



Mark Scheme

Specimen Papers Set 2

Pearson Edexcel GCSE (9 – 1)
In Statistics (1ST0)
Higher (Calculator) Paper 2H

Edexcel and BTEC Qualifications

Edexcel and BTEC qualifications are awarded by Pearson, the UK's largest awarding body. We provide a wide range of qualifications including academic, vocational, occupational and specific programmes for employers. For further information visit our qualifications websites at www.edexcel.com or www.btec.co.uk. Alternatively, you can get in touch with us using the details on our contact us page at www.edexcel.com/contactus.

Pearson: helping people progress, everywhere

Pearson aspires to be the world's leading learning company. Our aim is to help everyone progress in their lives through education. We believe in every kind of learning, for all kinds of people, wherever they are in the world. We've been involved in education for over 150 years, and by working across 70 countries, in 100 languages, we have built an international reputation for our commitment to high standards and raising achievement through innovation in education. Find out more about how we can help you and your students at: www.pearson.com/uk

General marking guidance

These notes offer general guidance, but the specific notes for examiners appertaining to individual questions take precedence.

- 1** All candidates must receive the same treatment. Examiners must mark the last candidate in exactly the same way as they mark the first.

Where some judgement is required, mark schemes will provide the principles by which marks will be awarded; exemplification/indicative content will not be exhaustive. When examiners are in doubt regarding the application of the mark scheme to a candidate's response, the response should be sent to review.

- 2** All the marks on the mark scheme are designed to be awarded; mark schemes should be applied positively. Examiners should also be prepared to award zero marks if the candidate's response is not worthy of credit according to the mark scheme. If there is a wrong answer (or no answer) indicated on the answer line always check the working in the body of the script (and on any diagrams), and award any marks appropriate from the mark scheme.

Questions where working is not required: In general, the correct answer should be given full marks.

Questions that specifically require working: In general, candidates who do not show working on this type of question will get no marks – full details will be given in the mark scheme for each individual question.

- 3** **Crossed out work**

This should be marked **unless** the candidate has replaced it with an alternative response.

- 4** **Choice of method**

If there is a choice of methods shown, mark the method that leads to the answer given on the answer line.

If no answer appears on the answer line then mark both methods **as far as they are identical** and award these marks.

- 5** **Incorrect method**

If it is clear from the working that the "correct" answer has been obtained from incorrect working, award 0 marks.

6 Follow through marks

Follow through marks which involve a single stage calculation can be awarded without working as you can check the answer, but if ambiguous do not award.

Follow through marks which involve more than one stage of calculation can only be awarded on sight of the relevant working, even if it appears obvious that there is only one way you could get the answer given.

7 Ignoring subsequent work

It is appropriate to ignore subsequent work when the additional work does not change the answer in a way that is inappropriate for the question or its context. (eg an incorrectly cancelled fraction when the unsimplified fraction would gain full marks).

It is not appropriate to ignore subsequent work when the additional work essentially makes the answer incorrect (eg incorrect algebraic simplification).

8 Probability

Probability answers must be given as a fraction, percentage or decimal. If a candidate gives a decimal equivalent to a probability, this should be written to at least 2 decimal places (unless tenths).

Incorrect notation should lose the accuracy marks, but be awarded any implied method marks.

If a probability fraction is given then cancelled incorrectly, ignore the incorrectly cancelled answer.

9 Range of answers

Unless otherwise stated, when an answer is given as a range (eg 3.5 – 4.2) then this is inclusive of the end points (eg 3.5, 4.2) and all numbers within the range.

Guidance on the use of abbreviations within this mark scheme

M	method mark awarded for a correct method or partial method
A	accuracy mark (awarded after a correct method; if no method or process is seen then full marks for the question are implied but see individual mark schemes for more details)
B	unconditional accuracy mark (no method needed)
oe	or equivalent
cao	correct answer only
ft	follow through (when appropriate as per mark scheme)
sc	special case
dep	dependent (on a previous mark)
indep	independent
awrt	answer which rounds to
isw	ignore subsequent working

