

# **Kettering Buccleuch Academy**

## **GCSE Statistics (1ST0)**

### **Higher Tier**

### **Scheme of Work**

Unit	Title	Estimated hours		
		1 yr	2 yr	
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	4	8
	b	Continuous data	4	8
	c	Tabulation	1	2
3	a	Measures of central tendency – mode, median and mean	3	6
	b	Measures of dispersion – range, quartiles, interquartile range, interpercentile range, interdecile range and standard deviation	4	8
	c	Box plots, skewness, calculating and representing outliers	4	8
4	a	Describing correlation by inspection, lines of best fit and Spearman's rank correlation coefficient, Pearson's product moment correlation coefficient	4	8
5	a	Calculating moving averages, seasonal and cyclic trends	3	6
6	a	Simple probability and theoretical probability	1	2
	b	Probability from two-way tables, sample space diagrams, tree diagrams and Venn diagrams	3	6
7	a	Interpreting index numbers in context and simple calculations	2	4
8	a	Binomial distribution	5	10
	b	Normal distribution and standardised scores	5	10
	c	Quality assurance	3	6
		Total hours	60	120

### SPECIFICATION REFERENCES

- 1a.01 know that a hypothesis can be tested only through the appropriate collection and analysis of data. formal 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 and **multivariate**
- 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 stated context
- 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 a. know appropriate sampling techniques in the context of the problem to avoid bias  
b. understand 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 sampling  
b. 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 sensitivity of the content matter, level of control 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 with problems 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 and the use of control groups
- 2h.02 **apply Petersen capture recapture formula to calculate an estimate of the size of a population**
- 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, raw data, grouped data, ungrouped data, investigation, factors, sample, sampling frame, population, census, stratified sampling, cluster sampling, systematic sampling, convenience sampling, quota sampling, random sampling, strata, quantitative, qualitative, categorical, ordinal, bivariate, multivariate, discrete, continuous, explanatory variables, response variables, primary data, secondary data, hypothesis, census, survey, pilot survey.

## **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.

### **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, rank, ordinal;
- 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 and multivariate data;
- Know the difference between independent and dependent variables.

### **GCSE MATHEMATICS LINKS**

A14, S6

### **POSSIBLE SUCCESS CRITERIA**

List the advantages and disadvantages of primary and secondary data.  
State the disadvantages of grouping data.

### **COMMON MISCONCEPTIONS**

Students 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 learning lots of new vocabulary and keywords. There are many fun quizzes available online which can test students' knowledge.

There are helpful ways for students to remember particular words:

The following are some helpful ways for students to remember particular words.

'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 (Unit 4a).

## OBJECTIVES

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

- Understand the meaning of the term 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;
- Apply and use Peterson's data capture technique to estimate population sizes and know the assumptions made.

## GCSE MATHEMATICS LINKS

S1, S5

## POSSIBLE SUCCESS CRITERIA

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

Explain what is meant by a random sample.

## COMMON MISCONCEPTIONS

Students have many misconceptions for the meaning of a sampling frame. Students tend to confuse this with sampling strategies. Feedback from Higher tier 2015 Q4, when students should have given the answer 'sampling frame' – "Whilst many candidates gave examples of sampling frames in part (a), only few were able to come up with the appropriate statistical term."

Feedback from Higher tier 2014 Q11, when students are asked about the assumptions made when using the data capture technique – "In part (b) many candidates described the method of calculation rather than the underlying assumptions. It was common for two or more comments to refer to there being no change in the population e.g. 'no fish joined the canal, no fish left the canal, none were born' etc., gaining just one mark. A constant population was the most common assumption to gain credit. Unfortunately there were a number of candidates who did not read the question carefully, who seemed to state problems rather than assumptions, and so did not score (e.g. 'some fish may have died' rather than 'no fish died')."

## NOTES

Discuss why the national census is only taken every 10 years and in a year ending with a 1.

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).

## **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;
  - stratified sampling (note this could be by more than one category).

## **GCSE MATHEMATICS LINKS**

S1

## **POSSIBLE SUCCESS CRITERIA**

Calculate the sample size for a stratified sample using more than one category.

