Mark Scheme (Results)

November 2021

Pearson Edexcel GCE
In Statistics (9ST0)
Paper 01: Data and Probability

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

November 2021
Question Paper Log Number P66655A
Publications Code 9ST0_01_2111_MS

General Marking Guidance

Total marks

The total number of marks for the paper is 80.

Mark types

The Edexcel Statistics mark schemes use the following types of marks:

- **M** **Method** marks,
  awarded for 'knowing a method and attempting to apply it',
  unless otherwise indicated.
- **A** **Accuracy** marks can only be awarded if the relevant method (M) marks have been earned.
- **B** **Unconditional accuracy** marks are independent of M marks
- **E** **Explanation** marks

NOTE: Marks should not be subdivided.

Abbreviations

These are some of the marking abbreviations that will appear in the mark schemes.

- ft follow through
- PI possibly implied
- cao correct answer only
- cso correct solution only
  (There must be no errors in this part of the question)
- awrt answers which round to
- awfw answers which fall within (a given range)
- SC special case
- nms no method shown
- oe or equivalent
- dep dependent (on a given mark or objective)
- dp decimal places
- sf significant figures
- ∗ The answer is printed on the paper

Further notes

- All candidates must receive the same treatment. Examiners must mark the first candidate in exactly the same way as they mark the last.
- Mark schemes should be applied **positively**. Candidates must be rewarded for what they have shown they can do rather than penalised for omissions.
- Examiners should mark according to the mark scheme not according to their perception of where the grade boundaries may lie.
- There is **no ceiling** on achievement. All marks on the mark scheme should be used appropriately.
- All the marks on the mark scheme are designed to be awarded. Examiners should always award full marks if deserved, i.e. if the answer matches the mark scheme. Examiners should also be prepared to award zero marks if the candidate's response is not worthy of credit according to the mark scheme.
- Where some judgement is required, mark schemes will provide the principles by which marks will be awarded and exemplification may be limited.
- When examiners are in doubt regarding the application of the mark scheme to a candidate's response, the team leader must be consulted.
- All A marks are 'correct answer only' (cao), unless shown, for example, as A1ft to indicate that previous wrong working is to be followed through.
- All M marks are 'possibly implied' (PI) unless specifically stated otherwise in the 'Notes' column.
- After a **misread**, the subsequent A marks affected are treated as A1ft, but manifestly absurd answers should never be awarded A marks.
- **Crossed out** work should be marked UNLESS the candidate has replaced it with an alternative response.
- If **two solutions** are given, each should be marked, and the resultant mark should be the mean of the two marks, rounded down to the nearest integer if needed.

<table>
<tr><th>Qu</th><th>Scheme</th><th>Marks</th><th>AO</th><th>Notes</th></tr>
<tr><td rowspan="2">1(a)(i)</td><td>$\frac{58+42}{136}$</td><td>M1</td><td>1.1</td><td>PI
Numerator correct (Accept each value ± 2, or total ± 4)
or
Denominator correct (may be seen in two fractions)</td></tr>
<tr><td>$= \frac{100}{136} = \frac{25}{34} = 0.735$</td><td>A1</td><td>1.1</td><td>awfw 0.71~0.76
Actual: 0.7352941…</td></tr>
<tr><td rowspan="8">1(a)(ii)</td><td>[X = number of errors on a page]</td><td></td><td></td><td></td></tr>
<tr><td>np method</td><td></td><td></td><td></td></tr>
<tr><td><table><tr><th>x</th><th>n</th><th>p</th><th>np</th></tr><tr><td>0</td><td>58</td><td>0.4265</td><td>0.0000</td></tr><tr><td>1</td><td>42</td><td>0.3088</td><td>0.3088</td></tr><tr><td>2</td><td>20</td><td>0.1471</td><td>0.2941</td></tr><tr><td>3</td><td>8</td><td>0.0588</td><td>0.1765</td></tr><tr><td>4</td><td>2</td><td>0.0147</td><td>0.0588</td></tr><tr><td>5</td><td>4</td><td>0.0294</td><td>0.1471</td></tr><tr><td>6</td><td>0</td><td>0.0000</td><td>0.0000</td></tr><tr><td>7</td><td>2</td><td>0.0147</td><td>0.1029</td></tr></table></td><td>M1</td><td>1.1</td><td>PI
Attempt to find np</td></tr>
<tr><td></td><td></td><td></td><td></td></tr>
<tr><td>$\mathrm{E}(X) = \sum np$</td><td>M1</td><td>1.1</td><td>PI
Correct method to find E(X)</td></tr>
<tr><td>$= 1.09$ (3 s.f.)</td><td>A1</td><td>1.1</td><td>awfw 0.97~1.19</td></tr>
<tr><td></td><td></td><td></td><td></td></tr>
</table>

