Scheme of Work

Cambridge O Level
Statistics 4040

For examination from 2018

Contents

Introduction

This scheme of work has been designed to support you in your teaching and lesson planning. Making full use of this scheme of work will help you to improve both your teaching and your learners' potential. It is important to have a scheme of work in place in order for you to guarantee that the syllabus is covered fully. You can choose what approach to take and you know the nature of your institution and the levels of ability of your learners. What follows is just one possible approach you could take.

Suggestions for independent study **(I)** and formative assessment **(F)** are also included. Opportunities for differentiation are indicated as **basic** and **challenging**; there is the potential for differentiation by resource, grouping, expected level of outcome, and degree of support by teacher, throughout the scheme of work. Timings for activities and feedback are left to the judgment of the teacher, according to the level of the learners and size of the class. Length of time allocated to a task is another possible area for differentiation.

Guided learning hours

Guided learning hours give an indication of the amount of contact time you need to have with your learners to deliver a course. Our syllabuses are designed around 130 hours for Cambridge O Level courses. The number of hours may vary depending on local practice and your learners' previous experience of the subject. The table below give some guidance about how many hours we recommend you spend on each topic area.

Topic	Suggested teaching time (%)	Suggested teaching order
Unit 1: Data and its collection	This unit should take about 9% of the course.	1.1, 1.2, 1.3, 1.4, 1.5
Unit 2: Summary representation of data [part 1]	This unit should take about 8% of the course.	2.1, 2.2 (part 1), 2.3, 2.4
Unit 3: Formation of data into ungrouped or grouped frequency distributions	This unit should take about 7% of the course.	3.1, 3.2
Unit 4: Formation of frequency distributions into cumulative frequency distributions	This unit should take about 4% of the course.	4.1, 4.2
Unit 5: Statistical measures, their interpretation and appropriate use	This unit should take about 16% of the course	5.1, 5.2, 5.3, 5.4
Unit 2: Summary representations of data [part 2]	This unit should take about 6% of the course	2.2 (part 2)

Topic	Suggested teaching time (%)	Suggested teaching order
Unit 6: Transformations involving mean and standard deviation	This unit should take about 5% of the course	6.1, 6.2
Unit 11: Elementary ideas of probability	This unit should take about 14% of the course	11
Unit 7: Crude and standardised rates and their appropriate use	This unit should take about 6% of the course	7
Unit 8: Index numbers	This unit should take about 6% of the course	8
Unit 9: Bivariate distributions and their representation by scatter diagrams	This unit should take about 7% of the course	9.1, 9.2
Unit 10: Time series	This unit should take about 6% of the course	10.1, 10.2
Unit 12: Probability distributions	This unit should take about 6% of the course	12.1, 12.2

Resources

The up-to-date resource list for this syllabus, including textbooks endorsed by Cambridge, is listed at **http://www.cambridgeinternational.org**
Endorsed textbooks have been written to be closely aligned to the syllabus they support, and have been through a detailed quality assurance process. As such, all textbooks endorsed by Cambridge for this syllabus are the ideal resource to be used alongside this scheme of work as they cover each learning objective.

Textbooks referred to in this scheme of work:

O Level Statistics
Author: Chalmers, Dean James
Published in 2016 (2nd Edition)
Published by Cambridge University Press
http://education.cambridge.org/

Statistics: A First Course
Author: Walker, James A, McLean, Margaret M and Matthew, James W
Published in 1993
Published by Hodder and Stoughton
www.hodderheadline.co.uk/

Success in Statistics
Author: Caswell, Fred
Published in 1995
Published by John Murray
www.johnmurray.co.uk/

School Support Hub

The School Support Hub **www.cambridgeinternational.org/support** is a secure online resource bank and community forum for Cambridge teachers, where you can download specimen and past question papers, mark schemes and other resources. We also offer online and face-to-face training; details of forthcoming training opportunities are posted online. This scheme of work is available as PDF and an editable version in Microsoft Word format; both are available on the School Support Hub **www.cambridgeinternational.org/support** If you are unable to use Microsoft Word you can download Open Office free of charge from **www.openoffice.org**

Websites

This scheme of work includes website links providing direct access to internet resources. Cambridge International Examinations is not responsible for the accuracy or content of information contained in these sites. The inclusion of a link to an external website should not be understood to be an endorsement of that website or the site's owners (or their products/services).

The website pages referenced in this scheme of work were selected when the scheme of work was produced. Other aspects of the sites were not checked and only the particular resources are recommended.

How to get the most out of this scheme of work – integrating syllabus content, skills and teaching strategies
We have written this scheme of work for the Cambridge O Level Statistics 4040 syllabus and it provides some ideas and suggestions of how to cover the content of the syllabus. We have designed the following features to help guide you through your course..

The **Topic Area** helps your learners by making clear the knowledge they are trying to build. Pass these on to your learners by expressing them as 'We are learning about…'.
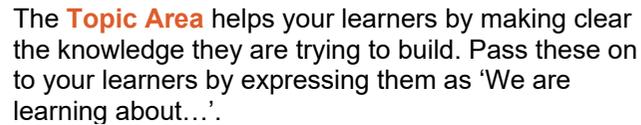
Suggested teaching activities give you lots of ideas about how you can present learners with new information without teacher talk or videos. Try more active methods which get your learners motivated and practising new skills.

Extension activities provide your more able learners with further challenge beyond the **basic** content of the course. Innovation and independent learning are the basis of these activities. **Challenging** activities are identified.

Independent study (I) gives your learners the opportunity to develop their own ideas and understanding with direct input from you.

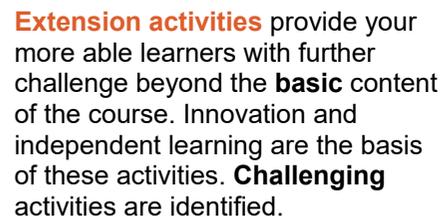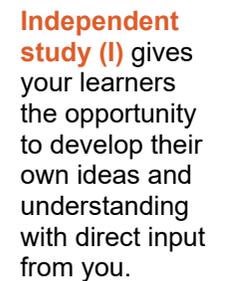| Syllabus ref	Topic Area	Suggested teaching activities
2.2	Stem-and-leaf diagrams and box-and-whisker diagrams.	**E** Learners should be given data in a variety of forms and asked to produce box-and-whisker diagrams. They should also be asked to compare distributions using a pair of box-and-whisker diagrams **(I)**. A simple exercise where lists of data are provided and learners are asked to find the median and quartiles before plotting box-and-whisker diagrams on grids that are provided can be found at (www.tes.com/teaching-resource/median-quartiles-and-box-plots-worksheets-6343880) **(basic, F)**. As an extension to this work, the more able learners could be shown how outliers can be calculated and illustrated on a box-and-whisker diagram **(challenging).** A very clear explanation of how to draw a box-and-whisker diagram can be found at (www.purplemath.com/modules/boxwhisk.htm), which goes on to describe how to calculate and display outliers.

Past and specimen papers

Past/specimen papers and mark schemes are available to download at **teachers.cie.org.uk**

Nov 2011	Paper 11 Q1
Jun 2012	Paper 12 Q1
Nov 2012	Paper 12 Q1 (a) and (b)

Past Papers, **Specimen Papers** and **Mark Schemes** are available for you to download at: **teachers.cie.org.uk**

Using these resources with your learners allows you to check their progress and give them confidence and understanding.

