GCSE Statistics

Getting ready to teach event

Nicola Woodford-Smith
July 2023

Pearson
Edexcel

Agenda

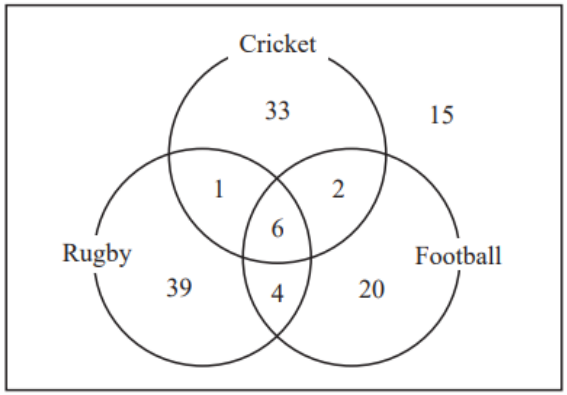
- Why choose GCSE Statistics
- Assessment structure
- Key Resources
- Key Content
- Key Topics for CPD
- Statistical Enquiry Cycle

Maths or Stats?

Stats 2022 1F

10 120 adults were each asked if they enjoy watching any of the sports Football, Cricket or Rugby.

The Venn diagram gives information about the numbers of these adults who enjoy watching these sports.



(a) How many of these adults enjoy watching all three sports?

(1)

One of these adults is chosen at random.

(b) Find the probability that this adult enjoys watching Cricket.

(1)

(c) Find the probability that this adult enjoys watching Rugby or Football.

Maths or Stats?

Stats 2022 1F

14 Weronika works for a road traffic organisation. One day she is investigating the speeds of cars and the speeds of motorcycles along a motorway.

Here is part of the spreadsheet that Weronika used to record her results.

Speed (s miles per hour)	Percentage of cars	Percentage of motorcycles
$30 \leqslant s < 40$	3	0
$40 \leqslant s < 50$	18	two
$50 \leqslant s < 60$	42	7
$60 \leqslant s < 70$	27	124
$70 \leqslant s < 80$	6	56
$80 \leqslant s < 90$		11
Total	**100**	**100**

(a) Give a reason why Weronika will need to clean the data.

(1)

Weronika concludes that the value of 124 in the spreadsheet must be wrong.

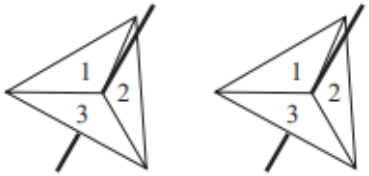
(b) Explain why.

(1)

Maths or Stats?

Maths 2020 3H

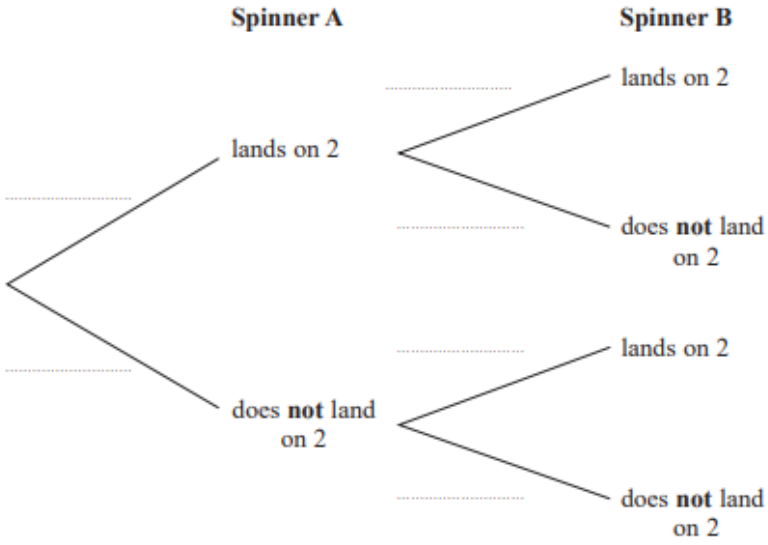
5 Amanda has two fair 3-sided spinners.



Spinner A **Spinner B**

Amanda spins each spinner once.

(a) Complete the probability tree diagram.



(2)

(b) Work out the probability that Spinner A lands on 2 and Spinner B does **not** land on 2

Maths or Stats?

Stats 2022 1H

There are 8 coloured discs in bag **X** and 9 coloured discs in bag **Y**.

In bag **X**, there are 3 red discs and 5 yellow discs.
In bag **Y**, there are 5 red discs and 4 yellow discs.

Sonia takes at random one disc from bag **X** and she takes at random one disc from bag **Y**.

Calculate the probability that the two discs taken by Sonia are the same colour.

Why choose GCSE Statistics?

Why choose GCSE Statistics?

