AQA Qualifications

GCSE

STATISTICS

(4310)

Assessment Guidance

GCSE

Specification

STATISTICS (4310)

Assessment Guidance

Our specification is published on our website (www.aqa.org.uk). We will let centres know in writing about any changes to the specification. We will also publish changes on our website. The definitive version of our specification will always be the one on our website, this may differ from printed versions.

You can get further copies of this Teacher Resource from:
The GCSE Mathematics Department
AQA
Devas Street
Manchester
M16 6EX

Or, you can download a copy from our All About Maths website (http://allaboutmaths.aqa.org.uk/).

Copyright © 2012 AQA and its licensors. All rights reserved.

AQA retains the copyright on all its publications, including the specifications. However, registered centres for AQA are permitted to copy material from this specification booklet for their own internal use.

AQA Education, is a company limited by guarantee registered in England and Wales (company number 3644723), and a registered charity 1073334.
Registered address:
AQA
Devas Street,
Manchester
M15 6EX

Contents

Planning a Strategy	**4**
Hypothesis	4
Planning an Investigation	4
Deciding between a survey / experiment	6
Data Collection	**7**
Types of Data	7
Obtaining Data	8
Census Data	8
Sampling	9
Conducting a survey / experiment	12
Tabulation and Representation	**20**
Tabulation	20
Diagrammatic Representation	21
Data Analysis	**28**
Measures of Location	28
Measures of Spread	30
Other Summary Statistics	33
Time Series	38
Quality Assurance	40
Correlation and Regression	40
Estimation	44
Probability	**45**
Probability	45
Data Interpretation	**49**
Data Interpretation	49

3.1.1 Planning a Strategy

Statement from Specification	Assessment Guidance (including examples)
3.1.1a Hypothesis	
Foundation and Higher tiers	
Specifying a hypothesis to be tested	A scenario will be described and candidates will be expected to write an appropriate hypothesis that could be tested given this scenario. **For example** Sally wonders whether talking to her plants will help them grow. She splits her plants into 2 groups. One group she ignores, the other she talks to every day. Specify a suitable hypothesis for Sally's experiment. *Plants that are spoken to grow taller than those not spoken to.* **Teaching Tip** Hypothesis needs to be testable. 'plants will grow bigger' is loosely defined so the above answer is better.
3.1.1b Planning an Investigation	
Foundation and Higher tiers	
Determine the data needed to address hypotheses and selecting an appropriate method for obtaining the data	A scenario or hypothesis will be given and candidates will need to choose both the data needed and the method they would use to obtain it. **For example** In the plants example above what data should be collected? *Sally will need to collect data for the heights of the plants in the two groups.* **Teaching Tip** An even better answer might be to collect data about the increase in heights of the plants in the two groups as they may not all start at the same height.

Statement from Specification	Assessment Guidance (including examples)
Higher tier only	
Justify the choice of method by comparing with possible alternatives	In most cases the alternatives will be made available and candidates will be expected to give a critical analysis of the different methods of obtaining the data (in much the same way as comparisons between sampling methods are often made).
Foundation and Higher tiers	
Specifying a research question to be investigated and breaking it down into sub-questions as necessary.	Similar questions to those where a hypothesis is to be specified except that the statements are to be questions. **For example** Specify a suitable research question that Sally could investigate. *Do the plants I talk to grow better than those I do not talk to?* This could then be broken down into sub-questions such as *Do the plants get taller?* *Do the plants give more flowers? and so on* **Teaching Tip** Candidates need to be clear about the differences between a hypothesis and a research question.

Statement from Specification	Assessment Guidance (including examples)
3.1.1c Deciding between a survey / experiment	
Foundation and Higher tiers	
Awareness of possible problems including 1. Identifying the population 2. Questionnaire distribution and collection 3. Non-response 4. Errors in recording answers 5. Missing data	Candidates should be able to identify situations where some of these issues may arise and be able to relay strategies for dealing with these problems in practical situations. 1. Surveying a sample of a population of a town – try to not use a method which excludes some of the population and is therefore biased eg, collecting data from one area of town when doing a survey of the whole town. 2. For example, weighing up cost and time issues for interviews against using postal questionnaires. 3. To increase response rate it is a common misconception that more questionnaires (oe) need simply be given out – this does not affect the **rate** it will be the distribution method that will need looking at. 4. What if a piece of data is recorded as ‘540’ when every other value is under 100? (State that this piece is to be ignored). 5. Dealing with omissions in data sets eg, in a sample stratified by gender what do you do if the person has forgotten to state their gender?

3.1.2 Data Collection

Statement from Specification	Assessment Guidance (including examples)
3.1.2a Types of Data	
Foundation and Higher tiers	
Raw data Primary and secondary data sources Qualitative and quantitative variables Categorical data Discrete and continuous data Grouped and ungrouped data Bivariate data	These definitions for different types of data can be tested in general terms **For example** Is the height of people a discrete or continuous variable? Give a reason for your answer. **Teaching Tip** Possibly more importantly if students have a genuine appreciation of the nature of the data they are faced with they are more likely to have a fuller understanding of the appropriate diagrams and techniques which may or may not be used on that data. **For example** Understanding that a pie chart should be used for categorical data. It is unlikely that the term bivariate data will be used at Foundation Tier but students should recognise that when data is bivariate then a scatter diagram and measures of correlation are appropriate.
Classification of data, class limits and intervals	Students at both tiers should be able to work with class intervals (usually equal widths at Foundation) and have an understanding (basic at Foundation) of the lower and upper class bounds for different types of class intervals
Higher tier only	
Implications of grouping for loss of accuracy in presentation and calculation	Higher students should be aware of the necessary balance between having lots of class intervals which requires excessive computation and time consuming completion of diagrams with having too few class intervals whereby detail of the shape of a distribution and any accuracy for the estimation of the values for measures is severely hampered.

Statement from Specification	Assessment Guidance (including examples)
3.1.2b Obtaining Data	
Foundation and Higher tiers	
Obtaining data by counting or measuring; accuracy of such measures	Students could be asked to consider how a practical set of data has been measured and whether the accuracy given is appropriate. Obviously incorrect values within data may be expected to be identified.
Design and use efficient methods of recording data; appropriate to the purpose it will be used	Specification mentions 'to include an awareness of data logging and data collection sheets'. Students may be asked to design data collection sheets (may be termed observation sheets) or to criticise already designed ones. Students may be asked to identify a situation where data logging could be used, or give advantages or disadvantages of data logging compared to other possible data collection methods. See also 3.1.2e
3.1.2c Census Data	
Foundation and Higher tiers	
A census obtains information about every element of the population	Students should be aware of the national census held every 10 years and have a basic knowledge of the reasons for it as well as some of the variables measured. Students should also know that the word 'census' is used more generally for the situation where every member of a population is sampled.
Obtaining information from well-defined populations	The importance of identifying accurately the population from which sampling is taken. A failure to do so seriously affects the accuracy and validity of the data collected. **For example** Taking data from street interviews of all people when the data is supposed to be from single people only.