Formative assessment (F) is on-going assessment which informs you about the progress of your learners. Don't forget to leave time to review what your learners have learnt, you could try question and answer, tests, quizzes, 'mind maps', or 'concept maps'. These kinds of activities can be found in the scheme of work.
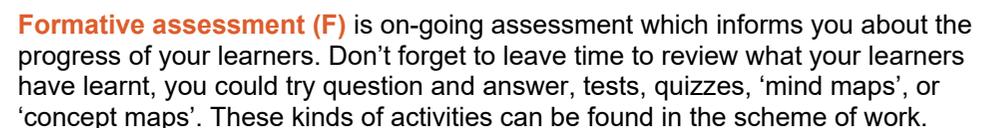
Unit 1: Data and its collection

Recommended prior knowledge
Candidates beginning this course are not expected to have studied Statistics previously.

Teaching time
Based on a total time allocation of 130 contact hours for this Cambridge O Level course, it is recommended that this unit should take 9% of the course.

Syllabus ref	Topic area	Suggested teaching activities
1.1	General ideas of sampling, including knowledge of the terms: population, census, sample, representative sample.	* A good initial activity might be for the learners to collect some continuous data of their own. This could be done in groups or as a class with the data kept and used again as new statistical techniques are introduced. In this way the learners will see the full statistical process from the collection to the recording, representation and analysis of data, providing a clear purpose to their studies.

A question involving comparison would provide a good starting point. When this initial data, collected by the learners, is referred to again in this scheme of work, the suggestion will be marked with an asterisk*.

One example of a question that could be posed is 'Are students' reaction times affected by background music?' Reaction times data can be collected quickly using an app on a mobile phone, such as the 'wait now – reaction time test' from the Google app store or a simple ruler catching test. A description of how to conduct a ruler catching test can be found at (www.topendsports.com/testing/tests/reaction-stick.htm) together with a table for converting distances on the ruler, in cm, to reaction times, in seconds. There is also a reaction timer on the NRICH website (http://nrich.maths.org/6044) together with further ideas for comparisons, such as a comparison of left and right hands rather than with and without background music.

In the suggested example, data can be collected from individuals initially without and then with background music playing and reaction times recorded, thus providing bivariate data. One idea would be to take three reaction times without and three reaction times with background music from each individual, and record the middle value in each case. (In a future lesson this could lead to a discussion about the relative merits of the mean and the median as a summary value).

Other examples of questions involving comparisons might be 'Do boys have larger hand spans than girls?', or a comparison from the sports field such as times to run a specific distance or the lengths of throws or jumps with a comparison of age groups or genders considered.

The comparison should be kept simple at this stage and ideally will involve the collection of continuous data, as this can then be used again as more advanced statistical techniques are introduced. |

Syllabus ref	Topic area	Suggested teaching activities
		Learners could be encouraged to think about the population for the chosen question and the terms census and sample can be introduced. How to collect a representative sample can be considered (see below). Advantages and disadvantages of taking a sample as opposed to a census should be discussed. There is a good discussion of the importance of sampling in Walker, McLean and Matthew, *Statistics a first course*, chapter 3. The importance of a sample being representative of the population should be discussed.

Syllabus ref	Topic area	Suggested teaching activities
1.2	Types of sampling, including knowledge of random, systematic, stratified and quota sampling methods, and the use of random numbers.	Having established the importance of sampling, a good explanation of some of the different types of sampling methods can be found on YouTube at (www.youtube.com/watch?v=be9e-Q-jC-0). Learners should know the definitions of each of the sampling methods and that they fall into two categories; namely random sampling, such as simple, systematic and stratified and non-random sampling such as quota. They should also be able to use random number tables to find the various random samples. A description of how to find the strata sizes for a stratified sample can be found on YouTube at (www.youtube.com/watch?v=rqhDEWUPF3M). Many textbooks also include descriptions of each of these sampling methods, for example Caswell, *Success in Statistics*, Unit 2, Walker, McLean and Matthew, *Statistics a first course*, chapter 3 and Chalmers, *O Level Statistics*, chapter 3. * Learners can consider how to collect a representative sample to answer the question posed at the start of the course. The various methods of sampling can be considered and how, for example, a random sample might be selected. A numbered list of the students at the school would be required. If genders are being compared, for example, it might be appropriate to consider stratifying each gender by year group. In reality a small convenience sample may prove to be the most practical to use, but this real-life context could provide opportunities for discussions about the alternatives. Indeed, if a convenience sample has been settled upon, then this can help with discussions about bias (see below). An exercise testing understanding of the various sampling methods, including the use of random number tables can be found in Chalmers, *O Level Statistics*, chapter 3 **(F)**.
1.3	Bias: how it arises and is avoided.	* As an introduction to the ideas of the possibilities of bias in a sampling method, there can be a class discussion about the reliability of the data collected at the start of the course, particularly if, say, a small convenience sample has been used. The avoidance of bias in a sampling method by having random samples in which every member of the population has an equal chance of being selected should be discussed as an ideal. There is a good description of the problems of bias in Caswell, *Success in Statistics*, Unit 2.

Syllabus ref	Topic area	Suggested teaching activities
1.4	General ideas of surveys, including the use of closed and open questions in questionnaires.	There is a good discussion about survey questions on the Maths is Fun website (www.mathsisfun.com/). Click on the 'Data' tab and scroll down to 'Survey Questions'. There are also two activities on this website: 'Asking Questions' and 'Improving Questions', which introduce the ideas of ensuring that questions on questionnaires are precise so that any results obtained can be considered reliable **(I)**. Also Walker, McLean and Matthew, *Statistics a first course*, chapter 1 provides clear guidelines for questionnaire design.
1.5	Types of data and variable, including knowledge of the terms: qualitative, quantitative, discrete, continuous.	There is a good explanation of the terms qualitative, quantitative, discrete and continuous on the Maths is Fun website (www.mathsisfun.com/). Click on the 'Data' tab and scroll down to 'What is Data?' There is also a set of self-marking questions which check understanding and provide feedback **(I, F)**. A useful example to include when explaining the difference between discrete and continuous data is UK shoe sizes, which can be 3.5, 4, 4.5, 5, 5.5, …, for example, but not any values in between these. This provides an example of discrete data which includes decimal values, illustrating the important feature of discrete data, namely that it can only take specific values, but removing the common misconception that discrete data can only take whole number values. There is an exercise in Chalmers, *O Level Statistics*, chapter 4 which tests understanding of the different types of data **(F)**.

Past paper and specimen papers
Nov 2013 paper 13 question 6 Nov 2013 paper 23 question 10 Nov 2014 paper 12 question 3 Nov 2014 paper 13 question 3 Nov 2014 paper 22 question 1, question 11 Nov 2014 paper 23 question 10 Nov 2015 paper 12 question 1 Nov 2015 paper 13 question 3, question 6(i) Nov 2015 paper 22 question 7 Nov 2015 paper 23 question 5 Specimen paper 1 question 2 Specimen paper 2 question 6

Unit 2: Summary representation of data [part 1]

Recommended prior knowledge
Candidates beginning this course are not expected to have studied Statistics previously.
It is advisable, although not essential, for learners to have studied unit 1, on data and its collection, prior to starting on unit 2.

Context
Unit 2 should be separated into two parts, beginning with simple pictorial and diagrammatic representations and leaving stem-and-leaf diagrams and box-and-whisker diagrams until after medians and quartiles have been studied.

Teaching time
Based on a total time allocation of 130 contact hours for this Cambridge O Level course, it is recommended that this unit should take 8% of the course.