## **COMMON MISCONCEPTIONS**

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

Students easily confuse the words 'systematic sampling' and 'stratified sampling'. (Higher tier June 2015 Q8)

## **NOTES**

Take a stratified sample with more than one category, e.g. gender and age group.

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

(1a.01, 1a.02, 1a.03, 1d.01, 1d.02, 1d.03, 1d.04, 1d.05, 1d.06, 1d.07)

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;
- Know the purpose of pilot surveys
- Know how random response is used for sensitive questions
- Understand the techniques used to deal with possible problems associated with the collection of data (including issues of sensitivity);
- Understand why control groups are used in questioning and testing and the system of matched pairs to avoid bias.
- 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

Write a plan to test a hypothesis including reasons for choices.

Give two reasons why you may do a pilot survey.

Design a suitable question for a questionnaire.

### COMMON MISCONCEPTIONS

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

### NOTES

Matched pairs design is when you have different participants in two different conditions, but you match them according to certain variables, such as age, personality, gender, IQ etc.

For example, you may have the following, where the control group is the group that does not experience the experimental condition:

Experimental condition	Control condition
Participant 1 – Age 24, Male	Participant 2 – Age 24, Male
Participant 3 – Age 46, Male	Participant 4 – Age 46, Male

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.

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

## UNIT 2: Processing, representing and analysing data

[Return to Overview](#)

### 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, **comparative pie chart**, **comparative 2D representations**, **comparative 3D representations**
- 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.04 **calculate and use frequency density to draw histograms (unequal class width) and interpret and compare data sets displayed in histograms (unequal class width)**
- 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 **or the misuse of formula when calculating the frequency densities of histograms**
- 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, 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).

## **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;
- Understand the distinction between well-presented and poorly presented data;
- Understand the potential for visual misuse, by omission or misrepresentation;
- Transform from one presentation to another;
- Understand how to discover errors in data and recognise data that does not fit a general trend or pattern;
- Use Comparative 2D and 3D representations;
- 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**

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 from June 2015 Higher tier Q2 – “Common mistakes here were to omit a key or carelessly omit one of the leaves.”

## **NOTES**

Bar charts, pictograms and Venn diagrams are all part of the GCSE Foundation Mathematics course, while composite bar charts and stem and leaf diagrams are not. Stem and leaf is a recent change to the Mathematics specification.

## 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 and unequal class intervals.
  - Frequency polygons;
  - Cumulative frequency diagrams;
  - Population pyramids;
  - Choropleth maps;
  - Box plots;
- Transform from one presentation to another;
- Understand how to discover errors in data and recognise data that does not fit a general trend or pattern;
- 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

Compare comparative pie charts.

Construct cumulative a frequency graph and box plot from a frequency table.

Calculate frequency density and draw a histogram.

## COMMON MISCONCEPTIONS

Feedback 2014 Q15 extract, where students were asked to construct a histogram – “For part (a) many students failed to find the frequency densities correctly but calculated  $f \times x$  instead. Others simply constructed a bar chart for the given frequencies. There were a number of successfully completed tables of frequency density although this did not necessarily transfer to correct or completed histograms, as candidates either struggled to deal with the scale for the histogram or were poor with plotting accuracy. There were lots of unlabelled graphs whilst many incorrectly had frequency for the y axis. Only about one in six managed full marks for the histogram.”

Feedback 2012 Q13 extract – “It was apparent that a small number of candidates were unaware of comparative pie charts as they said the different sizes were irrelevant or that it was misleading.”

From 2011 Foundation tier Q2(e), where the question was “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 examiner 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. This misconception has also been seen at the Higher tier when students have been asked which is the best diagram to use.

## NOTES

Students need to be able to construct histograms with equal and unequal class widths  
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.

### **OBJECTIVES**

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

- Construct, draw, use and understand:
  - Two-way tables and tally charts and any other data represented in a table format.

### **GCSE MATHEMATICS LINKS**

S2, P6

### **POSSIBLE SUCCESS CRITERIA**

If given a partly filled in two-way table, be able to complete it and answer questions from it.

### **COMMON MISCONCEPTIONS**

A common question which students are asked is why does the row or column not add to 100%. This is usually because the values have been rounded.