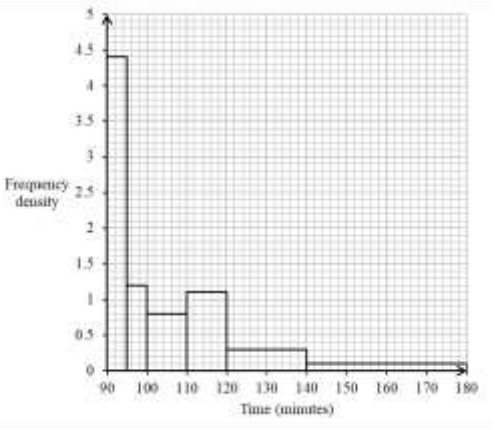
Question	Answer	Additional guidance	Mark
1	B1 Qu 1 is leading/biased B1 Qu 2 does not contain a time frame B1 Qu 3 is a sensitive question (people may not want to say why they were in hospital) OR there are lots of missing options so lots of people will tick other	B3 for a complete assessment that the questions are not appropriate and the reason why for each question (B2 for an incomplete response with 2 out of 3 bullet points correct B1 for an incomplete response with 1 out of 3 bullet points correct)	(3)

Question	Answer	Additional guidance	Mark
2 (a)	B1 89, 220, 303, 365, 382	B1 for all correct cumulative frequencies	(1)
(b)	<p>B1 for correct horizontal plots B1 ft for correct vertical plots B1 ft for correct cumulative frequency graph allow with straight lines or curve</p>	<p>B1 for correct horizontal plots B1 ft for correct vertical plots ($\frac{1}{2}$ square tolerance for plots)</p> <p>SC if B0 B0 then six correct points out of seven is B1</p> <p>B1 ft for correct cumulative frequency graph allow with straight lines or curve must be increasing curve for the ft</p>	(3)
(c)	B2 ft Hamish's conclusion is correct because the train median is higher than the median of the car which is 4.3	<p>B2 ft for correct conclusion from their graph with supporting figure for median Allow median in the range 4 to 5 or ft their graph providing their graph is increasing (B1 ft for median in range 4 to 5 or correct conclusion with incorrect supporting figure)</p>	(2)
(d)	B1 A cumulative frequency step polygon is more appropriate because data is discrete	B1 for any correct description why a cumulative frequency step polygon is more appropriate	(1)
(e)	<p>B1 Quota B1 Quota sampling is useful when</p> <ul style="list-style-type: none"> time is limited a sampling frame is not available the research budget is tight when detailed accuracy or randomness is not important 	<p>B1 for quota B1 for a correct explanation when it is appropriate to use quota sampling eg one of the bullet points</p>	(2)

Question	Answer	Additional guidance	Mark
3 (a)	B1 eg 'allows two data sets to be compared easily'	B1 for a suitable reason	(1)
(b)	B1 $a = 53$ B1 $b = 43$ B1 $c = 62$	B1 for each correct value found	(3)
(c)	B1 eg ' $\frac{1}{2}$ as sample median is likely to be the same as the population median since it is a random sample'	B1 for $\frac{1}{2}$ and correct supporting reason (accept $\frac{12}{25}$ from the stem and leaf diagram)	(1)
(d)	B1 ft Canada IQR = 17 and UK IQR = 19 or Canada range = 43 and UK range = 44 B1 ft IQR/range in UK is greater than IQR/range in Canada B1 There is a greater spread of ages in the UK parliament	B1 for identifying both IQRs or ranges (allow ft from part (b)) B1 for a correct comparison of measure of spread (allow ft from part (b)) B1 for a correct conclusion in context	(3)
(e)	B1 Any one from • used to select sample • used to identify the population	B1 for a correct use of a sample frame	(1)
(f)	B2 eg 'not a suitable sample frame since it does not include all members of the population'	B2 for assessing the suitability of the sampling frame with supporting reason (B1 for assessing the suitability of the sampling frame with incomplete reasoning)	(2)