Syllabus ref	Topic area	Suggested teaching activities
2.1	Classification and representation in tabular form, including two-way tables.	* As appropriate to the nature of the data collected by the learners at the start of the course, the data can be tabulated. If data has been collected by different groups of students it can be pooled and tabulated in an appropriate form by the class. Tally charts may be required, but learners must take care not to lose any information at this stage. For example, if bivariate data has been collected, then the pairs of data should not be separated in all the forms of tabulation. There may, for example, be a table of data for reaction times without background music and another table for reaction times with background music but there should also be a table where the pairs of data are retained. Data may also be grouped (see below) but again it will be useful to also retain the raw data. Data presented in two-way tables should be explained with learners given the opportunity to: interpret given tables; complete partially filled in two-way tables; and present given raw data by using a two-way table of their own. There are exercises involving the drawing and interpretation of two-way tables and frequency distributions in Chalmers, *O Level Statistics*, chapter 1 **(F)**.
2.2 [1]	Representation in pictorial or diagrammatic form, including pictograms, pie charts, comparative pie charts, Venn diagrams, bar charts, dual bar charts, sectional and percentage	Learners should be given the opportunity to draw for themselves various representations of data. The importance of labelling should be stressed so that any pictorial or diagrammatic representations have keys, are fully labelled, and have uniform scales that start at zero, as necessary. Diagrams missing these essential features could be used to illustrate the fact that the diagram becomes meaningless or misleading without them.

Syllabus ref	Topic area	Suggested teaching activities
	sectional bar charts.	If discrete data can be obtained that relates to the learners, perhaps regarding a local or school/college issue, then illustrating this in various pictorial or diagrammatic forms can prove particularly rewarding. Perhaps the number of learners achieving each examination grade in a previous year can be provided for the learners and they can then represent these data in a pictogram, a bar chart and a pie chart. Learners could work in small groups producing each of these pictorial and diagrammatic representations and they could be displayed in the classroom.

Data such as this could then be broken down by gender, for example, and displayed as dual, sectional and percentage sectional bar charts. If the numbers of males and females are different, then comparative pie charts could also be used.

Another suggestion for data that could be used for display in this way is that the learners could use the skills learnt in section 1.4 to conduct a survey; perhaps they want to investigate the modes of transport used by the population of learners at their school/college to get to and from school/college. This may also involve sampling, bringing in the skills learnt in section 1.2. Again working in groups, discrete data could be collected, pictorial and diagrammatic representations produced, and the results displayed in the classroom. The data could be broken down by gender or age group so that comparisons can be made.

An alternative example which provides discrete data might be a comparison of word or sentence length between two authors or between two different newspapers or article types. Again learners could work in groups to collect the data. A form of convenience sampling may be the most practical with each member of the group selecting a different page (perhaps at random) from the book and counting letters/words from, say, the first 20 words/sentences. Collected data can then be pooled, tabulated and displayed in various diagrammatic forms.

The importance of leaving gaps between the bars, if any kind of a bar chart is being used to illustrate discrete data, should be emphasised.

Bar and pie charts can also be created with your own data on the Maths is Fun website (www.mathsisfun.com/) under the 'Data' tab.

Scrolling down to 'Bar Graphs', you will find descriptions of how to draw and interpret simple bar charts with a set of simple self-marking questions **(basic, I, F)** as well as the opportunity to click on a link that allows you to create your own graphs.

Scrolling down to 'Pie Charts' you find a clear explanation of how to construct a pie chart, including an animated section on how to use a protractor **(basic)**. Learners can then 'have a go themselves' at measuring angles using an onscreen protractor that can be dragged and then rotated using the arrow keys. There are |

Syllabus ref	Topic area	Suggested teaching activities
		also self-marking questions at the end of this section **(I, F)**. Questions involving completing Venn diagrams can be found at (www.tes.com/teaching-resource/structured-venn-diagram-questions-11168691). Further questions on Venn diagrams and other pictorial representations of data are in Chalmers, *O Level Statistics*, chapter 1. These questions also involve interpretation of given pictorial representations of data (see section 2.4) and consideration of the advantages and disadvantages of the various forms of display (see section 2.3) **(F)**.
2.3	The purpose and use of various forms of representation, their advantages and disadvantages.	The advantages and disadvantages of the various forms of representation can be discussed with reference to the charts, produced by the learners, on display in the classroom. For example, the learners will see that: simple bar charts retain the original data whereas pie charts compare proportions; dual bar charts allow for easy comparison whereas sectional bar charts clearly display totals; percentage sectional bar charts allow proportions to be compared, but original data is lost. The fact that bar charts are useful for displaying discrete data, and that alternative methods of display for continuous data will be met later, can be mentioned at this point.
2.4	Interpretation of data presented in tabular, pictorial or diagrammatic form.	Real-life data presented in tabular, pictorial or diagrammatic form, perhaps from a local newspaper, can be used for interpretation. Learners can also interpret the data on display in the classroom. They can look at both the data that they/their group collected and also look at data collected by others and interpret the pictorial and diagrammatic forms accordingly. They may notice that more boys than girls got a grade B, say, or that more boys than girls walk to school or that there were more long words in the sports section than the international news section of the newspaper. The limitation of some forms of display, such as the fact that original values are lost in a percentage sectional bar chart, should come out of this discussion.

Past paper and specimen papers
Nov 2013 paper 13 question 2, question 3, question 4 Nov 2013 paper 23 question 5(i)–(iv) Nov 2014 paper 12 question 4(i), question 8(i)–(iv), (vi), (vii) Nov 2014 paper 13 question 5, question 6 Nov 2014 paper 22 question 5 Nov 2014 paper 23 question 5 Nov 2015 paper 12 question 3, question 9(i)(a)–(d), (ii), (iv) Nov 2015 paper 13 question 1(i)–(ii), question 2, question 4, question 6(ii) Nov 2015 paper 22 question 8(a) Nov 2015 paper 23 question 4 Specimen paper 1 questions 3(i)–(ii), question 6 Specimen paper 2 question 1

Unit 3: Formation of data into ungrouped or grouped frequency distributions

Recommended prior knowledge
Candidates beginning this course are not expected to have studied Statistics previously.
It is advisable, although not essential, for learners to have studied units 1, on data and its collection, and the first part of unit 2, on summary representation of data, prior to starting on unit 3.

Teaching time
Based on a total time allocation of 130 contact hours for this Cambridge O Level course, it is recommended that this unit should take 7% of the course.

Syllabus ref	Topic area	Suggested teaching activities
3.1	Class measures for grouped frequency distributions. Class limits, boundaries and mid-points, class intervals.	* If the data collected at the start of the course can be grouped, then this would form a useful introduction to this section of work. Class sizes can be discussed and the different ways of expressing those classes can be considered. Each of the terms 'class limit', 'class boundary', 'mid-point' and 'class interval' should be defined for the learners. It is particularly useful to consider the class boundaries in the context of drawing histograms (see below), so that the learners understand that the upper class boundary of one class is equal to the lower class boundary of the next class. It would be useful to use a variety of types of interval notation in any grouped frequency distributions, for example, a class representing measurements to the nearest cm might be expressed as '50–59' or '49.5 to under 59.5' or '$49.5 \leqslant x < 59.5$'. This will equip learners to be able to deal with any notation that they might meet. Examples, such as 'age in completed years', where, in the above example, the lower and upper class boundaries for the 50–59 class would be 50 and 60 respectively, should also be included in any worked examples and exercises given to the learners.
3.2	Representation in frequency polygons and histograms.	It may be easiest to teach the drawing of histograms before frequency polygons. In both cases, and as with all statistical drawings, the importance of labelling should be stressed. In particular for each of these representations a continuous scale along the horizontal axis should be used (and not interval notation). Correctly establishing the class boundaries and class intervals or widths should be the first step. Histograms with equal class widths (or 'frequency diagrams') can be dealt with first before introducing the idea that in a histogram area is proportional to frequency, and moving on to histograms with unequal class intervals.

