Mark Scheme (Results)

November 2021

Pearson Edexcel GCE

In Statistics (9ST0)

Paper 02: Statistical Inference

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 P66656A
Publications Code 9ST0_02_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.

Qu	Scheme	Marks	AO	Notes
1(a)	H_0: Independent H_1: Not independent	B1	1.3	oe Allow one or two-tailed **Accept** H_0: $\rho = 0$ H_1: $\rho \neq 0$ or $\rho > 0$ **Condone** use of r
	(5% significance) cv = 0.4973 (2-tail)	M1	1.3	For obtaining cv awrt 0.497 1-tail: cv = 0.4259 awrt 0.426
	0.716 > 0.4973 So reject H_0	A1dep	2.1b	Conclusion correct dep 'correct' cv (either) 1-tail: 0.716 > 0.4259
	There is significant evidence that …the population correlation coefficient is different from (or greater than) zero	E1dep	2.1a	**or** …that Ramon's belief is correct. dep previous M1 A1 1-tail: Conclusion must be consistent

Qu	Scheme	Marks	AO	Notes
1(b)	**Possible explanations (not exhaustive)**			
	The data does not appear to have a bivariate normal distribution…			
	The x-values are not normally distributed…			Must see x
	The data is not linear…			
	The data is in two clusters…			oe
	The x-values are bimodal…			
	The correlation looks different for each cluster…			
	…so the assumptions of the PMCC test have been violated.			Reasonable attempt to explain why the above may be troublesome. dep on sensible reason given **Condone:** 'therefore it may not be valid' oe
		E1, E1	3.1b, 3.1b	Any two sensible comments Must not be contradictory
	Total	**6**		

Qu	Scheme	Marks	AO	Notes
2(a)	$182.9 \pm 1.96 \times \frac{22}{\sqrt{16}}$	M1	1.3	Use of $\frac{22}{\sqrt{16}}$ PI
		B1	1.3	z = 1.96 PI
	CI is (172, 194)	A1	1.3	awrt
2(b)	175 is in the confidence interval	E1ft	2.1b	ft their (a) provided reasonable attempt to find CI
2(c)	$182.9 - 1.96 \times \frac{22}{\sqrt{n}} > 175$	M1	1.3	PI Correct expression **Accept** = sign used **Condone** one small slip
	$182.9 - 175 > 1.96 \times \frac{22}{\sqrt{n}}$			
	$7.9\sqrt{n} > 43.12$			
	$\sqrt{n} > 5.458$			
	$n > 29.8$	A1	1.2	PI **Accept** = sign used
	Mohamed needs to sample (at least) 30 courgettes.	E1	2.1a	
	Total	**7**		

Qu	Scheme	Marks	AO	Notes
3(a)	[X = Number of business leaders supporting a four-day working week]			
	$X \sim B(85, p)$	B1	2.1b	**Accept** binomial with no parameters given May be implied by correct use elsewhere in the question
3(b)(i)	H_0: p = 0.5, H_1: p > 0.5	B1	1.3	oe
3(b)(ii)	**Binomial method**			
	Under H_0, $X \sim B(85,0.5)$			
	$P(X \geq 49) = 1 - 0.9036 = 0.0964$ $P(X \geq 50) = 1 - 0.9358 = 0.0642$ $P(X \geq 51) = 1 - 0.9590 = 0.0410$	M1	2.1b	PI Relevant attempt at $P(X \geq n)$
	Critical region: $X \geq 51$	A1	1.3	oe **Condone** $X \geq 50$
	Normal approximation method			
	Under H_0, $X \sim B(85,0.5)$			
	$X \approx Y \sim N(42.5,21.25)$			
	$P(Y > y) = 0.05$			**or** $P(Y < y) = 0.95$
	$\frac{Y - 42.5}{\sqrt{21.25}} > 1.6449$	(M1)		PI Effort at z inequality with 1.6449 (to at least 3sf accuracy) **Accept** = sign used
	$Y > 50.08$			
	Critical region: $X \geq 51$	(A1)		oe **Condone** $X \geq 50$

Qu	Scheme	Marks	AO	Notes
3(c)	**Binomial method**			
	P(Type I error) = P($X \geq 51$) = 0.0410	B1	1.3	awrt 0.041 **Condone** 0.0642 for $X \geq 50$ 0.05 scores B0 awrt 0.064
	Normal approximation method			
	P(Type I error) = P($X \geq 51$) = 0.0326	(B1)		awrt 0.033 **Condone** 0.0519 for $X \geq 50$ awrt 0.052 0.05 scores B0