- Performance measures (Bucket 3)
- Straight-forward assessment structure
- Development of transferable skills
- Clear approach and content
- Accessible exam papers
- Comprehensive support

Assessment Structure

Assessment Structure

Paper 1 – 1ST0/1F, 1ST0/1H	Paper 2 – 1ST0/2F, 1ST0/2H
Compulsory paper Externally assessed 50% of the total GCSE 1 hour 30 minutes 80 marks Content: 1. The collection of data 2. Processing, representing and analysing data 3. Probability	Compulsory paper Externally assessed 50% of the total GCSE 1 hour 30 minutes 80 marks Content: 1. The collection of data 2. Processing, representing and analysing data 3. Probability

Assessment Objectives

Assessment Objectives		Weightings
AO1	Demonstrate knowledge and understanding, using appropriate terminology and notation, of standard statistical techniques used to: • collect and represent data • calculate summary statistics and probabilities	55%
AO2	Interpret statistical information and results in context and reason statistically to draw conclusions	25%
AO3	Assess the appropriateness of statistical methodologies and the conclusions drawn through the application of the statistical enquiry cycle	20%

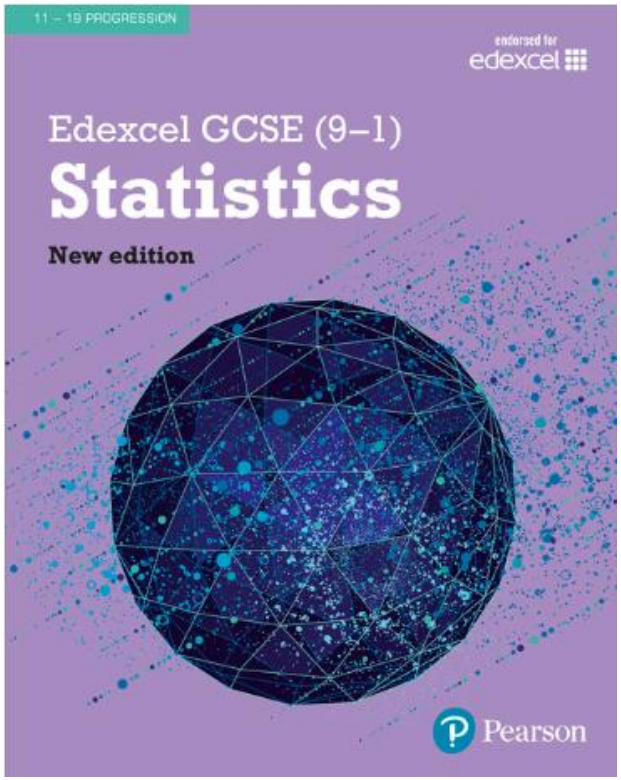
Grade Boundary Analysis

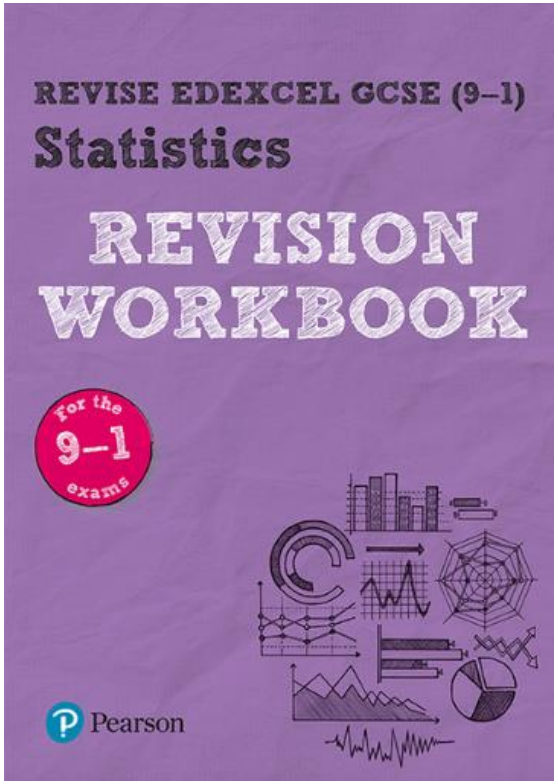
Stats	9	8	7	6	5	4	3	2	1	U
Found					59%	48%	36%	23%	11%	
Higher	76%	66%	55%	43%	31%	19%	13%			
Cum Gra %	**3.7%**	**10.4%**	**20.1%**	**34.3%**	**54.1%**	**73.3%**	**87%**	**95.1%**	**98.5%**	**100%**