Statement from Specification	Assessment Guidance (including examples)
3.1.2d Sampling	
Foundation and Higher tiers	
Purpose of sampling; variability between samples	An awareness of the reasons for sampling - usually based on either cost or time issues or that a product is used up by being tested so it is not feasible for all items from a population to be in a sample.
Randomness. Random numbers from tables. calculators and computers	Knowledge that random digits can be obtained from a variety of sources. In questions set in examinations students will never be asked to obtain their own random digits from their calculator but instead will be given a table of random digits to work with.
Sampling from a well-defined population	A catch all specification reference to intimate that students should be aware that to sample effectively it is important to know the extent of a population otherwise it is not possible to sample from it eg, cannot list or number whole population.
Sample frame	The list of available or known members of a population. eg, the sample frame for the UK population will be all those who could appear in a sample and could exclude, for example, people on holiday abroad or in prison etc.
Simple random sampling: the condition that all members of a population are equally likely to be included in the sample.	The second condition that all possible random samples are equally likely is not expected to be known for this GCSE specification. Candidates may be required to demonstrate the ability to obtain a random sample by using a given table of random digits. Issues such as repeated members and numbers above the likely numbering system may occur. It is not necessary to be able to number items a multiple number of times to use the maximum number of possible values.

Statement from Specification	Assessment Guidance (including examples)
Foundation and Higher tiers	
Use of stratification in sample design using a single category. Awareness of the dangers of convenience sampling.	To confirm – only named methods at Foundation are random, stratified and convenience sampling. How to stratify (calculation of correct numbers) and why this is often a useful idea can be tested at Foundation as well as Higher. At Foundation only one category will be used eg, gender or school year. Students should be aware that convenience or judgement sampling should be avoided wherever possible as it is very difficult to ensure that an unbiased sample can be achieved using that method. However, using named methods is not, in itself, a guarantee that a sample is free from bias.
Biased samples arising from sampling from the wrong population or non-random choice of individual elements	This bias can arise from situations such as opportunity sampling where people are chosen by interviewers without consideration of their profile. Also situations such as on-line surveys which exclude part of the general population.
Higher tier only	
Stratified sampling with no more than two sets of categories.	**For example** Gender and school year. The issue of rounding errors may occur in calculations where the sample size is not as required. This will require one of the sample values to be rounded or truncated to not the nearest integer – this should be chosen to cause the minimum rounding error.
Cluster sampling and quota sampling with particular reference to its use in conducting large scale opinion polls.	Definitions and advantages (or disadvantages) of these methods should be known. Use of quota sampling or random sampling in opinion polls – some of the criteria for filling the quota should be known, eg age, gender and income.

Statement from Specification	Assessment Guidance (including examples)
Higher tier only	
An awareness of multi-stage sampling	Knowledge that multi-stage sampling is a form of cluster sampling where it is hoped that the cluster chosen represents fairly the overall population. The typical example is in region to town to district and its advantages are around time and cost, however it should be appreciated that this method can be prone to error and the error increases with each additional stage of sampling.
Strengths and weaknesses of the various sampling methods including the dangers of convenience sampling.	Including systematic sampling – how to do it and the dangers in certain data sets of patterns affecting a systematic sample.
The criteria used for selecting sample members in national opinion polls: geographical area, sex, age, social and economic background. Associated sources of bias.	Some overlap here with the specification two boxes higher. Also awareness of the sample size being an important issue, not only that small samples will usually provide an unreliable result but also that an already large sample will not be made much more reliable by increasing it to a very large sample. It is easy to forget other issues such as design of the data collection method for example, can have just as much of an effect on reliability.

Statement from Specification	Assessment Guidance (including examples)
3.1.2e Conducting a survey / experiment	
Foundation and Higher tiers	
Obtaining primary data by questionnaire	See specific details below
Use and reasons for pilot studies and pre-testing	Pilot studies are a small scale practice of the full questionnaire. Pre-testing is a more generic term given to any small scale practice of a data collection method (not necessarily a questionnaire). Advantages usually centre around the idea of testing wording of questions, ease of understanding for participants and suitability of response options where appropriate. Less well known is that these can also be used to test the sample method to be used to verify it achieves representativeness.
Problems of design, wording, biased questions, definitions, obtaining truthful answers.	Students are often given a trial survey question and are asked to identify why it is a biased question. **For example** *Do you agree that it is a good idea to recycle as much as possible?* Students should be talking about either the 'do you agree' aspect or the use of the words 'good idea'. Students can also be asked to rewrite questions to free them from bias. **For example** *Should people recycle – as much as possible / when convenient / not at all / don't know* It is interesting that students, often having identified correctly a bias within a question, write a new question containing very similar bias. **Teaching Tip** It is an inherent part of asking for a question to be written that a response section is offered. (see open / closed questions below) Students even use the response section space to answer the question sometimes!

Statement from Specification	Assessment Guidance (including examples)
Foundation and Higher tiers	
The advantages and disadvantages of closed and open questions	Closed – specific response boxes offered making for more control over the responses with calculations and interpretations of results much easier. Open – answer space only – so responses are unpredictable and often difficult to collate. When choosing response boxes for numerical data students are expected to be able to achieve exhaustive and non-overlapping class intervals. Any appropriate labelling system is acceptable though double inequalities tend to be used most effectively by successful students. **Example** All Becki's CDs last between 40 and 70 minutes. Design a frequency table with 6 class intervals so that Becki can collect data about the lengths of CDs in her collection.
Higher tier only	
The use of opinion scales	Students are expected to know about 2 different types of opinion scales 1. A point (discrete) scale such as strongly agree / agree / neither agree nor disagree / agree / strongly agree. 2. A continuous scale responses being placed anywhere on the scale by the respondent.
The technique of random response, in its simplest form, for obtaining truthful answers to sensitive questions.	Used to find out, for example, smoking, drinking or under-age sex rates amongst students. The basic premise is to ask participants to toss a coin (or similar) and tick "yes" if they get heads and answer the question truthfully if they get tails. This means that on average any surplus "yes" responses above half of the panel are answering yes to the question. Clearly there are error possibilities especially if the sample size is not that large. Potential to carry out a simple experiment with your class with careful managing of questions and responses.

