Mark Scheme

Specimen Papers Set 1

Pearson Edexcel GCSE (9 – 1)
In Statistics (1ST0)
Foundation (Calculator) Paper 2F

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

All the material in this publication is copyright
© Pearson Education Ltd 2018

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

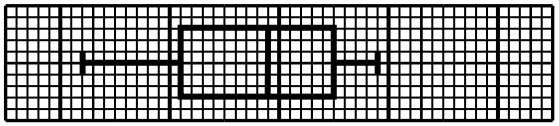
Question	Answer	Additional guidance	Mark
1 (a)	B1 for impossible		**(1)**
(b)	B1 for evens		**(1)**
(c)	B1 for a cross marked at $\frac{1}{4}$	B1 for a cross indicated between $\frac{1}{6}$ and $\frac{2}{6}$ inclusive	**(1)**
(d)	B1 for a cross marked at $\frac{3}{4}$	B1 for a cross indicated between $\frac{4}{6}$ and $\frac{5}{6}$ inclusive	**(1)**
2 (a)	B1B1B1 for three correct things identified eg • No title • No key • No label for one of the columns • 3D • Difficult to read • No label for the horizontal axis • Shading	B1 ×3 for demonstrating understanding of key features of a dual bar chart	**(3)**
(b)	M1 3 5 7 7 8 10 11 11 11 12 15 A1 10	M1 for ordering or using $(n + 1)/2$ A1 cao	**(2)**

3	**(a)**	M1 $\frac{8+8+9+9+6+10+10+4+10+10}{10}$ A1 8.4	M1 for appropriate addition and division by 10 (condone one error) A1 cao	**(2)**
	(b)	M1 $10-4$ A1 6	A1 cao	**(2)**
	(c)	B1 ft • Rohan has a greater mean than Ananya B1 ft • Rohan has a greater range than Ananya B1 ft eg • Rohan had longer words to spell • Rohan had words with less consistent lengths to spell	B1B1 for correct statistical reasoning comparing means and ranges (follow through their values in (a) and (b)) B1 for contextual interpretation of a comparison of means or ranges (follow through their values in (a) and (b))	**(3)**
4	**(a)**	B1 Advantage: • eg Convenient • eg Easy B1 Disadvantage • eg Not representative • eg Biased • eg Students arriving early may all travel to school using the same transport	1st B1 any one correct advantage 2nd B1 any one correct disadvantage	**(2)**
	(b)	B1B1B1 for each of three aspects from: • Number the students in the database / use the position in the database • Generate numbers using a calculator / computer / random number table • Students with the matching number are selected • Ignore repeated numbers or numbers outside the range of those given to the students	B1B1B1 for demonstrating understanding of how to select a random sample	**(3)**

5	**(a)**	B2 3 | 3 4 8 4 | 2 5 5 | 1 3 9 6 | 0 1 9 7 | 0 4 9 8 | 3 6 7 8 9 9 | 10 | 2 B1 Key: 4 | 2 represents 42 shots	B2 cao **or** if B2 not earned B1 for unordered diagram or ordered diagram with at most 2 errors B1 for correct key	**(3)**
	(b)	M1 (20 + 1)/2 A1 65	M1 for evidence of their 10.5th value being considered eg circling the 10th and 11th terms	**(2)**
	(c)	B1 There is no mode	B1 for correct statement explaining why the mode is not possible	**(1)**
	(d)	B1 3		**(1)**
	(e)	B2 for Jay is not necessarily correct because 57% is only an average figure for all 20 footballers and the top 3 may have a higher average than the top 20 Alternatively: M1 $\frac{57}{100} \times (88 + 89 + 102)$ or $\frac{159}{88+89+102} \times 100$ A1 = 159 so Jay is correct or = 57% so Jay is correct	B2 for correct interpretation and conclusion of the information given or B1 for a partially correct statement eg "the top 3 may have a higher average" with no conclusion M1 for correct interpretation of the data A1 for correct conclusion and 159(.03) or 56.9(8)%	**(2)**

6	**(a)**	B1 7		**(1)**
	(b)	M1 $11 - 5$ A1 6		**(2)**
	(c)	M1 $\frac{5\times3+6\times3+7\times5+8\times2+9\times3+10\times2+11\times2}{20}$ A1 7.65	M1 for $\sum fx \div 20$ condone one error A1 for an answer of 7.65 or 7.7	**(2)**
	(d)(i)	B1 Greater than	B1 for correct conclusion	**(2)**
	(d)(ii)	B1 The new reading is higher than the mean oe	B1 for statistical reasoning	
7	**(a)**	B2 for a complete answer, eg • Not suitable **AND** a personal question / people may not be willing to answer • Not suitable **AND** no option boxes / an open question • Not suitable **AND** there is no time frame OR if B2 not earned… B1 for an incomplete answer ie giving reasons but no conclusion	B2 for assessing the appropriateness of the given question and reaching the correct conclusion OR if B2 not earned… B1 for an incomplete attempt at assessing the appropriateness of the given question.	**(2)**
	(b)	B1 for eg the range is affected by extreme values		**(1)**
	(c)	B2 for no AND a correct reason, eg • reference to the extreme value in the earnings • reference to the median not being affected by extreme values OR if B2 not earned… B1 for an incomplete answer eg no with an attempt at reason, OR correct reason without a conclusion OR if B1 not earned… sc B1 for yes AND the mean uses all of the data	B2 for a complete answer assessing the appropriateness of the choice of average. OR if B2 not earned… B1 for an incomplete answer assessing the appropriateness of the choice of average.	**(2)**

