the occupations of 50 female adults.
Chapter 13

1 Central tendency is a statistical measure that identifies a single score as representative of an entire distribution of scores. The goal of central tendency is to find the single score that is most typical or most representative of the entire distribution. Unfortunately, there is no single, standard procedure for determining central tendency. The problem is that there is no single measure that will always produce a central, representative value in every situation. There are three main measures of central tendency: the arithmetical mean, the median and the mode.

2 $350 - 250 = 100$ employees produced more than 200 units. $310 - 250 = 60$ employees produced between 200 and 300 units.

3 The correct answer is A. An arithmetic mean might be distorted by extremely high or low values. For example, the mean of 3, 4, 4 and 6 is 4.25, but the mean of 3, 4, 4, 6 and 15 is 6.4. The high value, 15, distorts the average and in some circumstances the mean would be a misleading and inappropriate figure.

4 The correct answer is D. With the mean of our first exam = 77 and the median = 82 this would indicate a negatively skewed distribution.

5 The correct answer is False. Measures of dispersion include the range, mean deviation, variance and standard deviation. This tells us how spread out, or dispersed, scores are in a data set.

6 The correct answer is B. When data is normally distributed, the mean, median, and mode are equal.

7 The correct answer is A. Just over 68 per cent of scores fall within the one standard deviation interval around the mean (-1 SD to +1 SD).

8 The correct answer is A. $e^{-10.75} = 0.2636$.

9 The correct answer is B. A normal curve would not be skewed to the right.

10 When referring to the shape of frequency or probability distributions, ‘skewness’ refers to asymmetry of the distribution. A distribution with an asymmetric tail extending out to the right is referred to as ‘positively skewed’ or ‘skewed to the right,’ because most of the variance from the mean is in the positive direction. If the opposite is the case, the distribution is negatively skewed, or skewed left.
Chapter 14

1. The frequency histogram gives the number of times the data occur in the particular group or interval, while the relative frequency histogram gives the fraction of times the data occur in the particular group or interval.

2. A fair coin has two sides, and we assign the same probability to each because of the physical symmetry of the coin. Likewise, we assign the same probability to each of the six sides of a fair die because of symmetry of the die. There are lots of other situations where symmetry implies that the probabilities of the outcomes should be equal. However, the symmetry of the situation is required to conclude that the probability is uniform. For instance, if you throw a ball at the moon, there are two outcomes - either it hits the moon or it doesn’t. But the two outcomes are not symmetric, and it would be foolish to assume that the probability that your ball hits the moon is the same as the probability of not hitting the moon.

3. In probability theory, the **law of large numbers (LLN)** describes the result of performing the same experiment a large number of times. According to the law, the average of the results obtained from a large number of trials should be close to the expected value, and will tend to become closer as more trials are performed.

   For example, a single roll of a six-sided die produces one of the numbers 1, 2, 3, 4, 5, 6, each with equal probability. Therefore, the expected value of a single die roll is \((1+2+3+4+5+6)/6 = 3.5\).

   According to the law of large numbers, if a large number of dice are rolled, the average of their values (sometimes called the sample mean) is likely to be close to 3.5, with the accuracy increasing as more dice are rolled.

   Similarly, when a fair coin is flipped once, the expected value of the number of heads is equal to one half. Therefore, according to the law of large numbers, the proportion of heads in a large number of coin flips should be roughly one half. In particular, the proportion of heads after \(n\) flips will almost surely converge to one half as \(n\) approaches infinity.

4. The relative frequency of white counters is 0.3, and there are 20 counters in the bag, so, as an estimate, \(0.3 \times 20 = 6\) white counters.

5. The probability that person 1 AND person 2 AND person 3 will take public transportation is: \(.6 \times .6 \times .6 = 0.216\)

6. The probability of NOT getting a 4 on any roll is: \((5/6)(5/6)(5/6)(5/6) = .48\), so the probability of rolling at least one 4 is: \(1 - .48 = 0.52\).

7. \(P(\text{spade on the first draw}) = 13/52 = .25, P(\text{spade on the second draw | spade on the first draw}) = 12/51 = .235, .25 \times .235 = 0.0588\)

8. True.

9. There are 7 + 3 = 10 favourable outcomes out of a possible 15. 10/15 = 0.667

10. Two events are **independent** if the first one does not influence the second. For example, if a bag contains 2 blue balls and 2 red balls and two balls are selected randomly, the events are:

    - independent if the first ball is replaced after being selected
    - not independent if the first ball is removed without being replaced. In this instance, there are only three balls remaining in the bag so the probabilities of selecting the various colours have changed.

   Two events are independent if (and only if): \(P(A\cap B) = P(A) \times P(B)\)

11. The correct answer is A: 77/100.
Chapter 15

1. The null hypothesis says that any apparent effect is due to chance, so in this case, the null hypothesis would be that the two population means for the 1st and 2nd graders are equal. The null hypothesis is usually not the researcher's hypothesis.

2. \( H_0: \mu = \mu_0 \) \( H_a: \mu < \mu_0 \) (since Maureen thinks she makes less)
   \( \mu_0 = 100 \)
   \( \bar{x} = 93 \)
   standard deviation = 15
   \( n = 5 \)
   The sample is small; however the question has stated that the z score statistic should be used.
   \[ z = \frac{100 - 93}{15/\sqrt{5}} = -1.043 \]
   The critical z score value is -1.64 which is higher. We therefore do not reject the null hypothesis.
   Note. In this test, the p value would be 0.148. Since this value is greater than the stated level of significance (0.05), we fail to reject the null hypothesis.
   In other words, there is not sufficient evidence to support Maureen's claim.

3. C
   \( H_0: \mu < 3 \) years
   so even if \( \alpha = 0.01 \) (1 per cent level), the P-value is less.
   Therefore reject the null hypothesis and assume three or more years on average between major services.

4. C
   \( H_0: p = 0.4 \)
   \( H_a: p < 0.4 \)
   The test is one sided, and \( H_a \) is assumed true for smaller values of \( P \), which are to the left.
   \( H_0 \) could have been written as \( P \geq 0.4 \)

5. C
   If P-value = 0.2456, even if \( \alpha = 0.05 \) (5 per cent level), the P-value is still greater.
   So, fail to reject the null hypothesis.

6. B
   The null hypothesis is that the statement could be true by chance alone even if the true mean is < 3 years. The test is right tailed, so can have \( H_0: \) population mean < 3 years.

7. C
   \( H_0: \) average age a child moves out (\( \mu \)) > 23
   Type I error is rejecting the null hypothesis when it is true.
   i.e. assuming \( \mu \leq 23 \), when it is >23.

8. A
   \( H_0: \) average age a child moves out (\( \mu \)) > 23
   Type II error is not rejecting the null hypothesis when it is false.
   i.e. assuming \( \mu > 23 \) when it is \( \leq 23 \).
Examples of variables which might be correlated include a person's height and weight or the distance of a journey and the time it takes to make it.

The correlation coefficient of 0.9 is very close to 1 and so there is a very strong relationship between \( x \) and \( y \).

Correlation coefficient = \( r = 0.95 \)

Coefficient of determination = \( r^2 = 0.95^2 = 0.9025 \) or 90.25 per cent

This tells us that over 90 per cent of the variations in the dependent variable (\( Y \)) can be explained by variations in the independent variable, \( X \).

The equation of a straight line has the form \( y = a + bx \)

where: \( x \) and \( y \) are related variables
\( x \) is the independent variable
\( y \) is the dependent variable
\( a \) is the intercept of the line on the vertical axis
\( b \) is the gradient of the line.

The independent variable is denoted by \( X \) and the dependent one by \( Y \).

The variable to be forecast must always be \( Y \). Option A is therefore incorrect.

