Mark Scheme (Results)

Summer 2019

Pearson Edexcel GCE

In Mathematics (9ST0) Paper 2 Statistics

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 2019
Publications Code 9ST0_02_1906_MS
All the material in this publication is copyright
© Pearson Education Ltd 2019

General Marking Guidance

- 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.
- Crossed out work should be marked UNLESS the candidate has replaced it with an alternative response.

Paper 2: Statistical Inference Mark Scheme

Question	Scheme	Marks	AO	Notes
1(a)	$\bar{x} = 4.4 \quad s = 2.3 \quad n = 36$ $4.4 \pm 2.5758 \times 2.3/\sqrt{36} =$ Using t, CI is (3.36, 5.44) Using z, CI is (3.41, 5.39)	M1 B1 A1	1.3 1.3 1.3	Use of 4.4 and $2.3/\sqrt{36}$ (possibly implied by correct interval) PI $t_{35} = 2.72(4)$ or $z=2.58$ used (possibly implied by correct interval) PI Accept CI: (awrt 3.4, awrt 5.4) If z or t values seen, must be 2.58 or 2.72
1(b)	Because the sample is “large”… **or** n > 30… … so the Central Limit Theorem applies.	E1 E1	3.1a 3.1a	 allow CLT
1(c)	It should be a **random sample** of dolphins. **or** The population of dolphins in that area should be large.	E1	3.1a	If mention independence must also see dolphins
	Total	**6**		

<table>
<tr><th>Question</th><th>Scheme</th><th>Marks</th><th>AO</th><th>Notes</th></tr>
<tr><td>2</td><td>H_0: No association (between data sets)
H_1: Negative association (between data sets)</td><td>B1</td><td>1.3</td><td>oe
Condone ‘correlation’ or Condone ‘independent’ in H_0 only</td></tr>
<tr><td></td><td>Use of Spearman’s rank correlation coefficient</td><td>M1</td><td>2.1a</td><td>PI
Sight of SRCC, Spearman’s any ranks</td></tr>
<tr><td></td><td>Rank each separately:
<table>
<tr><td>OT</td><td>7</td><td>1</td><td>4</td><td>3</td><td>5</td><td>6</td><td>2</td></tr>
<tr><td>P</td><td>1</td><td>7</td><td>5</td><td>2</td><td>3</td><td>4</td><td>6</td></tr>
</table>
Ranks reversed
<table>
<tr><td>OT</td><td>1</td><td>7</td><td>4</td><td>5</td><td>3</td><td>2</td><td>6</td></tr>
<tr><td>P</td><td>7</td><td>1</td><td>3</td><td>6</td><td>5</td><td>4</td><td>2</td></tr>
</table></td><td>M1</td><td>1.3</td><td>PI</td></tr>
<tr><td></td><td>ts: $r_s = (–)0.75$</td><td>A1</td><td>1.3</td><td>Cao
Ignore sign</td></tr>
<tr><td></td><td>5% one-tailed cv = (–)0.6786</td><td>B1</td><td>1.3</td><td>Ignore sign</td></tr>
<tr><td></td><td>–0.75 < –0.6786 so reject H_0 at 5% level</td><td>M1ft</td><td>2.1b</td><td>Comparing ts (allow small slip) and cv with same signs.</td></tr>
<tr><td></td><td>There is evidence that when OT’s are relatively abundant, Peacocks are relatively uncommon.
or
Yes it does support the view.</td><td>A1</td><td>2.1a</td><td>In context, not too definite
Requires correct r_s and cv</td></tr>
<tr><td></td><td>Total</td><td>7</td><td></td><td></td></tr>
</table>

Question	Scheme	Marks	AO	Notes
3(a)	$H_0: \mu = 343$ $H_1: \mu \neq 343$	B1	1.3	Both correct
	$\bar{x} = 339.33$ $s = 3.01$ (or $s^2 = 9.07$)	B1	1.2	PI Both correct $\bar{x}$ awrt 339 s awrt 3.0 Condone σ=2.75
	Test statistic, $(t) = \frac{\text{"339.33"} - 343}{\frac{\text{"3.01"}}{\sqrt{6}}}$	M1	1.3	PI Method for ts Condone σ=2.75 if $\sqrt{5}$ used
	= –2.98	A1	1.3	AWFW(-3.26,-2.98)
	5% critical value (5df) = $(\pm)2.571$	B1	1.3	PI Alt: p-value = 0.01535 (or 0.0307…) CR: $\bar{X} \leq 339.8$ AWRT 340 $(\bar{X} \geq 346.2)$ CI (342.5,336.2)
	-2.98 < -2.571 so reject H_0	M1	2.1b	Clear correct comparison of like with like. Alt: 0.01535 < 0.025 (or 0.0307 < 0.05) Alt: 339.33>339.8
	Conclude that there is evidence to suggest that, on **average**, **measurements** of the speed of sound in dry air **differ** significantly from the known speed of 343 m/s.	A1dep	2.1a	Must be in context and conclusion not too definite using words in bold oe Dep on all previous marks except 1st B1