However, a similar question was asked in 2011 (Q3): "The answer to part (d) is not 100%. Suggest a reason why?" Many students answered that it was due to rounding. However, the part (d) answer was 91.6% – this is clearly not close enough to 100% to be due to rounding. Instead, the answer needed to be along the lines of non-response making up the remainder.

### **NOTES**

Higher tier students should find the concept of two-way tables and tally charts straightforward. They may have difficulties interpreting some probability questions that follow on from the table, in particular conditional probability questions.

## UNIT 3: Summarising data: measures of central tendency and dispersion

[Return to Overview](#)

### 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 **and by calculation**

$$\text{use of: skew} = \frac{3(\text{mean} - \text{median})}{\text{standard deviation}}$$

**formula will be given in the formulae sheet**

- 2a.10 interpret a distribution of data in terms of skewness identified from inspection **or calculation**
- 2b.01 calculate averages for discrete and grouped data: mode, median, arithmetic mean, **weighted mean, geometric mean, mean seasonal variation** the term 'mean' should be understood to be 'arithmetic mean' **unless 'geometric mean' is stated**
- 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, interpercentile range, interdecile range and standard deviation
- 2c.02 identify outliers by inspection **and using appropriate calculations**
- 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 **and standard deviation**
- 2c.05 use calculated or given median and interquartile range (IQR) **or interpercentile range or interdecile range or mean and standard deviation** 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 mean use 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, mode, median, range, interquartile range, percentile, interpercentile range, modal class interval, interpolate, interdecile range, standard deviation, outliers.

### 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.

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

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

**3a. Measures of central tendency – mode, median and mean**  
(2b.01, 2b.02, 2b.03, 2h.01)

**Teaching time**  
3–6 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) including linear interpolation for the median;
- 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;
- Calculate the geometric mean and weighted mean.

**GCSE MATHEMATICS LINKS**

S4

**POSSIBLE SUCCESS CRITERIA**

Be able to state the median, mode, mean and range from a small data set.

Extract the averages from a stem and leaf diagram.

Estimate the mean from a table.

Estimate the mean from a histogram.

**COMMON MISCONCEPTIONS**

When calculating the mean from a table, students often try to divide by the number of rows instead of the frequency. Feedback from June 2014 Q6 – “the common errors being dividing either  $\Sigma fx$  or the total frequency by 6.”

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.

**NOTES**

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  $y$ .

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?

Geometric mean is new to the specification.

**3b. Measures of dispersion – range, quartiles, interquartile range, interpercentile range, interdecile range and standard deviation**  
(2b.01, 2c.01, 2c.04, 2c.05)

**Teaching time**  
4–8 hours

**OBJECTIVES**

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

- calculate the range, quartiles, percentiles and interquartile range for discrete and continuous data;
- calculate the interpercentile range and interdecile range;
- calculate the standard deviation;
- compare data samples and to compare sample data with population data when given measures of dispersion.

**GCSE MATHEMATICS LINKS**

S4

**POSSIBLE SUCCESS CRITERIA**

Calculate the 10th to 90th percentile from a cumulative frequency graph.

Calculate the standard deviation when given summary statistics.

**COMMON MISCONCEPTIONS**

Students are known to confuse range with interquartile range. When given a box plot on June 2016 Q10 and asked to find the IQR the most common incorrect answer was that students calculated the range.

Students also struggle to categorise measures of dispersion with measures of centrality, they often think that the range is an average 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 June 2015 Q2) The most common incorrect answer was 'range'. This misconception is also seen at Higher tier.

**NOTES**

One example of interdecile range is from 10% to 90% of the data. The interpercentile could be any particular percentage of the data.

Students are expected to calculate the standard deviation for both a list and presented in a table. Students are also expected to be confident in notation.

Formulae for standard deviation will be on the formula sheet.

### 3c. Box plots, skewness calculating and representing outliers

(2a.09, 2a.10, 2a.03, 2b.03, 2c.02, 2c.03, 2c.04)

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;
- To calculate outliers using the formulae  
Small outlier is  $< LQ - 1.5 \times IQR$   
Large outlier is  $> UQ + 1.5 \times IQR$   
Or outlier is outside  $\mu \pm 3\sigma$
- To show outliers on box plots and comment with reference to the original data;
- To determine skewness by inspection and calculation;
- To make interpretations in context;
- To use box plots as a method to compare two (or more) sets of data for dispersion, measure of central tendency and skewness;
- Identify simple properties of the shape of distributions of data including symmetry, positive and negative skew.