<table>
<tr><th>Qu</th><th>Scheme</th><th>Marks</th><th>AO</th><th>Notes</th></tr>
<tr><td rowspan="5">1(a)(ii) cont.</td><td>Raw data method</td><td></td><td></td><td></td></tr>
<tr><td>
<table>
<tr><th>x</th><th>n</th><th>xn</th></tr>
<tr><td>0</td><td>58</td><td>0</td></tr>
<tr><td>1</td><td>42</td><td>42</td></tr>
<tr><td>2</td><td>20</td><td>40</td></tr>
<tr><td>3</td><td>8</td><td>24</td></tr>
<tr><td>4</td><td>2</td><td>8</td></tr>
<tr><td>5</td><td>4</td><td>20</td></tr>
<tr><td>6</td><td>0</td><td>0</td></tr>
<tr><td>7</td><td>2</td><td>14</td></tr>
</table>
</td><td>(M1)</td><td></td><td>Attempt to find xn
PI</td></tr>
<tr><td></td><td></td><td></td><td></td></tr>
<tr><td>$\mathrm{E}(X)=\dfrac{\sum xn}{136}=$</td><td>(M1)</td><td></td><td>PI
Correct method to find E(X)</td></tr>
<tr><td>$=\dfrac{148}{136}=\dfrac{37}{34}=1.09$ (3 s.f.)</td><td>(A1)</td><td></td><td>awfw 0.97~1.19</td></tr>
<tr><td>1(b)</td><td>$1.09\times 223=243$ (3 s.f.)</td><td>B1ft</td><td>1.2</td><td>awfw 216~266
ft their (a)(ii)</td></tr>
<tr><td rowspan="4">1(c)</td><td>Possible reasons (not exhaustive)</td><td></td><td></td><td></td></tr>
<tr><td>The novels are by the same author (so you would expect the distribution to be similar)</td><td></td><td></td><td></td></tr>
<tr><td>They are both novels</td><td></td><td></td><td></td></tr>
<tr><td></td><td>E1</td><td>3.1a</td><td>Any sensible reason</td></tr>
</table>

Qu	Scheme	Marks	AO	Notes
1(d)	**Possible reasons (not exhaustive)**			
	The author's spelling, punctuation, and grammar may have improved since the last novel.			
	The author may have learnt how to use a spellchecker since the last novel.			
	There may be a different number of words per page in the new novel.			
		E1	3.1a	Any sensible reason
	Total	**8**		