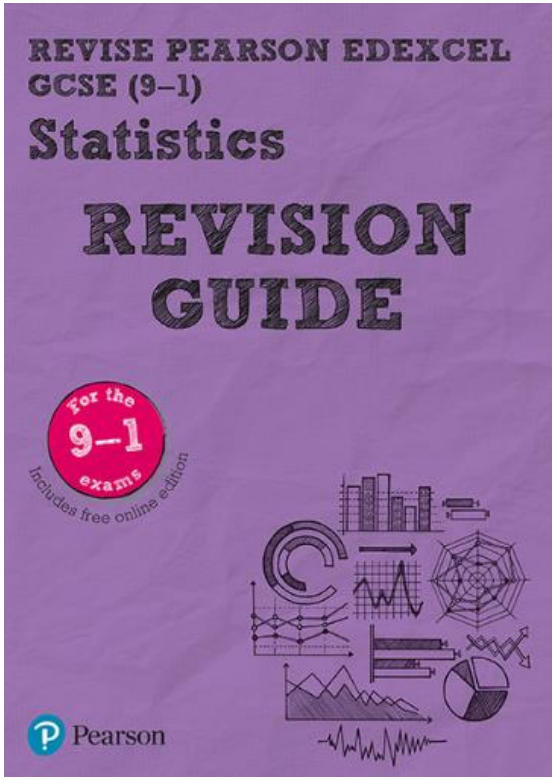
Maths	9	8	7	6	5	4	3	2	1	U
Found					77%	62%	46%	30%	15%	
Higher	83%	70%	57%	45%	33%	22%	16%			
Cum Gra %	**3.6%**	**8.6%**	**16.2%**	**25.5%**	**40.1%**	**60.2%**	**79.9%**	**91.5%**	**97.8%**	**100%**

Key Resources

Published Resources - link







Scheme of work

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 s
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

Mapping document

	A	B	C	D	E
	HIGHER				
	Ref	1MA1 (2015)	1ST0 Spec refs (2017)	Notes	
	5 Probability				
	P1	record, describe and analyse the frequency of outcomes of probability experiments using tables and frequency trees [S]	3p.01 Use collected data to calculate estimates of probabilities. [S]		
	P2	apply ideas of randomness, fairness and equally likely events to calculate expected outcomes of multiple future experiments [S]	3p.01 Use collected data to calculate estimates of probabilities. [S] 3p.03 Use probability values to calculate expected frequency of a specified characteristic within a sample or population. [S] 3p.05 Compare experimental data with theoretical predictions to identify possible bias within the experimental design. [S]		
	P3	relate relative expected frequencies to theoretical probability, using appropriate language and the 0-1 probability scale [S]	3p.01 Use collected data to calculate estimates of probabilities. [S] 3p.02 Compare the probability of different possible outcomes using the 0-1 or 0-100% scale and statements of likelihood. [S]		
	P4	apply the property that the probabilities of an exhaustive set of outcomes sum to one; apply the property that the probabilities of an exhaustive set of mutually exclusive events sum to one [S]	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. [U]	1ST0 guidance column includes reference to exhaustive events.	
	P5	understand that empirical unbiased samples tend towards theoretical probability distributions, with increasing sample size [U]	3p.05 Compare experimental data with theoretical predictions to identify possible bias within the experimental design. [S] 3p.06 Recognise that experimental probability will tend towards theoretical probability as the number of trials increases when all variables are random. [S]		
	P6	enumerate sets and combinations of sets systematically, using tables, grids, Venn diagrams [S] and tree diagrams [U]	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. [U]		
	P7	construct theoretical possibility spaces for single and combined experiments with equally likely outcomes and use these to calculate theoretical probabilities [S]	3p.07 Use … sample space diagrams … to represent all the different outcomes possible for at most three events. [U]		
	P8	calculate the probability of independent and dependent combined events, including using tree diagrams and other representations, and know the underlying assumptions [U]	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. [U] 3p.08 Know and apply the formal notation for independent events. [U]	1ST0 requires general addition law.	
	P9	calculate and interpret conditional probabilities through representation using expected frequencies with two-way tables, tree diagrams and Venn diagrams [B]	3p.07 Use two-way tables, … tree diagrams and Venn diagrams to represent all the different outcomes possible for at most three events. [U] 3p.09 Know and apply the formal notation for conditional probability. [U]		
	6 Statistics				

Exam practice materials

- 2019-2023 Past papers
- 2019-2020/22 Shadow papers
- Sample assessment material
- Specimen papers sets 1 and 2
- Exemplar answers and marking guidance
- GCSE Maths – Stats mapping document
- SEC Teacher guide and questions

- Coming soon…

*NEW Themed papers

Key Content

Foundation Tier

Forming a Statistical Hypothesis to evaluate a problem

- 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.
- 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.
- Determine proactive strategies to mitigate issues that might arise during the statistical enquiry process.

Foundation Tier

Types of Data

- Know and apply the terms explanatory (independent) variables and response (dependent) variables. Including appropriate choice of axes for scatter diagrams

Population and Sampling

- Know that 'population' can have different meanings within a stated context.
- Know reasons for employing judgement sampling or opportunity (convenience) sampling, and the associated risks of bias when these techniques are used.

