Mark Scheme (Results)

Autumn 2020

Pearson Edexcel GCE In A Level Statistics
(9ST0/02)
Paper 2: 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

Autumn 2020
Publications Code 9ST0_02_2010_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.
- 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)	$6.3 \pm z \times \frac{0.74}{\sqrt{14}}$	M1	1.3	PI (z or t)
	z = 1.96 (or 2.0 to 2 s.f.)	B1	1.3	PI or t=2.160
	CI is (5.91, 6.69)	A1	1.3	awrt (5.9, 6.7)
1(b)	5.8 is not in the confidence interval	B1ft	2.1b	oe comparison of 5.8 with their CI
	There is significant evidence that the mean weight of baby girls at 3 months of age is different from 5.8 kg	E1dep	2.1a	oe Dep on previous B1
1(c)	It is a **random sample** of babies weights at 3 months of age in Rose's clinic			oe
	The babies' **weights** at 3 months of age in Rose's clinic are **normally distributed**			oe
	The standard deviation of the babies' weights at 3 months of age in Rose's clinic is the same as the standard deviation for UK babies			oe
		E1, E1	3.1a, 3.1a	Any 2 for 2 marks
	Total	**7**		

Question	Scheme	Marks	AO	Notes
2	**Exact binomial method**			
	H_0: $\pi = 0.75$ H_1: $\pi > 0.75$	B1	1.3	Both correct Condone use of ‘p’
	n = 45 so let X ~ B (45, 0.75)	B1	1.3	Use of Binomial with n=45
	P ($X \geq 36$)	M1	1.3	
	= 1 – P($X \leq 35$)	M1	1.3	
	= 0.27997…	A1	1.3	AWRT 0.28 **or** CR:$X \geq 39$
	(0.28 > 0.05) so do not reject H_0	A1dep	2.1b	PI Dep on correct ts/ cv oe
	Julia’s survey result does not provide evidence that the **percentage preferring** unwrapped cucumbers is **higher than 75%** **or** There is **insufficient evidence** members of her gardening club **have a greater preference** for unwrapped cucumbers.	E1dep	2.1a	Dep on correct ts/ cv oe

Question	Scheme	Marks	AO	Notes
	Critical Region method			
	H_0: $\pi = 0.75$ H_1: $\pi > 0.75$	(B1)		Both correct Condone use of 'p'
	n = 45 so let X ~ B (45, 0.75)	(B1)		Use of Binomial with n=45
	$P(X \geq 39) = 0.044$	(M1)		Attempt to evaluate $P(X \geq x)$ to find CR
		(M1)		$P(X \geq 39)$ calculated condone $P(X \geq 38)=0.094$
	CR: $X \geq 39$	(A1)		CR:$X \geq 39$
	(36 < 39) so do not reject H_0	(A1dep)		PI Dep on correct ts/ cv oe
	Julia's survey result does not provide evidence that the **percentage preferring** unwrapped cucumbers is **higher than 75%** **or** There is **insufficient evidence** members of her gardening club **have a greater preference** for unwrapped cucumbers.	(E1dep)		Dep on correct ts/ cv oe

Question	Scheme	Marks	AO	Notes
2 (cont)	**Normal approximation method**			
	$H_0: \pi = 0.75$ $H_1: \pi > 0.75$	(B1)		Both correct Condone use of 'p'
	$\hat{p} = \frac{36}{45} = 0.8$	(B1)		
	$test\ statistic = \frac{0.8 - 0.75}{\sqrt{\frac{0.75 \times 0.25}{45}}}$	(M1)		PI Correct numerator Allow actual numbers rather than proportions
		(M1)		Correct denominator
	$test\ statistic = 0.775$ **or** *p*-value = 0.219…	(A1)		ts AWFW 0.77 - 0.78 **or** *p*-value
	(0.775 <1.6649 **or** 0.219… > 0.05) so do not reject H_0	(A1dep)		PI Dep on correct ts/ cv oe
	Julia's survey result does not provide evidence that the **percentage preferring** unwrapped cucumbers is **higher than 75%** **or** There is **insufficient evidence** members of her gardening club **have a greater preference** for unwrapped cucumbers.	(E1dep)		Dep on correct ts/ cv oe
	Total	**7**		

Question	Scheme	Marks	AO	Notes
3(a)	α = 0.05 OR df = n – 1 = 24	B1	1.3	Mark for use of α=0.05 if *p*-value method used **or** for correct df if critical value method used PI
	p-value = 0.00752… used with 0.05 **or** t = 2.619 used with $t_{24}(0.05) = 1.71$	M1	1.3	
	0.00752 < 0.05 **or** $2.619 ... > t_{24}(0.05) = 1.71$ ∴ reject H_0	A1dep	2.1b	Dep B1 M1
	There is evidence at the 5% level that the children did **better**, on average, in the test **after** practising the skill.	E1dep	2.1a	cso Dep on all previous marks
3(b)	Type I error is rejecting H_0 when H_0 is true	E1	3.1a	Possibly stated in context
	P(type I error) = 0.05	A1	2.1a	

Question	Scheme	Marks	AO	Notes
3(c)	**The conclusion may not be reliable as…**	E1	3.1b	Any sensible reason
	The differences in scores may not be normally distributed			
	The improvement in scores may not have been due to the computer program.			
	The pupils may not be representative of the population as a whole.			
	The conclusion is likely to be reliable as…			
	The 25 children are a random sample (of program users)			
	Total	**6**		