Syllabus ref	Topic area	Suggested teaching activities
		Taking a frequency distribution with equal class intervals and producing the histogram and then regrouping the classes so that the class intervals become unequal can help to illustrate why frequency density is used. If the three representations, one with equal class widths and 'frequency', one with unequal class widths and 'frequency' and one with the same unequal class widths and 'frequency density' are put together then it can be seen that the use of unequal class widths and frequency does not produce a good representation of the data. Problems involving both taking a grouped frequency distribution with unequal class intervals and representing it as a histogram, and taking a given histogram and working out the frequencies, should be provided for the learners **(F)**. A clear worked example showing both how to construct a histogram with unequal class widths and how to calculate frequencies from that histogram can be found in the first part of a video on YouTube at (www.youtube.com/watch?v=wtECBdpSyDQ). There are also PowerPoint presentations and worksheets on both drawing and interpreting histograms at (www.tes.com/teaching-resource/histograms-lessons-6308853) **(F)**. The use of frequency polygons in comparing sets of data should be stressed. Learners should be given exercises involving both drawing pairs of frequency polygons on the same set of axes and interpreting given frequency polygons to make comparisons between data sets **(F)**. * If it has been appropriate to group the data collected at the start of the course, the reaction times, say, then this could be illustrated using either a histogram or a pair of frequency polygons, and appropriate comparisons made. Exercises involving the construction of frequency polygons and histograms can be found in Chalmers, *O Level Statistics*, chapter 4 **(F)**.

Past paper and specimen papers
Nov 2013 paper 13 question 5 Nov 2013 paper 23 question 1 Nov 2014 paper 12 question 6 Nov 2014 paper 13 question 9 Nov 2014 paper 23 question 6 Nov 2015 paper 12 question 8 Nov 2015 paper 13 question 10(ii)–(iii) Nov 2015 paper 22 question 8(b) Nov 2015 paper 23 question 1 Specimen paper 1 question 7(i)–(iv) Specimen paper 2 question 8(iii)

Unit 4: Formation of frequency distributions into cumulative frequency distributions

Recommended prior knowledge
Learners will need to study unit 3 before they embark upon unit 4 as an understanding of upper class boundaries is required.

Context
The work in this unit is required in unit 5 for the estimation of medians and quartiles from a cumulative frequency curve or polygon. Some teachers may prefer to teach the basic elements of unit 5, statistical measures, such as measures of central tendency and dispersion for sets of numbers and ungrouped frequency distributions first. They could then teach the work below, before moving on to finding medians, quartiles and interquartile ranges for grouped frequency distributions using cumulative frequency curves and polygons.

Teaching time
Based on a total time allocation of 130 contact hours for this Cambridge O Level course, it is recommended that this unit should take 4% of the course.

Syllabus ref	Topic area	Suggested teaching activities
4.1	Representation in tabular form of cumulative frequency distributions.	* Data collected by the learner, such as reaction times data, that was grouped in section 3.1 can now be represented as a cumulative frequency distribution. It should be explained to the learners that the cumulative frequency gives the total frequency up to the upper class boundary of each class. These tables will be used in section 4.2, below, to give the values to plot a cumulative frequency curve and they will also be used later, in section 5.3 for estimating medians, quartiles and interquartile ranges using linear interpolation.
4.2	Representation in graphical form. Cumulative frequency curves and polygons for continuous data.	* Data collected by the learner, such as reaction times data, that was tabulated as a cumulative frequency distribution in section 3.1 can now be illustrated as a cumulative frequency curve or polygon. Learners should be provided with tables of grouped continuous data and shown how to produce, first, the cumulative frequency distribution and then the cumulative frequency curve or polygon. Clear labelling and continuous scales on the horizontal axis should be stressed. The fact that the curve or polygon should always be increasing and the distinctive shape should be highlighted. Learners should also be provided with cumulative frequency curves and polygons for interpretation. At this stage they should be able to estimate frequencies below, above and between given values of the variable.

Syllabus ref	Topic area	Suggested teaching activities
		They could work in pairs asking each other questions, such as 'how many girls had a height less than 136 cm?' or 'what percentage of the people spent between 5 and 10 minutes on the telephone?' **(I)**. Later, this section can be revisited when medians, quartiles and the interquartile range have been introduced. There is a worksheet available on (www.tes.com/teaching-resource/cumulative-frequency-worksheets-6414929) **(F)**. An exercise in Chalmers, *O Level Statistics*, chapter 4, on cumulative frequency distributions and cumulative frequency curves does not require knowledge of medians and quartiles **(F)**.

Past paper and specimen papers
Past paper questions tend to also include the finding of medians, quartiles and interquartile ranges, so these have been included after the next unit of work. Nov 2014 paper 13 question 4 (i), (ii), (iii)(b)

Unit 5: Statistical measures, their interpretation and appropriate use

Recommended prior knowledge
Learners will need to have studied unit 3, formation of data into ungrouped or grouped frequency distributions, before embarking upon this unit.

Context
Learners may study parts of this unit before unit 4, formation of frequency distributions into cumulative frequency distributions, and then study unit 4 as it becomes a necessary tool for finding the median and quartiles of grouped frequency distributions. Some teachers may therefore choose to teach unit 4 alongside this unit. The calculation of the mean from this unit is a pre-requisite for units 6, 7, 8, 9, 10 and 12. The calculation of the median and quartiles from this unit is a pre-requisite for box-and-whisker diagrams from unit 2.

Teaching time
Based on a total time allocation of 130 contact hours for this Cambridge O Level course, it is recommended that this unit should take 16% of the course.

Syllabus ref	Topic area	Suggested teaching activities
5.1	Measures of central tendency: mean, median, mode and modal class, including calculation or estimation from a set of numbers, an ungrouped or grouped frequency distribution.	The measures of central tendency, namely the mean, median and mode, should be introduced and calculated initially, from small sets of data.
		From a small set of data, particular care must be taken in locating the position of the median, with the data put in order before the middle value is located. Learners should be shown that with an odd number of pieces of data a middle value exists, but with an even number of pieces of data the middle value lies between two numbers in the list. For ungrouped data, the formula $\frac{1}{2}(n + 1)$ is used to find the position of the median.
		* If reaction times data was collected at the start of the course and the median of three readings recorded, then then the merits of both using a measure of central tendency and the merits of selecting the median as the chosen measure could be discussed at this stage.
		An interesting way to practise these ideas, and to introduce calculation of the measures of central tendency from frequency distributions, is to use the activity idea at (https://www.stem.org.uk/rxvse). The idea is that one learner leaves the room whilst the rest are very briefly shown a picture, on a PowerPoint presentation provided, containing a large number of frogs. The picture is removed and the remaining students try to estimate the number of frogs they have seen. The learner returns to the room and is allowed to ask one person in the room what their estimate is. They are then offered the option of trying to guess at the number of frogs based on the one estimate or ask five more learners for their estimates (sample size). These estimates are written on the board and the mean, median and mode calculated. The true number is then revealed. The