Foundation Tier

Collecting Data

- Know that data can be collected from different sources: **experimental (laboratory, field and natural), simulation, questionnaires, observation, reference**, census, population and sampling**. Know that sources of secondary data should be acknowledged.**
- Understand random, **systematic, and quota sampling**.
- Determine factors that may lead to bias
- 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.
- Know and demonstrate understanding of techniques used to deal with problems that may arise with collected data.
- Know why data may need to be 'cleaned' before further processing
- Know the importance of identifying and controlling extraneous variable.

Foundation Tier

Processing, representing and analysing data

- Interpret and compare data sets displayed pictorially: population pyramid, choropleth map,
- Recognise where errors in construction lead to graphical misrepresentation, including but not limited to incorrect scales, truncated axis, distorted sizing.

Measures of central tendency

- 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.**
- Determine skewness from data by inspection.
- Interpret a distribution of data in terms of skewness identified from inspection

Foundation Tier

Measures of dispersion

- Calculate different measures of spread: range, quartiles, interquartile range (IQR), percentiles.
- Compare different data sets using appropriate calculated or given measure of spread: range, **interquartile range (IQR), percentiles.**
- Use calculated or given median and **interquartile range (IQR)** to compare data samples and to compare sample data with population data.

Further summary statistics

- 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).

Foundation Tier

Correlation

- Determine line of best fit by eye, by drawing through a calculated double mean point $(\overline{x}, \overline{y})$.
- Interpret given Spearman's rank correlation coefficient in the context of the problem. Calculation not required.

Time series

- Identify trends in data through inspection **and by calculation of 4 point moving averages.** Interpret seasonal and cyclic trends in context.

Foundation Tier

Estimation

- 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.**

Probability

- Use collected data and calculated probabilities to determine and interpret relative risks and absolute risks, and express in terms of expected frequencies in groups.
- Know and apply the formal notation for conditional probability.

Higher Tier

Black text - content already in Foundation Tier GCSE Mathematics

Red Text - Higher Tier GCSE Mathematics content

Higher - Probability	
P9	calculate and interpret conditional probabilities through representation using expected frequencies with two-way tables, tree diagrams and Venn diagrams

Higher - Statistics	
S3	construct and interpret diagrams for grouped discrete data and continuous data, i.e. histograms with equal and unequal class intervals and cumulative frequency graphs, and know their appropriate use
S4	interpret, analyse and compare the distributions of data sets from univariate empirical distributions through: • appropriate graphical representation involving discrete, continuous and grouped data, including box plots • appropriate measures of central tendency (median, mean, mode and modal class) and spread (range, including consideration of outliers , quartiles and inter-quartile range)

Formula Sheet

Higher Tier Formulae

You must not write on this page.

Anything you write on this page will gain NO credit.

$$\text{Skew} = \frac{3(\text{mean} - \text{median})}{\text{standard deviation}}$$

$$\text{Standard deviation} = \sqrt{\frac{1}{n}\sum(x - \bar{x})^2}$$

An alternative formula for standard deviation is

$$\text{standard deviation} = \sqrt{\frac{\sum x^2}{n} - \left(\frac{\sum x}{n}\right)^2}$$

Spearman's rank correlation coefficient

$$r_s = 1 - \frac{6\sum d^2}{n(n^2 - 1)}$$

$$\text{Rates of change (e.g. Crude birth rate} = \frac{\text{number of births} \times 1000}{\text{total population}})$$

Higher Tier

Types of Data

- 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.**
- Know the difference between primary and secondary data.
- Consideration of advantages/disadvantages and access issues.

Higher Tier

Population and Sampling

- Understand systematic, and quota sampling.

Processing, representing and analysing data

- population pyramid, choropleth map, comparative pie chart, comparative 3D representations
- Recognise where errors in construction lead to graphical misrepresentation,
- Determine skewness from data by inspection. and by calculation. Use of: Skew = 3(mean - median)/standard deviation. Formula will be given in the formulae sheet.

Higher Tier

Measures of central tendency

- Weighted mean, geometric mean, mean seasonal variation.

Measures of dispersion

- interpercentile range, interdecile range and standard deviation. (For standard deviation only the formulae for a set of values are given. Students will need to know how to apply these to grouped data.)
- Identify outliers by inspection **and using appropriate calculations.**
- Use calculated or given means and standard deviation to standardise and interpret data collected in two comparable samples. Formulae for SD will be given in the formulae sheet.

Higher Tier

Further summary statistics

- 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).
- 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.

Higher Tier

Correlation

- Apply formula to determine Spearman's rank correlation coefficient and interpret it in context. Values found using calculator functions will be permissible. Formula is given.
- Interpret given Pearson's product moment correlation coefficient (PMCC) in the context of the problem. The calculation of PMCC will not be required.

Higher Tier

Time series

- Identify trends in data through inspection and by calculation of 4 or other determined appropriate point moving averages.
- Interpret seasonal and cyclic trends in context. Use such trends to make predictions.