Qu	Scheme	Marks	AO	Notes
3(d)	**Binomial method**			
	Assuming $X \sim B\left(85, \frac{2}{3}\right)$			
	P(Type II error) = P($X \leq 50$)	M1ft	1.3	**Condone** P($X \leq 49$) ft their critical region in (b)
	= 0.0796	A1ft	1.3	awfw 0.07~0.08 **Condone** 0.0515 for P($X \leq 49$) awfw 0.044~0.052 ft their critical region
	Power = 1 - 0.0796 = 0.9204	A1ft	1.3	awfw 0.92~0.93 **Condone** awfw 0.948~0.956 ft their critical region
	Normal approximation method			
	Assuming $X \sim B\left(85, \frac{2}{3}\right)$			
	$X \approx Y \sim N\left(56\frac{2}{3}, 18\frac{8}{9}\right)$			
	P(Type II error) = P($X \leq 50$)	(M1)		**Condone** P($X \leq 49$)
	= 0.0625	(A1)		awfw 0.054~0.063 **Condone** 0.0389 for P($X \leq 49$) awfw 0.033~0.039
	Power = 1 - 0.0796 = 0.9204	(A1)		awfw 0.937~0.946 **Condone** awfw 0.961~0.967
	Total	**8**		

Qu	Scheme	Marks	AO	Notes
4(a)	[W = Read widely, F = Read fairly widely]			
	H_0: $\mu_W - \mu_F = 0$ H_1: $\mu_W - \mu_F > 0$	B1	1.3	
	$S_p = \sqrt{\frac{14 \times 5.11^2 + 14 \times 4.18^2}{28}} = \sqrt{21.79}$ $= 4.67$	M1	1.3	PI Attempt to use S_p formula
	$df = n_W + n_F - 2 = 28$			
	cv = $t_{28}(0.05) = 1.70$	B1	1.3	Correct cv awrt
	$t = \frac{32.2 - 29.4}{\sqrt{4.67^2 \times \left(\frac{1}{15} + \frac{1}{15}\right)}}$	M1	1.3	PI Attempt at standardising Must see correct numerator $(\pm)$ Must see $\frac{1}{15}$
	$= 1.643$			**or** correct t-value seen
	$1.643 < 1.70$ ∴ do not reject H_0	A1	2.1b	Comparing correct ts with cv **or** *p*-value: $0.0560 > 0.05$ for full marks
	Insufficient evidence that Eleanor's students who read widely have a higher average vocab score than those who read fairly widely.	E1dep	2.1a	cso Dep on all previous marks
	SC z-test scores max B1M1B0M1A0E0 ts = 1.6426, cv = 1.6449 M1 lost for each error seen			

Qu	Scheme	Marks	AO	Notes
4(b)	**Possible criticisms (not exhaustive)**			
	Reading habits are self-selected.			
	Definition of 'widely' is vague/subjective.			
	Variances may be different.			
	Sample sizes are small.			
	Eleanor only uses her own students.			
		E1, E1	3.1a, 3.1a	Any two sensible comments
	Total	**8**		

Qu	Scheme	Marks	AO	Notes
5(a)	H_0: the (population) median finishing times at the two cities are the same H_1: the (population)median finishing times at the two cities are different	B1	1.3	**or** from the same/different population Condone use of η **Condone** average **Do not accept** mean
	Attempt to rank **jointly** across both samples	M1	1.3	
	M 1 2 5 6 8 10 14 18 20 21 **105** **L** 3 4 7 9 11 12 13 15 16 17 19 **126** **Total**			
	Ranks all correct	A1	1.3	
	$W_M = 105$ or $W_L=126$	M1	1.3	PI Attempt to total their ranks
	ts: $U_M = 105 - \frac{1}{2} \times 10 \times 11 = 50$	A1	1.3	**or** $U_L = 126 - \frac{1}{2} \times 11 \times 12 = 60$
	Two tailed 5% critical value cv = 27	B1	1.3	**or** 83
	$50 > 27$ so do not reject H_0	M1ft	2.1b	**or** $60 < 83$ ft their ts with correct tail
	There is **no evidence** that the **average** race time for the Men's 800 metres in Monaco **is different** from that for the Men's 800 metres in London.	E1dep	2.1a	Must be in context. Dep ts and cv both correct
5(b)	The samples chosen are not independent. The samples chosen are not random.	E1	3.1a	Either comment