Syllabus ref	Topic area	Suggested teaching activities
		activity is repeated with a second picture containing a different number of frogs and this time all the estimates from the class are used and a frequency distribution constructed. The mean, median and mode are again calculated. Again the true value is then revealed. Finally a third picture containing a much larger number of frogs is used, and this time a grouped frequency distribution is needed to collect the data. This time a modal class can be found and ways of estimating the mean considered. The advantages and disadvantages of each measure of central tendency should also be discussed. There is a presentation on (www.tes.com/teaching-resource/which-average-is-best-6278226), which clearly sets out the advantages and disadvantages of each average. It is important that learners understand that the median, for example, is unaffected by extreme values. The Maths is Fun website (www.mathsisfun.com) includes a section on finding the mean from a frequency distribution; click on the 'Data' tab and scroll down to 'Calculate the Mean from a Frequency Table'. Here the sigma sign is introduced and there is also a set of self-marking questions, including a final question which considers the effect on the mean of adding a constant to each observation (see section 6.1 below) **(I, F)**. Most textbooks contain exercises for finding the mean, the median and the mode from sets of numbers and from ungrouped frequency distributions, such as Chalmers, *O Level Statistics*, chapter 5, which also contains a clear table showing the features of the various measures of central tendency **(F)**. Two interesting tasks to stretch the more able learners involving the mean, median and mode can be found at (http://nrich.maths.org/6267) and (http://nrich.maths.org/11281) **(challenging)**. For grouped frequency distributions, the modal class can be found. Estimates for the mean using midpoints is covered within this section of the work and, again, many textbooks contain suitable exercises for practising this **(F)**. Please note that estimates for the median from grouped data, using either a cumulative frequency curve or using linear interpolation are covered in section 5.3.
5.2	Measures of dispersion: range, interquartile range, variance and standard deviation, including calculation or estimation from a set of numbers, an ungrouped or grouped frequency distribution.	Sets of data with the same mean, but very different spreads, can be used to show that a measure of dispersion, as well as a measure of centrality, can be useful as a summary statistic. The range can be defined at this point. As with section 5.1, small sets of data should be used at first. Quartiles will need to be explained at this point. When a small set of data is listed in order, the lower quartile is in the position of the median of the values below the median and the upper quartile is in the position of the median of the values above the median.

Syllabus ref	Topic area	Suggested teaching activities
		Note that these are also the positions that should be used for the quartiles when being found from a stem-and-leaf diagram (see section 2.2 below).

For ungrouped frequency distributions it may be easier to use the formulae $Q_1 = \frac{1}{4}(n + 1)$ th value and $Q_3 = \frac{3}{4}(n + 1)$th value , which produce the same values as in the above definition when n is odd and are very close to the above definition when n is even.

The interquartile range can then be calculated and, as with the median, the fact that this measure of dispersion is unaffected by extreme values should be explained. The limitations of the range as a measure of spread can also be considered at this point.

Many textbooks contain exercises on finding the range and interquartile range of ungrouped data **(F)**.

The need for a measure of dispersion that takes all the values into account can be explained and the variance and standard deviation introduced. The two formulae for the standard deviation should be introduced. The most able learners can be shown how one formula is obtained from the other, although this is beyond the scope of this syllabus **(challenging)**.

The formulae can be applied to small data sets first, before replacing n with Σf and applying the formulae to frequency distributions for ungrouped data.

The Maths is Fun website (www.mathsisfun.com) provides a good explanation for squaring the differences, but references to the formula for the sample standard deviation are not required at this level. Click on the 'Data' tab and scroll down to 'Standard Deviation'. There is a set of self-marking questions, which only require use of the population standard deviation formula **(I, F)**. The last few questions also provide an introduction to the ideas of the effect on the standard deviation of adding a constant to each observation and multiplying each observation by a constant (see section 6.1 below).

Many textbooks contain exercises on finding the variance and standard deviation for sets of numbers and for ungrouped frequency distributions **(F)**.

For grouped frequency distributions, midpoints can be used to find estimates for the variance and standard deviation.
Again many textbooks contain suitable exercises on estimating the variance and standard deviation for |

Syllabus ref	Topic area	Suggested teaching activities
		grouped frequency distributions**,** including Chalmers, *O Level Statistics*, chapter 7, which also contains a clear table showing the features of the various measures of dispersion **(F)**. Please note that estimating the interquartile range from grouped data, using either a cumulative frequency curve or using linear interpolation, is covered in section 5.3 below.

Syllabus ref	Topic area	Suggested teaching activities
5.3	Quartiles and percentiles, including estimation of median, quartiles and percentiles from a cumulative frequency curve or polygon, and by linear interpolation from a cumulative frequency table.	This section covers the calculation of the median and the interquartile range from a grouped frequency distribution, either by using a cumulative frequency curve or polygon or by using linear interpolation. In each case a cumulative frequency table will be required (see section 4.1). Unlike with ungrouped data, the median from a grouped frequency distribution is the $\frac{n}{2}$ th value and the lower and upper quartiles are the $\frac{n}{4}$ th and $\frac{3n}{4}$ th values, respectively. Cumulative frequency curves or polygons were covered in section 4.2. Learners can now be shown how these can be used to estimate the values of the median and quartiles. Also percentiles can be defined. Many textbooks contain exercises involving estimating medians and quartiles from cumulative frequency curves or polygons **(F)**. Chalmers, *O Level Statistics*, chapter 7 contains an exercise on the use of linear interpolation to estimate values such as the median and quartiles **(F)**.
5.4	Measures for combined sets of data, including calculation of mean and standard deviation.	The problem called Bats Wings on the NRICH website (http://nrich.maths.org/505) provides an introduction to this section of work. In this case a piece of data is missing and the overall mean is known. It introduces the idea of multiplying the mean by the number of values to obtain the sum of all the values. In Chalmers, *O Level Statistics*, chapters 6 and 7 there are exercises which included finding the mean and the standard deviation for combined sets of data.

Past paper and specimen papers
Nov 2013 paper 13 question 1, question 8(b)
Nov 2013 paper 23 question 3, question 7
Nov 2014 paper 12 question 1, question 2, question 4(ii), question 8(v), question 9
Nov 2014 paper 13 question 1, question 4(iii)(a)
Nov 2014 paper 22 question 2, question 4, question 10(a)
Nov 2014 paper 23 question 3, question 8
Nov 2015 paper 12 question 2, question 4, question 6, question 10(i)–(iv)
Nov 2015 paper 13 question 1(iii)–(iv), question 9
Nov 2015 paper 22 question 4(ii), question 6
Nov 2015 paper 23 question 6, question 8
Specimen paper 1 question 1, question 4, question 8(i) and (ii), question 10
Specimen paper 2 question 2, question 3(i), question 8

Unit 2: Summary representation of data [part 2]

Recommended prior knowledge
Learners will need to have met medians and quartiles, from unit 5, before they can draw box-and-whisker diagrams.

Context
Moving from stem-and-leaf diagrams onto box-and-whisker diagrams is a natural progression and thus stem-and-leaf diagrams have been left until this section of work. Alternatively stem-and-leaf diagrams could be introduced earlier, with the rest of unit 2.

Teaching time
Based on a total time allocation of 130 contact hours for this Cambridge O Level course, it is recommended that this unit should take 6% of the course.

Syllabus ref	Topic area	Suggested teaching activities
2.2	Stem-and-leaf diagrams and box-and-whisker diagrams.	Stem-and-leaf diagrams provide a useful method of display, because the original data is retained whilst the data is grouped into classes. Any quantitative data, whether discrete or continuous, can be displayed in this way. Simple stem-and-leaf diagrams should be introduced first, with an unordered stem-and-leaf diagram being produced before the final ordered version; the importance of the key should be stressed. Then back to back stem-and-leaf diagrams can be introduced to compare two sets of data. Again the importance of the key should be stressed. * If data such as reaction times was collected at the start of the course then this can now be displayed using a back to back stem-and-leaf diagram and comparisons can be made. A reaction time, found using the NRICH website (http://nrich.maths.org/6044), say, might provide data items such as 513 ms which could be recorded on a stem-and-leaf diagram as 51 in the stem and 3 as a leaf; 51	3. Learners can then be shown how the ordered list in the stem-and-leaf diagram can be used to find the median, quartiles and hence the interquartile range. As seen earlier, with ungrouped data, the lower quartile is in the position of the median of the values below the median. The upper quartile is in the position of the median of the values above the median. Exercises in producing stem-and-leaf diagrams from given raw data should be provided for the learners. They should also be asked to find medians, quartiles and interquartile ranges from data presented in this form. Clearly explained examples of stem-and-leaf diagrams can be found at (www.purplemath.com/modules/stemleaf.htm).