Statement from Specification	Assessment Guidance (including examples)
Foundation and Higher tiers	
Obtaining data by interview. Advantages and disadvantages of interviews compared with written questionnaires	Interview advantages include – opportunity to explain questions, opportunity to collect more than text responses, confidence of exactly who has completed the survey, more confident of truthful responses. Interview disadvantages – include time and cost of training interviewers, difficulty of reproducing exactly the same experience for respondents, people not wishing to discuss things face to face with someone

Statement from Specification	Assessment Guidance (including examples)
Foundation and Higher tiers	
Simulation. Use of, for example, dice, random numbers tables, ICT.	Only simple ideas will be tested with students often given the opportunity to carry out a small scale simulation using some dice or coin results given in the question. The example below is as complex as it can get. **Example** There are 4 polling stations in a town. At the last election Stations A, B and C were used by equal numbers of voters whilst Station D was twice as popular as any of the other 3. Run a simulation for the 1st 20 voters. Complete both tables. Use these dice rolls in your answer. 3 4 6 2 2 4 3 5 1 6 1 2 4 2 3 3 5 2 1 6 3 4 16 **Solution** The student must indicate clearly which dice value will equate to which polling station (this is the purpose of the 1st table) Then outcomes need to be given.

Dice Score	Polling Station
1	A
2	B
3	C
4	D
5	D
6	None

Simulation

3	4	6	2	2	4	3	5	1	6	1	2
C	D		B	B	D	C	D	A		A	B
4	2	3	3	5	2	1	6	3	4	1	6
D	B	C	C	D	B	A		C	D	A	

Statement from Specification	Assessment Guidance (including examples)
Foundation and Higher tiers	
Using secondary data; sources, reliability, accuracy, relevance and bias.	Students need to understand that if they have not physically collected data themselves it is not primary data. Therefore you are reliant, when using secondary data, on the quality of the sample and collection methods which you will probably know nothing about. A major aspect of the new specification is for students to be able to read and interpret data from published secondary sources.
Difference between census and sample data	Amply covered in sampling part of specification – nothing new arising here.
Designing and obtaining data from simple statistical experiments. Obtaining data from observation or experiments (laboratory, field or natural)	No detailed specialist knowledge required – simply that laboratory experiments give the investigator a high degree of control but the participants are not in their natural surroundings so their behaviour may be affected. Field experiments have a lower degree of control usually take place in a more natural environment but participants are still aware they are in an experiment. Natural experiments are usually simply observation with it being unlikely that participants are aware they are in an experiment – this gives the investigator very little control over the experiment giving opportunity for extraneous variables to become involved. **Example** An investigation is carried out into sleep patterns. 10 patients are brought into hospital and their heart rates monitored throughout the night. (a) Name this type of data collection. (b) A different investigation completes this at the patient's homes. Compare the two methods. **Solution** (a) A laboratory experiment. (b) First method – outcomes may be affected by people being in a hospital – unlikely to sleep well. Second method – outcomes may be affected by extraneous variables such as other people or external noise.

Statement from Specification	Assessment Guidance (including examples)
Foundation and Higher tiers	
Examples of extraneous variables	See above example – the word extraneous will be used in examinations.
Higher tier only	
Identification of extraneous variables and methods of controlling them : the need to hold extraneous variables constant for both groups	The additional work in Higher is the requirement to be able to make suggestions which should remove or at least minimise the effect of extraneous variables. **For Example** In the sleep experiment carried out in people's homes above, candidates could be asked to (i) Name a possible extraneous variable (ii) Suggest a way of minimising the effect of the variable you named in part (i) **Solution** (i) Being disturbed by the phone ringing in the night (ii) Unplug or turn off all phones **Teaching Tip** There are often many possible answers to these questions, better candidates should be thinking of (ii) when they make their choice for (i) as some would be far harder to deal with than others. eg, a common answer for (i) might be outside noise disturbing sleep, but with that it is very difficult to come up with a convincing answer for part (ii).
Matched pairs of groups; "before and after" experiments.	The idea of comparisons between groups in an experiment being most effective if the constituency of both groups is as similar as possible eg, testing two new drugs for flu – make sure patients for each drug have the same 'average' seriousness of symptoms. Before and after experiments include situations such as testing the effectiveness of a diet where it makes sense to have the same participants before and after to obtain measurable differences in weight.

Statement from Specification	Assessment Guidance (including examples)
Foundation and Higher tiers	
Issues of inter-observer bias	When collecting data by observation eg, queuing patterns in a supermarket, Ofsted observations of a teacher etc it is possible for different observers to make a different decision even though they have both witnessed the same "behaviours". This is inter-observer bias. It is best overcome with good training and activities such as joint observations to set standards. These issues are the limit of assessment in this area.
Explanatory and response variables; identification of the variables to be investigated	**Example** A new drug is tested to reduce blood pressure. 100 patients are divided into 4 groups of 25. Group 1 have 10mg per day Group 2 have 20 mg per day Group 3 have 30 mg per day Group 4 have 40 mg per day Their blood pressure is measured every day for one year. (a) Identify the explanatory variable (b) Identify the response variable (c) Identify a possible extraneous variable **Solution** (a) The explanatory variable is the amount of the drug given each day **Teaching Tip** This is the variable you can control – also called the independent variable (b) The response variable is the blood pressure of the patients **Teaching tip** This is the variable which is the outcome of the experiment – you cannot control it. It is also called the dependent variable.

Statement from Specification	Assessment Guidance (including examples)
Foundation and Higher tiers	
Use of a control group; use of a random allocation to experimental and control groups	The idea that it is not sound to simply have a group of patients given a new drug but to have a group of patients who **think** they are also getting the new drug to allow for psychological effects and similar. For example, in the example above a 5th group should be given a dummy pill (called a placebo) and their results should be analysed along with those of the 4 groups receiving various levels of the new drug. Patients should be allocated at random to each group (Higher students could be asked to consider the idea of matched pairs – see earlier) and patients should not know (of course) which group they're in (called a blind trial). Sometimes doctors and other medical staff also don't know which group of patients or which drugs are live or dummy. This would then be called a double blind trial. These words are not used in the specification but it would be useful for students to know their meaning.
Surveys	Use of the general term to indicate some of the data collection methods indicated above.

3.1.3 Tabulation and Representation

Statement from Specification	Assessment Guidance (including examples)
3.1.3a Tabulation	
Foundation and Higher tiers	
Construction of frequency tables by tallying raw data Use of five bar gates expected	In the earlier parts of the Foundation paper, the tables will be given to students and will often be discrete or qualitative data. Later in the Foundation paper and in Higher papers students may have to fill in the row headings for qualitative, discrete, grouped discrete or grouped continuous data. Full marks not possible for students who make a long list of tallies instead of using five bar gates.
Class intervals	Knowledge of different possible labelling systems for class intervals – the most common method used will be double inequalities as they lead to the easiest interpretations for drawing diagrams, finding midpoints for estimating means and upper class bounds for cumulative frequency diagrams. 10 – 19, 20 – 29, etc may be used in some circumstances. There is no intention to use labelling such as 10 – ,20 – etc; –10, –20, etc at Foundation and their use is also not expected to be common at Higher.
Higher tier only	
Open-ended classes	Only at Higher will students be expected to deal with a final class interval which is open such as $x > 100$. Students may have to make decisions about how to represent such classes or may be told within the question.
Foundation and Higher tiers	
Simplifying tables by combining categories and reducing the number of significant figures resulting effects on readability of masking of patterns/trends, loss of detail	See 3.1 .2a But reasoning will not expected to be complex at Foundation.
Problems of under and over simplification resulting from unsuitable choice of group size or number of significant figures	See 3.1 .2a Ideally a frequency table would consist of perhaps between 4 and 10 class intervals.