#### GCSE MATHEMATICS LINKS

S4

#### POSSIBLE SUCCESS CRITERIA

Make comparisons when given two box plots.

Calculate and interpreting skew if given the mean, median and standard deviation.

Justifying by calculation that a point on a box plot is an outlier.

#### COMMON MISCONCEPTIONS

Most common mistakes from examiner reports include:

- Using the word 'mean' or average when comparing two box plots. Feedback from June 2012 Q11 – "Correct descriptors are required, so for box plots we need to see *median* (not mean or average)";
- Failure to mention the IQR or range or use the correct terminology. Feedback from June 2011 Q7(c) – "Unfortunately many candidates used the word spread rather than IQR or range
- Students regularly confuse the direction of skew or fail to use the correct terminology by saying 'neutral' or 'even' skew (Higher tier June 2015 Q7);
- Confusing the word 'skew' with correlation;
- Students stating two values but not actually making a comparison or making comments about maximum and minimum values: "Some candidates wrote lengthy discussions which included comments on the lowest value/lower quartile/upper quartile/highest value which were not required and did not score any marks" (2013 Higher tier Q10).

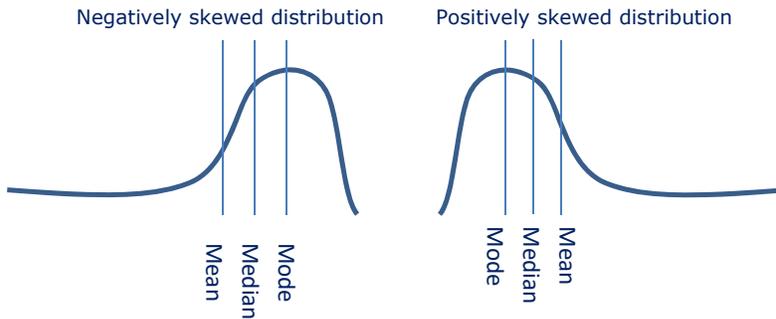
#### NOTES

Calculation of outliers is expected and students will need to remember the formulae.

Students could be expected to calculate a numerical value for skewness. The formula will be given in the formulae sheet.

The following shows two examples of calculations for determining skewness (there are more ways):

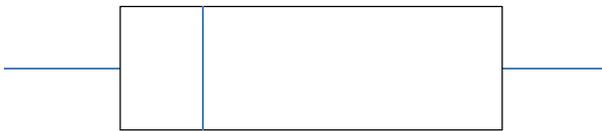
1) By comparing mean, mode and median



2) By comparing the quartiles

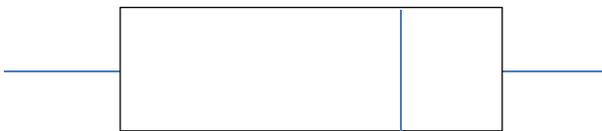
Positive skew

$$Q3 - Q2 > Q2 - Q1$$



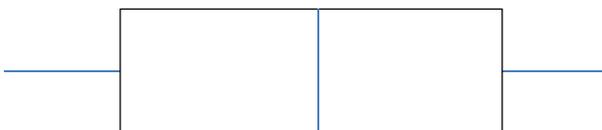
Negative skew

$$Q3 - Q2 < Q2 - Q1$$



No skew / Symmetrical

$$Q3 - Q2 = Q2 - Q1$$



## UNIT 4: Scatter diagrams and correlation

[Return to Overview](#)

### 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 and multiple factors may interact
- 2e.04 determine line of best fit by eye, by drawing through a calculated double mean point  $(\bar{x}, \bar{y})$  and by using the equation of the regression line
- 2e.05 **apply formula to determine Spearman's rank correlation coefficient. Values found using calculator functions will be permissible**
- 2e.06 interpret calculated or given Spearman's rank correlation coefficient in the context of the problem
- 2e.07 **interpret given Pearson's product moment correlation coefficient (PMCC) in the context of the problem**
- 2e.08 **understand the distinction between Spearman's rank correlation coefficient and Pearson's product moment correlation coefficient (PMCC)**