Syllabus ref	Topic area	Suggested teaching activities
		In Chalmers, *O Level Statistics*, chapter 4 there is an exercise on interpretation and construction of stem-and-leaf diagrams **(F)**. Using the medians and quartiles, together with the highest and lowest values, box-and-whisker diagrams (sometimes called box plots) can provide another very useful method of displaying data. The medians and quartiles might be obtained from a list of data, from a stem-and-leaf diagram or from a cumulative frequency curve or polygon, before the box-and-whisker diagram is produced. When two box-and-whisker diagrams are presented on the same diagram, with a single axis, they provide a very useful way of visually comparing both a measure of central tendency (the median) and a measure of dispersion (the interquartile range). * Again data previously collected by the learners can now be displayed as a pair of box-and-whisker diagrams. Comparisons can now clearly be made. Learners should be given data in a variety of forms and asked to produce box-and-whisker diagrams. They should also be asked to compare distributions using a pair of box-and-whisker diagrams. A simple exercise where lists of data are provided and learners are asked to find the median and quartiles before plotting box-and-whisker diagrams on grids that are provided can be found at (www.tes.com/teaching-resource/median-quartiles-and-box-plots-worksheets-6343880) **(basic, F)**. A task called 'Box Plot Match' on the NRICH website (http://nrich.maths.org/11002) provides six cumulative frequency curves and six box plots of the same data to match up. As an extension to this work, the more able learners could be shown how outliers can be calculated and illustrated on a box-and-whisker diagram **(challenging).** A very clear explanation of how to draw a box-and-whisker diagram can be found at (www.purplemath.com/modules/boxwhisk.htm), which goes on to describe how to calculate and display outliers.

Past paper and specimen papers
Specimen paper 2 question 7

Unit 6: Transformations involving mean and standard deviation

Recommended prior knowledge
Learners will need to have studied how to calculate the mean and the standard deviation from unit 5 before embarking on unit 6.

Teaching time
Based on a total time allocation of 130 contact hours for this Cambridge O Level course, it is recommended that this unit should take 5% of the course.

Syllabus ref	Topic area	Suggested teaching activities
6.1	Effect on mean and standard deviation of adding a constant to each observation and of multiplying each observation by a constant.	Two of the self-marking exercises on the Maths is Fun website (www.mathsisfun.com), under the headings, 'Calculate the Mean from a Frequency Table' and 'Standard Deviation', include questions that provide an introduction to the ideas of the effect on the mean and the standard deviation of adding a constant to each observation and of multiplying each observation by a constant. Through these, or other examples, learners can see that if a constant is added to each observation or if each observation is multiplied by a constant, then the mean will be affected in the same way as each observation. Indeed, as stated in Chalmers, *O Level Statistics*, chapter 8, this is the case for all the measures of central tendency. Learners can also be shown that the same is true of the standard deviation if each observation is multiplied by a constant, but that the standard deviation is not affected if a constant is added to each observation. Again in Chalmers, *O Level Statistics*, chapter 8, it is explained that this is true for all measures of dispersion; there is a clear illustration of this fact using the range as an example.
6.2	Linear transformation of data to a given mean and standard deviation.	The use of scaled data, for the purposes of comparison, should be explained. Learners should be able to transform an original reading from a data set with a known mean and standard deviation to a scaled value with a given mean and standard deviation. They should also be able to calculate an original reading given a scaled value. They may also be given both the original and scaled values and be asked for an unknown mean or standard deviation, or they may be asked to find the unknown value that is unchanged by the scaling process. Exercises can be found in Chalmers, *O Level Statistics*, chapter 8 covering both section 6.1 and section 6.2 of the syllabus **(F)**.

Past paper and specimen papers
Nov 2013 paper 13 question 11 Nov 2013 paper 23 question 4 Nov 2014 paper 13 question 8 Nov 2014 paper 22 question 10(b) and (c) Nov 2014 paper 23 question 4 Nov 2015 paper 13 question 10(iv)–(v) Nov 2015 paper 22 question 1, question 4 Nov 2015 paper 23 question 2 Specimen paper 2 question 3(ii)

Unit 11: Elementary ideas of probability

Recommended prior knowledge
Learners will need to be able to add and multiply fractions before beginning this unit.

Context
This unit should be studied before unit 12, probability distributions. It can otherwise be positioned at any point during the course as it does not depend on any of the other units. In this example scheme of work it has been placed part way through the course, with unit 12 at the end of the course. In this way the work on probability has been divided into two sections allowing the learners to return to the topic at a later stage on order to refresh their memories of it.

Teaching time
Based on a total time allocation of 130 contact hours for this Cambridge O Level course, it is recommended that this unit should take 14% of the course.

Syllabus ref	Topic area	Suggested teaching activities
11	Elementary ideas of probability including the treatment of mutually exclusive and independent events. Selections made with or without replacement.	Learners should be introduced to the basic ideas of probability and the language, such as 'outcome' and 'event' associated with it. They should be told that the probability of an event will take a value between 0 and 1. It can be useful to illustrate this on a scale with various events indicated along the scale, including impossible events at 0 and certain events at 1. Learners can be asked to place particular events onto the scale.
		If an event, E, consists of a number of equally likely outcomes then the probability of that event is given by, $$P(E) = \frac{\text{number of outcomes in } E}{\text{number of outcomes in the sample space}}.$$ An interesting lesson idea, called 'A Brief History of Probability', can be found at (http://nrich.maths.org/12153) which uses a historical context to teach the importance of considering equally likely outcomes when using the above formula.
		Learners should be introduced to the idea that probabilities can be expressed as fractions, decimals or percentages.
		On the Maths is Fun website (www.mathsisfun.com), click the 'Data' tab and scroll down to 'Probability'. Here you will find a very good introduction to the elementary ideas of probability and the language associated with it. There are also a set of simple self-marking questions **(basic, I, F)**.
		There is an exercise consisting of many questions on simple probability in Chalmers, *O Level Statistics*, chapter 2 **(F)**.

Syllabus ref	Topic area	Suggested teaching activities
		When two events *A* and B, say, cannot both occur at the same time they are mutually exclusive events. For mutually exclusive events P(*A* or *B*) = P(*A*) + P(*B*). Learners should also be introduced to the general addition law P(*C* or *D*) = P(*C*) + P(*D*) – P(*C* and *D*). These laws can be illustrated using Venn diagrams and learners should also be introduced to set theory notation, P(*C* ∪ *D*) = P(*C*) + P(*D*) – P(*C* ∩ *D*).
		On the Maths is Fun website (www.mathsisfun.com), click the 'Data' tab and scroll down to 'Mutually Exclusive Events. Here you will find a clear explanation, with examples, of what is meant by mutually exclusive events. You will also find a clear explanation of the addition law, with the set theory notation introduced.
		Learners should also understand that if all the possible mutually exclusive events associated with an experiment are considered, then the probabilities of all of them added together should total 1. They should also know that the probability of an event *A* not happening, denoted P(*A*') = 1 – P(*A*).
		An interesting lesson idea to introduce the idea of combining probabilities, called 'Chance Combinations', can be found at (http://nrich.maths.org/12156). Learners are invited to devise their own 3x3 bingo board for a game where two dice are to be thrown and the total called.
		If two events *A* and *B*, say, are independent, then the probability of one of the events occurring is not affected by the other event occurring. The probability of the combined event, P(*A* and *B*) = P(*A*) × P(*B*).
		Learners should be able to use the various formulae and be able to prove, for any particular pair of events with given probability facts, whether or not they are mutually exclusive and/or independent events.
		Learners should also be able to deal with problems where, if a bag contains coloured beads, say, and multiple selections are made, the beads are not replaced. In cases like this, subsequent probabilities depend on previous events. Learners may find tree diagrams helpful in problems of this sort.
		On the Maths is Fun website (www.mathsisfun.com), click the 'Data' tab and scroll down to 'Independent Events' for a clear description of what this means. Also if you scroll down to 'Complement', you will find examples where the simplest solution comes from considering the complement. There are also a set of self-marking questions. These also require an understanding of combined events both with and without replacement **(I, F)**.
		Further sources of examples for learners to try can be found in Walker, McLean and Matthew, *Statistics a first course*, chapter 9 or Caswell, *Success in Statistics*, unit 12 and Chalmers, *O Level Statistics*, chapter 9 **(F)**.