Statement from Specification	Assessment Guidance (including examples)
Foundation and Higher tiers	
Reading and interpreting data presented in tabular form	Many questions in the new specification will involve data being presented in a variety of types of tabular form, often using real data and often coming from well-known public websites such as www.statistics.gov.uk . Tabular data from other sources such as newspapers and magazines could be used.
Design of tables to summarise data effectively	Students may be told to design a table from a description of the data it will hold or from a list of the actual data.
Design and use of two-way tables	From a description of the data, a list of the data or similar. Questions will involved picking out the appropriate value(s) from a table and presenting them as values or probabilities

3.1.3b Diagrammatic Representation

Statement from Specification	Assessment Guidance (including examples)
Foundation and Higher tiers	
Qualitative data : bar and pie charts, pictograms. Multiple and composite bar charts	Students are expected to be able to draw and interpret these diagrams. They can also be expected to choose one of these diagrams as appropriate for a given data set. Note that the knowledge of the type of data for which diagrams are required is deemed important as it will also be very useful in the Controlled Assessment including the Written Assessment. Bar charts without gaps between the bars will be penalised. Pictograms will be limited to symbols and part symbols where it is clear how to divide them. Pie charts may use real data which leads to decimal values for angles. Percentage composite bar charts may be used.

Statement from Specification	Assessment Guidance (including examples)
Foundation and Higher tiers	
Dot plots for small data sets	Students should be able to draw and interpret a dot plot – a series of vertical columns of dots for a small set of qualitative or quantitative discrete data such as the score on a dice for 10 rolls. A vertical axis is not required on a dot plot and one dot will always represent one item. **Teaching Tip** Can be taught at the same time as bar charts as they are closely related to them in being similar forms of frequency diagram. Also could be considered as a sort of vertical pictogram though a single dot will only represent a single item. Offers an alternative diagram which could be used where appropriate in the Controlled Assessment Investigation or in response to questions in the written assessment which ask 'name another appropriate diagram you could have used'.
Higher tier only	
Comparative pie charts (area proportional to total frequency)	Students are expected to be able to draw and interpret these diagrams. They can also be expected to choose one of these diagrams as appropriate for a given data set. Some questions may simply ask students to calculate an appropriate radius for a chart given the radius of another chart and the two total frequencies.
Foundation and Higher tiers	
Discrete data : vertical line graphs	Students are expected to be able to draw and interpret these diagrams. They can also be expected to choose one of these diagrams as appropriate for a given data set.

Statement from Specification	Assessment Guidance (including examples)
Higher tier only	
Cumulative frequency step polygons	Students are expected to be able to draw and interpret these diagrams. They can also be expected to choose one of these diagrams as appropriate for a given data set. It is vital that students are clear that these are for discrete data.
Foundation and Higher tiers	
Continuous data: grouped frequency diagrams including histograms with equal class intervals and frequency polygons	Students are expected to be able to draw and interpret these diagrams. They can also be expected to choose one of these diagrams as appropriate for a given data set.
Higher tier only	
Histograms with unequal class intervals	Students are expected to be able to draw and interpret these diagrams. They can also be expected to choose one of these diagrams as appropriate for a given data set. Students may use other methods where possible but the method of frequency density must be known as this method may be specifically tested.
Foundation and Higher tiers	
Cumulative frequency graphs	Students are expected to be able to draw and interpret these diagrams. They can also be expected to choose one of these diagrams as appropriate for a given data set. The points in a cumulative frequency graph may be joined with either a curve or straight lines.
Population pyramids	Students are expected to be able to interpret these diagrams and make comparisons – drawing of them would be limited to part of one diagram in most cases.

Statement from Specification	Assessment Guidance (including examples)
Foundation and Higher tiers	
Output gap charts	Students should be able to explain what an output gap chart shows in very simple language eg, how actual output of a country compares with its GDP (which can be viewed as the expected value of all goods and services which the country can produce.) No complex interpretation required – interpretation will be limited to two cases (i) The chart has a line above zero (positive output gap) which indicates output above the GDP and may indicate a boom in an economy. (ii) The chart has a line below zero (negative output gap) which indicates output below the GDP and may indicate a recession in an economy. Students are not to be expected to draw these charts. We only expect to set "small" questions on this topic when set probably in conjunction with a question about GDP (at Higher) or other indices. **Teaching Tip** Can be covered at the same time as GDP and other index numbers. An output gap chart is also a form of time series graph but it fits better with economic index work rather than when time series is taught
Stem and leaf diagrams	Students are expected to be able to draw and interpret these diagrams. Stem and leaf diagrams should always have a key which if not provided should be given by the student completing the diagram. This would usually be prompted.
Choropleth maps	Students are expected to be able to draw and interpret these diagrams. If asked to name one of these diagrams the term 'shading map' would also be acceptable.

Statement from Specification	Assessment Guidance (including examples)
Foundation and Higher tiers	
Transforming data presentation from one form to another	For those types of data where more than one diagram is available questions could be set which ask this. **Example** Data presented in a bar chart. Draw a pictogram to represent the data (symbol and value may or may not be given). Likely follow up questions – give one advantage of the bar chart over the pictogram (or vice versa or both).
The shapes and simple properties of frequency distributions; symmetrical, positive and negative skew	Only the simple identification of the three possible situations to be considered at Foundation Tier. Degrees of skew could be tested at Higher by consideration of measures of skew (See 3.1.4b)
Higher tier only	
The shape and simple properties of the Normal frequency distribution	Though the specification is very specific about the values it is sufficient for students to know that within about 2 standard deviations of the mean lies 95% of the data and that virtually all the data lies within 3 standard deviations of the mean for a Normal distribution. Recognition of the symmetrical 'bell-shape' nature of the Normal distribution can be tested by asking for sketches of distributions. Furthermore in two sketches of Normal distributions students should • show that both curves have approximately the same area underneath • draw a higher peak for the distribution with the smaller standard deviation • draw both curves to about the mean plus and minus three standard deviations (however it is not crucial whether or not the student incorrectly accidentally joins the curve to the axis)