In calculating the correlation coefficient, it does not matter which variable is \( X \) and which is \( Y \), but a totally different regression line equation will result if \( X \) and \( Y \) are interchanged. Option B is therefore incorrect.

The scatter diagram is used to show whether or not there is a relationship between \( X \) and \( Y \) and it does not matter which variable is associated with a particular axis. Option D is therefore incorrect.
Glossary of terms
Part one: Economics

**Absolute advantage.** A country has an absolute advantage in the production of a good if it can produce more of the good with a fixed amount of resources than can any other country.

**ACCC.** Australian Competition and Consumer Commission.

**Accelerator principle.** The theory that investment changes disproportionately in response to change in output.

**Aggregate demand (AD).** The total demand in the economy for goods and services.

**Aggregate supply (AS).** The total supply of goods and services in the economy.

**APRA.** Australian Prudential Regulatory Authority.

**Average cost (AC).** Average cost for a given level of output is simply the total cost divided by the total quantity produced.

**Barriers to entry.** Factors which make it difficult for suppliers to enter a market.

**Broad money.** 'It provides an indicator of the private sector's holdings of relatively liquid assets – assets which could be converted with relative ease and without capital loss into spending on goods and services' (*UK Treasury Economic Progress Report*).

**Central bank.** A bank which acts on behalf of the government.

**Command economy.** Sometimes referred to as state controlled. In this type of economy decisions are taken collectively, usually by central planning committees.

**Comparative advantage.** The advantage in the production of a product enjoyed by one country over another when that product can be produced at a lower opportunity cost.

**Complements.** Goods that tend to be bought and used together, so that an increase in the demand for one is likely to cause an increase in the demand for the other.

**Cost push inflation.** Inflation resulting from an increase in the costs of production of goods and services.

**Deflation.** The name given to an decrease in price levels generally.

**Demand.** The quantity of that good that potential purchasers would buy, or attempt to buy, if the price of the good were at a certain level.

**Demand pull inflation.** Inflation resulting from a persistent excess of aggregate demand over aggregate supply.

**Deregulation.** The removal or weakening of any form of statutory (or voluntary) regulation of free market activity.

**Economies of scale.** Factors which cause average cost to decline in the long run as output increases.

**Exchange rate.** An exchange rate is the ratio at which two currencies are traded. The price of one currency in terms of another.

**Externalities.** Effects of a transaction which extend beyond the parties to the transaction. The differences between the private and the social costs, or benefits, arising from an activity are externalities.

**Fiscal policy.** Government policy on taxation, public borrowing and public spending.

**Fixed costs.** Costs which do not change when levels of production change, for example, the rent of premises.

**Free market economy.** Sometimes called capitalism. In this type of economy most decisions are taken through the operation of the market mechanism.

**Gini coefficient.** Based on the Lorenz curve, it is the ratio of between 0 and 1 commonly used to measure income inequality within a population.

**Gross domestic product.** The value of the goods and services produced by an economy in a given period.
**Gross national income.** GDP plus income accruing to domestic residents from investments abroad less income accruing to foreign residents from investments in the domestic economy.

**Inflation.** The name given to an increase in price levels generally.

**Interest rate.** A rate that is charged or paid for the use of money.

**Laffer curve.** A curve depicting the relationship between tax revenue and the average tax rate, designed to illustrate the thesis that there is an optimal tax rate at which tax revenues are maximised.

**Law of diminishing returns.** If one or more factors of production are fixed, but the input of another is increased, the extra output generated by each extra unit of input will eventually begin to fall.

**Liquidity preference.** People's preference for holding on to their savings as money (in liquid form) rather than investing it.

**Long-run.** A period sufficiently long to allow full flexibility in all the inputs used.

**Lorenz curve.** A graph depicting income distribution across a population.

**Macroeconomics.** The study of the aggregated effects of the decisions of economic units. It looks at a complete national economy or the international economic system as a whole.

**Marginal cost (MC).** The addition to total cost of producing one more unit of output.

**Marginal utility.** The satisfaction gained from consuming one additional unit of a good or the satisfaction forgone by consuming one unit less.

**Market.** A situation in which potential buyers and potential sellers (suppliers) of a good or service come together for the purpose of exchange.

**Market demand.** The total quantity of a product that all purchasers would want to buy at each price level.

**Market failure.** When a free market mechanism fails to produce the most efficient allocation of resources.

**Market supply curve.** The aggregate of all the supply curves of individual firms in the market.

**Merit goods.** Considered to be worth providing in greater volume than would be purchased in a free market, because higher consumption is in the long-term public interest.

**Microeconomics.** The study of individual economic units; these are called households and firms.

**Mixed economy.** An economy where there is a balance between market forces and state intervention.

**Monetarism.** A theory that advocates that the level of national income is determined by the quantity of money in circulation (the quantity theory of money).

**Monetary policy.** Government policy on the money supply, the monetary system, interest rates, exchange rates and the availability of credit.

**Monopolistic competition.** A market structure in which firms' products are comparable rather than homogeneous.

**Monopoly.** In a monopoly, there is only one firm: the sole producer of a good which has no closely competing substitutes.

**Monopsony.** A market structure where there is one buyer and many sellers.

**Multiplier.** The ratio of the total increase in national income to an initial increase.

**Narrow money.** 'Money balances which are readily available to finance current spending, that is to say for transactions purposes' (UK Treasury Economic Progress Report).

**National income.** The sum of all incomes which arise as a result of economic activity, that is from the production of goods and services.

**National income accounting.** The system through which activity and national scale is measured.

**Net present value.** The value obtained by discounting all cash outflows and inflows of a capital investment project by a chosen target rate of return or cost of capital.

**Nominal rates of interest.** Rates expressed in money terms.

**Oligopoly.** A market structure where a few large suppliers dominate.
Oligopsony. A market structure where there are few buyers and many sellers.

Opportunity cost. The cost of an item measured in terms of the alternatives forgone.

Perfect competition. A theoretical market structure in which no supplier has an advantage over another.

Phillips curve. A graphical illustration of the historic inverse relationship between the rate of wage inflation and the rate of unemployment.

Price cartel. Created when a group of oligopoly firms combine to agree on a price at which they will sell their product to the market.

Price ceiling. Government-imposed limit on the maximum price that can be charged or paid for a product.

Price elasticity of demand. A measure of the extent of change in the market demand for a good in response to a change in its price.

Price elasticity of supply. Indicates the responsiveness of supply to a change in price.

Price floor. A government-imposed limit on the minimum price that can be charged or paid for a product.

Private good. A good which must be exclusive and rivalrous. Exclusivity means that it is reasonable to prevent a class of consumers (or firms) from consuming the good. Rivalrous means that consumption by one party prevents simultaneous consumption by another.

Privatisation. The transfer by government of state owned activities to the private sector.

Product differentiation. Gives the products some market power by acting as a barrier to entry.

Production. The process and method employed to transform tangible and intangible inputs into goods and services.

Productivity. A measure of the efficiency with which output has been produced.

Profit. Equal to total revenue minus total cost of any level of output.

Quantity theory of money. The theory which holds that changes in the level of prices are caused predominantly by changes in the supply of money. Also known as the classic theory of money.

Real rates of interest. The rates of return that investors get from their investment, adjusted for the rate of inflation.

Regulation. Any form of state interference in the operation of the free market.

Reserve Bank of Australia. The independent central bank of Australia responsible for monetary policy.

Scarcity. The excess of human wants over what can actually be produced.

Short-run. The short-run is a time period in which the amount of at least one input is fixed.

Social goods. Some goods, by their very nature, involve so much ‘spill-over’ of externalities that they are difficult to provide except as social goods whose production is organised by the government.

Specific tax. Tax charged as a fixed sum per unit sold.

Stagflation. A time of slow economic growth, inflation and relatively high unemployment.

Subsidy. A payment to the supplier of a good by the government.