3(b)	Acceptance region is $1481 \pm t\frac{1.25}{\sqrt{10}}$	M1	2.1	Use of 1481 and $\frac{1.25}{\sqrt{10}}$
	Use of t = 2.26(2)	B1	1.3	
	$(\bar{x} > 1481.89\ and\ \bar{x} < 1480.11)$	A1	1.3	AWRT 1480 & AWRT 1482
	So critical region is $(\bar{x} > 1481.9\ and\ \bar{x} < 1480.1)$	A1	1.3	Both parts required to exactly 1dp accuracy
	Total	**11**		

Question	Scheme	Marks	AO	Notes
4	H_0: p = 0.5 H_1: p > 0.5	B1	1.3	p or π For both
	$z = \dfrac{0.51 - 0.5}{\sqrt{\dfrac{0.5 \times 0.5}{1025}}}$	M1	1.3	Correct formula
	$= 0.64(03)$	A1	1.3	awrt 0.64
	5% cv = 1.6449	B1	1.3	1.64 ~ 1.65 (Or p = 0.261)
	0.64 < 1.64 so cannot reject H_0	M1	2.1b	Or 1.28 for 1.64 Or 0.261 > 0.05 (accept 0.10 or higher sig. level if < 0.261) Must score previous B1
	so cannot reject H_0 or Accept H_0	A1	2.1a	
	There is no evidence that the majority support nuclear energy. **or** The test shows the poll does not convincingly support the advocacy group's claim.	E1	3.1b	Does not need all previous marks. In context, not too strong.
	Alternative (exact binomial)			
	H_0: p = 0.5 H_1: p > 0.5	(B1)		p or π For both
	Calculating np	(M1)		implied by 522.75
	B(1025,p)	(A1)		any p
	P(522 or 523)=0.266 or 0.287	(B1)		
	comparing p-value with 0.05	(M1)		must score previous B1
	Accept H_0 oe	(A1)		

	There is no evidence that the majority support nuclear energy. **or** The test shows the poll does not convincingly support the advocacy group's claim.	(E1)		Do not need previous marks In context, not too strong.
	Comments: The group were **not necessarily** wrong.	E1	3.1b	oe
	It is possible that there is a majority supporting nuclear energy, but the sample just failed to provide enough evidence. A larger sample may have found support.			Recognition of pure chance sampling effects. (Allow mention of type II error in context.)
	Small sample relative to size of USA adult population Also, the sample may have been biased as it was only conducted by telephone (or some other comment on the non-randomness of the sample.)			Recognition of possible bias effects.
		E1	3.1b	E1 for sensible comment
		Total	**9**	

Question	Scheme	Marks	AO	Notes
5 (a)	Wilcoxon rank-sum test used or stated	M1	2.1a	Condone Mann Whitney U
	H_0 : no difference in population medians H_1: population median score for B < population median score for A	B1		Both oe or using η (Not η_d=0) or samples from identical populations
	Ranks given below	M1	1.3	Attempt at ranks as one group.
		A1	1.3	At least four ranks correct.

Rank (A)	7	12	8.5	5.5	14	4	11	15
Rank (B)	8.5	1	2	3	5.5	10	13	

Rank (A)	9	4	7.5	10.5	2	12	5	1
Rank (B)	7.5	15	14	13	10.5	6	3	

Scheme	Marks	AO	Notes
$(T_A = 7 + 12 + \cdots + 15 = 77$ $T_B = 8.5 + 1 + \cdots + 13 = 43)$ $U_A = "77" - \frac{1}{2}(8 \times 9) = 41$ $U_B = "43" - \frac{1}{2}(7 \times 8) = 15$	M1	1.3	Attempt at either U, ft their totals but *n* and m correct
Test statistic = 15 (or 41)	A1	1.3	cao
critical value = 13 (or 43)	B1	1.3	cao
ts = 15 > cv = 13 (or ts = 41 < 43) so cannot reject H_0	M1dep	2.1b	Comparison of their ts & correct cv in **same tail** Dependent on previous M1 PI
Thus there is **no evidence** (at the 5% sig. level) that the **median score** for **version B** is **lower** than the median score for version A….**OR** …There is	A1	2.1a	Must be in context and conclusion not too definite