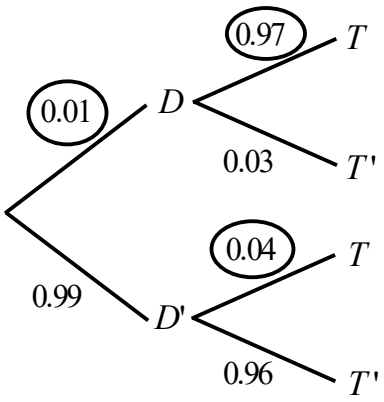
Qu	Scheme	Marks	AO	Notes
2(a)(i)	Probability = 0	B1	1.1	
2(a)(ii)	$(0.5-0.28)$ [×1]	M1	1.2	PI Clear attempt to find correct area or interval size on x-axis
	$= 0.22$	A1	1.2	cao SC 0.44 scores M1A0
2(b)	0.1×2 [×1]	M1	1.2	PI **or** $(0.1-(-0.1))$ Clear attempt to find correct area or interval size on x-axis
	$= 0.2$	A1	1.2	cao
2(c)	[X = Number of readings showing a minor error in 30 seconds]			
	$X \sim B(30, 0.2)$	B1	2.1a	PI Binomial distribution used
	$\mathrm{P}(X \geq 10)$	M1	2.1b	PI
	$= 0.0611$ (3 s.f.)	A1	1.2	awrt 0.061 Actual: 0.061087…
2(d)	Measurements are independent of one another.	E1	3.1a	**Accept:** The probability of a minor error remains the same throughout the 30-second interval. **Do not accept:** Only two possible outcomes (as this is evident in the question) oe
	Total	**9**		

Qu	Scheme	Marks	AO	Notes
3(a)	**Possible criticisms (not exhaustive)**			
	The horizontal axis labels have unequal differences.			
	There is no horizontal axis title.			**Condone:** No axis titles (or labels) for E1 only
	The vertical axis title is in an unusual place.			**Do not accept:** No vertical axis title
	The £ unit is only included on one label on the vertical axis.			
	It is hard to read off values on the graph.			
	It is very crowded.			**or** busy oe
	The pictures of the coins are unnecessary.			
	There appear to be horizontal gaps between some parts of the curve.			
				Do not accept: lines are disconnected oe
		E1, E1	3.1a, 3.1a	E1 for each sensible criticism (Max E2)
3(b)	2	B1	1.1	awfw 1.7~2
	Total	**3**		

Qu	Scheme	Marks	AO	Notes
4(a)	**(A) Disproportional**…	B1	1.1	
	…**stratified** sample	B1	1.1	
	(B) Cluster sample	B1	1.1	**Accept** opportunity sample **Accept** quota sample
	(C) Judgemental sample	B1	1.1	**Accept** quota sample
4(b)	**Example process**			
	Import the data into a **spreadsheet**.			**Accept** database **Accept** named spreadsheet or database (e.g. Excel, Access)
	Number each tree (1–917).			**or** 'add ID field' oe **or** 0–916 ?
	Use a **random number generator** to find a number between 1 and 917.			Accept between 1 and 91/92 or 0–90/91
	Generate 9 more numbers by **adding on 91** each time, (cycling back to 1 after 917 is reached).			Accept 92
	Filter/delete/hide all rows/records/trees which do not correspond to the generated numbers.			**or** use LOOKUP/ VLOOKUP function to find all of the trees corresponding to the generated numbers
		E1, E1, E1, E1	1.1, 1.1, 1.1, 1.1	E1 for each relevant step (max E4)
	SC Systematic sample not correctly applied scores E3 max			

Qu	Scheme	Marks	AO	Notes
4(c)	**Example advantages (not exhaustive)**			
	This is most likely to be representative of the population of trees.			
	Kit could specify the type of sample of trees he needs.			
	It would be easier for Kit, as the manager is selecting the sample.			**Do not accept** quicker
		E1	3.1a	E1 for sensible advantage described **Condone** no context
	Example disadvantages (not exhaustive)			
	The manager may not have sufficient knowledge of the trees.			
	Asking someone's opinion about the trees may introduce bias.			**or** the manager may choose her favourite trees which have similar characteristics oe **Do not accept** 'bias' alone
	It may be difficult to get time with the manager, as she may be very busy.			**or** cannot be completed by Kit alone
	The trees may be awkward to get to (or find).			
	The sample of trees is not random.			
		E1	3.1a	E1 for sensible disadvantage described **Condone** no context
	SC1: Max E1 if advantage and disadvantage not made clear			
	Total	**10**		