### 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, prediction, correlation

### 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, Pearson's product moment correlation coefficient**  
(1b.03, 2e.01, 2e.02, 2e.03, 2e.04, 2e.05, 2e.06, 2e.07, 2e.08)

**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 through  $(\bar{x}, \bar{y})$  to the points on a scatter diagram and to find the equation of the regression line;
- Understand the pitfalls of interpolation and extrapolation;
- Interpret data presented in the form of a scatter diagram;
- Calculate and interpret Spearman's rank correlation coefficient;
- Interpret Pearson's product moment correlation coefficient;
- Understand the distinction between Spearman's and Pearson's correlation coefficients
- 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.

**GCSE MATHEMATICS LINKS**

S4, S6, A9, A10, A12

**POSSIBLE SUCCESS CRITERIA**

Calculate Spearman's rank correlation coefficient if given ranked data.

When given the equation of a regression line (e.g.  $y = 0.2x + 52.5$ ) draw it on a scatter diagram.

Interpret the value of the gradient of a regression line.

**COMMON MISCONCEPTIONS**

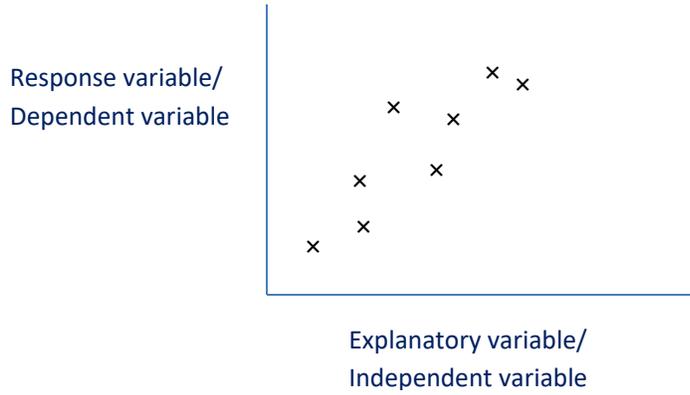
The meaning of negative values of Spearman's rank correlation coefficient. (In Higher tier 2015 Q9, students were asked interpret why  $-1.25$  was not possible: the examiner report said many students "incorrectly stated that it 'cannot be negative'.")

Feedback from Higher tier 2011 Q12 – "In part (a) most candidates were able to correctly rank the numbers and start to use the formula. It is perhaps surprising that quite a number of candidates were unable to copy the formula correctly from the formula page. A common error in calculating the  $d^2$  values was usually due to dealing incorrectly with the squaring of negative integers.

"In part (b), of those candidates with a correlation coefficient to comment upon, the majority were able to identify their answer as positive, negative or no correlation. A number of candidates lost a mark for only describing the strength of the correlation, e.g. weak correlation. Only a minority were able to place their Spearman value in the context of the question. A surprising number of candidates described the relationship between the variables as a positive correlation (which we accepted) despite the low value of the coefficient."

**NOTES**

Students are allowed to use built in calculator functions to calculate Spearman's rank. Tied ranks will not be assessed. The formula will be provided in the formula sheet. Students need to be able to interpret correlation in more detail than simply saying positive, negative or no correlation. They need to give context to the answer. Calculation of PMCC is not required. Students need to know the  $y$  axis is the dependent/response variable. The  $x$  axis is the explanatory/ independent variable.



### SPECIFICATION REFERENCES

- 2b.01 calculate averages for discrete and grouped data: mode, median, arithmetic mean, **weighted mean, geometric mean, mean seasonal variation** the term 'mean' should be understood to be 'arithmetic mean' **unless 'geometric mean' is stated**
- 2f.01 identify trends in data through inspection and by calculation of 4 or other determined appropriate point moving averages
- 2f.02 interpret seasonal and cyclic trends in context  
**use such trends to make predictions**

### PRIOR KNOWLEDGE

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

### KEYWORDS

Trend, seasonal trend, cyclic trend, extrapolate, average, mode, median, arithmetic mean, weighted mean, predictions.

### 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.

### **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 or other specified appropriate moving average;
- Find the mean seasonal variation and extrapolate the data to make predictions for future years.

