**GCSE Statistics (1ST0)**

**Foundation Tier**

**Scheme of Work**

|  |  |  |
| --- | --- | --- |
| Unit | Title | Estimated hours |
|  |  | **Condensed course** | **Full course** |
| 1 | a | Types of data | 3 | 6 |
| b | Population and sampling | 3 | 6 |
| c | Sampling methods | 4 | 8 |
| d | Planning and collecting data | 4 | 8 |
| 2 | a | Qualitative and discrete data | 6 |  12 |
| b | Continuous data | 6 |  12 |
| c | Tabulation | 2 | 4 |
| 3 | a | Measures of central tendency – mode, median and mean  | 4 | 8 |
| b | Measures of dispersion – range, quartiles, interquartile range and percentiles | 3 | 6 |
| c | Box plots, skewness and representing outliers | 4 | 8 |
| 4 | a | Describing correlation by inspection, lines of best fit and Spearman’s rank correlation coefficient | 4 | 8 |
| 5 | a | Calculating moving averages, seasonal and cyclic trends | 4 | 8 |
| 6 | a | Simple probability and theoretical probability | 4 | 8 |
| b | Probability from two-way tables, sample space diagrams, tree diagrams and Venn diagrams | 5 |  10 |
| 7 | a | Interpreting index numbers in context and simple calculations | 4 | 8 |
|  |  | Total hours |  60 | 120 |

|  |
| --- |
| **UNIT 1: The collection of data** |

 [Return to Overview](#Foundn0)

**SPECIFICATION REFERENCES**

|  |  |
| --- | --- |
| 1a.01 | know that a hypothesis can be tested only through the appropriatecollection and analysis of dataformal use of null hypothesis will not be required |
| 1a.02 | know the constraints that may be faced in designing an investigation to test a hypothesis including factors such as time, costs, ethical issues, confidentiality, convenience |
| 1a.03 | determine proactive strategies to mitigate issues that might arise during the statistical enquiry process |
| 1b.01 | know and apply terms used to describe different types of data that can be collected for statistical analysis: raw data, quantitative, qualitative, categorical, ordinal, discrete, continuous, ungrouped, grouped, bivariate |
| 1b.02 | know the advantages and implications of merging data into more general categories, and of grouping numerical data into class intervals |
| 1b.03 | know and apply the terms explanatory (independent) variables and response (dependent) variables |
| 1b.04 | know the difference between primary and secondary data |
| 1c.01 | know the difference between population, sample frame and sample |
| 1c.02 | know that ‘population’ can have different meanings within a statedcontext |
| 1c.03 | know reasons for employing judgement sampling or opportunity (convenience) sampling, and the associated risks of bias when these techniques are used |
| 1c.04 | know appropriate sampling techniques in the context of the problem to avoid biasunderstand random, systematic, and quota sampling |
| 1c.05 | know the key features of a simple random sample and demonstrate understanding of how different techniques, both physical and electronic, are used to select random members from a population: including, but not limited to, dice, cards, random number lists, and calculator functions |
| 1c.06 | use stratification and know when this is appropriate before sampling takes place |
| 1d.01 | a. know that data can be collected from different sources: experimental (laboratory, field and natural), simulation, questionnaires, observation, reference, census, population and samplingb. know that sources of secondary data should be acknowledged |
| 1d.02 | know the importance of reliability and validity with regard to collected data |
| 1d.03 | determine factors that may lead to bias, including issues of sensitivityof the content matter, and know how to minimise data distortion |
| 1d.04 | know the key features to be considered when planning data collection: leading questions, avoiding biased sources, time factors, open/closed questions, different types of interview technique |
| 1d.05 | know and demonstrate understanding of techniques used to deal withproblems that may arise with collected data |
| 1d.06 | know why data may need to be ‘cleaned’ before further processing, including issues that arise on spreadsheets and apply techniques to clean data in context |
| 1d.07 | know the importance of identifying and controlling extraneous variables |
| 2h.03 | know that sample size has an impact on reliability and replication |

**PRIOR KNOWLEDGE**

Students will have appreciation of data and the various places data can be found.

Students should have some understanding of the difficulties with working with data.

Students should have experience with collecting data (possibly in other curriculum areas).

Students should have an awareness that data can be presented in a large variety of different forms.

**KEYWORDS**

Random, sample, population, experimental, raw data, census, stratified sampling, cluster sampling, systematic sampling, convenience sampling, random sampling, quota sampling, bias, quantitative, qualitative, ordinal, categorical, bivariate, continuous, discrete, grouped, ungrouped, variables, hypothesis, census, pilot survey, primary data, secondary data, sample frame, hypothesis, explanatory variables, response variables

**STATISTICAL ENQUIRY CYCLE OPPORTUNITIES**

1. Students need to be given the opportunity to learn the planning stage of the statistical enquiry cycle through:

* defining a question or hypothesis (or hypotheses) to investigate
* deciding what data to collect and how to collect and record it giving reasons
* developing a strategy for how to process and represent data giving reasons.

2. Students need to be aware of the constraints at the planning stage:

* when designing collection methods for primary data
* when researching sources for secondary data, including from reference publications,
* the internet and the media
* through appreciating the importance of acknowledging sources
* recognising where issues of sensitivity may influence data availability.

Students can be given practical exercises of collecting data. See previous controlled assessments for examples of tasks where students can practise the first stage the statistical enquiry cycle.

Present students with a set of data that has items that need to be ‘cleaned’. For example, entries with:

* no entry
* incorrect format
* missing decimal place
* incorrect entry
* extreme outlier.