Statement from Specification	Assessment Guidance (including examples)
Foundation and Higher tiers	
Bivariate data: scatter diagrams	See 3.1.4f
Time series : line graphs	Students can be asked to plot and / or interpret time series graphs. See 3.1.4d
Other diagrammatic representations for comparisons of data using length	This reference simply allows a diagram to be used which may not be named above. If this is done it will be clear and may well invite students to criticise its use – for example if it is taken from the media.
Higher tier only	
Use of area and volume. Comparison of the various diagrammatic representations using area or volume including their advantages and disadvantages.	Discussion of different diagrams with consideration of their effectiveness and whether an alternative diagram might have been better.
Foundation and Higher tiers	
Visual misrepresentation: misuse of omission of origin or scale. Broken, incorrect or changed scales. Incomplete definitions and labelling.	This very common feature of real life (eg, media) use of statistics which gives a misleading impression of data will be regularly tested. Students will need to inspect provided diagrams carefully for the features described and be prepared on occasions to correct or redo the diagram correctly.
Simple misuse of area and volume	Just an awareness no calculations required at Foundation Tier
Higher tier only	
Misuse of length area and volume in pictorial comparison	Diagrams may be drawn in 2 or 3 dimensions and students may be asked to identify incorrect sizing or calculate the appropriate area or volume for a second diagram given the frequencies for both. This may entail very simple knowledge of area and volume calculations such as squares, rectangles, cubes and cuboids.

Statement from Specification	Assessment Guidance (including examples)
Foundation and Higher tiers	
Read or interpret information presented in diagrammatic form; distinction well and poorly presented data	General interpretation of diagrams often drawn from the media, Government or other statistical sources. Also even though diagrams may be accurate and appropriate, some will be better than others in different circumstances possible according to the nature of the interpretation required **Example** Data showing absences of boys and girls each day from a school. Which would be better out of a multiple or composite bar chart to identify (a) comparisons between the boys and girls absent each day (b) comparisons between the total number absent each day?
Spotting possible errors in a data set by recognising outliers that do not fit a general pattern	Students should be aware when values in a given data set seem to be erroneous due to their value compared with the rest of the data. A simple consideration of whether these values should or should not be excluded. See also 3.1 .4b – Outliers

AQA

3.1.4 Data Analysis

Statement from Specification	Assessment Guidance (including examples)
3.1.4a Measures of Location	
Foundation and Higher tiers	
Mean, mode and median for raw data	To include calculating these measures and interpreting them, possibly in a given context.
Higher tier only	
Use of a change of origin when calculating the mean. Effect on the mean of linear transformations	**Example** The mean of the numbers 1003, 1005, 1006 and 1009 is 1000 plus the mean of 3, 5, 6 and 9. Students can be instructed to use these methods or might decide themselves to use them. A linear transformation could also see each value of a set being divided or multiplied by a constant. **Example** If the mean of a set of data, x, is c, what is the mean of the data set y, where $y = ax + b$?
Foundation and Higher tiers	
Mean, median and mode for discrete frequency distributions	To include calculating these measures and interpreting them, possibly in a given context. The formula for the mean will be given but use of fx is not expected at Foundation – however students are of course welcome to use it if they wish. Modes and medians may be identified from diagrams.
Modal class for grouped frequency distributions	Simple identification from tables or diagrams.
Median for grouped frequency distributions	Identification of the group within which the median lies at Foundation or estimation of the value of the median by interpolation at Higher. In theory this latter option is also on the Foundation specification but is unlikely to be tested at that tier. Alternatively estimating the value from a suitable diagram.

Statement from Specification	Assessment Guidance (including examples)
Foundation and Higher tiers	
Mean for grouped frequency distributions	To include calculating the measure and interpreting, possibly in a given context. Use of midpoints required, this may or may not be prompted.
Higher tier only	
Use of $\sum fx$ notation	Values for the statistic $\sum fx$ may be given.
Foundation and Higher tiers	
Advantages and disadvantages of each of the three measures of location in a given situation	**For example** Seeing that there is a particular value much larger than the rest in a set of data and so indicating the mean would not be the best choice. **For example** Seeing that the mode also happens to be the lowest value and so indicating the mode would not be the best choice.
Higher tier only	
Reasoned choice of a measure of location appropriate to the nature of the data and the purpose of the analysis.	At Higher students may have to choose the measure of average which seems most suitable (including possibly use of the geometric mean (see below)).
Geometric mean	To include calculating the measure and interpreting, possibly in a given context.

Statement from Specification	Assessment Guidance (including examples)
3.1.4b Measures of Spread	
Foundation and Higher tiers	
Range	To include calculating the measure and interpreting, possibly in a given context. The range must be given as a single value not, for example, as 4 – 12.
Quartiles for discrete data	Worked out from ordered lists, stem-and-leaf diagrams or cumulative frequency diagrams.
Quartiles and percentiles for grouped frequency distributions	Worked out from ordered lists, stem and leaf diagrams or cumulative frequency diagrams.
Higher tier only	
Deciles	Worked out from ordered lists, stem and leaf diagrams or cumulative frequency diagrams.
Foundation and Higher tiers	
Interquartile range for discrete and continuous data	Worked out from ordered lists, stem and leaf diagrams or cumulative frequency diagrams. **Teaching Tip** Though not required in the specification students often find use of Q1, Q2 and Q3 for the lower quartile, median and upper quartile respectively useful in remembering their positions and the formula for interquartile range.
Higher tier only	
Interpercentile ranges. Interdecile ranges.	Worked out from ordered lists, stem-and-leaf diagrams or cumulative frequency diagrams. **Teaching Tip** Consider use of P27 for the 27th percentile for example or D9 for the 9th decile. Thus the Interdecile range could be given as D9 – D1 which is probably going to be easier to remember for students.

Statement from Specification	Assessment Guidance (including examples)
Higher tier only	
Variance and standard deviation. Candidates should know how to use the formula for finding the variance and standard deviation. The formula will be given.	Worked out from ordered lists, stem and leaf diagrams or cumulative frequency diagrams. Students will not be expected to calculate either the variance or the standard deviation from a frequency table. Instead summary statistics will be given such as $\sum fx^2$ or equivalents will be given. However please be aware that the ability to work out these measures from an initial table could arise in the Controlled Assessment once data has been collected. Students are encouraged to use their calculator to evaluate these measures but working should still be shown otherwise an isolated wrong answer can only score 0. Students are expected to be using the divisor of n for variance and standard deviation but would not be penalised if they used $n - 1$.
Foundation and Higher tiers	
Advantages and disadvantages of each of these measures of spread	It is the presence or otherwise of outliers / extreme values that usually dictates whether range is a suitable measure compared with the IQR (at Foundation) or whether standard deviation is better than IQR or IDR at Higher. Reasons for using the IDR above the IQR may be tested at Higher (only excluding 20% of the data rather than 50%).
Construction of box and whisker plots	Students are expected to be able to draw and interpret these diagrams. No outliers at Foundation Tier.
Higher tier only	
Use of box and whisker plots to identify outliers	An outlier is defined as an observation less than Q1 – 1.5 (Q3 – Q1) or greater than Q3 + 1.5 (Q3 – Q1) They should be identified with a cross on a box and whisker plot with the whiskers going as far as the lowest (or highest) value which is not an outlier.