Qu	Scheme	Marks	AO	Notes
5(c)	**Possible suggestions (not exhaustive)**			
	Do a randomly selected **paired sample** of the same athletes' performances in different years at the two venues.			
	Choose **larger independent random samples** using data from different years at the same venues.			
	Add more locations [and complete an ANOVA test].			
	Investigate races other than the men's 800m.			
	Add a blocking factor, such as age of athlete.			Any sensible blocking factor
	Investigate different factors, such as weather and stadium design, separately.			
		E1, E1	3.1a, 3.1a	Any two sensible comments
	Total	**11**		

Qu	Scheme	Marks	AO	Notes
6(a)	53.77	B1	1.3	awrt 53.8
	5.98	B1	1.3	awrt 6.0
6(b)	H_0: the weights are consistent with N(5.01, 1.30^2) H_1: the weights are not consistent with N(5.01, 1.30^2)	B1	1.3	Both correct but H_1 need not refer to N(5.01, 1.30^2)
	[Contribution to χ^2] 0.964, 0.960, 0.481, 0.187, 1.576, 0.157, 0.086, 1.341, 0.142, 0.175	M1	1.3	Attempt at any $\frac{(O-E)^2}{E}$ PI
	ts: $\chi^2 = \frac{(15-19.31)^2}{19.31} + \cdots + \frac{(7-5.98)^2}{5.98}$	M1	1.3	Intention to sum PI
	$= 6.07$	A1	1.3	awfw 6.02~6.12 Actual: 6.0667…
	Degrees of freedom = 10 – 3 = 7	B1	1.3	PI **Condone** df = 9 B1 if cv = 2.167 or 3.325 seen (wrong tail used)
	cv: χ_7^2 = 14.067	B1	1.3	awrt 14.1 **Condone** χ_9^2 = 16.919 awrt 16.9
	6.07 < 14.067 so do not reject H_0	M1ft	2.1b	PI **or** *p*-value: $0.532 > 0.05$ ($\nu = 7$) $0.733 > 0.05$ ($\nu = 9$) for full marks ft their ts/cv

<table>
<tr><th>Qu</th><th>Scheme</th><th>Marks</th><th>AO</th><th>Notes</th></tr>
<tr><td>6(b) cont.</td><td>There is no significant evidence to doubt that the weights of the lambs are consistent with N(5.01, 1.30^2)</td><td>E1dep</td><td>2.1a</td><td>oe
Dep on ts and cv or p-value correct</td></tr>
<tr><td rowspan="3">6(c)</td><td>H_0: the population mean (breed) weights are the same (for all four breeds)</td><td></td><td></td><td>oe
Accept
$\mu_i = \mu_j$ for all i, j
$\mu_i = \mu$ for all i</td></tr>
<tr><td>H_1: at least two population mean (breed) weights are different</td><td></td><td></td><td>oe
Accept
$\mu_i \neq \mu_j$ for some i, j
Do not accept
$\mu_i \neq \mu$ for some i</td></tr>
<tr><td></td><td>B1</td><td>1.3</td><td>Both correct</td></tr>
<tr><td>6(d)</td><td colspan="5"><table>
<tr><td>Source of variation</td><td>Sum of squares</td><td>Degrees of freedom</td><td>Mean square</td><td>F ratio</td></tr>
<tr><td>Between breeds</td><td>578.521</td><td>3</td><td>192.84</td><td>7.25</td></tr>
<tr><td>Within Breeds</td><td>148147.218</td><td>5567</td><td>26.61</td><td></td></tr>
<tr><td>Total</td><td>148725.739</td><td>5570</td><td></td><td></td></tr>
</table></td></tr>
<tr><td></td><td>Degrees of freedom</td><td>B1</td><td>1.3</td><td>Between breeds = 3</td></tr>
<tr><td></td><td>Mean squares</td><td>M1</td><td>1.3</td><td>MS=SS/df for between breeds and within breeds</td></tr>
<tr><td></td><td>$F = \frac{192.84}{26.61} = 7.25$</td><td>A1</td><td>1.3</td><td>awfw 7~8
Actual: 7.2465…</td></tr>
<tr><td rowspan="3">6(e)</td><td>Critical value $F_{\infty}^{3}\ (0.05) = 2.605$</td><td>B1</td><td>1.3</td><td>AWFW 2.60- 2.61
or $p = 0.00015$</td></tr>
<tr><td>$7.25 > 2.605$ so reject H_0</td><td>M1ft</td><td>2.1b</td><td>ft their F
or $0.00015 < 0.05$</td></tr>
<tr><td>There is significant evidence that at least two population mean (breed) weights are different.</td><td>A1</td><td>2.1a</td><td>Conclusion in context</td></tr>
</table>