Past paper and specimen papers
Nov 2013 paper 13 question 7 Nov 2013 paper 23 question 8 Nov 2014 paper 12 question 5, question 11 Nov 2014 paper 13 question 7 Nov 2014 paper 22 question 3, question 6 Nov 2014 paper 23 question 9 Nov 2015 paper 12 question 5, question 10(v)–(vi) Nov 2015 paper 13 question 7 Nov 2015 paper 22 question 2, question 9 Nov 2015 paper 23 question 3, question 7 Specimen paper 1 question 3(iii)–(iv), question 8(iii) Specimen paper 2 question 5, question 11

Unit 7: Crude and standardised rates and their appropriate use

Recommended prior knowledge
Learners should have studied unit 5 prior to starting this unit.

Context
The calculation of a standardised rate involves finding a weighted average and therefore this unit should come at some point after unit 5. It is not a pre-requisite for any of the remaining units.

Teaching time
Based on a total time allocation of 130 contact hours for this Cambridge O Level course, it is recommended that this unit should take 6% of the course.

Syllabus ref	Topic area	Suggested teaching activities
7	Crude and standardised rates and their appropriate use, including application to death rates, fertility rates, accident rates.	Learners should be introduced first to the crude rate, calculated by finding the number of events in a given time period divided by the total population size (often, but not always, expressed as per 1000 of the population, depending on the relative sizes of the numbers involved). The fact that the calculation of a crude rate requires only the total number of events and the population size should be emphasised. The 'events' being considered may be accidents, births, deaths, illnesses, etc. and learners should meet a variety of different contexts. Crude rates for two populations, say two towns or two hospitals, should then be compared. It should be explained that these crude rates take none of the other factors that might contribute to the rates into account. A particular town, for example, may have a large proportion of elderly people, and thus the death rate might be higher than for a town with a smaller proportion of elderly people. Thus comparing the crude death rates for these two towns, in order to decide which has the healthier environment, may be considered to be unfair. If the ages of the residents in the two towns were also taken into account, then a fairer comparison could be made. Learners can then be introduced to the standardised rate, where a potentially contributory factor to the rate in question is also considered. In the example above, the age profile of each town should be considered, and initially the death rate for each individual age group should be calculated. A single standardised rate for each town can then be obtained by using a common standard population distribution. This standard population is used as weights for the individual death rates for each age category, and a weighted average found. The standardised rates obtained can then be compared, without the differing age profiles of the populations of the two towns distorting the figures. Learners should be presented with examples where crude rates might suggest one interpretation and the

Syllabus ref	Topic area	Suggested teaching activities
		standardised rates suggest another. In particular it might be helpful to consider examples where the crude rates, say, for two populations are the same, but the standardised rates different. This might help learners to understand how standardising has created a more useful rate for comparison. A useful explanation of crude and standardised rates can be found on YouTube at (www.youtube.com/watch?v=3mN987K-7u4), where death rates are compared between California and New York. In this example, the rates used are per 100 000. A good source of suitable questions to test understanding is Chalmers, *O Level Statistics*, chapter 6 **(F)**.

Past paper and specimen papers
Nov 2013 paper 13 question 9 Nov 2014 paper 12 question 7 Nov 2014 paper 13 question 10 Nov 2015 paper 12 question 11 Nov 2015 paper 13 question 11 Specimen paper 1 question 9

Unit 8: Index numbers

Recommended prior knowledge
Learners should have studied unit 5 prior to starting this unit.

Context
The calculation of a weighted aggregate cost index involves finding a weighted average and therefore this unit should come at some point after unit 5. It is not a pre-requisite for any of the remaining units.

Teaching time
Based on a total time allocation of 130 contact hours for this Cambridge O Level course, it is recommended that this unit should take 6% of the course.

Syllabus ref	Topic area	Suggested teaching activities
8	Index numbers, including price relatives and weighted aggregate index numbers. Use and limitations of weighted aggregate index numbers.	Index numbers should be defined for the learners as a measure which shows the relative change in a quantity expressed as a percentage of the original quantity, but without the percentage sign (%). Price relatives can then be introduced as index numbers which show the change in the price of an item over an interval of time. The term 'base year', being the year from which the percentages are calculated and the year when the price relative is 100, should also be introduced. Learners should be able to calculate a price relative for an item given its price at the time the price relative is to be calculated and its price in the base year. They should also be able to interpret a given price relative. So for example, given a price relative of 124, they should be able to say that the price of the item has increased by 24% in the time period between the base year and the year in which the price relative was taken; or for a price relative of 94, that the price of the item has decreased by 6% in the given time period. Finally, given a price relative and either the price of the item in the base year or the price in the year of the given price relative, they should be able to calculate the missing price. Exercises on price relatives can be found in Chalmers, *O Level Statistics*, chapter 6 **(F)**. To introduce the idea of weighted aggregate cost index, an example of a company heavily dependent upon oil could be cited. Such a company will have a number of different costs each with their own price relative. An aggregate (combination) of these price relatives can give an overall picture of the company's costs. Oil prices fluctuate a lot and if a large part of a company's expenditure is on oil then the fluctuating prices will have a

Syllabus ref	Topic area	Suggested teaching activities
		great effect on that company. Another company which spends a smaller proportion of its expenditure on oil will be less affected by the fluctuating oil prices. Thus a weighted rather than an unweighted aggregate cost index gives a more useful value. The weights can be based on expenditure in the base year. Learners need to know that the formula to calculate a weighted aggregate cost index is $$\frac{\Sigma(\text{weight} \times \text{price relative})}{\Sigma \text{weights}}$$ Learners can then be shown how to use the weighted aggregate cost index, together with the costs in the base year to find an estimate for the costs in the year under consideration. It is important that learners appreciate that the figure obtained is an estimate and that they understand why it is an estimate. Learners need to be reminded that the weights used were based on expenditure in the base year. Using the company heavily dependent upon oil as an example, the class could discuss whether the companies spending patterns might have changed, if, say, the price of oil had risen sharply. The question that could be asked is: What steps might the company take to try and reduce its costs? The fact that the price of oil is given and cannot be changed, but that the quantity of oil used could be altered by the company, should come out of such a discussion. It is important to stress to learners that reasons why an estimated figure may be incorrect should always be given in the context of the question posed. Exercises on weighted aggregate index numbers can be found in Chalmers, *O Level Statistics*, chapter 6 **(F)**. Some of these questions also include interpretation of the figures obtained and some ask for reasons why the estimates might be incorrect.

Past paper and specimen papers

Nov 2013 paper 23 question 2, question 6
Nov 2014 paper 22 question 8
Nov 2014 paper 23 question 7
Nov 2015 paper 22 question 10
Nov 2015 paper 23 question 9
Specimen paper 2 question 10

Unit 9: Bivariate distributions and their representation by scatter diagrams

Recommended prior knowledge
Learners should have studied unit 5 prior to starting this unit.