The Mayfield data set is a possible example of this. Discuss how to work around these problems with students.

|  |  |
| --- | --- |
| **1a. Types of data**(1b.01, 1b.02, 1b.03, 1b.04) | **Teaching time**3–6 hours |

**OBJECTIVES**

By the end of the sub-unit, students should be able to:

* Recognise that data can be obtained from primary and secondary sources;
* Recognise the difference between quantitative and qualitative variables;
* Recognise the difference between discrete and continuous data;
* Recognise and use scales of measurement – categorical, ordinal, rank;
* Categorise data through the use of well-defined, precise definitions or class boundaries;
* Understand, use and define situations for grouped and ungrouped data;
* Understand the meaning of bivariate data;
* Know the difference between independent and dependent variables.

**GCSE MATHEMATICS LINKS**

A14, S6

**POSSIBLE SUCCESS CRITERIA**

Give the advantages and disadvantages of primary and secondary data.

State the disadvantages of grouping data.

**COMMON MISCONCEPTION**

Students are not confident with using the word ‘variable’. Examiners have reported confusing types of variable as independent and dependent. (Foundation tier June 2015 Q3(b))

Students also struggle to categorise some continuous and discrete variables. They like to think that any decimals are continuous and whole numbers are discrete. Obviously shoe size does not follow this rule. Be careful with definitions and explanations.

**NOTES**

This section of the course requires the students to learn a lot of new vocabulary. There are many fun quizzes available online which can test students’ knowledge. The following are some helpful ways for students to remember particular words.4

‘Quantitative’ can be memorised by thinking of a quantity therefore numerical. So qualitative is the opposite and non-numerical.

‘Secondary data’ can be thought of second hand, so primary is first hand.

‘Bivariate’ can be memorised by ‘bi’ meaning two and ‘variate’ for variable (so two variables). This topic, along with discussing independent and response variables, could be left until learning about scatter diagrams (2a.03).

|  |  |
| --- | --- |
| **1b. Population and sampling**(1c.01, 1c.02, 2h.03) | **Teaching time**3–6 hours |

**OBJECTIVES**

By the end of the sub-unit, students should be able to:

* Understand the meaning of the terms population and sample;
* Understand the word ‘census’ with regard to small scale and large scale populations;
* Understand the reasons for sampling and that sample data is used to estimate values in a population;
* Understand that sample size has an impact on reliability and replication;
* Understand, design and use a sampling frame.

**GCSE MATHEMATICS LINKS**

S1, S5

**POSSIBLE SUCCESS CRITERIA**

State a population for an investigation.

Explain why a sample may not be representative of a whole population.

**COMMON MISCONCEPTIONS**

Students have many misconceptions for the meaning of a sampling frame. Students tend to confuse this with sampling strategies. Feedback from Foundation June 2015 Q13(a): when students were asked “A list of all voters in the town is known as …” and they should have given the answer ‘sampling frame’, the most common incorrect answers were ‘census’ and ‘population’.

**NOTES**

Discuss why the national census is only taken every 10 years.

Ask students what the sampling frame would be for:

* the school (list of all students, and teachers)
* people with voting rights (a list of all voters – the electoral register)
* car owners in the county (a list of all cars in the county)
* patients at the local surgery (a list of all patients at the surgery).

|  |  |
| --- | --- |
| **1c. Sampling methods**(1c.03, 1c.04, 1c.05, 1c.06) | **Teaching time**4–8 hours |

**OBJECTIVES**

By the end of the sub-unit, students should be able to:

* Understand the terms random, randomness and random sample;
* Understand the use of random numbers and some of the methods of generating these:
	+ random number tables;
	+ random number function on the calculator;
	+ picking random numbers from a hat;
* Be able to select a random sample, or a stratified sample, by one category as a method of investigating a population;
* Appreciate how bias in a sampling procedure might occur and how it might be minimised;
* Know the difference between:
	+ opportunity (convenience) sampling;
	+ systematic sampling;
	+ quota sampling;
	+ judgement sampling.

**GCSE MATHEMATICS LINKS**

S1

**POSSIBLE SUCCESS CRITERIA**

Work out the number of items needed for a stratified sample.

Describe how to take a random sample from a population.

Explain why a stratified sample would be better than a random sample.

**COMMON MISCONCEPTIONS**

Students tend to think that population is only to do with people.

Students often think that ‘picking numbers out of a hat’ is suitable for selecting random numbers for large sample sizes. Feedback from Foundation tier June 2015 Q14 – “candidates need to realise however that ‘picking from a hat’ is not appropriate except for small populations.”

**NOTES**

Show students how random numbers can be generated. Random number tables can easily be obtained by searching the internet.

Alternatively, random numbers can be generated from a scientific calculator.

|  |  |
| --- | --- |
| **1d. Planning and collecting data**(1d.01, 1d.02, 1d.03, 1d.04, 1d.05, 1d.06, 1d.07, 1a.01, 1a.02, 1a.03) | **Teaching time**4–8 hours |

**OBJECTIVES**

By the end of the sub-unit, students should be able to:

* Understand that there are different methods to collect primary data from different sources;
* Identify appropriate sources of secondary data;
* Extract data from secondary sources, including those based on ICT;
* Understand that data needs to be ‘cleaned’ before being used;
* Understand the aspects of accuracy, reliability, relevance and bias as related to secondary data;
* Understand the techniques used to deal with possible problems associated with the collection of data (including issues of sensitivity);
* Form a hypothesis, and know the appropriate strategies to test this hypothesis;
* Be aware of factors involved with testing a hypothesis (including time, costs, ethical issues, confidentiality and convenience);
* Identify problems that may arise with the statistical enquiry cycle (e.g. non response of surveys, difficulty estimating the population or unexpected outcomes) and come up with strategies to help overcome these.

**GCSE MATHEMATICS LINKS**

No links

**POSSIBLE SUCCESS CRITERIA**

When given a hypothesis to be able to write and appropriate plan and give reasons for their appropriate plan.