Substitute goods. Goods that are alternatives to each other, so that an increase in the demand for one is likely to cause a decrease in the demand for another. Switching demand from one good to another ‘rival’ good is substitution.

Supply. The quantity of a product that existing or would-be suppliers would want to produce at a given price.

Total cost (TC). Comprises total fixed cost (TFC) and total variable cost (TVC) for a given level of output.

Total utility. The total satisfaction that people derive from spending their income and consuming goods.

Utility. The pleasure or satisfaction or benefit derived by a person from the consumption of goods.

Variable costs. Costs which change according to the level of output, for example, raw material costs.

Withdrawals. Movements of funds out of the cycle of income and expenditure between firms and households.
Part two: Statistics

Alpha ($\alpha$). The probability of making a Type I error.

Alternative hypothesis. Denoted by $H_1$ or $Ha$ is the hypothesis that sample observations are influenced by some non-random cause. The alternative hypothesis is the hypothesis that the researcher expects to support.

Arithmetic mean. Calculated from the sum of values of items divided by the number of items. The arithmetic mean of a variable $x$ is denoted by $\bar{x}$ ('x bar').

Attribute. Something an object either has or does not have. It cannot be measured. For example, an individual is either male or female. There is no measure of how male or how female somebody is: the sex of a person is an attribute. These types of data are qualitative.

Bar graph. Also known as the bar chart, is one of the most common methods of presenting data in a visual form. Bar graphs are a family of charts that display quantitative information by means of a series of rectangles that can be displayed horizontally or vertically. Each rectangle represents a data element in a data series and a complete set of bars represents a data series.

Beta ($\beta$). The probability of making a Type II error.

Binomial distribution. Measures the probabilities of the number of successes over a given number of trials with the same chance of success in each try.

Census. A survey of an entire population, as opposed to a sample survey.

Central tendency (location) of data. Where data tend to fall, as measured by the mean, median, and mode.

Class interval. A range of values of a variable; an interval used in dividing the scale of the variable for the purpose of tabulating the frequency distribution of a sample.

Cluster sampling. A non-random sampling method that involves selecting one definable subsection of the population as the sample, that subsection taken to be representative of the population in question.

Coefficient of determination ($r$-squared). The square of the correlation coefficient. Its value may vary from zero to one. It has the advantage over the correlation coefficient in that it may be interpreted directly as the proportion of variance in the dependent variable that can be accounted for by the regression equation.

Component bar graph. A bar graph that gives a breakdown of each total into its components. The total length of each bar and each component on a component bar graph indicates magnitude (a bigger amount is shown by a longer bar). This is sometimes called a stacked bar graph.

Conditional probability. Denoted by $P(A|B)$. This phrase is read: the probability that $A$ will occur given that $B$ is known to have occurred.

Consumer panels. Panels with personal visits are called home audit panels and panels which send data by post are called diary panels. For example, a panel of households might keep a purchase diary of the goods they have bought, and submit this diary regularly to the market research company. Panels might be established for a long-term or short-term period.

Consumer panels (test panels). Some research firms have created consumer panels consisting of a representative cross-section of consumers who have agreed to give regular information about their attitudes or buying habits through personal visits or mail questionnaires.

Continuous variable. A variable which can assume an infinite number of values. Between any two measures of weight (e.g., 50 to 51 kg) lie an infinite number of possible values (e.g., 50.1, 50.2, 50.21...).

Correlation. A standardised index of the strength and direction of the relationship between two variables. The range for the possible correlation between any two variables is from -1.00 (a perfect inverse relationship) to +1.00 (a perfect positive relationship).

Cumulative distribution (or cumulative frequency table). Can be used to show the total number of times that a value above (or below) a certain amount occurs.
Data. The raw material for data processing. Data consists of numbers, letters and symbols and relates to facts, events, and transactions.

Decile. A subset of adjacent scores in a distribution representing 10% of a sample or a population. A ‘decile score’ is a raw score corresponding to the 10th, 20th, or 30th etc. percentage score.

Decision-making. The selection of a course of action from among alternatives. A decision is a choice whereby a person forms a conclusion about a situation.

Dependent (or conditional) events. Events where the outcome of one event depends on the outcome of the other.

Descriptive statistics. Ways of describing, organising, summarising, and presenting large sets of quantitative (numerical) data.

Discrete data. Data are said to be discrete when they can only take on specific fixed values, e.g. the actual number of vehicles through a car wash per day could be 35 but not 35.3, a shoe size could be 5½ but not 5.193 and 9 people could enter a supermarket between 14.30 and 15.00 but not 9.999.

Dispersion. The ‘spread’ of a data set; the departure from central tendency.

Distribution. In a distribution, the horizontal axis (x-axis) represents the variable being described. The density of the smooth curve over the x-axis represents the probability of occurrence for each of the values on the x-axis.

Event. Any collection of outcomes from a probability experiment. An event may consist of one or more simple events. Events are denoted using capital letters such as E. The probability of an event, denoted P(E), is the likelihood of that event occurring.

Experiment. An activity that is either observed or measured, such as tossing a coin, or drawing a card. It is any process that can be repeated in which the results are uncertain.

Exponential distribution. Frequently described as the waiting time distribution since it can be used to model the time until some event happens or something stops working. For example, the distances traveled by a bus between major mechanical breakdowns or the lengths of time between earthquakes with a magnitude greater than seven in California.

External data. Relates to data collected from outside or the ‘environment’ of the organisation. Data relating to the environment of an organisation might be classified as political, economic, social, technological and competitive (such as the behaviour of customers, suppliers and rivals).

Frequency distribution. A summary of the values obtained in a survey and the frequencies with which these values have occurred.

Grouped frequency distributions. If there is a large set of data or if every (or nearly every) data item is different, it is often convenient to group frequencies together into bands or classes.

Histogram. A data presentation method for (usually) grouped data of a continuous variable. Visually similar to a bar chart but frequencies are represented by areas covered by the bar rather than their height.

Hypotheses. A set of two or more mutually exclusive and often exhaustive statements. The goal of hypothesis testing is to determine which is true.

Hypothesis testing. A procedure, based on sample evidence and probability, used to test claims regarding a characteristic of one or more populations. The characteristic is usually population mean or population proportion.

Independent events. Events where the outcome of one event in no way affects the outcome of the other events.

Independent samples t-test. In hypothesis testing, this is the procedure used to compare the means of two different samples. As is true for all t-tests, the standard error is not known and is estimated from sample data.

Information. Data that has been processed in such a way as to be meaningful to the person who receives it. Data is subjected to a ‘value-added’ process (data processing or information processing) where its form is aggregated, manipulated and organised or its content is analysed and evaluated and is placed in a proper context for a human user.
**Intercept.** The point at which a straight line crosses the y-axis.

**Internal data.** Relates to activities or transactions performed within the organisation, e.g. administrative tasks such as correspondence or payroll calculations, the production of products and services, or the sale of those products. Often these activities generate costs and revenues; so much of the internal data collected will be quantitative. **multiple bar graph (or compound bar graph)** is a bar graph in which two or more separate bars are used to present sub-divisions of data. This is sometimes called a side-by-side bar graph.

**Inter-quartile range.** The difference between the values of the upper and lower quartiles and hence shows the range of values of the middle half of the population, i.e. upper quartile (Q3) – lower quartile (Q1).

**Interval data.** Data that possess magnitude (one value can be judged greater than, less than, or equal to another) and a constant distance between intervals (units of measurement are the same on the scale regardless of where the unit falls). Temperature is an example of interval data: the difference between 100 degrees and 99 degrees is the same as the difference between 40 degrees and 39 degrees. Interval data do not necessarily have an absolute zero point (i.e., a temperature of zero degrees does not indicate that there is no temperature).

**Joint probability.** Measures the likelihood that two or more events will happen concurrently. It is denoted by \( P(A \text{ and } B) \) or \( P(A \cap B) \). To become eligible for the joint probability, both events A and B must occur.