Qu	Scheme	Marks	AO	Notes
6(f)	Variances of weights of all breeds should be the same.			
	Weights of lambs should be normally distributed.			
	The lambs should be a random sample.			
		E1, E1	3.1a, 3.1a	Any two correct responses in context
	Total	**19**		

Qu	Scheme	Marks	AO	Notes
7(a)	**Possible reasons (not exhaustive)**			
	As n is large…			**Accept** $n \geq 30$
	…so by the central limit theorem, sample means are Normally distribution (CLT).			
	Variance of sample mean is $\frac{\sigma^2}{n}$ (providing samples independent).			**Accept** $\frac{s^2}{n}$ **Accept** mention of standard error with this formula
	Large samples so we can use sample variances to represent population variances.			
	Because the means are not known.			
		E1, E1	3.1b, 3.1b	Any two sensible reasons
7(b)	[B = Trains from Brighton S = Trains from Southend-on-sea]			
	H_0: $\mu_S = 0$ H_1: $\mu_S < 0$	B1	1.3	both
	cv $= -1.6449$	B1	1.3	**Condone** 1.6449 at least 3sf accuracy
	ts $= \frac{-0.709}{\frac{2.45}{\sqrt{55}}}$	M1	1.3	Use of $\sqrt{55}$
	$= -2.15$	A1	1.3	awrt Actual: -2.146157 ... **or** p-value = awrt 0.016 Actual: 0.0159
	$-2.15 < -1.6449$ so reject H_0	M1ft	2.1b	**or** 0.0159 < 0.05
	There is **significant evidence** that trains from Southend-on Sea arrive **early on average**.	E1dep	2.1a	oe Dep on ts/ cv or p-value correct

Qu	Scheme	Marks	AO	Notes
7(c)	**Hypotheses** $H_0: \mu_B - \mu_S = 3$ $H_1: \mu_B - \mu_S > 3$	B1	1.3	oe Both correct
	Critical region $z > 1.6449$	B1	1.3	**Condone** just 1.6449 seen, provided it is made clear this is the cv (or cr) at least 3sf accuracy
	Evidence 2.074 > 1.6449			
	Significant evidence against H_0	A1dep	2.1b	oe dep on previous B1
7(d)	*p*-value = P(z > 2.074) = 0.019	B1	1.3	
	Possible explanations (not exhaustive)			
	Using each method results in rejecting H_0, so there is no advantage to the *p*-value method.			
	The *p*-value shows that the outcome observed is much less probable than 0.05 and is a preferable measure.			
		E1	2.1b	Any sensible explanation

Qu	Scheme	Marks	AO	Notes
7(e)	H_0: $\pi_B - \pi_S = 0$ H_1: $\pi_B - \pi_S > 0$	B1	1.3	condone p
	$\hat{p} = \frac{12}{88} = \frac{3}{22}$	M1	1.3	PI
	$ts = \frac{\frac{8}{44} - \frac{4}{44}}{\sqrt{\frac{3}{22} \times \frac{19}{22}\left(\frac{1}{44} + \frac{1}{44}\right)}}$	M1	1.3	Numerator or denominator correct
	$= 1.24$	A1	1.3	awrt 1.24 Actual: 1.2425 ...
	5% critical value = 1.6449			
	1.24 < 1.6449 so do not reject H_0	M1	2.1b	Comparison of their ts with 1.6449 (to at least 3sf accuracy) **or** p-value:$0.107 > 0.05$ for full marks
	There is no significant evidence that there are more likely to be days when at least one train from Brighton is cancelled than from Southend-on-Sea.	E1dep	2.1a	Must be in context Dep on correct ts/cv or p-value
7(f)	**Possible assumptions (not exhaustive)**			
	The sample of days should be random.			
	The normal approximation for proportions can be used.			
	Cancellations are independent of day.			
	Cancellations are independent of location.			
		E1, E1	3.1a, 3.1a	Any two sensible assumptions
	Total	**21**		

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