Statement from Specification	Assessment Guidance (including examples)
Higher tier only	
Calculation, interpretation and use of standardised scores	For Normal distributions the standardised score is calculated as Z where: $Z = \frac{value - mean}{s.d}$ It is important to preserve the sign of Z after a calculation. **Example** A typical question may have 2 values from a Normal distribution and the student is asked to work out which is the more likely. This will be the one with the standardised score closer to zero. eg, if the standardised score of A is – 1.6 and the standardised score of B is (+) 1.7, A is more likely. A common error is that the larger value is more likely.
Foundation and Higher tiers	
Use of tabulated data, diagrams, measures of location, measures of spread and skew to compare data sets	Questions are often set which require students to make comparisons. Usually a good comparison involves a measure of spread and a measure of location. At Higher measures of skew are also available.
Higher tier only	
Calculation, interpretation and use of measures of skewness. Candidates should know how to use Pearson's measure. The formula for this will be given.	Though other measures of skew exist and could be used (with the formula always given), the most likely measure of skew to be used is Pearson's. The formula which is $\frac{3(mean - median)}{s.d}$ will always be given.

Statement from Specification	Assessment Guidance (including examples)
3.1.4c Other Summary Statistics	
Foundation and Higher tiers	
Simple index numbers	Problems based on using the base index of 100 and comparing prices or finding other indices for related prices. **Example** In 2009 petrol cost 104 pence per litre. In 2010 petrol cost 124 pence per litre. Using 2009 as base, calculate an index number for 2010. **Solution** $\frac{124}{104} \times 100 = 1.19230769 \times 100$ $= 110.230769$ $= 119 \ \ (3 \text{ sf})$ Note the advised use of full calculator displays at ALL times followed by rounding. (If rounding is not specifically requested it need not be carried out, nor will incorrect rounding be penalised as long as a fuller correct answer has been seen earlier). **Example** In 2009 bread cost £1 for a loaf. Using 2009 as base the 2010 cost has an index number of 115. Write down the 2010 cost of the equivalent loaf of bread. **Solution** £1.15

Statement from Specification	Assessment Guidance (including examples)
Higher tier only	
Weighted index numbers	This is effectively a weighted average of different indices (though weighted averages per se are not on the specification and will not be tested). Candidates will usually be given a table with different costs or indices and then be asked to produce a weighted index based on a given weight for the items in the table. Or changes or values may be expressed as percentages. The formula $\frac{\sum(\text{index number} \times \text{weight})}{\sum \text{weights}}$ may be helpful for some candidates. **Example** Brian buys either water or milk to go with his lunch. The cost of his water was 60 p in 2000 and 80 p in 2010. The cost of his milk was 45 p in 2000 and 70 p in 2010. Given that 80% of the time he chooses milk, calculate a weighted index number for 2010 using 2000 as base. **Solution** For water, index number is $\frac{80}{60} \times 100 = 133.333\ldots$ for milk index number is $\frac{70}{45} \times 100 = 155.555\ldots$ Weighted number = $\frac{(133.333... \times 20) + (155.555... \times 80)}{20 + 80} = 151.111\ldots$ **Teaching Tip** Leave the full calculator display showing throughout the calculation to avoid premature approximation problems and resulting loss of accuracy. A challenging question could be to give the weighted index but have one of the individual prices missing – this would be a good test for A* candidates and would enhance their understanding.

Statement from Specification	Assessment Guidance (including examples)
Higher tier only	
Chain based numbers	This gives an index number using the previous value as the base rather than always referring back to a base that could be some time ago. Calculations or advantages can be asked for. Advantages include the feature of tracking percentage increases or decreases each period through the number obtained. **Example** The cost of a particular computer game is recorded to the nearest £ over 4 years. 2007 £40 2008 £30 2009 £25 2010 £18 (a) Work out chain base numbers for each year. (b) In which year was the largest percentage decrease? **Solution** (a) 2008 $\frac{30}{40} \times 100 = 75$ (ie, a 25% decrease) 2009 $\frac{25}{30} \times 100 = 83.3\ldots$ (a 16.7% decrease) 2010 $\frac{18}{25} \times 100 = 72$ (a 28% decrease) (b) 2010 has the largest percentage decrease as it has the lowest chain base index number.

Statement from Specification	Assessment Guidance (including examples)
Higher tier only	
General Index of Retail Prices (RPI) General Index of Consumer Prices (CPI) Indices to measure GDP and Retail Sales	Specialist or detailed knowledge of these measures is **not** expected in GCSE Statistics. The level of knowledge should be that of • they are measures of inflation (or deflation) for the economy based on the cost of different goods and services and how these costs change • they are weighted index numbers as different goods and services play a larger part in family budgets than others • the differences between the two centre on the CPI **not** including housing related costs such as mortgage payments. This means that different bodies may choose one or other to make a case for, say, a pay increase, depending upon the nature of the changing values of the 2 indices. • the GDP is a measure of a country’s wealth as it is a measure of the income and output of a country’s economy • no other measures can be tested without being defined.

Statement from Specification	Assessment Guidance (including examples)
Foundation and Higher tiers	
Crude rates	Crude rate calculations work out the number of instances of such as a birth or a death per 1000 of the population. For example, the crude death rate is: $\dfrac{\text{Number of deaths recorded}}{\text{Total population}} \times 1000$ **Example** Healthytown has 25 000 residents on Jan 1st 2009. Jim says the crude death rate was 4.5 in 2009. Explain why Jim must be wrong. **Solution** Using the above formula $\dfrac{\text{Number of deaths recorded}}{25000} \times 1000 = 4.5$ so number of deaths recorded $= \dfrac{4.5 \times 25000}{1000} = 112.5$ It is not possible to have .5 of a death so Jim must be wrong.
Higher tier only	
Standardised rates	Candidates are **not** expected to calculate these rates (this is a change in the new specification) but should be able to understand and interpret such measures. Issues candidates should appreciate: • many rates are affected by the age distribution of the town or area being considered • standardised rates take into account these age distributions whereas crude rates do not • a standardised rate is a weighted average of the individual crude rates for each age group.