**COMMON MISCONCEPTIONS**

Students need to remember that population is not only to do with people. For example: if a study is of size of leaves on a tree, then the population would be all tree leaves.

**NOTES**

Investigate methods of collecting data in the real world, e.g. tax return, passport application, national census.

Most data sources will have data that has already been cleaned (i.e. no formatting errors, missing entries, incorrect entries). Therefore it is a good idea to give students an example of data with these problems. A good example is the original Mayfield data set which is available on the Mathematics Emporium at [www.edexcelmaths.com](http://www.edexcelmaths.com/). Alternatively, a different data set could be created.

Students can be given the opportunity to design questionnaires. These can be passed to fellow students, parents or teachers to critique and offer suggestions of improvement.

|  |
| --- |
| **UNIT 2: Processing, representing and analysing data**  |

[Return to Overview](#Foundn0)

**SPECIFICATION REFERENCES**

|  |  |
| --- | --- |
| 1b.02 | know the advantages and implications of merging data into more general categories, and of grouping numerical data into class intervals |
| 2a.01 | represent data sets pictorially using calculated key values as necessary, and interpret and compare data sets displayed pictorially: tabulation, tally, pictogram, pie chart, stem and leaf diagram, Venn diagram |
| 2a.02 | interpret and compare data sets displayed pictorially: population pyramid, choropleth map**.** |
| 2a.03 | represent data sets graphically using calculated key values as necessary, and interpret and compare data sets displayed graphically: bar charts, line graphs, time series, scatter diagrams, bar line (vertical line) charts, frequency polygons, cumulative frequency (discrete and grouped) charts, histograms (equal class width), and box plots |
| 2a.05 | justify the appropriate format and produce accurate visualisation of data |
| 2a.06 | recognise where errors in construction lead to graphical misrepresentation, including but not limited to incorrect scales, truncated axis, distorted sizing |
| 2a.07 | extract and calculate corresponding values in order to compare data sets that have been presented in different formats and be able to present the same information in multiple formats |
| 2a.08 | select and justify appropriate form of representation with regard to the nature of data |
| 2h.01 | use calculated or given summary statistical data to make estimates of population characteristics. Use samples to estimate population mean. Use sample data to predict population proportions |

**PRIOR KNOWLEDGE**

Students should have prior knowledge of some of these topics, as they are encountered at Key Stage 3. In particular:

* bar charts;
* pie charts;
* stem and leaf diagrams.

Students should also have an awareness that graphs can be used to better represent data.

Knowledge of how to use a protractor.

**KEYWORDS**

Frequency, tally, total, merging, pictogram, pie chart, stem and leaf diagram, Venn diagram, histogram, frequency density

**STATISTICAL ENQUIRY CYCLE OPPORTUNITIES**

3. Students need to understand ways that data can be processed and presented, including:

* organising and processing data, including an understanding how technology can be used
* generating diagrams and visualisations to represent the data, including an understanding of outputs generated by appropriate technology
* generating statistical measures to compare data, understanding the advantages of using technology to automate processing.

4. Students need to understand that results must be interpreted with reference to the context of the problem, including:

* analysing/interpreting the diagrams and calculations/measures
* reaching conclusions that relate to the questions and hypotheses addressed
* making inferences and/or predictions
* discussing the reliability of findings.

5. Students should be able to evaluate any statistical work including:

* identifying weaknesses in approach or representation
* suggesting improvements to processes or the presentation
* refining the processes to elicit further clarification of the initial hypothesis.

Students can be given practical experience of processing, representing and analysing data. This could be given to them (e.g. using the Mayfield data set).

|  |  |
| --- | --- |
| **2a. Qualitative and discrete data**(1b.02, 2a.01, 2a.02, 2a.03, 2a.05, 2a.06, 2a.07, 2a.08) | **Teaching time**6–12 hours |

**OBJECTIVES**

By the end of the sub-unit, students should be able to:

* Construct, draw, use and understand:
* Pictograms;
* Bar charts;
* Multiple or composite bar charts for qualitative and discrete data;
* Vertical line graphs;
* Stem and leaf diagrams;
* Venn diagrams;
* Box plots;
* Pie charts (including comparative pie charts);
* Cumulative frequency graphs;
* Understand the distinction between well-presented and poorly presented data;
* Understand the potential for visual misuse, by omission or misrepresentation;
* Select the appropriate representation for the data;
* Group data into class intervals and be aware of the advantages and implications of doing so.

**GCSE MATHEMATICS LINKS**

S2, S4, P6

**POSSIBLE SUCCESS CRITERIA**

Construct cumulative frequency graphs and box plots from frequency tables.

Draw an ordered stem and leaf diagram.

If given a graph say what is misleading/wrong with it.

**COMMON MISCONCEPTIONS**

Students regularly forget the key when drawing a stem and leaf diagram. Feedback June 2015 Q11 extract – “A key was often omitted, meaning full marks were not possible”.

Students should pay attention to carefully labelling axes and selecting appropriate scales when drawing diagrams. Feedback June 2015 Q12(b), when students were asked to draw a frequency polygon – “Some made a poor choice of scale which made it difficult to plot the points accurately. Completely correct polygons with a labelled axis were not that commonly seen.”

Feedback suggests that students struggle with questions when they are asked to present the best graph to represent information (see Q4 June 2012: although tally was not the correct answer, it was commonly given by students).

**NOTES**

Various worksheets are available on the internet giving students practice of the above diagrams, for example, [corbettmaths.com/more/practice-questions](https://corbettmaths.com/more/practice-questions). Students can also be given exposure to constructing the diagrams using ICT.