Qu	Scheme	Marks	AO	Notes
5(a)	[D = Event that adult is a user of the drug] [T = Event that Test A is positive]			
	Bayes' theorem method			
	$\mathrm{P}(D)=0.01$	B1, B1	1.2, 1.2	PI B1 for one correct B2 for all three
	$\mathrm{P}(T \mid D)=0.97$			
	$\mathrm{P}(T \mid D')=0.04$			
	$\mathrm{P}(D \mid T)$	M1	2.1b	PI Clear intent to find this probability
	$=\dfrac{\mathrm{P}(D)\times\mathrm{P}(T \mid D)}{\mathrm{P}(D)\times\mathrm{P}(T \mid D)+\mathrm{P}(D')\times\mathrm{P}(T \mid D')}$	M1	2.1b	PI Clear attempt at Bayes' theorem
	$=\dfrac{0.01\times0.97}{0.01\times0.97+0.99\times0.04}$	M1	1.2	Correct formula with no more than 1 error This expression **must** be seen Denominator = 0.0493
	= 0.1968 (4 d.p.)	A1*	1.2	Complete solution with no errors seen

Qu	Scheme	Marks	AO	Notes
5(a) cont.	**Tree diagram method**			
		(B1)		Correct diagram structure with D and T (oe) correctly placed.
		(B1)		Circled probabilities correct and correctly placed.
	$\mathrm{P}(D\cap T)=0.0097$ $[\mathrm{P}(D\cap T')=0.0003]$ $\mathrm{P}(D'\cap T)=0.0396$ (3 s.f.) $[\mathrm{P}(D'\cap T')=0.9504]$	(M1)		PI Clear attempt at multiplying probabilities along branches
	$\mathrm{P}(D\mid T)$	(M1)		PI Clear intent to find this probability
	$=\dfrac{\mathrm{P}(D\cap T)}{\mathrm{P}(D\cap T)+\mathrm{P}(D'\cap T)}$ $=\dfrac{0.0097}{0.0097+0.0396}$	(M1)		Correct formula or method (must be clear) This expression **must** be seen Denominator = 0.0493
	= 0.1968 (4 d.p.)	(A1*)		Complete solution with no errors seen
5(b)	The majority of people testing positive for the drug are not users of the drug.	E1	2.1a	**or** too many false positives oe

Qu	Scheme	Marks	AO	Notes
5(c)	**Situation**			
	Test A could be given to all adults in the facility…	E1	2.1b	**or** all people, most people etc oe
	…and Test B only given to people who test positive to Test A.	E1	2.1b	oe
	Benefits			
	Limiting the use of Test B will keep costs down, …			oe
	…and minimise the use of invasive procedures for people in the facility.			oe
	The addition of test B will improve the accuracy of the drug-testing scheme.			oe
		E1, E1	2.1a, 2.1a	E1 for each sensible benefit (max E2)
	Total	**11**		

Qu	Scheme	Marks	AO	Notes
6(a)	**Possible conditions**			
	Hiccups occur **randomly**.			
	Hiccups occur **independently** of one another.			
	Hiccups occur at a **constant (average) rate**.			**or** probability of an event occurring is proportional to the length of time.
	Hiccups occur **singly**.			
		E1, E1, E1	3.1a, 3.1a, 3.1a	E1 for each correct condition (max E3) Max E2 if no context
6(b)	[X = number of hiccups in the next minute]			
	Assume $X \sim \text{Po}(8.5)$	M1	2.1b	PI Clear use of Poisson distribution with $\lambda = 8.5$
	$\text{P}(X \leq 2) = 0.00928$	A1	1.2	awrt 0.0093 Actual: 0.00928324…