### **GCSE MATHEMATICS LINKS**

S4

### **POSSIBLE SUCCESS CRITERIA**

If given moving averages to be able to plot them on a time series graph and make predictions.

Find the mean seasonal effect from a time series graph and use this to make predictions.

### **COMMON MISCONCEPTIONS**

When discussing the trend of the data, it is important to state 'upwards trend' or 'downwards trend'. From the examiner report of Higher tier 2016 Q2 – "It must be noted that 'positive correlation' is an incorrect description. Most of the incorrect responses seen, however, described the falls and rises rather than the overall trend. Students are advised that the trend is the overall picture shown by the data ignoring fluctuations along the way and that a trend should be described as upwards (or rising) or downwards (or falling).

There are many feedback reports which complain of students not showing the calculation of the 4 point moving averages. Feedback from Higher tier 2014 Q13 – "Very disappointingly few candidates knew how to find or use seasonal variation, with two thirds failing to score a mark on this question. Most commonly candidates found the average of the three values for Quarter 2, whilst some in part (b) simply stated the trend line value. A few in (b) seemed to guess a value, often about 660; this was out of the acceptable answer range but the calculation was required to be seen on this question anyway. Those who knew what seasonal variation meant commonly scored all 4 marks."

### **NOTES**

Students are expected to know how to draw a trend line through the moving averages, calculate the average seasonal variation and apply this to the following year.

### 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
- 3p.10 **comment on the differences between experimental and theoretical values in terms of possible bias. Formal tests of significance will not be required**

### 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, 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, collect 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.07, 3p.08, 3p.09,  
3p.10)

**Teaching time**  
1–2 hours

**OBJECTIVES**

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

- Calculate estimates of probabilities and represent these as fraction, decimal or percentage;
- Interpret probability values;
- Understand the meaning of the words 'impossible', 'certain', 'highly likely', 'likely', 'unlikely', 'possible', 'evens', and present them on a likelihood and number scale;
- Compare expected frequencies and actual frequencies;
- Recognise that experimental probability will tend towards theoretical probability as the number of trials increases;
- Identify bias if experimental probability does not tend towards theoretical probability;
- To know and apply formulae conditional probability and independent events;
- Use collected data and calculated probabilities to determine and interpret relative risks and absolute risks, and express in terms of expected frequencies in groups;
- Comment on the differences between experimental and theoretical values in terms of possible bias.

**GCSE MATHEMATICS LINKS**

P1, P2, P3, P4, P5

**POSSIBLE SUCCESS CRITERIA**

To be able to calculate probabilities from a two-way table.

Explain what is meant by the word independent.

If ten cards each have a number wrote on them, to be able to calculate probabilities of picking certain numbers, e.g.  $P(2 \text{ or } 3)$ ,  $P(\text{even number})$ ,  $P(\text{not a } 5)$ .

**COMMON MISCONCEPTIONS**

Most common mistakes from examiner reports include:

- Obtaining answers greater than 1;
- 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"; Using a ratio when asked to give a probability (Higher tier June 2013 Q1);
- Using words such as 24 out of 120 (Higher tier June 2013 Q1).

**NOTES**

Conditional probability formula or the calculation to check for independence will not be given.

## 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  $P(A \text{ or } B) = P(A) + P(B)$  for two mutually exclusive events;
- Draw and use probability tree diagrams for independent events. (At most three events)
- Use Venn diagrams and two-way tables to represent all possible outcomes;
- Understand, use and apply the addition for mutually exclusive events, and multiplication laws for independent events;
- To calculate conditional probability following a tree diagram, two-way table or Venn diagram.

## GCSE MATHEMATICS LINKS

P6

## POSSIBLE SUCCESS CRITERIA

Calculate probabilities from a Venn diagram.

If given  $P(A \text{ and } B)$ ,  $P(A|B)$  and  $P(B|A)$ , deciding if  $A$  and  $B$  are independent events.

Complete a tree diagram and find probabilities from it.

## COMMON MISCONCEPTIONS

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. Feedback Higher tier 2013 Q9, when students had to complete a blank Venn diagram – “Even more common was to see the region outside the two circles left blank.”

When selecting probability for a particular event, students only pick the one number and not everything in the circle.