**Kurtosis.** The degree of flatness or peakedness of a graph of a frequency distribution. The relatively flat distributions are described as platykurtic. Distributions with medium curvature are mesokurtic (note: a normal distribution is mesokurtic). The most peaked distributions are leptokurtic.

**Least squares method of linear regression analysis.** Provides a technique for estimating the equation of a line of best fit. The term ‘squares’ in ‘least squares regression analysis’ refers to the squares of the differences between actual values of the dependent variable (Y) and predicted values given by the regression line of best fit. These differences are referred to as residuals or residual errors. ‘Least squares’ means that the line of best fit that is calculated is the one that minimises the sum of the squares of all the residuals.

**Leptokurtic.** A distribution that is more peaked than a normal distribution. This indicates that there are more cases concentrated close to the mean than in a normal distribution.

**Line of best fit (or least squares fit).** The least squares fit procedure allows us to reduce the scatterplot to a single straight line described by a linear equation. It minimises the square of the vertical distance between each point and the regression line.

**Mail surveys.** Relatively low in cost. As with any other survey, problems exist in their use when insufficient attention is given to getting high levels of co-operation. Mail surveys can be most effective when directed at particular groups, such as subscribers to a specialised magazine or members of a professional association.

**Marginal probability (or unconditional probability).** The probability of an event, regardless of the results of any other events. It is denoted \( P(A) \), where A is the event.

**Mean.** A measure of central tendency calculated by dividing the sum of the scores in a distribution by the number of scores in the distribution. This value best reflects the typical score of a data set when there are few outliers and/or the dataset is generally symmetrical.

**Median.** The value of the middle member of a distribution once all of the items have been arranged in order of magnitude. (i.e. an equal number of scores lie above and below it). As a measure of central tendency, it is largely unaffected by extreme values.

**Mode.** The value which indicates the most frequently occurring value.

**Multiple bar graph (or compound bar graph).** A bar graph in which two or more separate bars are used to present sub-divisions of data. This is sometimes called a side-by-side bar graph.

**Mutually exclusive outcomes.** Outcomes where the occurrence of one of the outcomes excludes the possibility of any of the others happening.

**Negative skew.** Asymmetry in a distribution in which the scores are bunched to the right side of the centre. With a negatively skewed distribution, the mean generally falls to the left of the median and the median usually lies to the left of the mode.
**Nominal level data.** A nominal measurement scale used for variables in which participants or observations are put into mutually exclusive categories. For example, categorising study participants into 'male' and 'female' categories demonstrates that 'sex' is measured on a nominal scale.

**Normal distribution.** A continuous, symmetric, bell-shaped distribution of a variable.

**Null hypothesis.** Denoted by $H_0$, it is usually the hypothesis that sample observations result purely from chance. It is the prediction that the researcher believes will be ‘nullified’ i.e. the researcher believes this prediction is not true.

**Observation.** Used as a means of obtaining sample data where quantitative data are required. For example, if data are needed about the volume of traffic passing along a road at a certain time of day, observers (either people or recording equipment) can be placed so as to count the traffic as it passes by. Observation can also be used to study consumer behaviour, although this is usually within a controlled experiment.

**One-tailed tests.** The rejection region of the test is dependent upon the sign of the test statistic and the null hypothesis stated. Generally, the rejection region for positive critical values encompass values greater than the critical value. Negative critical values generally are associated with rejection regions less than the critical value (i.e. more negative in value). Two-tailed tests relate to both tails of the distribution, hence values in the rejection region encompass the area of the distribution greater than the critical value and less than the negative of the critical value.

**Ordinal level data.** Indicates something about the rank ordering of study participants. The categories for an ordinal set of data have a natural order. For example, suppose a group of people were asked to taste varieties of biscuit and classify each biscuit on a rating scale of 1 to 5, representing strongly dislike, dislike, neutral, like, strongly like. A rating of 5 indicates more enjoyment than a rating of 4, for example, so such data are ordinal. However, the distinction between neighbouring points on the scale is not necessarily always the same. For instance, the difference in enjoyment expressed by giving a rating of 2 rather than 1 might be much less than the difference in enjoyment expressed by giving a rating of 4 rather than 3.

**Outcome.** A list of possibilities, usually with a probability assigned to each.

**Outlier.** A value in a data set that is very different from most other values in the set.

**Pearson’s correlation coefficient.** Measures the degree of correlation between two variables, $r$. The nearer $r$ is to $+1$ or $-1$, the stronger the relationship.

**Percentile.** A measure that expresses position in terms of a percentage for the data set. **Percentile** is a value that exceeds a specific percentage of the distribution. Therefore, if the 63rd percentile score for a set of students on a verbal exam is 560, then 63 per cent of scores are at or below 560.

**Personal interviews.** May be conducted by market researchers in the interviewee’s home. In-person interviews in a respondent’s home or office are much more expensive than mail or telephone surveys. They may be necessary, however, especially when complex information is to be collected.

**Pie chart.** A chart which is used to show pictorially the relative size of component elements of a total. It is a circle which is divided into segments. Each segment represents a particular category. The area of each segment is proportional to the number of cases in that category.

**Platykurtic.** A distribution that is flatter than a normal distribution. This means that there are more cases in the tails of the distribution than in a normal distribution.

**Population.** The set of all possible data values that could be observed. It is the entire group of individuals that we want information about.

**Positive skew.** Asymmetry in a distribution in which the scores are bunched to the left side of the centre. With a positively-skewed distribution, the mean generally falls to the right of the median and the median usually lies to the right of the mode.

**Primary data.** Collected especially for a particular enquiry. Raw data are primary data which have not been processed at all, and which are still just a list of numbers. The main sources of primary data are personal investigation, teams of investigators, interviews, questionnaires and telephone surveys.

**Probability.** A measure of likelihood and can be stated as a percentage, a ratio, or more usually as a number from 0 to 1. In statistics, probabilities are more commonly expressed as proportions than as percentages.
**Probability matrices.** Tools for solving probability problems. A probability matrix displays the marginal probabilities and the intersection probabilities of a given problem. Union probabilities and conditional probabilities can be computed from the matrix. Generally a probability matrix is constructed as a two dimensional table.

**Probability sample.** Sampling in which each element within a study population has a known, nonzero chance of being selected into the sample.

**p-value.** The strength of evidence in support of a null hypothesis is measured by the p-value. Suppose the test statistic is equal to T. The p-value is the probability of observing a test statistic as extreme as T, assuming the null hypothesis is true. If the p-value is less than the significance level, we reject the null hypothesis.

**Qualitative data.** Cannot be measured numerically but may reflect distinguishing characteristics or attributes, e.g. town where a product is manufactured.

**Quantiles.** Quartiles, deciles, and percentiles and any other similar dividing points for analysing a frequency distribution are referred to collectively as quantiles. The purpose of quantiles is to analyse the dispersion of data values.

**Quantitative data.** Data that can be measured. Examples of quantitative data include the temperature, which can be measured in degrees Fahrenheit or Celsius, or the time it takes you to swim 50 metres in a swimming pool. This can be measured in minutes and seconds.

**Quartiles.** One means of identifying the range within which most of the values in the population occur. The lower quartile is the value below which 25 per cent of the population fall and the upper quartile is the value above which 25 per cent of the population fall.

**Quintile.** A subset of adjacent scores in a distribution representing 20 per cent of a sample or a population. A 'quintile score' is a raw score corresponding to the 20th, 40th, 60th, or 80th percentile score.

**Random sample.** A sample that contains observations which are selected from a population so that every member of the population has a known chance of selection for a sample.

**Random sampling.** Choosing a sample in such a way that at each step all members of the population not already selected stand an equal chance of being chosen.