Bar charts, pictograms and Venn diagrams are part of the GCSE Foundation Mathematics course, while composite bar charts and stem and leaf diagrams are not.

|  |  |
| --- | --- |
| **2b. Continuous data**(1b.02, 2a.01, 2a.02, 2a.03, 2a.04, 2h.01) | **Teaching time**6–12 hours |

**OBJECTIVES**

By the end of the sub-unit, students should be able to:

* Construct, draw, use and understand:
* Pie charts;
* Histograms with equal and unequal class intervals – including comparing data sets displayed in histograms;
* Frequency polygons;
* Cumulative frequency diagrams;
* Population pyramids;
* Choropleth maps;
* Box plots;
* Understand the distinction between well-presented and poorly presented data;
* Understand the potential for visual misuse, by omission or misrepresentation;
* Select the appropriate representation for the data;
* Group data into class intervals and be aware of the advantages and implications of doing so;
* Use calculated/given summary statistics for continuous data to make estimates of population characteristics, for example, samples to estimate the population mean.

**GCSE MATHEMATICS LINKS**

S2

**POSSIBLE SUCCESS CRITERIA**

Be able to describe and give a reason for where the greatest number of objects are on a choropleth map.

Complete a frequency table when given a histogram of equal class widths.

**COMMON MISCONCEPTIONS**

Feedback 2014 extract– “It should be noted that this diagram shows the percentage of population in each age group and not the number. Although responses referring to ‘number’ were condoned on this occasion, candidates should be aware that no information about either population size is given in the diagrams.”

Feedback 2012 extract– “Many would have benefited from showing the angle calculations in part (a), particularly when they did not have a protractor to accurately complete the diagram.”

Feedback 2011Q2 presented the problem: “A quarter of the employees walk to work. Write down the best diagram to show this: a pie chart or a bar chart. You must give a reason for your choice.” The feedback stated “many candidates found it difficult to express clearly the difference in purpose of a pie chart compared to a bar chart. For those who selected pie chart but did not score, it was usually when their reasons could equally apply to a bar chart.”

**NOTES**

Pie charts is the only item in the list which overlaps with the new Foundation tier GCSE Mathematics specification.

Both histograms and cumulative frequency overlap with the new Higher tier GCSE Mathematics specification.

Choropleth maps, population pyramids and frequency polygons do not overlap with the Mathematics specification. Frequency polygons have now been removed from the GCSE Mathematics specification in both Foundation and Higher tiers.

Students are expected to learn the formula for frequency density and the formula for calculating the angle for a sector.

Students need to remember that frequency polygons are drawn by plotting the midpoints with the associated frequencies. The points are then joined with straight lines.

Students will easily confuse histograms with bar charts. Histograms have no space horizontally between the bars.

Students will need practice drawing a pie chart from a table. They must first work out the calculation 360 divided by the frequency. This is then multiplied by each group to work out the angle size for each item.

Alternatively, students may be given a pie chart and asked to analyse and interpret the data.

For cumulative distribution curves, students need to remember to first calculate the cumulative frequencies then plot these values with the end points of the intervals. It is then recommended to join these with a smooth curve. Be careful as some learning resources may suggest joining the points with straight lines (this is needed for a cumulative step polygon, which is not covered).

|  |  |
| --- | --- |
| **2c. Tabulation**(2a.01) | **Teaching time**2–4 hours |

**OBJECTIVES**

By the end of the sub-unit, students should be able to:

* Construct, draw, use and understand:
* Two-way tables and tally charts.

**GCSE MATHEMATICS LINKS**

S2

**POSSIBLE SUCCESS CRITERIA**

Complete a tally chart when given the raw data.

Suggest a chart that could be drawn when given a completed tally chart.

**COMMON MISCONCEPTIONS**

Adding up rows and columns on a two-way table to get the final total in the bottom right (getting double the total, instead of just adding the total of the rows or the total of the columns). Feedback from June 2014 Q4, when the final total on a two-way table was 50 – “The majority understood that the total number of visitors was 50, however a significant amount of candidates still misinterpret the two-way table and add together the rows and the columns to obtain a total of 100.”

**NOTES**

Two-way tables and tally charts are part of the Foundation GCSE Mathematics specification.

It is possible to leave two-way tables until probability is covered. Students will be expected to calculate probabilities from a two-way table.

When students draw a tally diagram, it is important to encourage them to cross off items as they go. They should also check they have the same number of items as the total frequency.

|  |
| --- |
| **UNIT 3: Summarising data: measures of central tendency and dispersion**  |

[Return to Overview](#Foundn0)

**SPECIFICATION REFERENCES**

|  |  |
| --- | --- |
| 2a.03 | represent data sets graphically using calculated key values as necessary, and interpret and compare data sets displayed graphically: bar charts, line graphs, time series, scatter diagrams, bar line (vertical line) charts, frequency polygons, cumulative frequency (discrete and grouped) charts, histograms (equal class width), and box plots |
| 2a.09 | determine skewness from data by inspection |
| 2a.10 | interpret a distribution of data in terms of skewness identified from inspection |
| 2b.01 | calculate averages for discrete and grouped data: mode, median, arithmetic meanthe term ‘mean’ should be understood to be ‘arithmetic mean’ |
| 2b.02 | justify the rationale for selecting appropriate types of average in context |
| 2b.03 | compare different data sets using appropriate calculated or given measure of central tendency: mode, modal class, median and mean |
| 2c.01 | calculate different measures of spread: range, quartiles, interquartile range (IQR), percentiles |
| 2c.02 | identify outliers by inspection |
| 2c.03 | comment on outliers with reference to the original data |
| 2c.04 | compare different data sets using appropriate calculated or given measure of spread: range, interquartile range (IQR), percentiles |
| 2c.05 | use calculated or given median and interquartile range (IQR) to compare data samples and to compare sample data with population data |
| 2h.01 | use calculated or given summary statistical data to make estimates of population characteristics. Use samples to estimate population meanuse sample data to predict population proportions |

**PRIOR KNOWLEDGE**

Students should have met finding the mean, mode, median and range at Key Stage 3.