Qu	Scheme	Marks	AO	Notes
6(c)	**Exponential method**			
	[Y = time until next hiccup]			
	Assuming $Y \sim \text{Exp}(8.5)$	B1	2.1b	PI Clear use of Exponential distribution with $\lambda = 8.5$ or $\mu = \frac{1}{8.5} = \frac{2}{17} = 0.1176...$
	$\text{P}(Y \geq 0.5)$	M1	1.2	PI Clear attempt to find this probability
	$= 1 - \text{P}(Y \leq 0.5)$			
	$= 1 - \left(1 - e^{-8.5x}\right)$	M1	1.2	PI Correct use of cumulative formula
	$= 0.0143$	A1	1.2	awrt 0.014 Actual: 0.01426423…
	SC $\lambda = 4.25$ can score B1M1M1A0 max			
	Poisson method			
	[W = number of hiccups in the next half-minute] Assuming $W \sim \text{Po}(4.25)$	(B1)		PI Clear use of Poisson distribution…
		(M1)		PI …with $\lambda = 4.25$
	$\text{P}(W = 0)$	(M1)		PI Clear attempt to find this probability
	$= 0.0143$	(A1)		awrt 0.014 Actual: 0.01426423…

Qu	Scheme	Marks	AO	Notes
6(d)	30 seconds is not a sufficient amount of time to wait…	E1dep	3.1a	**Not** sufficient Dep on good attempt at explanation.
	…as more than 1 in 100 trials may show a false success…	E1	3.1a	The probability is not negligible.
	…and Yasmine plans to run 500 trials.	E1	3.1a	Lots of trials
	Alternative (E2 max)			
	30 seconds is a sufficient amount of time…	(E1dep)		Dep on good attempt at explanation.
	…as the probability in (c) is very low.	(E1)		
	SC Max E1 for sensible explanation based on ft from 6(c)			

Total	**12**

Qu	Scheme	Marks	AO	Notes
7(a)	17×0.55	M1	1.2	PI
	$=9.35$	A1	1.2	
7(b)	[X = number of frames won by Gordon in 17 frames]			
	$X\sim B(17,0.55)$	B1	2.1a	PI Clear use of binomial distribution with $n=17$
	$\mathrm{P}(X\geq 9)$	M1	1.2	PI
	$=1-\mathrm{P}(X\leq 8)$ $=0.663$ (3 s.f.)	A1	1.2	awrt 0.66 Actual: 0.662564…
7(c)	**Single binomial distribution method**			
	[Acknowledging that additional frames could theoretically be played after one player had won, but that they would have no effect on the outcome]			
	[Y = number of frames won by Gordon in 25 frames]			
	[$Y\sim B(25,0.55)$]			
	$\mathrm{P}(Y\leq 7)$ or $\mathrm{P}(Y\geq 18)$	M1	2.1b	PI Either seen
	$\mathrm{P}(Y\leq 7)=0.0058$ $\mathrm{P}(Y\geq 18)=0.0639$			**or** either correct probability seen **Condone** 0.936
	P(Finish in 3 sessions) $=\mathrm{P}(Y\leq 7)+\mathrm{P}(Y\geq 18)$	M1	2.1b	PI Attempt to sum both tails
	$=0.0697$ (3 s.f.)	A1	1.2	awrt 0.070

Qu	Scheme	Marks	AO	Notes
7(c) cont.	**Multiple binomial distribution method**			
	Clear attempt to use binomial distribution with multiple values of n between 18 and 25.	(M1)		
	Clear attempt to consider both tails of the binomial distribution	(M1)		Consideration of $Y \leq$ and $Y \geq$ or Consideration of two binomial distributions with $p = 0.55$ and $p = 0.45$
	P(Finish in 3 sessions) $= 0.0697$ (3 s.f.)	(A1)		awrt 0.070
7(d)	[There is a 7% probability that] the organisers will lose ticket and television revenue from the final session.	E1 ft	2.1a	Organisers may lose money.
	or			
	[There is a 7% probability that] spectators who have bought tickets to the 4th session will not be able to watch the match			