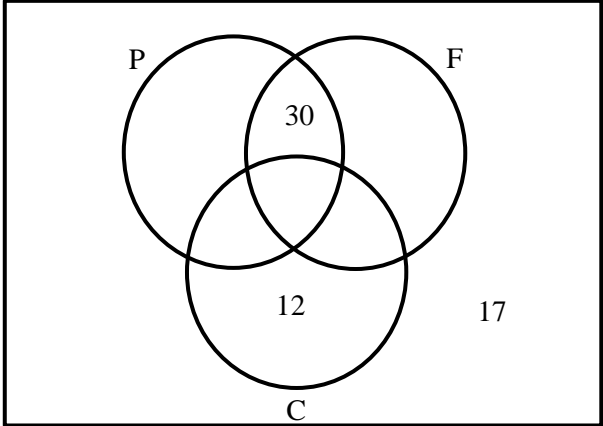
Question	Answer	Additional guidance	Mark																																				
4 (a)	<table border="1"> <thead> <tr> <th>Country</th><th>GDP rank</th><th>d</th><th>Σd^2</th></tr> </thead> <tbody> <tr><td>Norway</td><td>2</td><td>-1</td><td>1</td></tr> <tr><td>Denmark</td><td>4</td><td>-2</td><td>4</td></tr> <tr><td>Iceland</td><td>3</td><td>0</td><td>0</td></tr> <tr><td>Switzerland</td><td>1</td><td>3</td><td>9</td></tr> <tr><td>Finland</td><td>6</td><td>-1</td><td>1</td></tr> <tr><td>Netherlands</td><td>5</td><td>1</td><td>1</td></tr> <tr><td>Canada</td><td>7</td><td>0</td><td>0</td></tr> <tr><td>New Zealand</td><td>8</td><td>0</td><td>0</td></tr> </tbody> </table> <p>B1 GDP rank 2, 4, 3, 1, 6, 5, 7, 8</p> <p>M1 differences -1, -2, 0, 3, -1, 1, 0, 0</p> <p>M1 $1 - \frac{6 \times 16}{8 \times (8^2 - 1)}$</p> <p>A1 0.81</p>	Country	GDP rank	d	Σd^2	Norway	2	-1	1	Denmark	4	-2	4	Iceland	3	0	0	Switzerland	1	3	9	Finland	6	-1	1	Netherlands	5	1	1	Canada	7	0	0	New Zealand	8	0	0	<p>B1 for GDP rank correct (allow reversed ranks 7, 5, 6, 8, 3, 4, 2, 1)</p> <p>M1 for difference in ranks (condone one slip and allow \pm). Can be implied by $\Sigma d^2 = 16$ (or $\Sigma d^2 = 152$ for reversed ranks)</p> <p>M1 for demonstrating correct use of Spearman's formula</p> <p>A1 for awrt 0.81 (allow awrt - 0.81 for reversed ranks)</p>	(4)
Country	GDP rank	d	Σd^2																																				
Norway	2	-1	1																																				
Denmark	4	-2	4																																				
Iceland	3	0	0																																				
Switzerland	1	3	9																																				
Finland	6	-1	1																																				
Netherlands	5	1	1																																				
Canada	7	0	0																																				
New Zealand	8	0	0																																				
(b)	<p>B1 Positive correlation</p> <p>B1 The wealthier the country the happier the country (for the sampled people)</p>	<p>B1 for correct description of correlation</p> <p>B1 for correct interpretation of the correlation</p>	(2)																																				
(c)	B1 The result is outside the range $-1 \leq r \leq 1$	B1 for correct interpretation of the correlation coefficient, allow the result is bigger than 1	(1)																																				
(d)	<p>B1 both values show negative correlation</p> <p>B1 -0.8 shows stronger correlation than -0.5</p>	B1 for each correct comparison	(2)																																				

Question	Answer	Additional guidance	Mark
5 (a)	B1 (Use the summing feature of the spreadsheet to) add up each column to check it equals 100%	B1 for correct explanation how to use the spreadsheet to check if data needs to be cleaned	(1)
(b)	B1 for 176 circled and answer 17.6		(1)
(c)	B2 Ben's conclusion is not a good one because eg <ul style="list-style-type: none"> although this flavour is clearly dropping in popularity it is still one of the top flavours for percentage of total sales it would be better to drop Pickled Onion which is showing a drop in percentage of total sales and has the lowest percentage. 	B2 for complete assessment of validity of Ben's conclusion with a decision that it is not good and including an appropriate reason (B1 for a decision that it is not good with an incomplete reason)	(2)
(d)	B3 The chart should have <ul style="list-style-type: none"> a line showing the target weight at 510 g Lines showing warning limits at 550 g and 470 g Lines showing action limits at 570 g and 450 g The sample means (495 g and 558 g) are plotted B2 <ul style="list-style-type: none"> 495 g is inside the warning limits no action is required 558 g is between the warning limits and action limits another sample should be taken 	B5 for a complete explanation including the key values explaining how the quality control chart is drawn and used (Award B1 for each correct bullet point to a maximum of B3 for the first four bullet points)	(5)
(e)(i)	B1 approximately equal to		(1)
(e)(ii)	B1 greater than		(1)