**KEYWORDS**

Mean, median, mode, range, interquartile range (IQR), percentile, modal class interval, interpolate, outliers, skewness, population, standard deviation, interdecile range.

**STATISTICAL ENQUIRY CYCLE OPPORTUNITIES**

Students can be given practical experience of summarising data. Data could be given to them using the Mayfield data set or they could collect their own. They could then calculate measures of central tendency and dispersion and use their calculations to make comparisons in their data.

**NOTES**

This topic is used to carry out comparisons of two sets of data.

|  |  |
| --- | --- |
| **3a. Measures of central tendency – mode, median and mean**(2b.01, 2b.02, 2b.03, 2h.01) | **Teaching time**4–8 hours |

**OBJECTIVES**

By the end of the sub-unit, students should be able to:

* Calculate the mean, mode and median for a list of numbers;
* Calculate the mean, mode and median for discrete data listed in a table (grouped);
* Calculate the mean, mode and median for continuous data listed in a table (grouped);
* Understand the appropriateness, advantages and disadvantages of each of the three measures of central tendency;
* Understand the effect of transformations on the mean, mode, median.

**GCSE MATHEMATICS LINKS**

S4

**POSSIBLE SUCCESS CRITERIA**

Estimate the mean from a grouped frequency table.

When given a set of data, decide which average to use and give a reason why.

Extract the averages from a stem and leaf diagram.

If given a grouped frequency table explain why the mean would only be an estimate.

**COMMON MISCONCEPTIONS**

When calculating the mean from a table, students often try to divide by the number of rows instead of the frequency. The following feedback is from Q12 June 2015, when students had to calculate the mean from grouped data – “A large number did identify correct midpoints in (a) and often went on to multiply these by the frequencies. Fewer then totalled these values correctly but those who did often then divided by the number of classes rather than the total frequency.”

Some students confuse finding the mean from a table with a cumulative frequency or histogram question as the table is presented in the same way. From the examiner report June 2015 Q12, when students were asked to calculate the mean for grouped data – “Some candidates were mistakenly finding cumulative frequency.”

Students incorrectly find the mean for a median question and vice versa, this was seen on Foundation tier June 2015 Q11(b) when students were given a list of numbers and had to draw a stem and leaf diagram and then calculate the median. A small number of students found the mean instead.

**NOTES**

Transformations of mean, mode and median will be simple scalings such as:

* Everyone in the class results were reduced by 5 marks: how would this effect the mean, mode and median?

Calculating the mean for a list of numbers can often be examined as part of a scatter diagram question as the line of best fit needs to go through the mean value of *x* and *y*.

Students need to be aware of the appropriateness of the mean, mode and median, e.g. the median is a better measure for skewed data as it excludes extreme values.

Measures of central tendency is used to describe mean, mode and median. Students do not need to know this term for examination purposes.

|  |  |
| --- | --- |
| **3b. Measures of dispersion – range, quartiles, interquartile range and percentiles**(2b.01, 2c.01) | **Teaching time**3-6 hours |

**OBJECTIVES**

By the end of the sub-unit, students should be able to:

* Calculate the range for a list of numbers;
* Calculate the five number summary (minimum, lower quartile, median, upper quartile, highest value) for a list of numbers;
* Use interpolation to calculate the median;
* Calculate the interquartile range;
* Calculate the percentiles for a set of data.

**GCSE MATHEMATICS LINKS**

S4

**POSSIBLE SUCCESS CRITERIA**

Compare and interpret the spread of data when given the five number summaries for two sets of data.

**COMMON MISCONCEPTIONS**

Students are known to confuse range with interquartile range. This was seen on June 2014 Q12, when students were asked to calculate the IQR from a boxplot – “there was a fundamental lack of understanding of IQR by a lot of candidates, with some of them finding the range instead or using the median in place of one of the quartiles.”

Students also struggle to decide between measures of dispersion and measures of centrality, e.g. “A journalist is investigating the numbers of children in families in the town. Suggest a different average she could use instead of the mean.” (Foundation tier 2015 Q2(e)) The most common incorrect answer was range.

**NOTES**

Students will not be given the formula for interquartile range. (Interquartile range = upper quartile – lower quartile)

Calculation of standard deviation is not expected.

Dispersion is used to describe the spread of the data. Students do not need to know this term for examination purposes.

|  |  |
| --- | --- |
| **3c. Box plots, skewness and representing outliers**(2a.03, 2a.09, 2a.10, 2b.03, 2c.02, 2c.03, 2c.04, 2c.05) | **Teaching time**4–8 hours |

**OBJECTIVES**

By the end of the sub-unit, students should be able to:

* Construct, interpret and use box plots from summary statistics;
* Construct, interpret and use box plots from cumulative frequency graphs;
* Identify outliers by inspection;
* Show outliers on box plots and to be able to correctly interpret these;
* Determine skewness by inspection and to make interpretations in context;
* Use box plots as a method to compare two (or more) sets of data for dispersion, measure of central tendency and skewness;
* Given the median and interquartile range, make comparisons between different data samples of to compare the sample and population data;
* Identify simple properties of the shape of distributions of data including symmetry, positive and negative skew.

**GCSE MATHEMATICS LINKS**

S4

**POSSIBLE SUCCESS CRITERIA**

Compare data sets when given the median and IQR.

Given two box plots, be able to make comparisons between them, and in context.

