Mark Scheme (Results)

Summer 2022

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

Summer 2022
Question Paper Log Number 69600
Publications Code 9ST0_01_2206_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.

Question	Scheme	Marks	AO	Notes
1(a)	Median from box plot=35	A1	1.1	
	IQR from box plot=9	A1	1.1	Or range=23
	Median from graph=40	A1	1.1	Or comment such as "obviously greater than 35 as it starts at 35"
	IQR from graph=6 or 7	A1	1.1	Or range=16
	So Kings Cross has a higher average journey time	E1	1.1	Must have context
	The spread is larger for Waterloo	E1	1.1	Must have context
1(b)	The average journey times for Wimbledon and Stevenage commuters from their home to the train are similar.	E1	2.1a	e.g. the distributions are similar.
1(c)	**Possible sources (not exhaustive)** Waiting time at the station. Distance lived from the station. Traffic during journey to the station. Walking speed.	E1, E1	2.1a, 2.1a	Must be related to the first part of the journey.
		Total	**9**	

Question	Scheme	Marks	AO	Notes
2(a)	To check if the relationship is **linear** To check if correlation is positive or negative To see if the PMCC is an appropriate measure to use To get a rough idea of the PMCC before calculation as a check To check for anomalies To see if there is only one trend Any 2 reasons	E1,E1	1.1, 1.1	Condone type of correlation
2(b)	Chesapeake Bay Retriever As it doesn't fit trend of the other data on the **graph** or As it is the only breed with male lower than female in the **table**	E1 E1	1.1 1.1	
2(c)	PMCC is independent of units	B1	2.1b	oe e.g. the relationship
2(d)	It would be a value bigger than 0.988 but less than 1.	B1	2.1b	

2(e)	**Statements in support** Certain breeds of dogs may have a large minimum height and a small maximum height **Statements against** Breeds of show dog with a larger minimum desired height would be larger dogs in general. You would expect these to have a larger desired max height also. This would require the maximum heights to decrease as minimum heights increase which does not happen in our sample Sample has a positive correlation Minimum desired height can't be higher than maximum desired height which could happen with a negative correlation	E1,E1, E1	2.1a	Each of up to 3 statements (may include a mix of support/against)
	Alana is **incorrect**	E1	3.1b	Awarded for overall disagreement with Alana with attempt at a reason SC supporting statement and Alana is correct scores E1E0E0E1
	Total	**10**		

Question	Scheme	Marks	AO	Notes
3	Leo should choose a **large sample** of participants His sample of participants should be **randomly selected** He should show them one painting at a time and ask if they think it was done by a professional or by a child Mention of blocking generally Mention of a sensible specific blocking factor e.g. background in art Leo should collect paintings from several different artists/several different children Leo should collect an equal number of paintings from artists and children He could try to pair similar paintings from a child/professional artist Blind/double blind trial mentioned in context He should remove identifying information such as signatures from the artists' work He should keep participants separate so they do not discuss Use same paintings for all participants	E1, E1, E1, E1,	3.1a	Condone $n \geq 30$ or implied by context Or He should show them two paintings at a time and ask which they think is by a professional e.g. randomised block design Consideration of the bias caused by the painters One mark for each comment up to 7

		E1, E1, E1		
	Total	**7**		

Question	Scheme	Marks	AO	Notes
4(a)	Players with a spend of 0 give the data a heavy skew Revenue cannot be negative and normal is unbounded	E1	2.1a	Or reverse with whales to right
4(b)	0.1587	B1	1.2	awrt 0.159
4(c)	$\overline{X} \sim N\left(50, \frac{100}{12}\right)$ $P(\overline{X} < 40) = 0.0003$ **Alternative** $X_1 + \cdots + X_{12} \sim N(600,1200)$ $P(X_1 + \cdots + X_{12} < 480) = 0.0003$	M1 A1 (M1) (A1)	1.2 1.2	Divides variance by 12 or standard deviation by $\sqrt{12}$ awrt
4(d)	Because the standard deviation of $\overline{X}$ is smaller than that of X	E1	2.1b	Or 40 is more standard deviations from the mean
4(e)	Because it's a normal model Because $P(W = 40) = 0$ Because W is continuous Any of these	E1	2.1b	
4(f)	$(1 - \text{"0.1587"})^4$ 0.50096	M1 A1	1.2 1.2	$(1 - \text{'their (b)'})^4$ awrt 0.5