Context
The calculation of semi-averages involves finding means and therefore this unit should come at some point after unit 5. It is not a pre-requisite for any of the remaining units.

Teaching time
Based on a total time allocation of 130 contact hours for this Cambridge O Level course, it is recommended that this unit should take 7% of the course.

Syllabus ref	Topic area	Suggested teaching activities
9.1	Elementary ideas of correlation, including understanding of the terms: positive, negative, strong and weak correlation.	* If bivariate data, such as reaction times both with and without background music, were collected at the start of the course, then this can be displayed in a scatter diagram. The question that could be asked is 'do people with the quickest reaction times when there is no distraction also have the quickest reaction times when background music is playing?' On (www.bbc.co.uk/schools/gcsebitesize/maths/statistics/scatterdiagramsrev2.shtml) there is a clear explanation, with contextual examples, of the language associated with correlation. Working in pairs, learners could draw sketches of possible scatter diagrams and label one of the axes with a possible variable. They could then swap sketches, name the type of correlation, and suggest a possible variable for the other axis. A discussion between the pair about the suggestion could help the learners gain a deeper understanding of correlation and causation **(I)**. If bivariate data was not collected at the start of the course, then some suitable data could be collected at this point. A simple example might be to collect data from the class on heights and foot length, say. The learners could be asked if taller people have bigger feet and asked to think about how they could test their theory. Clear labelling of the axes, including units, should be stressed.
9.2	Lines of best fit, including the method of semi-averages, and the derivation of the equation of the fitted straight line in the form $y = mx + c$. Use and limitations of a line of best fit in prediction.	If bivariate data on heights and foot lengths has been collected, the learners could be told to imagine that a footprint of a certain size has been found at the scene of a crime. Can they use their scatter diagram to estimate the height of the suspect? How confident would they be able to be about this estimate? On the Maths is Fun website (www.mathsisfun.com) under the 'Data' tab scroll down to 'Scatter Plots' where you will find an explanation of scatter diagrams, lines of best fit and their equations (but not the method of

Syllabus ref	Topic area	Suggested teaching activities
		semi-averages). The use of lines of best fit in prediction is also considered, together with ideas of correlation. There is also a set of self-marking questions **(I, F)**. Chalmers, *O Level Statistics*, chapter 10 is a good source of questions on scatter diagrams and includes questions requiring interpretation in context **(F)**. It also includes a clear explanation of the method of semi-averages.

Past paper and specimen papers
Nov 2013 paper 13 question 10 Nov 2014 paper 12 question 10 Nov 2014 paper 13 question 11 Nov 2015 paper 12 question 7 Nov 2015 paper 13 question 8 Specimen paper 1 question 5, question 11

Unit 10: Time series

Recommended prior knowledge
Learners should have studied unit 5 prior to starting this unit.

Context
The calculation of moving averages and mean seasonal variation requires an understanding of the mean from unit 5. This unit is not a pre-requisite for any other units.

Teaching time
Based on a total time allocation of 130 contact hours for this Cambridge O Level course, it is recommended that this unit should take 6% of the course.

Syllabus ref	Topic area	Suggested teaching activities
10.1	Understanding of trend, including determination by calculation of moving averages, with centring, where appropriate.	Learners can be given the opportunity to display some data using a time series graph by plotting some values of a variable at regular time intervals. Plotted points should be joined with straight line segments, with time plotted on the horizontal axis. Repeating patterns in the variation can be discussed with possible explanations offered.

The idea of the general trend of the data should be introduced. Learners can then be introduced to the idea of smoothing out the variations in the data by calculating moving averages. They should consider carefully the number of consecutive values, n, that it would be appropriate to average in a given situation in order to best achieve the effect of smoothing out the variation.

When n is even, the idea of centring the moving average values so that calculated values occur at the same points in time as original data items should also be introduced.

These moving average values can then be plotted onto the time series graph and the smoothing out of the variation observed. A line of best fit (a trend line) can then be drawn through the moving average values, and the general trend more easily observed. Learners should be encouraged to describe the general trend in the context of the data presented.

Worked examples can be found at (www.tes.com/teaching-resource/time-series-and-moving-averages-teaching-resources-6219792), although these do not include centring when n is even.

Chalmers, *O Level Statistics*, chapter 11 includes examples and questions where the moving averages require centring **(F)**. |

Syllabus ref	Topic area	Suggested teaching activities
10.2	Understanding of seasonal variation, including calculation of mean seasonal variation. Use of a trend line and seasonal component in prediction.	Learners should be encouraged to consider the difference between the moving average values or trend line values and the original data values; these differences being positive when the original data lies above the trend line and negative when they lie below the trend line. For each 'season' the mean seasonal variation, called the seasonal component, can be calculated and then used in making predictions. If all the seasonal components are added together then they should total zero. The use of the trend line and the seasonal components to make predictions should be explained. The limitations of this method of prediction should also be considered. Chalmers, *O Level Statistics*, chapter 11 has an exercise which involves the calculation of the mean seasonal variation and its use in making predictions **(F)**.
Past paper and specimen papers		
Nov 2013 paper 23 question 11 Nov 2014 paper 22 question 7 Nov 2014 paper 23 question 11 Nov 2015 paper 22 question 5 Nov 2015 paper 23 question 10 Specimen paper 2 question 9		

Unit 12: Probability distributions

Recommended prior knowledge
Learners should have studied unit 11 on elementary ideas of probability, and means from unit 5 before embarking upon this unit.

Context
Learners often find the work contained within this unit challenging and it might therefore be appropriate to study it towards the end of the course. It is not a pre-requisite for any other units. It could come immediately after unit 11, although teachers may prefer to leave some time between units 11 and 12 so that the learners get the opportunity to return to the topic of probability and reinforce ideas met at an earlier stage of the course.

Teaching time
Based on a total time allocation of 130 contact hours for this Cambridge O Level course, it is recommended that this unit should take 6% of the course.

Syllabus ref	Topic area	Suggested teaching activities
12.1	Formation of the probability distribution of a discrete variable.	Learners need to know that a probability distribution is a table containing all the possible outcomes of an experiment together with their probabilities. On the Maths is Fun website (www.mathsisfun.com), click the 'Data' tab and scroll down to 'random variables' where you will find some examples of discrete random variables and their probability distributions. There is also a set of self-marking questions **(I, F)**.
12.2	Expectation, including expected profit and loss in simple games. Idea of a fair game.	For a probability distribution of a discrete random variable, the expected value or expectation is the value of the variable that on average you would expect to obtain. It may not correspond to an actual value that could occur.

It is calculated by summing the product of each probability with the corresponding value of *x*; so for a random variable, *X*, the expected value, $E(X) = \Sigma P(x = X) \times x$.

On the Khan Academy website (www.khanacademy.org/math/probability/random-variables-topic/expected-value) click on 'expected value - practice problems' for some **basic** questions or 'expected value with calculated probabilities – practice problems' for some **challenging** questions **(I, F)**.

When the concept of the expectation or expected value is put into the context of a simple game, and the idea of profit or loss is considered, it becomes more meaningful. Also the idea of a fair game can come out of such examples.

Chalmers, *O Level Statistics*, chapter 9 is a good source of questions on fair games and expected winnings **(F)**. |

Past paper and specimen papers
Nov 2013 paper 23 question 9 Nov 2014 paper 22 question 9 Nov 2014 paper 23 question 2 Nov 2015 paper 13 question 5 Nov 2015 paper 22 question 11 Nov 2015 paper 23 question 11 Specimen paper 2 question 4