**COMMON MISCONCEPTIONS**

The most common mistakes from examiner reports include:

* Using the word ‘mean’ when comparing two box plots. From the examiner feedback June 2013 Q9 – “Correct descriptors are required, so for box plots we need to see ‘median’ (not mean or average).”
* Failure to mention the IQR or range (2013 feedback): simply saying one graph is more spread is not sufficient;
* Students regularly confuse the direction of skew;
* Confusing the word ‘skew’ with ‘correlation’;
* Students stating two values but not actually making a comparison: “It should also be noted that stating values (e.g. zoo deer and wild deer medians) is not a comparison in itself; when values are stated there needs to be use of comparative language (e.g. ‘... which is larger than ...’)”. (2013 feedback)

**NOTES**

When comparing box plots, students are expected to compare a measure of spread (range or inter quartile range), the median and the skew. This should be followed with a comment in context.

Box plots can be generated from data or from a cumulative frequency graph. Students need to remember that the lower quartile, median and upper quartile are obtained from drawing lines from the y axis and reading the corresponding values from the x axis.

Calculation of outliers is not expected. Students need to know that outliers may be genuine unusual values and not always errors.

Students find the concept of skewness confusing. A YouTube clip (Skew: How to memorize, negatively and positively skewed distributions) gives students a good visual way to remember the concept.

|  |
| --- |
| **UNIT 4: Scatter diagrams and correlation** |

[Return to Overview](#Foundn0)

**SPECIFICATION REFERENCES**

|  |  |
| --- | --- |
| 1b.03 | know and apply the terms explanatory (independent) variables and response (dependent) variables |
| 2e.01 | know and apply vocabulary of correlation: positive, negative, zero, causation, association, interpolation and extrapolation |
| 2e.02 | describe and make comparisons of correlation by inspection: strong or weak |
| 2e.03 | know that correlation does not necessarily imply causation |
| 2e.04 | determine line of best fit by eye, by drawing through a calculated double mean point ($\overline{x} ,\overline{y}$) |
| 2e.06 | interpret given Spearman’s rank correlation coefficient in the context of the problem |

**PRIOR KNOWLEDGE**

Students will need to know how to calculate the mean to be able to draw the line of best fit.

Students should be confident in plotting coordinates in the positive *x* and *y* axes.

**KEYWORDS**

Axes, axis, bivariate, independent variable, dependent variable, explanatory variable, response variable, interpolate, extrapolate, reliable, unreliable, predict, positive, negative, zero, causation, association, interpolation, extrapolation

**STATISTICAL ENQUIRY CYCLE OPPORTUNITIES**

Given some example sets of data, students could experiment using a spreadsheet to investigate the relationships between two variables. Try to use information they will find engaging and be interested in seeing the strength of the relationship.

Encourage students to label both axes and name the graph: without these the graph means nothing to someone else reading the information.

Students can be given free license to research the information themselves but they may find it difficult to get information already in a spreadsheet format. Remember that collecting data is not the aim, rather it is to practise looking at relationships between two variables.

|  |  |
| --- | --- |
| **4a. Describing correlation by inspection, lines of best fit and Spearman’s rank correlation coefficient**(1b.03, 2e.01, 2e.02, 2e.03, 2e.04, 2e.06) | **Teaching time**4–8 hours |

**OBJECTIVES**

By the end of the sub-unit, students should be able to:

* Plot points as points on a scatter diagram;
* Recognise positive, negative and zero correlation by inspection;
* Understand the distinction between correlation and causality;
* Draw a line of best fit to the points on a scatter diagram through ($\overline{x} ,\overline{y}$) ;
* Understand the pitfalls of interpolation and extrapolation;
* Interpret data presented in the form of a scatter diagram;
* Know and apply the following words: positive, negative, zero, causation, association, interpolation, extrapolation, independent variable, explanatory variable, response variable, dependent variable;
* Describe and make comparisons of the strength of correlation by inspection;
* Interpret Spearman’s rank correlation coefficient in the context of a problem.

**GCSE MATHEMATICS LINKS**

S4, S6

**POSSIBLE SUCCESS CRITERIA**

Make estimations from a scatter graph by drawing a line of best fit.

When given the value of Spearman’s correlation coefficient decide what type of correlation the data has.

**COMMON MISCONCEPTIONS**

Students can say what type of correlation data shows but often fail to interpret the correlation. Feedback from June 2013 Q8 – “The majority of candidates scored 1 mark in (c) for correctly identifying positive correlation, or fewer for giving a correct interpretation, but not too many did both to score 2 marks. ‘The line slopes up’ is not interpretation but was commonly seen.”

Many students misunderstand the meaning of negative values of Spearman’s rank correlation coefficient. (Spearman’s rank was previously a Higher tier only topic. In 2015 Higher tier Q9, students were asked to interpret why –1.25 was not possible: the examiner report said many students “incorrectly stated that ‘it cannot be negative’.”)

**NOTES**

Calculation of Spearman’s rank is not necessary but students need to be able to correctly interpret the value.

Students need to be able to interpret correlation beyond stating positive, negative or no correlation – they need to give context to the answer.

Students need to know the *y* axis is the dependent/response variable, the *x* axis is the explanatory/independent variable.

Students need to practise: accurately plotting values, plotting the mean point, drawing the line of best fit through the mean point, interpolating and extrapolating values from the graph.

Students need to be clear on the reliability of interpolation and extrapolation.

|  |
| --- |
| **UNIT 5: Time series analysis** |

[Return to Overview](#Foundn0)

**LIST OF SPECIFICATION REFERENCES FOR UNIT**

|  |  |
| --- | --- |
| 2f.01 | identify trends in data through inspection and by calculation of 4 point moving averages |
| 2f.02 | interpret seasonal and cyclic trends in context |

**PRIOR KNOWLEDGE**

Students should be able to find the average (mean) of four numbers.