Quality assurance

- Know that a set of sample means are more closely distributed than individual values from the same population.
- Use action and warning lines in quality assurance sampling applications.

Higher Tier

Probability

- Use collected data and calculated probabilities to determine and interpret relative risks and absolute risks, and express in terms of expected frequencies in groups.
- Know and interpret the characteristics of a binomial distribution.
- Know and interpret the characteristics of a normal distribution.
- 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.

Key Topics CPD

CPD Topics

- Chain Based Index Numbers
- Geometric Mean
- Binomial Probabilities
- Absolute and Relative Risk
- Standardised Scores
- Conditional Probabilities
- Skewness
- Calculating the SRCC (& Interpreting the PMCC)
- Standard Deviation & Normal Distribution
- Action and Warning Lines
- Rates of Change

Chain Based Index Numbers

Chain Based Index Numbers

The table gives information about the change in the cost of a second class stamp from 2011 to 2017

Year	2011	2013	2015	2017
Cost of second class stamp (pence)	36	50	54	56
Chain base index number		138.9	108	

(Source: *www.2ndclassstamp.co.uk*)

By working out the geometric mean of 3 appropriate chain base index numbers, what can be deduced about the average two-yearly change in the cost of a second class stamp from 2011 to 2017?

Question	Answer	Additional guidance	Mark
	M1 $56/54 \times 100$ A1 103.7 M1 for $\sqrt[3]{138.9 \times 108 \times '103.7'}$ Al ft 115.9 B1 ft for (average) **rate** of 'increase' per **two years** B1 ft ... is '15.9'%	M1 for correct calculation of chain base index number A1 for awrt 103.7 M1 ft for correct calculation of the geometric mean of 3 chain base index numbers A1ft for awrt 115.9 B1 ft for correct contextual interpretation as rate of increasing per two years B1 ft for correct contextual interpretation of the value for their geometric mean	(6)

= 6 marks)

Chain Based Index Numbers

The table gives information about the monthly average price per litre, in pence, of diesel over a period of five months.

The table also gives some of the chain base index numbers, correct to one decimal place, for this information.

	May	June	July	August	September
Monthly average price (p)	109.1	111.8	112.7	111.2	113.1
Chain base index number		102.5	100.8		

(Source: *theaa.com*)

(a) Calculate the chain base index numbers for August and September and write them in the table.

Give each value correct to one decimal place.

(2)

(b) (i) Calculate the geometric mean of the four chain base index numbers.

You must show your working.
Give your answer correct to one decimal place.

(2)

(ii) Interpret your answer.

(2)

Binomial Probabilities

Binomial Probabilities

Jasper has 3 coins.

In an experiment, Jasper flips each of the 3 coins and records the total number of heads that he gets.

Jasper believes that each coin is biased so that the number of heads he gets can be modelled by the binomial distribution, B(3, 0.4).

(a) Show that P(0 heads) = 0.216

(1)

(b) Work out the probability that the outcome of the experiment is exactly 1 head.

(2)

Question number	Answer	Additional guidance	Mark
(a)	B1 0.6^3 [= 0.216]	B1 for a correct calculation of the given probability Allow ${}^3C\ (0.4)\ (0.6)_0{}^{0\ 3}$	(1)
(b)	M1 $3 \times 0.4 \times 0.6^2$ A1 0.432 oe	M1 for a correct expression (allow ${}^3C_1$ for 3) A1 for awrt 0.43	(2)

Binomial Probabilities

Jasper carries out the experiment 100 times.

The table shows information about the number of heads he gets for each experiment.

Number of heads	Frequency
0	24
1	41
2	30
3	5

(c) Determine whether or not the model B(3, 0.4) is suitable for Jasper's experiment.

(5)

(Total for question = 8 marks)

Binomial Probabilities

(c)	M1 $3 \times 0.4^2 \times 0.6$ (=0.288) or 0.4^3 (=0.064) A1 both awrt 0.29 and awrt 0.06 M1A1ft depB1ft 'B(3, 0.4) is a suitable model/Jasper is correct'	M1 for calculating either probability of obtaining 2 heads or 3 heads A1 for both probabilities calculated (awrt 0.29 and awrt 0.06) M1 for converting observed values and expected values into the same units (may be implied by at least one correct or correct ft) i.e. multiplying binomial probabilities by 100 or for converting observed frequencies into probabilities/percentages A1ft for all 4 expected frequencies correct or correct ft awrt 22, '43', '29' and '6' or 0.24, 0.41, 0.30, and 0.05 depB1ft (dependent upon both M marks) correct conclusion or correct ft conclusion SC: $\frac{24\times0+41\times1+30\times2+5\times3}{300}=\frac{29}{75}$ (= awrt 0.39) or $\frac{24\times0+41\times1+30\times2+5\times3}{100}=1.16$ and $3\times0.4=1.2$ scores M0A0M1A1B0	(5)

Expected frequencies	Observed frequencies
$100 \times 0.216 = 21.6$	[24]
$100 \times 0.432 = 43.2$	[41]
$100 \times 0.288 = 28.8$	[30]
$100 \times 0.064 = 6.4$	[5]


Risk

Risk

9 Katrina travels to work by train or by bus or by car.

The table gives some information about her 200 journeys to work last year.