Qu	Scheme	Marks	AO	Notes
7(e)	**Possible solutions (not exhaustive)**			
	Increase ticket prices to allow for this possibility.			
	Organise other snooker players to play a short match in the final session.			
	Show a replay on the television for the final session.			
	Make the snooker players play all 35 frames.			
	Put fewer frames in the first three sessions to increase the probability that a fourth session will be played.			
	Offer a discounted price for the 4^{th} session.			
	Do not sell 4^{th} session tickets until it is confirmed to take place.			
		E1, E1	2.1a, 2.1a	E1 for each sensible solution (Max E2)

Qu	Scheme	Marks	AO	Notes
7(f)	**Constant probability assumption**			
	This is unlikely to be valid…	E1	3.1b	This mark may be scored in the other assumption if not here **Condone** certainty (e.g. 'is not valid')
	Possible explanations (not exhaustive)			
	…as a player's mood/concentration level may change over time.	E1	3.1b	E1 for sensible explanation
	…as a player may have a bad session.			
	…as the player who takes the break may get an advantage/disadvantage.			
	Independent results assumption			
	This is unlikely to be valid…	(E1)		First E1 may be scored here if not scored above
	Possible explanations (not exhaustive)			
	…as a player may lose confidence if (s)he is losing	E1	3.1b	E1 for sensible explanation
	…as a player may be having an off day.			
	…as a player who is winning may feel more confident and play better.			
	…as a player may be more determined to win if (s)he is losing.			
	Total	**14**		

Qu	Scheme	Marks	AO	Notes
8(a)	Mode for each sensor is 0	B1	1.1	
8(b)	The sensors cannot detect very low levels of light…			
	…so all low levels of light will be recorded as 0…			
	…and so more 0s will be expected than any other 2dp measurement.			
		E1, E1	2.1a, 1.1	Any two of these
8(c)	Residual $r_A = y_A - (a + bx_A)$			
	$= 3.58 - (1.8392 + 1.0771 \times 3.59)$	M1	1.2	PI Correct use of regression equation
	$= -2.13$ (2 d.p.)	A1	1.2	awrt Actual: -2.125989
	SC $r_A = 2.13$ scores M1A0			
	Alternative			
	Sum of residuals = 2.12	(M1)		PI Evidence of summing residuals
	So $r_A = -2.12$	(A1)		cao

Qu	Scheme	Marks	AO	Notes
8(d)	**Example situations (not exhaustive)**			
	The (left) sensor was found to be faulty for this collision.			or turned off oe
	Experimental conditions had changed during this collision.			
	The measurement(s) had not been recorded/entered properly.			
	Point G may be outside the bounds of Therese's intended research.			
		E1	2.1b	E1 for sensible situation explained
8(e)(i)	$r = 0.8931$	B1	1.2	awrt 0.893
8(e)(ii)	$y = 0.7720 + 1.2098x$	M1	1.2	At least one correct coefficient seen awrt 1.21 awrt 0.772
		A1	1.2	Form and coefficients all correct
8(f)	*r* would remain the same…	E1	2.1b	***r* only**
	Possible reasons (not exhaustive)			
	…as the correlation between *x* and *y* is the same as the correlation between *y* and *x*.			
	…as a sample can't have two different values of *r*.			
	…as the formula for *r* is the same if you switch *x* and *y*.			Symmetric formula
		E1	2.1b	E1 for sensible reason described

Qu	Scheme	Marks	AO	Notes
8(g)	It is **too precise**…	E1	3.1a	**accept** accurate **or** it is not appropriate oe
	Possible reasons (not exhaustive)			
	…as the data is only given to 2 decimal places.			
	…as the sample is small, so the values are unlikely to be accurate for the population.			
		E1	3.1a	E1 for sensible reason described
	SC It is valid as the statistics will be more accurate scores E1E0			

Total	**13**

Pearson Education Limited. Registered company number 872828
with its registered office at 80 Strand, London, WC2R 0RL, United Kingdom