4(g)	$X \sim B(12, 0.8413)$ $P(X \geq 8) = \left(1 - P(X \leq 7)\right) = 0.970$	M1 M1 A1	1.2 1.2 1.2	Binomial with n=12 seen or used $p = 1 - 'their\ (b)'$ awfw 0.969-0.972
4(h)	$P(30 < X < 40) = 0.1359$ $\frac{P(30 < X < 40)}{P(X > 30)} = \frac{0.1359}{0.9772}$ 0.139	B1 M1 A1	1.2 1.2 1.2	 Divides by 0.9772
4(i)	**Not appropriate** The actual distribution would be discrete not continuous The normal distribution would not have an upper limit of £60 Whales may be inclined to spend £60 meaning the distribution is skew **Appropriate** Could model using e.g. $P(X = 5) = P(2.5 < W < 7.5)$ **Conclusion** Conclusion that normal may be appropriate/inappropriate Three comments	E1, E1, E1	2.1a	Or lower limit of £0 Continuity correction suggested Only valid if agrees with at least one of their reasons
	Total	**17**		

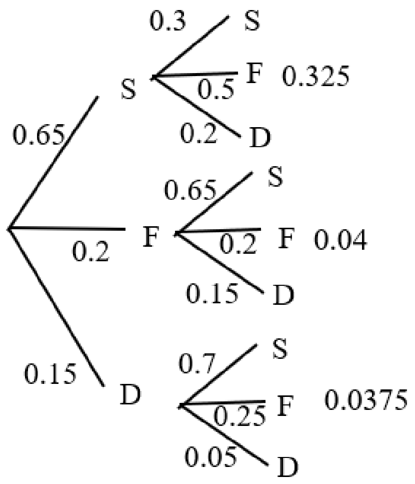
Question	Scheme	Marks	AO	Notes
5(a)(i)	X~B(120,0.35) or Y~B(300,0.15)	M1	2.1a	PI
	$X'\sim N(42,27.3)$ *or* $Y'\sim N(45,38.25)$	M1	2.1a	PI
	42+45=87	E1	1.2	Addition of means (no variances seen scores M1M0E1E0)
	27.3+38.25=65.55	E1	1.2	Addition of variances not standard deviations
5(a)(ii)	$P(T \geq 100) \approx P(T' > 99.5)$	M1	1.2	Continuity correction used
	0.0613	A1	1.2	sc 0.0542~0.0543 No continuity correction scores M0A1
5(b)	The binomial distributions X & Y can be approximated as a normal distribution as n is sufficiently large	E1	3.1a	Equivalently correct comment regarding np or the mean
	Two independent normal distributions can be added to give a normal distribution, and we assumed independence above	E1	3.1a	
5(c)	All tickets will sell – may not be justifiable as it's unlikely every single ticket will sell every night as some nights may be busier than others	E1	3.1a	Reason why assumption all tickets sell may be incorrect
	Independence of tickets being concessions – may not be justifiable as people may come in groups of friends of similar ages	E1	3.1a	Reason why assumption of independence may be incorrect e.g. not independent because people seeing one film can't see the other

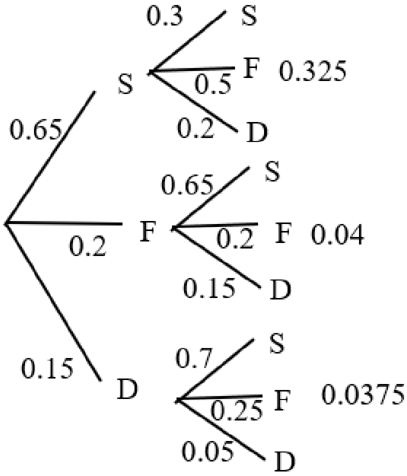
5(d)(i)	$420 - T$	B1	1.2	
5(d)(ii)	$P = 6(420 - T) + 4T[-200]$	M1	1.2	
	$P = 2320 - 2T$	M1	1.2	Or 2520
	$P \sim N(2146, \ldots$	A1	1.2	Mean (or 2346)
	$\ldots, 262.2)$	A1	1.2	Variance
	$P(P > 2125) = 0.9027$	A1	1.2	Or >2325
	Total	**16**		