Travel by	Number of journeys	Number of times late for work
Train	120	27
Bus	30	x
Car	50	15

One of the days that Katrina travelled to work last year is picked at random.

(a) Find the probability that she travelled by train and was late for work on that day.

.....

(1)

The absolute risk of Katrina arriving late for work last year when travelling by bus was 0.6

(b) Show why the value of x in the table is 18

Risk

(c) (i) Show that the relative risk of Katrina being late for work last year when she travelled by car compared with when she travelled by bus is 0.5

(2)

(ii) Interpret this relative risk.

.....

.....

(1)

Risk

Question number	Answer	Additional guidance	Mark
9 (a)	B1 $\frac{27}{200}$ (= 0.135)	B1 for exact equivalent fraction, decimal or percentage	(1)
(b)	B1 $\frac{x \text{ (or } 18)}{30} = 0.6$ or 0.6×30 (=18) or $\frac{18}{0.6} = 30$	B1 for correct use of absolute risk in a calculation. (NB Answer 18 is given)	(1)
(c)(i)	M1A1 $\frac{15}{50} \div 0.6$ (= 0.5) or $0.3 \div 0.6$ (= 0.5)	M1 for a probability $\div 0.6$ (or $\div \frac{18}{50}$) A1 for fully correct calculation (may be seen in stages) NB Answer 0.5 is given and need not be stated, or may be embedded. Accept e.g. $0.5 \times 0.6 = 0.3$ for M1A1	(2)
(c)(ii)	B1 e.g. Lateness is half as likely by car (than by bus) or Lateness is twice as likely by bus (than by car)	B1 for a correct contextual interpretation of relative risk	(1)

Standardised Scores

Standardised scores

Kirstin took tests in Maths, in Physics and in French.

The table shows information about the marks of all students who took the tests.

	Maths	Physics	French
Mean	53	69	48
Standard deviation	8	10	6

Kirstin scored 63 marks in Maths.

(a) Show that Kirstin's standardised score in Maths is 1.25

(1)

Kirstin scored 78 marks in Physics.

(b) Work out whether Kirstin did better in Maths or in Physics.
You must explain your answer.

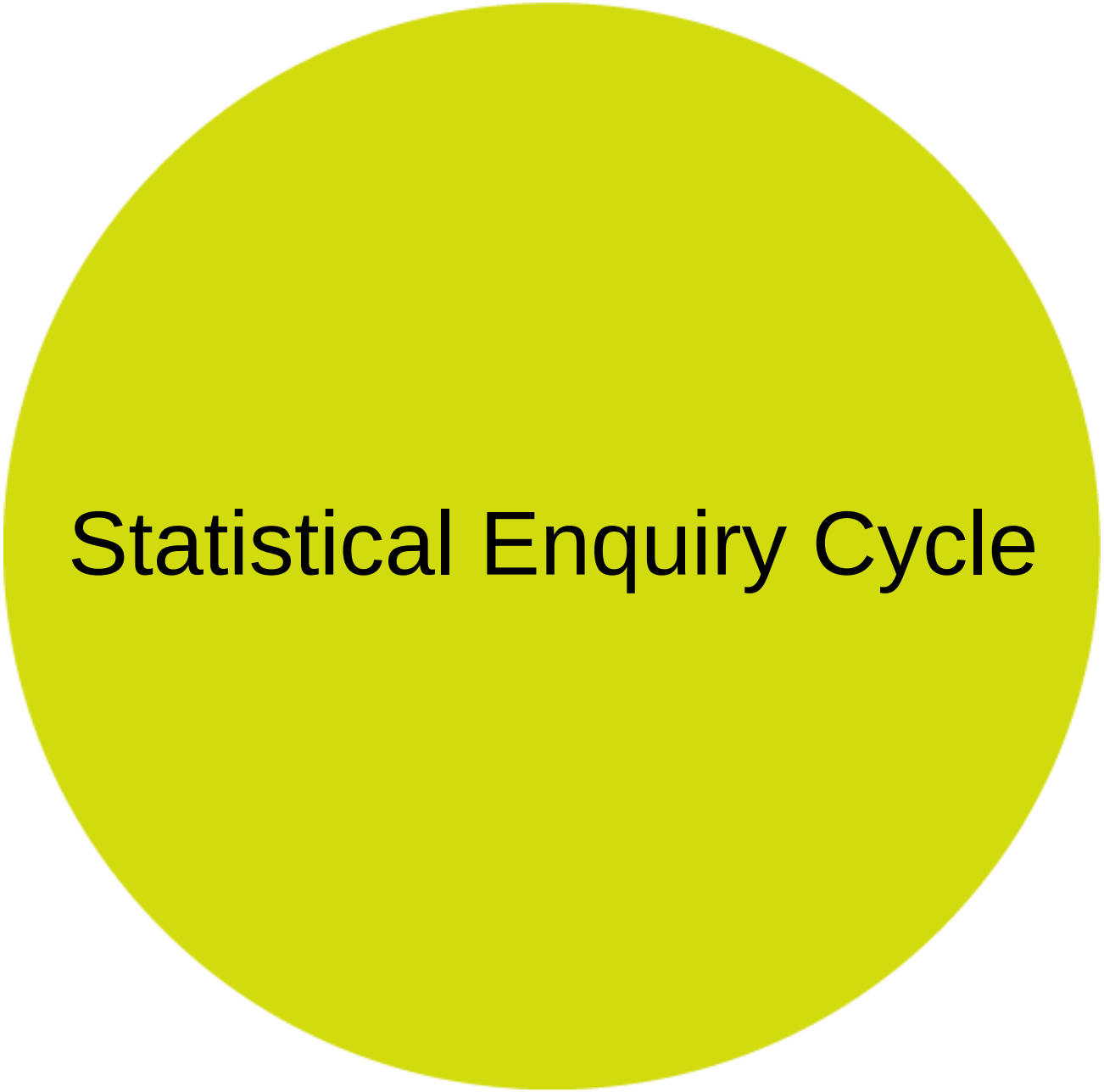
(3)