Question	Answer	Additional guidance	Mark
6 (a)	B2 Histogram with unequal class widths is best choice since data is continuous and unequal class widths will reduce the number of rectangles where frequency densities are low.	B2 for a decision of histogram with both supporting reasons (B1 for decision of histogram with just one supporting reason)	(2)
(b)	<p>M1 freq density = freq/cw $22 \div 5$ etc. A1 fd 4.4, 1.2, 0.8, 1.1, 0.3, 0.1</p> <p>A2</p> 	<p>M1 for attempt at a calculation of freq density = freq/cw implied by any correct fd or bar height 4.4, 1.2, 0.8, 1.1, 0.3, 0.1 (allow multiples of these) A1 all correct fd may be implied by graph</p> <p>A2 for correct histogram – tolerance $\frac{1}{2}$ square (A1 for at least 3 bars correct width and height)</p>	(4)
(c)	<p>M1 $11 + 6$ competitors between 110 and 140 minutes (area of bars four and five) A1 $\frac{1}{2}$ of the third bar/class is between 105 and 110 minutes $\frac{1}{4}$ of the last bar/class is between 140 and 150 minutes so $\frac{1}{2} \times 8 + 11 + 6 + \frac{1}{4} \times 4 = 22$</p>	<p>M1 for showing working and explanation for adding bars/classes four and five A1 for full complete explanation including working out the fraction of the bars/classes and adding all the values</p>	(2)

Question	Answer	Additional guidance	Mark
7 (a)	B1 name		(1)
(b)	B2 It is not an appropriate way to stratify because you should not stratify by the variable you are investigating	B2 for decision of not appropriate with a correct assessment as to why it is not appropriate (B1 for decision of not appropriate with any reason)	(2)
(c)	B2 The spread of values below the median is greater than the spread of values above the median because the skew is negative	B2 for complete correct reasoning and conclusion comparing spread of values above and below the median (B1 for an incomplete reason eg not mentioning the skewness)	(2)

Question	Answer	Additional guidance	Mark
8 (a)	<p>Problem 1 B3</p> <ul style="list-style-type: none"> binomial appropriate as <ul style="list-style-type: none"> fixed number of trials two possible outcomes – getting yellow/not getting yellow trials are independent $n = 3$ and $p = 0.1$ oe assume probability of yellow is constant <p>Problem 2 B3</p> <ul style="list-style-type: none"> binomial may not appropriate as <ul style="list-style-type: none"> although there is a fixed number of trials and two possible outcomes – five-set /not five-set trials are not independent, the people he is playing change (people in early matches are likely to be easier to beat in 3 sets than people in later matches) probability is unlikely to remain constant his form will not be constant <p>OR SC B2</p> <ul style="list-style-type: none"> binomial may be appropriate as <ul style="list-style-type: none"> fixed number of trials two possible outcomes – five-set match/not five-set match $n = 8$ and $p = \frac{6}{64}$ oe 	<p>B3 for fully correct assessment of suitability and assumptions with all 6 bullet points correct</p> <p>(B2 for partially correct answer, 4 bullet points correct B1 for partially correct answer, 2 bullet points correct)</p> <p>B3 for fully correct assessment of suitability and assumptions with all 5 bullet points correct</p> <p>(B2 for partially correct answer, 2 bullet points correct B1 for partially correct answer, 1 bullet point correct)</p> <p>B2 for complete assessment of suitability and assumptions with all 4 bullet points correct</p> <p>(B1 for partially correct answer, 1 bullet point correct)</p>	(6)
(b)	<p>M1 0.93^6 or $6 \times 0.93^5 \times 0.07$ M1 $1 - (0.93^6 + 6 \times 0.93^5 \times 0.07)$ A1 0.061</p>	<p>M1 for one correct probability M1 for a complete method A1 for awrt 0.061</p>	(3)

Question	Answer	Additional guidance	Mark
9 (a)	B1 3 people played the piano and played the clarinet but did not play the flute	B1 for correct interpretation of the region 3 is in	(1)
(b)	B2 	B2 for all three correct values (B1 for one correct)	(2)
(c)	M1 $3 + 7 + 9 + '30'$ or 49 A1 ft $\frac{7}{49}$ oe	M1 for addition of regions containing people that play at least two instruments A1 ft their Venn diagram	(2)

Question	Answer	Additional guidance	Mark
10	M1 $56/54 \times 100$ A1 103.7 M1 for $\sqrt[3]{138.9 \times 108 \times '103.7'}$ A1 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)

Question	Answer	Additional guidance	Mark
11 (a)	B2 Debbie is correct because <ul style="list-style-type: none"> 63 is the only outlier at the higher end as it is the only value larger than 60 the limit for lower outliers is negative and you can't have negative time so there will be no values lower than this therefore no outliers at the lower end 	B2 for complete correct explanation interpreting $LQ - 1.5 \times IQR$ and $UQ + 1.5 \times IQR$ including a decision and both bullet points (B1 for an incomplete explanation eg just one bullet point)	(2)
(b)	M1 $\frac{20+4}{1.5} [= IQR]$ oe A1 36	M1 implied by $IQR = 16$ or for full working eg $60 - \frac{20+4}{1.5} \times 1.5$, $60 - 24$	(2)