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.

Students need to make sure they read questions carefully with tree diagram questions to see if items have been replaced.

### SPECIFICATION REFERENCES

- 2d.01 use different types of index **and weighted 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, crude birth rate, standardised birth rate, chain based index numbers

### 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.
- The relationship between mortgage interest rates and inflation.

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**  
2–4 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**

Given a base year and figures, be able to calculate the retail price index for other years.

**COMMON MISCONCEPTIONS**

Feedback from Higher tier June 2015 Q10 – “Chain base index numbers remains a challenging topic for many candidates and just over  $\frac{1}{2}$  scored 0 marks here. Though many candidates found it difficult to interpret the index number in context, some were aware that below 100 represented a decrease. It was evident that they were not always certain what was decreasing. A few candidates showed where the figure of 92 came from rather than interpreting it.

“In part (b) it was clear that many candidates did not understand how to calculate chain base index numbers and candidates tended to score either 0 marks or 3. There were some instances of candidates losing the final mark due to incorrect truncation of the second answer to 103 without showing the full answer anywhere in the working out. Quite a significant proportion divided 27 and 28 by 26 i.e. not using chain base index numbers but taking 2008 as the base year.”

When students are able to correctly calculate index numbers they often lose marks by adding incorrect units. Feedback from Higher tier June 2014 Q4 – “Some with the correct calculation unfortunately gave their answer as £ or % losing the accuracy mark.”

**NOTES**

Formulae will be given for crude birth rate and standardised birth rate.

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.

## UNIT 8: Probability distributions

[Return to Overview](#)

### SPECIFICATION REFERENCES

- 2c.04 compare different data sets using appropriate calculated or given measure of spread: range, interquartile range (IQR), percentiles **and standard deviation**
- 2c.05 use calculated or given median and interquartile range (IQR) or interpercentile range or interdecile range or mean and standard deviation to compare data samples and to compare sample data with population data
- 2c.06 **use calculated or given means and standard deviation to standardise and interpret data collected in two comparable samples**  
**formulae for standard deviation will be given in the formulae sheet**
- 3p.11 **know and interpret the characteristics of a binomial distribution**
- 3p.12 **know and interpret the characteristics of a normal distribution**
- 3p.13 **know that, for a normal distribution, values more than three standard deviations from the mean are very unusual; know that approximately 95% of the data lie within two standard deviations of the mean and that 68% (just over two thirds) lie within one standard deviation of the mean**
- 2g.01 **know that a set of sample means are more closely distributed than individual values from the same population**
- 2g.02 **use action and warning lines in quality assurance sampling applications**

### PRIOR KNOWLEDGE

Students should have a good understanding of probability.  
Knowledge of tree diagrams when the probability remains the same.

### KEYWORDS

Event, outcome, percentage, decimal, normal distribution, binomial distribution, standard deviation, mean, symmetrical, frequencies.

### STATISTICAL ENQUIRY CYCLE OPPORTUNITIES

Students can compare data sets or to see if a data set can be modelled using a normal distribution. This could be from previous work they have looked at.

### OBJECTIVES

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

- Understand the notation  $B(n, p)$ ;
- Identify when a binomial distribution should be used and the conditions needed;
- Calculate probabilities using any standard method including use of calculator;
- Know the calculation for the mean of a binomial distribution is  $np$ ;
- Know the properties for the binomial distribution

### GCSE MATHEMATICS LINKS

No links

### POSSIBLE SUCCESS CRITERIA

Write down one condition so that the binomial distribution is a suitable model.

Calculate binomial probabilities for a given  $n$  and  $p$ .

### COMMON MISCONCEPTIONS

Most mark schemes identified binomial as one of the more challenging questions on the paper. From Higher June 2011 Q14 – “Only the best candidates were able to score full marks in this question. In part (a) most candidates managed to put in 0.9 and 0.2 correctly but few put in 0.05 and 0.95 correctly – values of 0.8 and 0.2 were often seen repeated. A number of candidates showed their lack of understanding on how to complete a tree diagram by putting values in each ‘pair’ that added to 1.