8	**(a)**	M1 $54 - 22$ A1 32 (%)	M1 for 54 and 22 seen	**(2)**
	(b)	B1 for eg we only know percentages and not the numbers of households	B1 for any reason which implies we only know percentages/proportions	**(1)**
	(c)	B1 for eg 'There are only two years of data'	B1 for appropriate reason why a time series would not be suitable for this data	**(1)**
9	**(a)**	M1 $\frac{697852 \times 1000}{65110000}$ A1 10.7	Accept answers in the range 10.7-10.8	**(2)**
	(b)	B1 for eg we know the birth rate per 1000 not the number of births	B1 for any reason which implies that these are relative to the total population of the country	**(1)**
10		M1 0.6×0.6 A1 $= 0.36$ M1 $2 \times 0.4 \times 0.6$ A1 $= 0.48$ depB1 ft for correct conclusion with comparison of probabilities	depB1 ft dependent on M1M1 (allow ft conclusion with a correct comparison of their probabilities)	**(5)**
11	**(a)**	B1 eg those age 12 do not know which box to tick / could tick more than one option	B1 for answer recognising the overlap (poor 'reliability' as different boxes may be ticked by people of same age) Accept 'does not include under 10 or over 16'	**(1)**
	(b)	B1 eg it is a leading question	B1 for equivalent wording recognising that it is a biased question.	**(1)**

	(c)	B1 eg it is an open question / answers will be difficult to handle	B1 for recognising it is open, or for explaining problems with using open questions Accept 'no time frame'	**(1)**
	(d)	B1 eg members may not wish to answer (as questionnaire is not anonymous)	B1 for answer recognising that it is a sensitive question or that 'truthful' answers may not be obtained. (poor 'validity')	**(1)**
12	**(a)**	 **Memory test score**	M1 for a box with two whiskers drawn with at least 3 values correct A1 fully correct	**(2)**
	(b)	B1 for eg median for teenagers greater than median for parents B1 for eg IQR/range for parents greater than IQR for teenagers B1 for eg both negative skew depB1 for eg "The teenagers did better on the memory test than their parents" OR "The teenagers were more consistent on the memory test"	B1 for a correct statistical statement comparing the medians B1 for a correct comparison of the IQRs or ranges B1 for a correct comparison of the skews depB1 for a correct contextual interpretation comparing medians or IQR/ranges (dependent upon previous B1 being scored)	**(4)**
	(c)	M1 56×0.75 A1 42	M1 for identifying 75% or $\frac{3}{4}$ or 0.75 A1 for 42	**(2)**

13	**(a)**	B1 eg Students who score highly in GCSE Mathematics also score highly in A-level Mathematics	B1 for a suitable hypothesis regarding relative performance in the two exams	**(1)**
	(b)	B1 GCSE is sat first It is plotted on the *x*-axis	B1 for an acceptable reason. Allow equivalent wording. Condone 'horizontal' axis.	**(1)**
	(c)	B2 ft The scatter graph shows <u>positive correlation</u> ... which supports the hypothesis	B2 for a correct conclusion for their hypothesis (ft) and mention of positive correlation. (Otherwise B1 for identifying positive correlation)	**(2)**
	(d)	B1 straight line with correct gradient B1 straight line through (578, 78)	1st B1 accept $0.6 < \text{gradient} < 0.8$ (not inclusive) If line does not extend at least from $x = 540$ to $x = 600$ then score max B1B0	**(2)**
	(e)	B1 ft eg for every extra mark at GCSE an extra 0.7 is scored for A-level	B1 ft for correct equivalent interpretation. Allow ft from their line. (eg 7 extra A-level marks for every extra 10 GCSE marks)	**(1)**
	(f)	B1 (Will not be reliable because) 540 is outside the range of data / it is extrapolation	B1 for assessing the appropriateness of the method	**(1)**

14	**Data collection** B1 for identifying one appropriate thing that should be included in the plan for data collection **and** B1 for explaining why this aspect is appropriate B1 for deciding what data to collect and/or how to collect and record it **and** B1 for an appropriate reason OR B1 for designing a collection method for primary/secondary data **and** B1 for an appropriate reason OR B1 for recognising where issues of sensitivity may influence data availability B1 for an appropriate reason	B1 for eg use distance to holiday to the nearest mile **and** B1 for eg this is sufficient as there will be a large range of distances B1 for eg use quota sampling **and** B1 for eg this ensures that you have data for each price range **OR** B1 for eg collecting data for prices of only one type of holiday, for example all inclusive **and** B1 for eg this means that the prices aren't affected by factors other than distance B1 for eg find prices from a holiday company website rather than asking people how much they paid for their holiday **and** B1 for eg as people may not want to say how much they spent on their holiday	**(6)**

	Presenting and interpreting diagrams B1 for identifying an appropriate method for presenting data in a diagram **and** B1 for explaining how this will be interpreted in the context of the investigation **Presenting and interpreting calculations** B1 for identifying an appropriate method to generate statistical measures to compare data **and** B1 for explaining how this will be interpreted in the context of the investigation	B1 for eg use a scatter diagram for the data **and** B1 for eg as this will allow you to see if there is a correlation (between the distance travelled and the price of the holiday) B1 for eg calculate the mean point to plot an accurate line of best fit **and** B1 for eg as you could use the line of best fit to make predictions about the cost of holidays of a certain distance **OR** B1 for eg calculate the equation of the line of best fit **and** B1 for eg to be able to see the relationship between distance travelled and price **OR** B1 for eg use a statistical package to calculate Spearman's rank correlation coefficient **and** B1 for eg to measure the strength of any correlation between distance travelled and price	