**Random variable.** The measurements of a random variable vary in a seemingly random and unpredictable manner. A random variable assumes a unique numerical value for each of the outcomes in the sample space of the probability experiment.

**Range.** A simple measure of dispersion, indicating the difference between the lowest and highest values observed.

**Ratio level data.** With interval scales, there is no absolute zero point. For this reason, it is inappropriate to express interval level measurements as ratios; it would not be appropriate to say that 60 degrees is twice as hot as 30 degrees. Ratio scales do have a meaningful and fixed zero point, which allows for such ratio comparisons. Weight, for example, has a definite zero (no weight) and 10 kg is twice as much as 5 kg.

**Raw data.** For example, a list of results from a survey. Data needs to be summarised and analysed to give it meaning. One of the most basic ways is the preparation of a table.

**Region of acceptance.** A range of values. If the test statistic falls within the region of acceptance, the null hypothesis is accepted. The region of acceptance is defined so that the chance of making a Type I error is equal to the significance level.

**Regression.** A statistical procedure that allows us to determine the extent to which we can predict a given observation's score on a dependent variable, given that observation's score on one or more independent variables.

**Regression analysis.** Concerned with how the values of a dependent variable depend on the corresponding values of the independent variable.

**Regression coefficient.** The slope of the regression line. It represents the change in y for every one unit change in x.
Regression line. A model that simplifies the relationship between two variables. By approximating a line through the centre of a scatterplot that represents the data, we create a two dimensional 'centre' for the data. The line summarises the data points in the same way that measures of central tendency do.

Sample. A collection of observations selected from a larger population. The sample is the part of the population that we actually examine in order to gather information.

Sample design. Making a plan regarding the size and selection of the sample, collection of the sample data and preparation of the final results based on the sample study.

Sample space. The set of all possible outcomes in an experiment. The sample space for the roll of a single die is 1, 2, 3, 4, 5, and 6.

Sampling error. The extent to which a sample distribution is different than the population distribution from which the sample is drawn.

Scatter diagrams (or scatter plots). Graphs which are used to exhibit data (rather than equations) in order to compare the way in which two variables vary with each other. A group of data points are plotted along x-axis and y-axis coordinates and every individual is represented as a data point, whereby a perpendicular line from the individual’s ‘X’ value intersects a horizontal line from the individual’s ‘Y’ value.

Secondary data. Data which has already been collected elsewhere, for some other purpose, but which can be used or adapted for the survey being conducted. For example from government, banks, newspapers, the Internet.

Shape. Described by symmetry, skewness, and kurtosis (peakedness) of data.

Significance. A statistical term that tells how sure you are that a difference or relationship exists. To say that a significant difference or relationship exists only tells half the story. We might be very sure that a relationship exists, but is it a strong, moderate, or weak relationship? After finding a significant relationship, it is important to evaluate its strength. Significant relationships can be strong or weak. Significant differences can be large or small, depending on the sample size.

Simple bar graph. A graph consisting of one or more bars, in which the length of each bar indicates the magnitude of the corresponding data item.

Simple regression. Used to examine the relationship between one dependent and one independent variable.

Single sample t-test. In hypothesis testing, this is the procedure used to compare the mean of one sample to a known population mean. As is true for all t-tests, the standard error is not known and is estimated from sample data.

Skewness. Asymmetry in a distribution in which scores are bunched on one side of the distribution.

Standard deviation. σ. A measure of dispersion describing the spread of scores around the mean. It is the square root of the variance.

Standard error. The standard deviation of a sampling distribution.

Standard error (SE) of the mean. How much the sample mean varies from sample to sample (it is the standard deviation of the sample mean given a particular sample size (n)).

Standard error of the mean of difference scores. The standard deviation of a sampling distribution of the mean of difference scores.

Statistics. A set of methods that are used to collect, analyse, present, and interpret data. They are the means by which data are interpreted to give meaningful information.

Stratified sampling. A variation on the random sampling method. This is the best method of choosing a sample in many situations. The sample frame must be divided into strata or categories. The strata may be regions, towns, streets, sexes, age groups, social classes, occupations and so on, depending on the nature of the enquiry.

Survey. A method of descriptive research used for collecting primary data based on verbal or written communication with a representative sample of individuals or respondents from the target population.

Symmetrical frequency distribution. Like the normal distribution, the mean, median, and mode are all the same value, M. Its two halves mirror each other.
Symmetry. Implied when data values are distributed in the same way above and below the middle of the sample.

Tables. A simple way of presenting information about two variables. A table is a matrix of data in rows and columns, with the rows and the columns having titles.

Tabulation. Putting data into tables.

Telephone interviews. An efficient method of collecting some types of data and are commonly used. They lend themselves particularly well to situations where timeliness is a factor and the length of the survey is limited.

T-test. A statistical test for the mean of a population.

Type I error. Occurs if we reject the null hypothesis when it is true.

Type II error. Erroneously failing to reject the null hypothesis: concluding that a sample came from the given population when in fact is from a different population.

Uniformity. When the observations in a set of data are equally spread across the range of the distribution, the distribution is called a uniform distribution. A uniform distribution has no clear peaks.

Union probability. Denoted by P(A or B) or P(A \cup B), where A and B are two events. P(A or B) is the probability that A will occur or that B will occur or that both A and B will occur.

Variable. Something that can be measured. For example, the height of a person is a variable which can be measured according to some scale (such as centimetres). These types of data are quantitative. Variables may be classified as discrete (can only take a finite or countable number of values within a given range) or continuous (may take on any value).

Variance. A measure of dispersion, indicating the mean of the squared deviations of a set of scores from the mean of the scores.

Venn diagram. A pictorial method of showing probability.

y-intercept. The point through which a line intersects the Y-axis. It is the value of y when x equals zero.

z-score. The number of standard deviations you are from the mean of 0 (recall that by subtracting the mean and dividing the result by the standard deviation you convert your data to the standard normal distribution which has a mean of 0 and standard deviation of 1). The z-score indicates the number of standard deviation units of the sample from the population mean.

z-test. Compares sample and population means to determine if there is a significant difference. It requires a simple random sample from a population with a normal distribution where the mean is known. A one-sample z-test can be used to assess whether a sample drawn at random from a population tends to have the same characteristics as the population from which it is taken.
Formulae
Statistics (Chapters 12 to 16)

The arithmetic mean of a frequency distribution.

Using the Σ sign to mean the sum of what follows, the arithmetic mean of a frequency distribution is:

\[ \bar{x} = \frac{\sum fx}{n} \] or \[ \frac{\sum fx}{\sum f} \] where \( n \) is the number of values recorded, or the number of items measured, \( f \) is the number of occurrences of each value, and \( fx \) is each value.

Variance = \( \sigma^2 = \frac{\sum (x - \bar{x})^2}{n} \) \( \sigma \) is the lower case Greek letter sigma. The variance is often called 'sigma squared'.

Standard deviation = \( \sigma = \sqrt{\frac{\sum (x - \bar{x})^2}{n}} \)

Sample standard deviation = \( s = \sqrt{\frac{\sum (x - \bar{x})^2}{n - 1}} \)

The probability density function (PDF) of an exponential distribution is:

\[ f(x; \mu) = \frac{1}{\mu} e^{-\frac{x}{\mu}} \text{, where } x > 0, \text{ and } \mu > 0 \]

where: 1/\( \mu \) = constant event rate, in events per unit of measurement, e.g. events per hour, per cycle, etc.

\( \mu \) = mean time between events, or to an event.

\( x \) = operating time, life or age, in hours, cycles, miles, actuations, etc.

The probability addition law

If A and B are events, the probability of obtaining either of them is:

\[ P(A \text{ or } B) = P(A) + P(B) - P(A \text{ and } B) \]

The probability multiplication law

The simple multiplication law for two independent events, A and B, is as follows:

\[ P(A \text{ and } B) = P(A \cap B) = P(A) \times P(B) \]

Note that \( P(A \text{ and } B) = 0 \) when A and B have mutually exclusive outcomes.