“Part (b), a question on conditional probability was badly done and was clearly beyond the capability of most candidates. An incorrect answer of 0.125 was common. In part (c)(i) few candidates recognised the correct distribution – normal, bimodal or a sampling method were common incorrect answers.

“In part (c)(ii) many candidates were able to give only one of the required reasons, usually the ‘probability of getting the allergy and not getting the allergy’. In part (c)(iii) many candidates identified correctly the need to calculate  $10p^3q^2$ , but most were unable to calculate this correctly. A surprising number of candidates used values of  $p$  and  $q$  which did not total 1, often with values of  $p$  and/or  $q$  each in excess of 1, e.g.  $p = 2$  and  $q = 3$ .”

### NOTES

For extension, students could be introduced to cumulative binomial tables.

Students need to be aware, the properties for the binomial distribution are:

- There are two possible outcomes: fixed chance of success ( $p$ ) and failure ( $1 - p$ );
- There is a fixed number of trials/ events;
- The trials are independent.

### OBJECTIVES

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

- Know the shape of a normal distribution curve and how this occurs;
- Understand the notation  $N(\mu, \sigma^2)$ ;
- Know the conditions that make the normal distribution model suitable;
- Know that 68% of data lies within one standard deviations of the mean, 95% of data lies within two standard deviations of the mean;
- Know how to draw two distribution curves on the same graph;
- Use standardised scores to compare two samples of data.

### GCSE MATHEMATICS LINKS

No links

### POSSIBLE SUCCESS CRITERIA

Work out the mean if given a standardised score and the standard deviation.

### COMMON MISCONCEPTIONS

If drawing two distribution curves on the same axes, the data which is more spread out should be wider and lower than the data which is not as spread.

June 2015 asked to select which variable is most likely to be modelled a normal distribution from gender, year group, number of pieces in the jigsaw, time taken to complete the puzzle and favourite subject. According to the examiner feedback, "Only the most able candidates were able to pick out time taken to be continuous and, hence, the only suitable variable."

Students also need to remember the key values for 1 and 2 standard deviations. From 2014 – "It was clear in part (c) that fewer than one in four candidates knew the  $\pm 2$  standard deviations property of a normal distribution. Some used 1sd or 3sd instead. There were a number left blank."

### NOTES

The formula for standardised scores will not be provided. Mean and standard deviation may be given or may need to be calculated for this type of question.

Students need to learn:

- 95% of the data lies between 2 standard deviations of the data
- 68% of the data lies between 1 standard deviations of the data.

(Slightly different values may be used from different learning resources – the exam uses these values for simplicity.)

Some of the conditions for a normal distribution curve are:

- mean = mode = median
- bell-shaped.

If data is skewed, a normal distribution curve would not be suitable.

For extension, students could be introduced to the normal distribution tables.

### OBJECTIVES

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

- Understand the process of quality assurance and see why this is necessary in the real world;
- Know how to calculate both warning limits and action limits;
- Know how to draw warning limits and action limits on a sample mean, median or range versus sample number graph;
- Understand how action and warning limits are used in the manufacturing process.

### GCSE MATHEMATICS LINKS

No links

### POSSIBLE SUCCESS CRITERIA

Draw action lines and warning lines on a control chart when given the mean and standard deviation.

Determine what action to take on a control chart when given a sample mean.

### COMMON MISCONCEPTIONS

Feedback extract from June 2013 Q15 – “There were many responses in (d) which described what warning and action limits are, rather than how they are used. For example, there were lots of discussions about things such as ‘2 standard deviations above and below the mean’ or ‘they are used to help monitor the process’. Poorly expressed answers generally prevented candidates from achieving all three marks in this part. Quite a few candidates obtained the mark for knowing another sample needed to be taken but many candidates did not clearly explain about the sample being **between** the warning and action limits, rather they said ‘outside warning limits’ which of course could include the region outside the action limits too.”

### NOTES

Students need to:

- Be able to graph sample number with the sample (mean). Students need to know how to add action limits and warning limits to the graph;
- Know that warning limits are set at  $\pm 2$  standard deviations and action limits are set at  $\pm 3$  standard deviations;
- Know that if sample is within warning limits take no action;
- Know that if a sample (mean) is between warning limits and action limit, a further sample is taken;
- Know that if a sample (mean) is outside the warning limits, then stop the process.