Kirstin's standardised score in French was –0.5

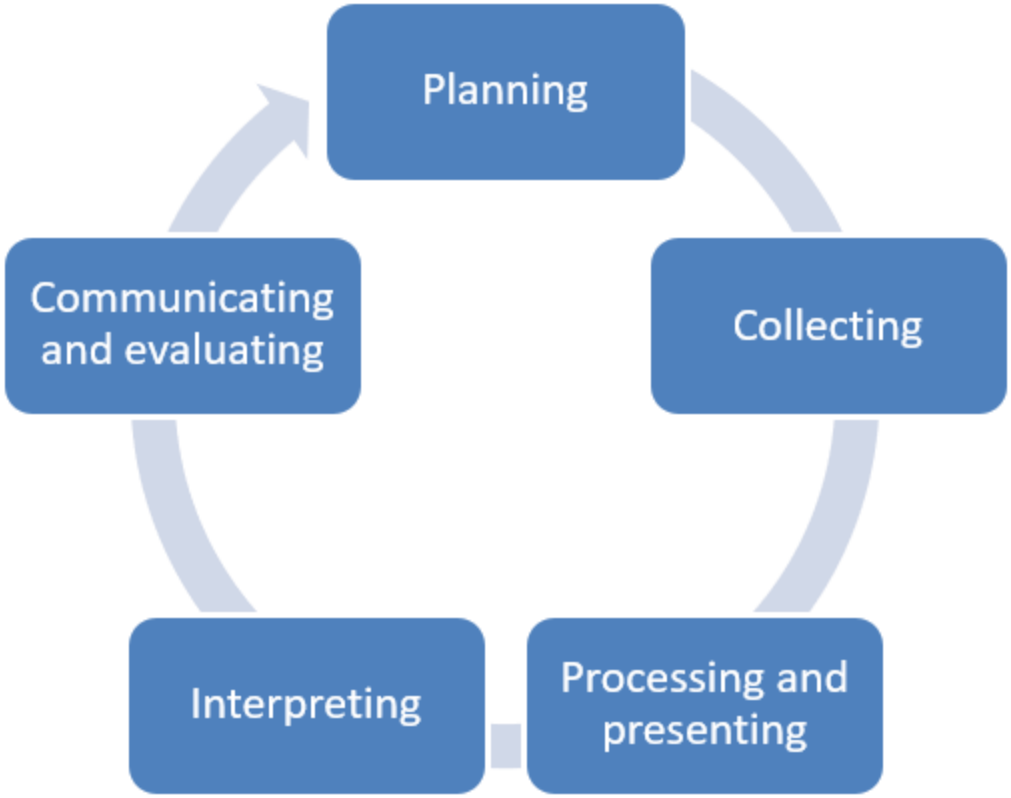
(c) Work out Kirstin's mark in French.

(2)

(Total for question = 6 marks)


Statistical Enquiry Cycle

The Concept


Planning
Collecting
Processing and presenting
Interpreting
Communicating and evaluating

The Concept

The papers will assess the key stages of the statistical enquiry cycle:

- Importance of initial planning
- Constraints in sourcing data; including effects of rounding
- Ways data can be processed and presented, including use of technology
- Results interpreted in the context of problems
- Clear, concise communication of key findings including
    - Importance of awareness of target audience
    - Importance of evaluating statistical work

Example Question (F)

3 Tina is investigating the performance of athletes at the Paralympics.
She has collected data for the distances thrown, in metres, by male and female athletes competing in F34 Javelin.