Question	Scheme	Marks	AO	Notes
6(a)	**Explanation of how Caroline chose her parameters** $n = 100\ 000$ represents number of hairs on a head $p = 0.001 = \frac{100}{100000}$ 100 hairs out of 100000 lost per day	 E1 E1	 1.3 1.3	
6(b)	**Reasonable** Assuming that values for Britain are similar to America $n = 100\ 000$ is reasonable from the article Caroline may have chosen 100 hairs (at the upper end) as she has mostly female customers who lose more hair			

	Unreasonable Assuming that values for Britain are different to America Assumes that everyone is treating their hair similarly e.g. similar products Probability is different for men and women as women tend to lose more hair $n = 100\ 000$ unlikely to be valid as likely to be substantial differences between individuals Caroline chose 100 hairs when that was an upper end of a range Family members may book appointments on the same day so unlikely to be independent events Probability doesn’t take into account effect of stress or health conditions so not valid if customer suffers from one of these The falling out of hairs may not be independent of one another (e.g. one hair falling out may weaken the surrounding hairs)			Or mention of how people differ
		E1, E1, E1	3.1b, 3.1b, 3.1b	Any 3 of these comments. Cannot score 3 without a comment about whether this is reasonable or not
	Total	**5**		

Question	Scheme	Marks	AO	Notes
7(a)	$\frac{7}{56}\left(=\frac{1}{8}\right)$	B1	1.2	oe
7(b)	$\frac{6}{9}\left(=\frac{2}{3}\right)$	B1	1.2	oe
7(c)	$21 \times 20 \times 16 \times 15$	M1	1.2	Numerator
	37 seen or used in denominator	B1	1.2	
	$37 \times 36 \times 35 \times 34$	M1	1.2	Denominator decreases
	$\times 6$	M1	1.2	6 orders (or $4C2$)
	$\frac{21}{37} \times \frac{20}{36} \times \frac{16}{35} \times \frac{15}{34} \times 6 = \frac{240}{629}$	A1	1.2	Or 0.382
7(d)(i)	Probability in (a) would be smaller as more species means more animals overall	E1	2.1b	
7(d)(ii)	(b) would be unchanged as it is conditional on animal being an alligator and the number of male and female alligators is the same in each game	E1	2.1b	
7(d)(iii)	The probability in (c) would be smaller because without replacement the probability gets smaller for each successive animal of each type	E1	2.1b	Or by explicit calculation

7(e)	$0.65 \times 0.5 + 0.2 \times 0.2 + 0.15 \times 0.25$ 0.4025 **Tree Diagram Method**  0.4025	M1 A1 (M1) (A1)	1.2 1.2	Tree with at least the 0.65, 0.2, 0.15 as the first layer or 0.5, 0.2 and 0.25 on the fluctuating branch of 2^{nd} layer

7(f)	$P(S\,W_2 \mid F\,W_3)\left(=\dfrac{P(F\,W_3 \mid S\,W_2)\times P(S\,W_2)}{P(F\,W_3)}\right)$	M1	1.2	Attempt at Bayes theorem seen
		B1	1.2	Use of 0.65×0.5
	$=\dfrac{0.65\times 0.5}{\text{"}0.4025\text{"}}$	M1	1.2	Denominator is 'their (e)'
	0.807	A1	1.2	Or $\frac{130}{161}$
	Tree Diagram Method			
		(M1)		Tree seen with correct section for "spike" in week 2
		(B1)		$0.65 \times 0.5 = 0.325$
	$\dfrac{0.325}{\text{"}0.4025\text{"}}$	(M1)		Denominator is their (e)
	0.807	(A1)		
	Total	**16**		

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