Statement from Specification	Assessment Guidance (including examples)
3.1.4d Time series	
Foundation and Higher tiers	
Drawing a trend line by eye and using it for prediction	The trend line should be a line of best fit for the time series plots but there is no double mean point requirement in a time series trend line plot.
Evaluating and plotting appropriately chosen moving averages	Students may be asked to determine appropriate moving averages or justifying why, for example, a four point moving average is appropriate for quarterly data. Moving averages should be plotted at the midpoint of the time period to which they relate.
Trend line based on moving averages	If moving averages are plotted then the trend line should be a line of best fit for the moving average plots but there is no double mean point requirement in a time series trend line plot.
Identification of seasonal variation	At Foundation simply the awareness that a seasonal pattern may exist in data and identifying it.
Higher tier	
Seasonal effect at a given data point	Defined as the difference between the actual data value and the value of the trend line for the moving averages as measured on the time series graph.
Average seasonal effect	The mean of all the seasonal effects for the corresponding values each timer period – for example if you have 3 years of quarterly data, the average seasonal effect for the first quarter would be the mean of the three values for the first quarter.
Prediction of future values	Using an extended trend line plus the value of the average seasonal effect to predict a future value for the given time.

Statement from Specification	Assessment Guidance (including examples)
Higher tier only	
'Z' charts	Students will be expected to be able to construct, read off values from and interpret 'Z' charts. All questions will only require that students are aware of the meaning of the 3 lines which constitute the 'Z' chart namely: (i) The base line as the actual data for the time period (ii) The rising line as the cumulative data for the time period (iii) The roof line which indicates the running total for the equivalent total period ending at that point. Examples of questions which may be asked. 'Use the Z chart to work out / write down (a) The value for March (b) The value for the year up to June (c) The value for the 12 months ending in October. (d) Two patterns in the data Any drawing of a Z chart is basically a plotting exercise which can be done fairly easily as long as students recognise what the three lines represent and how to join them up (in Z shape) **Teaching Tip** These graphs are a type of time series but will also need to be taught after cumulative frequency graphs have been covered.

Statement from Specification	Assessment Guidance (including examples)
3.1.4e Quality Assurance	
Higher tier only	
Plotting sample means, medians or ranges over time to view consistency and accuracy against a target value.	Use of control charts to see the movement over time of these measures in a production process. Often action limits (dotted lines on the graph) show the values above which the production process is deemed to be unacceptable but also trends within a control chart which at the current time remain inside action limits may also be a cause for concern in the production process. Typically a machine will have had measurements taken and students will be asked to discuss the performance of the machine in line with either action limits or the context of the question. Students may also be asked to plot a control chart based on given data.
3.1.4f Correlation and Regression	
Foundation and Higher tiers	
Scatter diagrams. Recognition by eye of positive correlation, negative correlation and no correlation.	Students will be expected to be able to plot a scatter diagram, identify one of the types of correlation listed as appropriate and interpret this correlation in the context of the problem if appropriate.
The distinction between correlation and causality.	Using simple ideas at Foundation students should have awareness that just because it may appear that there is a strong connection between two sets of data it does not mean that (for example) the increase in one set of data is **caused** by the increase in the other. It could be the other way around or it could be a third variable related to both which results in the seen correlation.
Interpret values of Spearman's correlation coefficient in the context of a problem.	NO calculations of Spearman's at Foundation simply a basic understanding of the –1 to +1 scale and where appropriate comments in context.

Statement from Specification	Assessment Guidance (including examples)
Higher tier only	
Spearman's rank correlation coefficient as a measure of agreement; its calculation and limitation in interpretation	Frequently it is expected that the data will already be ranked for candidates and that often the value of $\sum d^2$ will be given so that students are not calculating from scratch. However bear in mind that for the Controlled Assessment the skill of working from raw data may be needed. Interpretations of the values will be fairly general so that students do not need to be overly worried about when a value is strong, moderate or weak. For example a value of 0.7 could be interpreted as strong or moderate (but not weak). A basic understanding that this measure only relates to the ranking of the sets of data and not their actual values is expected (see PMCC). The formula will always be given. Tied ranks are possible.
Foundation and Higher tiers	
Interpret bivariate data presented in the form of a scatter diagram	Nothing new arising from this reference.
Higher tier only	
Comparison of the degree of correlation between two or more pairs of data sets with reference to scatter diagrams.	A simple understanding either by eye or by making calculations of which of a set of scatter diagrams may show the greatest degree of correlation.

Statement from Specification	Assessment Guidance (including examples)
Higher tier only	
Product Moment Correlation Coefficient, and it's interpretation	No calculations of the PMCC will ever be set. Students will need to be aware of the following issues: (i) PMCC is a measure of correlation using the same –1 to +1 scale as Spearman's Rank Correlation coefficient and students will need to be able to interpret values in the same way as they should be able to interpret values of SRCC (in context when required) (ii) PMCC is a more sophisticated measure of correlation than SRCC in that it uses the actual size of the data rather than its rank. This means that a + (or minus) 1 value indicates a perfect straight line on a scatter graph whereas a + (or minus) 1 for SRCC simply means a continually increasing (or decreasing) set of points. **Teaching Tip** Teach the SRCC, its calculation and interpretation. It may then with able groups be possible to show them different sets of data which all would have a SRCC of +1 but they would look different on a scatter diagram. This could then lead to discussions about whether using ranking is that useful – noting the availability of the PMCC as an alternative using the actual values rather than the ranks.
Foundation and Higher tiers	
Fitting a straight line of best fit by eye through $(\bar{x}, \bar{y})$ to the plotted points on a scatter diagram	It is often clearly set up in a question on correlation that students should be plotting a double mean point when drawing a line of best fit but many omit to do this. It is also important that students remember to draw the line as far as the lowest and highest data values in the x direction.

Statement from Specification	Assessment Guidance (including examples)
Higher tier only	
Obtaining the regression equation in the form $y = mx + c$; the interpretation of the regression coefficients m and c	Students need to know about the gradient and intercept and how these values are used in the equation $y = mx + c$. It is also important that students can give practical interpretations of these values in the context of the problem. Often the intercept provides a value which in the context of the question is not possible for $x = 0$ and awareness of this can be tested. (eg, the charge for a taxi travelling 0 miles might be £3 according to a graph or equation)
Non-linear data	A simple awareness that just because there may be no linear correlation between two variables does not necessarily mean there is no relationship at all between the two variables. The only cases that could be suggested are x^2, $\frac{1}{x}$, or $\sqrt{x}$ would simply need to note that there was a connection but it was not a linear one.
Foundation and Higher tiers	
Interpolation and extrapolation	Students need to know the meaning of these two terms and to be able to carry out both using a line of best fit (extending it if necessary for extrapolation). Knowledge that in most cases extrapolation is unwise is expected as well as using a line of best fit for estimation on a scatter diagram with weak correlation.