**KEYWORDS**

Trend, seasonal trend, cyclic trend, extrapolate, moving averages

**STATISTICAL ENQUIRY CYCLE OPPORTUNITIES**

Students can be presented with some real life time series graphs, such as the FTSE 100 Index. Discuss how this data can be collected so students are aware of how time series graphs could form part of the statistical enquiry cycle. Time series graphs for example data sets from previous examinations are also useful.

|  |  |
| --- | --- |
| **5a. Calculating moving averages, seasonal and cyclic trends**(2f.01, 2f.02) | **Teaching time**4–8 hours |

**OBJECTIVES**

By the end of the sub-unit, students should be able to:

* Plot points as a time series;
* Draw a trend line by eye and use it to make a prediction;
* Interpret seasonal and cyclic trends in context;
* Calculate and use a 4 point moving average.

**GCSE MATHEMATICS LINKS**

S4

**POSSIBLE SUCCESS CRITERIA**

Calculate and plot 4 point moving averages on a time series graph.

Describe and interpret trends from a time series graph.

**COMMON MISCONCEPTIONS**

When discussing the trend of the data, it is important to state ‘upwards trend’ or ‘downwards trend’. (From the examiner report of Foundation tier 2011 Q14:

“Identifying trend was poor for many candidates. There was no upward or downward trend of any significance. Some stated that it ‘goes up and down’, which may well have been referring to the seasonal variation in quarterly figures rather than moving averages.”)

There are many feedback reports which complain of students not showing the calculation of the 4 point moving averages. This was seen on the feedback from Foundation tier June 2013 Q10 – “Whilst nearly half of candidates found the correct value in (a), gaining 3 marks, those who did not usually scored zero as working was commonly not shown.”

**NOTES**

Students will only be given 4 point moving averages to calculate.

The first moving average (of a 4 point moving average) is plotted in between the 2nd and 3rd value. It is important to show students this and explain what the moving averages are and why they are useful to find trends.

|  |
| --- |
| **UNIT 6: Probability** |

[Return to Overview](#Foundn0)

**SPECIFICATION REFERENCES**

|  |  |
| --- | --- |
| 3p.01 | use collected data to calculate estimates of probabilities |
| 3p.02 | compare the probability of different possible outcomes using the 0–1 or 0–100% scale and statements of likelihood |
| 3p.03 | use probability values to calculate expected frequency of a specified characteristic within a sample or population |
| 3p.04 | use collected data and calculated probabilities to determine and interpret relative risks and absolute risks, and express in terms of expected frequencies in groups |
| 3p.05 | compare experimental data with theoretical predictions to identify possible bias within the experimental design |
| 3p.06 | recognise that experimental probability will tend towards theoretical probability as the number of trials increases when all variables are random |
| 3p.07 | use two-way tables, sample space diagrams, tree diagrams and Venn diagrams to represent all the different outcomes possible for at most three events |
| 3p.08 | know and apply the formal notation for independent events |
| 3p.09 | know and apply the formal notation for conditional probability |

**PRIOR KNOWLEDGE**

Students should have an awareness of the probability scale.

Students should know that probability can be expressed as fractions, decimals and percentages.

Students should be able to represent a basic probability of something occurring numerically and using words.

**KEYWORDS**

Certain, impossible, likely, equal chance, fair, biased, random, theoretical, 50-50, chance, trials, independent, conditional, outcome, sample space diagram, experimental

**STATISTICAL ENQUIRY CYCLE OPPORTUNITIES**

There are a large range of tasks that can be investigated in the classroom using dice or coins. Students can make their hypotheses relating to probability, collecting the data, and completing the cycle. This can also bring in the concept of theoretical and experimental probability.

|  |  |
| --- | --- |
| **6a. Simple probability and theoretical probability**(3p.01, 3p.02, 3p.03, 3p.04, 3p.05, 3p.06, 3p.08, 3p.09) | **Teaching time**4–8 hours |

**OBJECTIVES**

By the end of the sub-unit, students should be able to:

* Calculate probabilities from selected data and interpret
* Understand the meaning of the words ‘impossible’, ‘certain’, ‘highly likely’, ‘likely’, ‘unlikely’, ‘possible’, ‘evens’, and present them on a likelihood and number scale.
* Use probability to calculated expected frequency for a population.
* Compare expected frequencies and actual frequencies. To recognise that experimental probability will tend towards theoretical probability as the number of trials increases.
* Use collected data and calculated probabilities to determine and interpret relative risks and absolute risks, and express in terms of expected frequencies in groups.
* To know and apply formulae conditional probability and independent events.

**GCSE MATHEMATICS LINKS**

P1, P2, P3, P4, P5

**POSSIBLE SUCCESS CRITERIA**

Mark events on a probability scale.

Estimate probabilities when given a frequency table.

**COMMON MISCONCEPTIONS**

Most common mistakes from examiner reports include:

* Obtaining answers greater than 1 (Foundation tier June 2015 Q6 );
* Using the words ‘even chance’: “Care needs to be taken by candidates to check what they have written: ‘each number has a chance (or an even chance) of coming up’ was not uncommon, when they should have said an equal chance” (Foundation tier 2015 Q6(a));
* Using probability words instead of a probability (Foundation tier June 2016 Q2(b)(ii));
* Using a ratio when asked to give a probability (Foundation tier June 2015 Q8);
* Using words such as 24 out of 120.

**NOTES**

Conditional probability is a new addition to the Foundation tier. The formulae for conditional probability and for checking if events are independent will not be given.

|  |  |
| --- | --- |
| **6b. Probability from two-way tables, sample space diagrams, tree diagrams and Venn diagrams**(3p.07, 3p.08, 3p.09) | **Teaching time**5–10 hours |

**OBJECTIVES**

By the end of the sub-unit, students should be able to:

* Produce, understand and use a sample space;
* Understand the terms mutually exclusive and exhaustive and to understand the addition law P(*A* or *B*) = P(*A*) + P(*B*) for two mutually exclusive events;
* Draw and use probability tree diagrams, Venn diagrams and two-way tables;
* To calculate probabilities from tree diagrams, Venn diagrams and two-way tables including conditional probabilities;
* Understand, use and apply the addition for mutually exclusive events, and multiplication laws for independent events.