	no evidence to support Robert's suspicion.			
5(b)	Yes it is reasonable to assume this because:	E1dep	3.1a	dependent on following E1
	Any one of the following • Different people are used in each group • A random sample was initially used • Random assignment of versions of the test was used	E1	3.1a	
		E1	3.1a	Fully explained in context
	SC1 comment on groups independent because each is doing a different version of the test. E1E1E0 max			

Total	**12**

<table>
<tr><th>Question</th><th>Scheme</th><th>Marks</th><th>AO</th><th>Notes</th></tr>
<tr><td>6(a)</td><td>Mean = 1.1
Use of $P(X = x) = \frac{\lambda^x}{x!}e^{-\lambda}$
P(3) = 0.0738, P(4) = 0.0203
P(5+) = 0.0054</td><td>B1
M1
A1
A1</td><td>1.2
1.2
1.2
1.2</td><td>cao
Their λ. PI implied by correct answer
Any 1 to 4 dp accuracy
All 3 to 4 dp accuracy
Note P(5)=0.0045 A0</td></tr>
<tr><td>6(b)</td><td>H_0 : X can be modelled by Poisson distn
H_1 : Poisson model cannot be used for X</td><td></td><td></td><td>Do not have to be stated</td></tr>
<tr><td></td><td><table>
<tr><th>No. of customers</th><th>O</th><th>E</th></tr>
<tr><td>0</td><td>73</td><td>59.922</td></tr>
<tr><td>1</td><td>50</td><td>65.916</td></tr>
<tr><td>2</td><td>30</td><td>36.252</td></tr>
<tr><td>3</td><td>20</td><td>13.284</td></tr>
<tr><td>4</td><td>7</td><td>3.654</td></tr>
<tr><td>5+</td><td>0</td><td>0.972</td></tr>
</table></td><td>M1

M1</td><td>1.3

1.3</td><td>Correct exp freqs, ft their probs × 180
(Allow 1 dp accuracy for E’s.)
Note ft P(5) Exp(5)=0.804
Last three classes combined correctly</td></tr>
<tr><td></td><td>New O’s and E’s:
<table>
<tr><td>O</td><td>73</td><td>50</td><td>30</td><td>27</td></tr>
<tr><td>E</td><td>59.922</td><td>65.916</td><td>36.252</td><td>17.910</td></tr>
</table></td><td>A1</td><td>1.3</td><td>At least 3 E’s correct to 1dp</td></tr>
<tr><td></td><td>$\sum\frac{(O-E)^2}{E}$ = 12.389
2.9+3.8+1.1+4.6</td><td>M1</td><td>1.3</td><td>Correct attempt at $\sum\frac{(O-E)^2}{E}$
At least two seen</td></tr>
<tr><td></td><td>12.4</td><td>A1</td><td>1.3</td><td>AWFW (12,13)</td></tr>
<tr><td></td><td></td><td></td><td></td><td>Alternatively, p=0.002(0) Correct p implies previous M1A1</td></tr>
<tr><td></td><td>df = 4 – 1 – 1 = 2
then 5% cv χ^2_2 is 5.99(1)</td><td>B1
B1ft</td><td>1.3
1.3</td><td>
ft their df</td></tr>
<tr><td></td><td>ts = 12.389 > 5.991

Evidence that the Poisson distribution is not a suitable model for the number of</td><td>M1</td><td>2.1b</td><td>Comparing their ts and correct cv with correct conclusion</td></tr>
</table>

	customers visiting the ATM during the evening. Lara's assumption is not reasonable.	A1	2.1a	**Alternatively**, comparison using *p*-value, $p = 0.002(0) < 0.05$ Correct conclusion in context
	Notes **(i)** No pooling gives $\sum \frac{(O-E)^2}{E} = 14.7 \sim 15.3$ for max M1 M0 A1 M1 A0 B0 B1ft M1 A0 (5/9) **(iii)** If E's taken to nearest whole number then : if pooled, $\chi^2 = 12.2$ for M1 M1 A0 M1 A0 B1 B1 M1 A0 (6/9) if not pooled, $\chi^2 = 14.7$ for (4/9) scored as in **(i)**.			
6(c)	(Customers do not appear to be arriving) at random/indep of each other (Customers do not appear to be arriving) at a constant average rate. Most of the time nobody arrives. Four or more customers very unlikely Some relevant comparison of O's and E's in context… eg More observed than expected in first and last categories suggests there are more 'busy' and 'quiet' times than a constant rate through the evening would suggest.	E1, E1 E1	3.1b 3.1b 3.1b	E1 for each sensible comment (max E2) For referencing **customers/people** in **context**
	Total	**16**		