The general rule of multiplication for two dependent events, A and B is as follows.

\[ P(A \text{ and } B) = P(A) \times P(B|A) \]

The conditional probability of an event, such as A, occurring given that another event, such as B, has occurred is expressed as:

\[ P(A|B) = \frac{P(A \text{ and } B)}{P(B)} \]

The binomial formula

\[ P(x) = \frac{n!}{x!(n-x)!} \pi^x (1-\pi)^{n-x} \]

Correlation coefficient, \( r = \frac{n\sum XY - \sum X \sum Y}{\sqrt{[n\sum X^2 - (\sum X)^2][n\sum Y^2 - (\sum Y)^2]}} \)

Where: \( X \) and \( Y \) represent pairs of data for two variables \( X \) and \( Y \).

\( n \) = the number of pairs of data used in the analysis.
The **equation of a straight line** has the form $y = a + bx$

where:
- $x$ and $y$ are related variables
- $x$ is the independent variable
- $y$ is the dependent variable
- $a$ is the intercept of the line on the vertical axis
- $b$ is the gradient of the line.

The **least squares method of linear regression analysis** involves using the following formulae for $a$ and $b$ in $Y = a + bX$.

$$b = \frac{n\sum XY - \sum X \sum Y}{n\sum X^2 - (\sum X)^2}$$

$$a = \bar{Y} - b \bar{X}$$

where:
- $n$ is the number of pairs of data
- $\bar{X}$ is the mean $X$ value of all the pairs of data
- $\bar{Y}$ is the mean $Y$ value of all the pairs of data

A **z-score** is the number of standard deviations that a given $x$ value is above or below the mean. If $z$ represents the z-score for a given $x$ value then

$$z = \frac{x - \mu}{\sigma}$$

where $\mu$ is the population mean.

For a sample, the **z-score** or **z statistic** is:

$$z = \frac{\bar{X} - \mu_x}{\sigma/\sqrt{n}}$$

where $\bar{X}$ is the sample mean, $\mu_x$ is the population mean defined in the hypothesis, $\sigma$ is the population standard deviation, $n$ is the sample size.

The **t-test statistic** is given as:

$$t = \frac{\bar{X} - \mu_x}{S/\sqrt{n}}$$

where $t$ is the test statistic, $\bar{X}$, $\mu$ and $n$ are as previously defined and $S$ is the sample standard deviation.

**Two sample test of proportions**

To compare proportions (or probabilities, $p_1$ and $p_2$) found in samples (of size $n_1$ and $n_2$) from two independent populations,

$$z = \frac{p_1 - p_2}{\sqrt{\frac{p_1(1 - p_1)}{n_1} + \frac{p_2(1 - p_2)}{n_2}}}$$

**Two-tail testing of two sample means from independent populations**

These tests assume the populations are approximately normal with equal variances.

$$s^2_w = \frac{(n_1 - 1)s^2_1 + (n_2 - 1)s^2_2}{n_1 + n_2 - 2}$$

$s^2_1$ is the variance of sample 1 and $s^2_2$ is the variance of sample 2.

$n_1$ and $n_2$ are the sample sizes.
The t-statistic is then

\[ t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_w^2}{n_1} + \frac{1}{n_2}}} \]  

where \( \bar{x}_1 \) and \( \bar{x}_2 \) are the sample means.
Index
A
Absolute advantage, 10
Accelerator Principle, 188
Acceptability, 214
Acceptable error, 392, 393
Accounting profits, 77
Ad valorem tax, 235
Addition law, 367, 378
Aggregate demand, 174
Aggregate supply, 174
Allocative efficiency, 102, 261
Allocative inefficiency, 110, 263
Alternative hypotheses, 389

Alternative, 389
Arc elasticity of demand, 47
Arithmetic average of the squared deviations, 337
Arithmetic mean, 322, 336, 338
Arithmetic mean of combined data, 324
Assigning probability, 362
Asymmetric distribution, 340
Asymmetrical frequency distributions, 330
Attribute, 281
Australian census, 293
Average, 332
Average costs, 75
Average propensity to consume, 178
Average revenue, 70
Averages, 322

B
Balance of payments, 207
Bank of England, 249
Bar graph, 296
Barriers to entry, 104, 110
Benchmark, 249
Bimodal distributions, 340
Binomial distribution, 343, 349
Black marketeers, 35
Books, 292
Breakeven analysis, 73
Broad money, 215
Budget, 230
Budget deficit, 231
Budget surplus, 231
By-product data, 284, 293

C
Capital, 7
Capital-output ratio, 189
Cartels, 115
Cash reserve ratio, 244
Census, 285
Central bank, 248, 249
Central tendency, 319, 321
Centrally planned economy, 6, 269
Ceteris paribus, 25
Circular flow of income, 154
Classical approach, 361
Classical probability, 361, 362
Classical probability approach, 362
Classical theory of money, 218
Cluster sampling, 288
Coefficient of determination (r^2), 425, 435
Coincident indicator, 243
Collectively exhaustive events, 363
Collusion, 115
Collusive oligopoly, 118
Command economy, 6, 130, 269
Comparative advantage, 10
Competition Policy, 267
Complementary outcome, 366, 367
Complements, 25
Component bar graph, 298

Compound bar graph, 300
Conditional events, 370
Conditional probability, 373
Conditions of demand, 26
Constant returns to scale, 82
Consumer panels, 291
Consumer Prices Index (CPI), 207
Consumer surplus, 32
Consumer watchdog bodies, 266
Consumption, 176
Consumption function, 178
Contingency table, 371, 372
Continuous data, 282
Convenience, 214
Correlation, 420
Correlation and causation, 426
Correlation coefficient, 423
Correlation in a time series, 424
Cost accounting, 77
Cost of production, 74
Cost push inflation, 209
Credit controls, 245
Cross elasticity of demand, 55
Crowding out, 181
Cumulative frequency, 329
Cumulative frequency table, 328
Cyclical or demand-deficient unemployment, 203

D
Deadweight burden, 108
Deciles, 333, 334
Decision rules, 402

Decision-making, 280
Deflationary gap, 192, 210
Degrees of correlation, 421
Demand, 22, 46
Demand curve, 22
Demand deficient unemployment, 204
Demand management, 174, 210
Demand pull inflation, 208
Demand schedule, 22
Demerit good, 134
'Demultiplier', 185
Dependent events, 370
Deregulation, 261
Derived demand, 12
Descriptive statistics, 321, 339
Diary panels, 291
Dimensional economies of scale, 84
Diminishing returns, 78
Direct tax, 235
Discrete data, 282
Diseconomies of scale, 82, 83, 85
Disincentive effect, 237
Dispersion, 321, 332
Distribution of income, 26
Dumping, 37
Duopoly, 118
Durability, 214

Economics of scale, 82
Elasticity, 46
Elasticity of demand, 46
Elasticity of supply, 56
Elasticity of supply, factors affecting, 57
Empirical approach, 362
Empirical method, 364
Enterprise, 7
Entrepreneurship, 82
Environment, 283
Equilibrium price, 31, 32
Event, 364
Exchange rate, 243
Expectational inflation, 209
Expectations augmented Phillips curve, 211
Expenditure approach to measuring national income, 159

Experiment, 363
Explicit costs, 77
Export multiplier, 184
Exports, 156
External balance, 230
External data, 283
External economies of scale, 85
Externalities, 131, 133
Extrapolation, 435

Factor incomes, 152
Factors of production, 7, 74
Failure to reject, 391
Firms, 12
Fixed costs, 74, 79
Floor price, 36
Forecasting, 430

Formal data, 283
Free market, 269
Free riders, 132, 134
Frequency distribution, 304, 321
Frictional unemployment, 203
Full employment, 230
Full employment national income, 205