**GCSE MATHEMATICS LINKS**

P6

**POSSIBLE SUCCESS CRITERIA**

Calculate probabilities when given a Venn diagram.

If given the P(*A*) and the P(*B*), find the P(*A* and *B*).

**COMMON MISCONCEPTIONS**

There is no prior feedback on Venn diagrams as this is new to Foundation tier. However, from Higher tier feedback, the most common problem for students when constructing a Venn diagram is not starting in the middle and, following from this, not subtracting the centre if a number is given for events *A* and *B*. Feedback from Higher tier June 2013 Q9 – “Most candidates were able to gain at least one mark for their Venn diagram in part (a) usually for placing the 5 correctly in the centre. Failure to subtract meant that progress was limited for some and it was fairly common to see 20 and 40 instead of 15 and 35.”

Students need to remember to always have a rectangle around the Venn diagram and that a number needs to be present outside the two/three circles.

Misconceptions students have with tree diagrams include knowing the two probabilities on the branches add to one and knowing to multiply along the branches. The following feedback is from Foundation tier 2013 Q5.

The question required students to completing a tree diagram and comment why getting two sixes is unlikely: “More than two thirds of candidates correctly completed the tree diagram for part (a). However tree diagrams are often difficult for some candidates; common incorrect answers were putting $\frac{1}{6}$ on each branch or sometimes$ \frac{1}{4}$. 90% of candidates scored just one of the two available marks in part (b). Usually this was by agreeing with the statement but either referring to a six being unlikely on just one die, or by having the common misconception that the likelihood of double six was$ \frac{2}{12}$ (as they would need the two sixes from twelve available numbers as they perceived it). To score the second mark a small number of candidates were able to give a coherent argument along the lines of six being unlikely on each die, but only the strongest candidates used the more rigorous approach of evaluating the product of probabilities from the tree diagram to get $\frac{1}{36}$.”

Assessing the understanding of sample space diagrams appeared in Foundation tier 2014 Q13. Students were asked to complete a sample space diagram for a dice and a spinner numbered 1 to 4. Students were confident completing the totals although it was not uncommon to see ordered pairs or products instead of totals. The report follows that the most discriminating question on the paper was knowing how to find the probability of getting a total of 3 or 4. Note that students tend to confuse the word ‘and’ with ‘or’.

**NOTES**

Deeper understanding can be gained by showing how sample space diagrams, Venn diagrams and tree diagrams can all be used to show the same information.

Students need to be confident in constructing and interpreting probabilities from two-way tables, sample space diagrams, tree diagrams and Venn diagrams.

|  |
| --- |
| **UNIT 7: Index numbers** |

[Return to Overview](#Foundn0)

**SPECIFICATION REFERENCES**

|  |  |
| --- | --- |
| 2d.01 | use different types of index numbers in context, including but not limited to, retail price index (RPI), consumer price index (CPI) and gross domestic product (GDP) |
| 2d.02 | interpret data related to rates of change over time (including, but not limited to, percentage change, births, deaths, house prices, and unemployment) when given in graphical form. Calculate and interpret rates of change over time from tables using context specific formula |

**PRIOR KNOWLEDGE**

Students should have some experience of calculating percentage increases and decreases.

Students should have experience of substituting numbers into formulae.

**KEYWORDS**

Consumer price index (CPI), retail price index (RPI), gross domestic product (GDP), rate of change, percentage increase, percentage decrease

**STATISTICAL ENQUIRY CYCLE OPPORTUNITIES**

The Pearson Edexcel legacy specification GCSE Statistics (2ST01) controlled assessment for May 2012 included an option on the topic of money. See the Money controlled assessment brief, which is available on the Mathematics Emporium at [www.edexcelmaths.com](http://www.edexcelmaths.com/), for ideas of how the topic of index numbers could be incorporated into the statistical enquiry cycle. Ideas that students could investigate are:

 • How adult earnings have changed over time;

 • How earnings affect the buying of new consumer items;

 • The make-up of the retail price index (RPI) and how it has changed over time.

Secondary data may be collected from the internet or students may collect their own primary data.

|  |  |
| --- | --- |
| **7a. Interpreting index numbers in context and simple calculations**(2d.01, 2d.02) | **Teaching time**4–8 hours |

**OBJECTIVES**

By the end of the sub-unit, students should be able to:

* Have an understanding of the retail price index (RPI), consumer price index (CPI) and gross domestic product (GDP) and other index numbers in context;
* Calculate and interpret simple index numbers;
* Calculate and interpret rates of change over time including, but not limited to, births, deaths, house prices, unemployment and percentage change.

**GCSE MATHEMATICS LINKS**

R9

**POSSIBLE SUCCESS CRITERIA**

Describe how increase in price compares with RPI over a period of time.

**COMMON MISCONCEPTIONS**

Students tend to just subtract the two given figures for years and not calculate the index number. Feedback from Foundation tier June 2014 Q14 – “Foundation candidates continue to find index numbers a demanding topic and this was once again no exception. At this level, a subtraction of the two given figures led to the most common incorrect answer of £145.”

Students loose accuracy marks when they have the correct answer by then adding % or £ to their answer. (Feedback from Foundation tier June 2014 Q14)

**NOTES**

Chain based index and weighted index numbers are not included.

Crude birth rate and standardised birth rate formulae will be given.

Students need to practise substituting numbers into the formula to generate index numbers.

Students need to practise interpreting index numbers: numbers below 100 mean a decrease and above 100 an increase.