Question	Scheme	Marks	AO	Notes
7(a)	$H_0: \mu_A = \mu_B = \mu_C$ H_1: at least 2 of the means differ			
	Total $SS = 57555 - \frac{1007^2}{18}$ $= 57555 - 56336.06 = 1218.94$	M1	1.3	SS Total. PI.
	Drinks SS = $\frac{320^2+322^2+365^2}{6} - 56336.06$ = 215.44	M1	1.3	SS between columns (condone small slip). PI.
	Subj SS = $\frac{178^2+144^2+\cdots+187^2}{3} - 56336.06$ = 874.28	M1	1.3	SS between rows (condone small slip) PI.
	Error SS= 1218.94 – 215.44 – 874.28 = 129.22	M1dep	1.3	Condone small slip - not if negative. Dep previous 3 M's. PI
	(ANOVA table: **ss**, **df**, **ms**) **Drinks**: 215.44, 2, 107.72 **Subjects**: 874.28, 5, 174.86 **Error**: 129.22, 10, 12.922 **Total**: 1218.94, 17	B1 M1dep	1.3 1.3	df correct. PI. MS=SS/df for Error and Drinks. PI. Dep previous M's
	$F = \frac{107.72}{12.922} = 8.34$	A1	1.3	**AWRT** 8.3 or p = 0.0074
	Critical value $F_{10}^{2}(0.05) = 4.103$ $OR\ F_{10}^{2}(0.01) = 7.559$	B1	1.3	Either cv Or p= 0.0074 < 0.05 or 0.01
	(8.34 > cv so) reject H_0	A1dep	2.1b	Comparison and correct conclusion. Dep all correct.
	There is significant evidence to suggest a difference between (at least two of the) mean endurance times recorded for the three energy drinks	E1dep	2.1a	For conclusion in context, not too definite. Dep all previous marks

Note: Use of 1-factor ANOVA here (as outlined in (d) below) scores M1M1M0M1B1M1A0B1A0E0 for max 6/10

Between	215.44	2	107.72	1.610
Error	1003.5	15	66.9	
Total	1218.94	17		

Critical value $F_{15}^{2}(0.05) = 3.682\ OR\ F_{15}^{2}(0.01) = 6.359$

p-value=0.233

7(b)	• Drink C appears to result in longer times (greater endurance) than either of the other two drinks SC1 E1E0 Drinks A and B are indistinguishable in terms of their effects on endurance	E1 E1	2.1a 2.1a	drink C referencing fact it had highest average
7(c)	Randomised Block (Design) **Advantages (examples):** • Any effect of differences between the subjects (volunteers) used in the experiment is accounted for • The test for differences between the 3 drinks is more powerful/sensitive than a completely randomised design would be • Error variance is reduced. • Possible to analyse differences between subjects (volunteers) • Every volunteer is tested with every drink • Reduces experimental error	E1 E1, E1	1.1 3.1a	Do not allow completely randomised block One mark for each separate point, max 2
7(d)	The blocking factor does seem to have been effective… …because there seems to be a difference between the **subjects** used in terms of endurance. Backed up with some **numerical** justification using the ANOVA table	E1dep E1 M1	3.1a 3.1a 3.1a	For "effective" dependent upon a reason oe PI by numerical justification For attempting one of these three numerical justifications

	eg MS between subjects (174.86) is the largest (ie subjects are the largest source of variation in times) OR F test for difference between subjects gives F =13.5 which is highly significant (1% CV = 5.636, 5% CV = 3.326) OR demonstrating that a completely randomised analysis obtained by pooling gives new error MS of 1103.5/15 = 66.9 and new ts for drinks of F=1.61. This is not significant so difference between drinks is then not detected. **Special Case** If numerical justification does **not** use the ANOVA table, ie uses only the totals/means for each subject then max E1 E1 M1 for 3/4	A1	3.1a	For completely correct numbers or calculations
	Total	**19**		

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