Game theory, 118
‘GDP deflator’, 163, 167
General rule of addition, 368
Giffen goods, 53
Gini coefficient, 268, 270
Government agencies, 292
Government data, 293
Government spending, 156
Government spending multiplier, 184
Gradient, 427
Gross domestic product (GDP), 152
Gross value added, 153
Grouped frequency distributions, 304
Grouped frequency distributions of continuous variables, 305

Histograms, 307
Histograms of frequency distributions with unequal class intervals, 307
Home audit panels, 291
Homogeneity, 215
Households, 12
Hyperinflation, 207
Hypothesis, 387, 400
Hypothesis testing, 388, 393

Imperfections in a market, 131
Implicit costs, 77
Import cost-push inflation, 209
Imports, 156
Income approach to measuring national income, 157, 160
Income effect, 53
Income elasticity of demand, 54
Independent events, 369, 378
Indirect tax, 235
Indirect taxation, 52, 135
Inelastic demand, 48
Inferential statistics, 284, 339, 361
Inferior goods, 26, 55
Inflation, 163, 206, 210, 221, 230, 246
Inflationary expectations, 247
Inflationary gap, 191, 210
Informal data, 283
Information, 281
Injections into the circular flow of income, 155
In-person interviews, 291
Interaction of demand and supply, 20
Interest, 7
Interest rates, 217, 242, 244, 246
Internal data, 283
Internal economies of scale, 85
International trade, 10
Internet, 292
Internet research, 292
Interpolation, 435
Inter-quartile Range, 336
Intersection, 367
Interval level data, 294
Investment, 156, 180
Investment multiplier, 183
IS-LM curve, 193

K
Keynesians, 172
Kinked oligopoly demand curve, 116
Kurtosis, 347

L
Labour, 7
Laffer curve, 236
Lagging indicator, 243
Land, 7
Law of diminishing returns, 79
Law of Statistical Regularity, 285
Leading indicator, 243
Least squares method of linear regression analysis, 430
Leptokurtic distributions, 348
Liberalisation, 261
Liquidity preference, 217
Long run, 58, 74
Long run costs, 81
Lorenz curve, 268, 270

M
Macroeconomics, 7
Mail surveys, 291
Managerial economies, 93
Marginal costs, 75
Marginal efficiency of capital, 182
Marginal propensity to consume (MPC), 177
Marginal propensity to save (MPS), 177
Marginal revenue, 70
Marginal utility, 12, 13, 14
Market, 12
Market clearing price, 32
Market demand, 26
Market demand curve, 24
Market economy, 6
Market failure, 130, 260
Market period, 58
Market research, 292
Market supply curve, 29
Maximum prices, 35
Measures of dispersion, 332
Measures of variability, 332
Merit goods, 134
Mesokurtic distributions, 348
Microeconomics, 7
Microsoft Excel, 397
Minimum efficient scale, 86
Minimum price legislation, 36
Minimum prices, 36
Minimum wages, 38
Mixed economies, 269
Mixed economy, 6
Mode, 325
Mode of a grouped frequency distribution, 326
Monetarists, 172, 209, 221
Monetary policy, 241, 244, 245
Money stock, 215
Money supply, 221, 241
Monopolies, 131
Monopolistic competition, 111
Monopoly, 102
Monopsony, 119
Monopsony/monopsonists, 131
Moral suasion, 245
Multiple bar graph, 300
Multiplication law, 369
Multiplier, 183
Mutually exclusive events, 358, 362, 363, 367, 368, 369, 376, 378, 380
Mutually exclusive outcomes, 367