Male	14.24	15.00	15.29	16.96	17.88	18.16	18.42	21.86
Female	13.84	27.70	27.83	30.36	31.31	33.42	34.15	36.65

Tina wants to compare the average distance thrown by male athletes with the average distance thrown by female athletes.

Tina thinks that she should use the mean to compare the average distance thrown.

(a) Give one reason why the mean would not be an appropriate average to use to represent the average distance thrown by female athletes. (1)

Tina plans to use a scatter diagram to compare the distances thrown by the male athletes with the distances thrown by the female athletes.

(b) Discuss whether or not a scatter diagram would be a suitable diagram to use. (2)

(Total for Question 3 is 3 marks)

edexcel

Example Question (F)

	sample		
3(a)	B1 for a correct reason, e.g. • reference to the extreme value in the female distances thrown • reference to the mean being affected by extreme values	B1 for assessing the appropriateness of the mean as a choice of average	(1)
3(b)	B2 Not suitable as the data is not bivariate / in related pairs oe OR if B2 not earned... B1 Not suitable with an attempt at a reason	B2 for a complete answer assessing the appropriateness of the choice of diagram OR if B2 not earned... B1 for an incomplete answer assessing the appropriateness of the choice of diagram	(2)

Pearson

Example Question (H)

1* Mr Jones is investigating the amount of time that students at his school spend on their homework each week.

He writes the following hypothesis for his investigation:

"Older st
than you

Mr Jones decic
He is going to
on their home

Mr Jones then
complete an o

Some of the d

	Age	Time spent on homework in the last week
1	14	50 min
2	15 years old	10
3	11 ½	90 minutes
4	13	Don't know
5	13years 4months	3 hours
6	12	40
7	14y 7m	2.5 hours

(a) Give two reasons why Mr Jones must clean the data before processing it. (2)

(b) Discuss how Mr Jones' data collection plan could affect the reliability of his conclusions. (2)

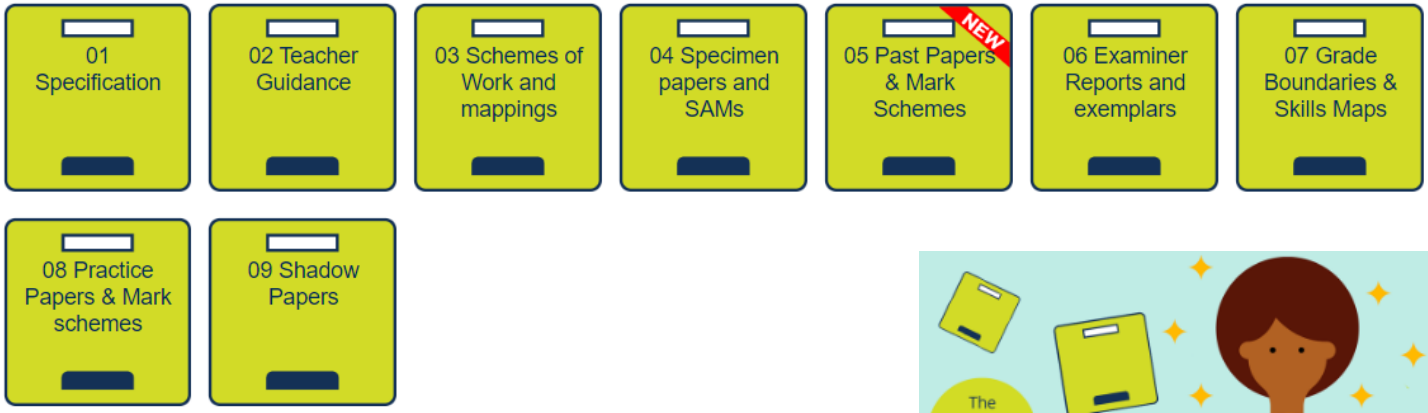
(Total for Question 1 is 4 marks)

Example Question (H)

Question number	Answer	Additional guidance	Mark
1(a)*	B1B1 for two correct reasons, e.g. • ages given in different formats (nearest year / year and month, minutes / hours) • remove extraneous symbols • remove missing data ('don't know')	B1 for each correct reason for the need to clean data on the database prior to processing it	(2)
1(b)*	B1B1 for two correct aspects, e.g. • large sample size increases reliability • issues with consistency in recording time taken for homework • issues with how the data collection is carried out (may be errors in data entry) • potential for students to lie about how much time they spend on homework • non-response decreases reliability	B1B1 for two correct comparisons assessing the reliability of the conclusions drawn	(2)

Website and more resources

- GCSE Statistics website
- The Maths Emporium
- Exam Wizard



Find out more

For more courses see our Pearson Professional Development Academy.

Pearson