Statement from Specification	Assessment Guidance (including examples)
3.1.4g Estimation	
Foundation and Higher tiers	
Estimation of population mean from a sample	Simply knowing that the sample mean can be used to estimate the population mean (assuming the sample is representative of the population).
Estimation of a population proportion from a sample; the use of this method of estimation in opinion polls.	Simply knowing that the proportion of a sample having a feature can be used to estimate the proportion of the population having the same feature (assuming the sample is representative of the population). Knowing that this method is used in opinion polls to make prediction about things such as proportion of votes in an election.
Variability in estimates from different samples and the effect of sample size	At Foundation Tier students should be aware that different samples will almost certainly give slightly different values for measures and therefore for estimates of population measures. Also a basic understanding that the larger the sample size the more reliable estimates made from that sample will be (all other things being equal).
Higher tier only	
Estimation of population size based on the capture / recapture method. Conditions for this method to be appropriate.	Students will be expected to be able to carry out suitable calculations to work out population sizes using this technique and to be aware of the practicalities of carrying the method out such as, with animal populations, ensuring marked animals have had time to mix thoroughly with the whole population and that marking them does not make them behave differently to the rest of the population.
An elementary quantitative appreciation of appropriate sample size	Apart from the general idea from two rows above that the larger the sample size then generally the better the estimates, the only other required knowledge is the concept that to halve the variability in an estimate four times the sample size is required.
Understanding what affects the accuracy of the estimates	Apart from issues already covered, then, depending on the context, other issues affecting accuracy such as the data collection method.

3.1.5 Probability

Statement from Specification	Assessment Guidance (including examples)
3.1.5a Probability	
Foundation and Higher tiers	
Probability of an event, impossible events, certain events	Knowing that different events have different likelihoods of happening. Knowledge of zero for impossible events and 1 for certain events.
Use of words such as possible, likely	Other words can be used by students such as unlikely, evens or even chance and use of very where applicable.
Putting events in order of probability	Either by an appreciation of their likelihood or by evaluation of their probabilities. Students will be expected to be able to order decimals successfully and where appropriate convert between fractions and decimals.
Probability on a scale from 0 to 1	Knowing that all probabilities should be in this inclusive range. **Teaching Tip** Stress to students that probabilities should never be expressed using words or ratios and that these methods are not accepted in the examination. Students will find it useful to use P(....) instead of writing 'the probability of .... ' but this is not a requirement of the specification.
Probability as the limit of relative frequency as the number of observations increases. Equally likely events.	Students should know that in some situations probabilities are theoretical using equally likely outcomes and the classical definition of probability (number of 'successes' divided by total number of outcomes) such as the probability of a 6 on a fair dice. They should also know that there are situations where probabilities cannot be calculated and have to be estimated either by carrying out experiments (such as the probability of a 6 on a biased dice) or by collecting data (such as the probability of a white Christmas). In these latter two cases the frequency of the desired event out of the total number of outcomes is known as the relative frequency. The more experiments or data collected the better the estimate of probability should be and in the end with enough data the value should 'arrive' at the actual probability.

Statement from Specification	Assessment Guidance (including examples)
Foundation and Higher tiers	
Sample space: pictorial representation; probability by counting. Use of Venn diagrams, tables and Cartesian grids.	A sample space is a list of all possible outcomes for a given situation. It can be shown as a list, diagram, (Cartesian) grid or table. Students should be able to illustrate all the possible outcomes for a given situation in any of these forms and use their or a provided diagram to answer questions about the probability of events related to the given situation.
Exhaustive events	Events which cover all possible outcomes. This includes the idea that the probability that an event does not happen is 1 – the probability the same event does happen (as happening and not happening are exhaustive). **Teaching tip** This and the next few parts of the specification on probability may be easier to teach if some basic set notation is used such as P(A or B) etc – it is not a requirement of the specification but for many students should make some of the work easier to express.
Mutually exclusive events; the addition law	Mutually exclusive events cannot happen at the same time and so their total probabilities can be obtained by adding their individual probabilities. This is known as the addition law and is straight forward for mutually exclusive events, the sum of all mutually exclusive events being 1. Often the word 'OR' is associated with adding probabilities though there are dangers in this as revealed by the next part of the specification.

Statement from Specification	Assessment Guidance (including examples)
Higher tier only	
The General Addition Law	**Teaching Tip** Some events have overlapping elements. For example, on a dice the probability of an even number is $\frac{3}{6}$ and the probability of a number less than three is $\frac{2}{6}$. However the probability of an even number or a number less than three is **not** $\frac{3}{6} + \frac{2}{6} = \frac{5}{6}$ as the two events are not mutually exclusive. In this case the answer is $\frac{4}{6}$ as there are 4 outcomes which meet the requirements 1, 2, 4 and 6. Getting $\frac{5}{6}$ means the outcome 2 has been counted twice - Venn diagrams are often useful means of showing what is going on to students.
Foundation and Higher tiers	
Independent events; the multiplication law	Students should be aware that probabilities of two independent events should be multiplied if they are required to find the probability of both happening (one AND the other). This is known as the multiplication law.
Higher tier only	
The general multiplication law	Dependent events can also be multiplied as long as the values for the numerator and denominator have taken into account the previous event's outcome. This is effectively use of conditional probability and is often seen when tree diagrams are used.
Foundation and Higher tiers	
Tree diagrams (two stage only) (independent or 'with replacement' only)	These restrictions on tree diagrams apply to Foundation Tier only.
An intuitive approach to conditional probability eg, two-way tables of Venn diagrams	For example, in a two way table for boys and girls and being left-handed or right –handed working out the probability that one of the girls is left-handed is conditional probability.

Statement from Specification	Assessment Guidance (including examples)
Foundation and Higher tiers	
Expected frequencies	The result of multiplying the probability of a particular outcome with the number of experiments to be carried out. Students may, depending upon the context, be expected to round expected values to an integer value.
Comparison of actual frequencies with expected frequencies	Once expected frequencies have been evaluated they can be compared with the actual frequencies as a means of, for example, testing the appropriateness of an assumption made about the nature of the data or the value of a probability.

3.1.6 Data Interpretation

Statement from Specification	Assessment Guidance (including examples)
3.1.6a Data Interpretation	
Foundation and Higher tiers	
	An appreciation of limitations and conclusions – nothing particularly new to add here other than to reiterate the emphasis the specification places on good statistical interpretation ahead of excessive calculation and representation. Also referenced by some of the likely questions in the Written Assessment.

GCSE Statistics from 2009 onwards

Qualification Accreditation Number: 500/4473/4

For updates and further information on any of our specifications, to find answers or ask us a question, register with Ask AQA at:

aqa.org.uk/askaqa

Free launch meetings are available in 2011 followed by further support meetings through the life of the specification. Further information is available at:

http://events.aqa.org.uk/ebooking

Copyright © 2013 AQA and its licensors. All rights reserved.

AQA Education (AQA) is a registered charity (number 1073334) and a company limited by guarantee registered in England and Wales (number 3644723). Our registered address is AQA, Devas Street, Manchester M15 6EX.