N
NAIRU (Non-accelerating inflational rate of unemployment), 210
Narrow money, 215
National Debt, 249
National income, 152, 167
Nationalised industries, 265
Natural monopoly, 103
Natural rate hypothesis, 247
Natural rate of unemployment, 210
Negative correlation, 422
Negative kurtosis, 347
Negatively skewed distribution, 347
Net national product, 152
Net present value, 182
New classical school, 248
New quantity theory of money, 220
Nominal level data, 294
Nominal rates of interest, 217
Non-linear relationships, 426
Non-price competition, 112
Normal distribution, 340
Normal distribution test (Z-test), 397
Normal goods, 25, 55
Normal profit, 74, 77, 98
Null, 389
Null hypothesis, 387, 389
<table>
<thead>
<tr>
<th>Term</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Observation</td>
<td>291</td>
</tr>
<tr>
<td>Observational study</td>
<td>284</td>
</tr>
<tr>
<td>Official Cash Rate</td>
<td>249</td>
</tr>
<tr>
<td>Oligopoly</td>
<td>114</td>
</tr>
<tr>
<td>Oligopsony</td>
<td>119</td>
</tr>
<tr>
<td>One-tailed significance tests</td>
<td>400</td>
</tr>
<tr>
<td>One-tailed tests</td>
<td>396</td>
</tr>
<tr>
<td>On-line databases</td>
<td>292</td>
</tr>
<tr>
<td>OPEC</td>
<td>116</td>
</tr>
<tr>
<td>Open market operations</td>
<td>249</td>
</tr>
<tr>
<td>Opinion polls</td>
<td>292</td>
</tr>
<tr>
<td>Opportunity cost</td>
<td>9, 11</td>
</tr>
<tr>
<td>Ordinal level data</td>
<td>294</td>
</tr>
<tr>
<td>Ostentation, goods of</td>
<td>53</td>
</tr>
<tr>
<td><strong>Outcome</strong>, 363</td>
<td></td>
</tr>
<tr>
<td>Output approach to measuring national income</td>
<td>157</td>
</tr>
<tr>
<td>Partial correlation</td>
<td>421</td>
</tr>
<tr>
<td>Pearson’s correlation coefficient, r</td>
<td>422</td>
</tr>
<tr>
<td>Percentage component bar graph</td>
<td>299</td>
</tr>
<tr>
<td>Percentiles</td>
<td>333</td>
</tr>
<tr>
<td><strong>Perfect competition</strong>, 98, 130</td>
<td></td>
</tr>
<tr>
<td>Perfect correlation</td>
<td>421</td>
</tr>
<tr>
<td>Personal interviews</td>
<td>291</td>
</tr>
<tr>
<td>Phillips curve</td>
<td>210</td>
</tr>
<tr>
<td>Pie chart</td>
<td>301</td>
</tr>
<tr>
<td>Platykurtic distributions</td>
<td>348</td>
</tr>
<tr>
<td>Point elasticity of demand</td>
<td>47</td>
</tr>
<tr>
<td>Pollution</td>
<td>132, 135</td>
</tr>
<tr>
<td>Pollution policies</td>
<td>135</td>
</tr>
<tr>
<td>Population</td>
<td>285, 339</td>
</tr>
<tr>
<td>Population mean</td>
<td>391, 401</td>
</tr>
<tr>
<td>Population standard deviation</td>
<td>339, 349</td>
</tr>
<tr>
<td>Population variance</td>
<td>339</td>
</tr>
<tr>
<td>Positive correlation</td>
<td>422</td>
</tr>
<tr>
<td>Positive kurtosis</td>
<td>347</td>
</tr>
<tr>
<td>Precautionary motive</td>
<td>219</td>
</tr>
<tr>
<td>Price and output determination</td>
<td>20</td>
</tr>
<tr>
<td>Price ceiling</td>
<td>35</td>
</tr>
<tr>
<td>Price discrimination</td>
<td>106</td>
</tr>
<tr>
<td>Price elasticity of demand</td>
<td>46</td>
</tr>
<tr>
<td><strong>Price elasticity of supply</strong>, 56</td>
<td></td>
</tr>
<tr>
<td>Price leadership</td>
<td>117</td>
</tr>
<tr>
<td>Price legislation</td>
<td>135</td>
</tr>
<tr>
<td>Price mechanism</td>
<td>31, 131</td>
</tr>
<tr>
<td>Price regulation</td>
<td>35</td>
</tr>
<tr>
<td>Price theory</td>
<td>13</td>
</tr>
<tr>
<td>Primary data</td>
<td>282</td>
</tr>
<tr>
<td>Private benefit</td>
<td>132</td>
</tr>
<tr>
<td>Private cost</td>
<td>132</td>
</tr>
<tr>
<td>Private costs</td>
<td>131</td>
</tr>
<tr>
<td>Private goods</td>
<td>131, 132</td>
</tr>
<tr>
<td><strong>Privatisation</strong>, 262</td>
<td></td>
</tr>
<tr>
<td>Probability</td>
<td>358, 361</td>
</tr>
<tr>
<td>Probability matrices</td>
<td>374</td>
</tr>
<tr>
<td>Probability matrix</td>
<td>375</td>
</tr>
<tr>
<td>Product differentiation</td>
<td>112, 114</td>
</tr>
<tr>
<td>Production</td>
<td>8</td>
</tr>
<tr>
<td><strong>Production possibility frontier</strong>, 8</td>
<td></td>
</tr>
<tr>
<td>Production quotas</td>
<td>37</td>
</tr>
<tr>
<td>Productive inefficiency</td>
<td>263</td>
</tr>
<tr>
<td>Productivity</td>
<td>8</td>
</tr>
<tr>
<td>Profit</td>
<td>8, 70, 77</td>
</tr>
<tr>
<td><strong>Profit maximisation</strong>, 70, 87</td>
<td></td>
</tr>
<tr>
<td>Profit maximising position</td>
<td>72</td>
</tr>
<tr>
<td>Progressive tax</td>
<td>234</td>
</tr>
<tr>
<td>Proportion</td>
<td>404</td>
</tr>
<tr>
<td>Proportional tax</td>
<td>233</td>
</tr>
<tr>
<td>Public sector net cash requirement (PSNCR)</td>
<td>180</td>
</tr>
<tr>
<td>P-values</td>
<td>397</td>
</tr>
<tr>
<td>Quality</td>
<td></td>
</tr>
<tr>
<td>Qualitative controls</td>
<td>244</td>
</tr>
<tr>
<td>Qualitative data</td>
<td>281</td>
</tr>
<tr>
<td>Quantiles</td>
<td>333</td>
</tr>
<tr>
<td>Quantitative controls</td>
<td>244</td>
</tr>
<tr>
<td>Quantitative data</td>
<td>281</td>
</tr>
<tr>
<td>Quantitative regression</td>
<td>420</td>
</tr>
<tr>
<td>Quartiles</td>
<td>333, 334</td>
</tr>
<tr>
<td><strong>Random</strong>, 286</td>
<td></td>
</tr>
<tr>
<td>Random number tables</td>
<td>286</td>
</tr>
<tr>
<td>Random sampling</td>
<td>286, 404</td>
</tr>
<tr>
<td>Ratio</td>
<td>332</td>
</tr>
<tr>
<td>Rationality</td>
<td>13</td>
</tr>
<tr>
<td>Rationing</td>
<td>35</td>
</tr>
<tr>
<td>Raw materials costs</td>
<td>28</td>
</tr>
<tr>
<td>Real rates of interest</td>
<td>217</td>
</tr>
<tr>
<td>Real wage unemployment</td>
<td>204</td>
</tr>
<tr>
<td>Redistribution of wealth</td>
<td>135, 206</td>
</tr>
<tr>
<td>Regression</td>
<td>420</td>
</tr>
<tr>
<td>Regression analysis</td>
<td>426</td>
</tr>
<tr>
<td>Regression coefficients</td>
<td>433</td>
</tr>
<tr>
<td>Regressive tax</td>
<td>233</td>
</tr>
<tr>
<td>Regulation of markets</td>
<td>260</td>
</tr>
<tr>
<td>Regulatory bodies</td>
<td>292</td>
</tr>
<tr>
<td>Rejection region</td>
<td>396</td>
</tr>
<tr>
<td>Relative Frequencies</td>
<td>365</td>
</tr>
<tr>
<td>Relative frequencies approach</td>
<td>361</td>
</tr>
<tr>
<td>Relative probability</td>
<td>364</td>
</tr>
<tr>
<td>Rent</td>
<td>7</td>
</tr>
<tr>
<td>Research and development</td>
<td>93</td>
</tr>
<tr>
<td>Reserve Bank of Australia (RBA)</td>
<td>248</td>
</tr>
<tr>
<td>Reserve requirements</td>
<td>244</td>
</tr>
<tr>
<td>Residual errors</td>
<td>431</td>
</tr>
<tr>
<td>Retail Prices Index</td>
<td>207</td>
</tr>
<tr>
<td>Rounding errors</td>
<td>296</td>
</tr>
<tr>
<td><strong>Sample</strong>, 285, 339</td>
<td></td>
</tr>
<tr>
<td>Sample correlation coefficient</td>
<td>422</td>
</tr>
<tr>
<td>Sample design</td>
<td>286</td>
</tr>
<tr>
<td><strong>Sample space</strong>, 363</td>
<td></td>
</tr>
</tbody>
</table>
Index

Sample standard deviation, 339, 340, 349
Sample variance, 339, 340
Sampling, 285
Sampling frame, 286
Sampling method, 404
Savings, 156, 176
Scarcity, 7
Scarcity of resources, 7
Scatter diagrams, 309, 420
Scattergraph method, 429
Seasonal unemployment, 203
Secondary data, 283
Secular period, 58
Self-regulation, 260
Set-aside, 37
Short run, 58, 74
Short run average cost (SAC) curve, 79
Short run costs, 75
Sigma (Σ), 323
Sigma (σ), 338
Significance level, 399
Simple bar graph, 297
Simple linear regression, 428
Simple regression, 420
Skewed distributions, 331, 345
Skewness, 340, 345
Social benefit, 132
Social cost, 131, 132
Social goods, 134
Specialisation, 11
Speculative motive, 219
Stagflation, 192
Standard deviation, σ, 336, 337, 338, 401
Statistical analysis, 280
Statistical hypothesis, 389
Statistical test procedure, 390
Statistics, 281
Statutory minimum wage, 38
Store of value, 214
Stratified sampling, 287
Structural unemployment, 203
Subjective approach, 362
Subjective probabilities, 365
Subsidies, 135, 139
Substitutes, 25, 53
Substitution effect, 53
Sunk costs, 77
Supernormal profits, 104
Supply, 28
Supply curve, 28
Supply schedule, 28
Survey, 285
Symmetric distribution, 340
Symmetrical frequency distribution, 330
Symmetrical outcomes, 361
Symmetry, 340

T
Tables, 295
Taxation, 156, 232
Technical efficiency, 99
Technical inefficiency, 110, 263
Technological developments, 29
Technological progress, 109
Technological unemployment, 203
Telephone interviews, 291
Term structure of interest rates, 217
Time horizon, 53
Total costs, 75
Total revenue, 70
Trade audits, 291
Trade cycles, 175
Transactions motive, 219
Transfer payments, 155
Transformation curve, 8
Trend line, 310
T-statistic, 392
Two-tailed significance tests, 400
Two-tailed tests, 396

Underlying rate of inflation, 208
Unemployment, 203
Uniform distribution, 340
Unimodal distributions, 340
Union, 367
Union probability, 372
Unit elasticity of demand, 49
Utility, 13

Value added method of measuring national income, 161
Variability, 321, 322
Variable, 281
Variable costs, 74, 79
Variance, 336, 338
Venn diagram, 366, 367, 368
Voluntary unemployment, 204

Wage-price spiral, 209
Wages, 7
Withdrawals from the circular flow of income, 155

X
X-Inefficiency, 110

Z
Z score, 388, 404