<table>
<tr><th>Question</th><th>Scheme</th><th>Marks</th><th>AO</th><th>Notes</th></tr>
<tr><td rowspan="3">4(a)</td><td rowspan="2">
<table>
<tr><td>$\frac{5\times57}{95}=\mathbf{3}$</td><td>2</td></tr>
<tr><td>18</td><td>12</td></tr>
<tr><td>3</td><td>2</td></tr>
<tr><td>1.8</td><td>1.2</td></tr>
<tr><td>7.8</td><td>5.2</td></tr>
<tr><td>23.4</td><td>15.6</td></tr>
<tr><td>57</td><td>38</td></tr>
</table>
</td><td>M1</td><td>1.3</td><td>Attempted method for calculating at least one expected value - PI</td></tr>
<tr><td>A1</td><td>1.3</td><td>At least one relevant expected value</td></tr>
<tr><td>Expected values < 5 for
Australia & South Africa, Far East and South America
So these regions need to be combined with each other (or others)</td><td>E1</td><td>3.1a</td><td></td></tr>
<tr><td rowspan="3">4(b)</td><td>H_0: no association (between region of the world and sex of player)
H_1: an association (between region of the world and sex of player)</td><td>B1</td><td>1.3</td><td>Both correct</td></tr>
<tr><td colspan="4">Observed frequencies:
<table>
<tr><th>Region of the World</th><th>Men</th><th>Women</th></tr>
<tr><td>Eastern Europe</td><td>11</td><td>19</td></tr>
<tr><td>USA & Canada</td><td>8</td><td>5</td></tr>
<tr><td>Western Europe</td><td>29</td><td>10</td></tr>
<tr><td>Other Regions of the World</td><td>9</td><td>4</td></tr>
</table>
</td></tr>
<tr><td></td><td>M1</td><td>1.3</td><td>Observed frequencies required for combined (Other regions) class only
PI</td></tr>
</table>

Question	Scheme	Marks	AO	Notes
4(b) (cont)	**Expected frequencies:** **Region of the World** | **Men** | **Women** Eastern Europe | 18 | 12 USA & Canada | 7.8 | 5.2 Western Europe | 23.4 | 15.6 Other Regions of the World | 7.8 | 5.2			
		A1	1.3	Expected values for all classes but may appear in 4(a) above PI awrt values in table
	Contribution to χ^2: **Region of the World** | **Men** | **Women** Eastern Europe | 2.722 | 4.083 USA & Canada | 0.005 | 0.007 Western Europe | 1.340 | 2.010 Other Regions of the World | 0.185 | 0.277			
		M1	1.3	Attempt at $\frac{(O-E)^2}{E}$ PI
	Test stat = $\frac{(11-18)^2}{18} + \cdots + \frac{(4-5.2)^2}{5.2}$	M1	1.3	Intention to sum PI
	$\chi^2 = 10.63$	A1	1.3	awrt 10.6

Question	Scheme	Marks	AO	Notes
4(b) (cont)	*p*-value = 0.0139 **or** cv of χ^2 at 5% level = 7.81	M1	1.3	*p*-value must be compared with $\alpha = 0.05$
	10.63 > 7.81 **or** 0.0139 < 0.05 so reject H_0	A1dep	2.1b	Dep on ts/ cv or *p*-value correct PI
	There is significant evidence of an **association** between **region** of the world and **sex** of player	E1dep	2.1a	Dep on ts/ cv or *p*-value correct
4(c)	The greatest contribution (4.08) to the association is from women and 'Eastern Europe" where more women were observed (19) than would be expected (12) from Eastern Europe to earn at least $1 million in 2018.	E1ft	2.1a	Or fewer men from Eastern Europe No numerical justification
		E1ft	2.1b	Full numerical justification
	Total	**14**		

Question	Scheme	Marks	AO	Notes
5(a)	Sign Test	B1	2.1a	
	H_0: population median = -0.02 H_1: population median > -0.02	B1	1.3	Condone use of η or two-tail H_1 May be awarded if seen in (b)
	$P(x \leq 2) = 0.0547 > 0.05$ Do not reject H_0	E1	2.1b	
5(b)	H_0: **population** average = -0.02 H_1: **population** average > -0.02	B1	1.3	One-tail H_1 used Allow mean or median oe μ/η
	Ranks of |X|: 2 1 3 4 5 6 7 8 9 10	M1	1.3	Only first two required check table
	W = 1 + 2 = 3	A1	1.3	Or 52
	cv = 11 (for α = 0.05, one-tailed)	B1	1.3	Or 44
	3 < 11 so reject H_0 (one-tailed)	M1	2.1b	Same tail comparison
	Sufficient evidence that the average **performance for SW** London is **better** than the average for England.	E1	2.1a	Allow mean or median Dep on correct ts/cv
5(c)	If the ranks are consistent with a symmetric population then Wilcoxon can be used. (Otherwise the sign test is needed).	E1	3.1a	
	Use Wilcoxon if possible because it is a more powerful test OR Use Wilcoxon if possible as it takes account of the sizes of the differences rather than just the +/- signs	E1	3.1a	
	Total	**11**		

Question	Scheme	Marks	AO	Notes
6(a)	For a double-blind trial neither the patients nor the doctors/researchers should know which treatment has been assigned to any patient.	E1	1.1	
	Patients would be aware of which diet/treatment they were receiving so it couldn't be a double-blind trial.	E1	1.1	
6(b)	H_0: $\mu_A = \mu_B$ H_1: $\mu_A \neq \mu_B$	B1	1.3	both
	Test stat = $\frac{10.7-3.1}{\sqrt{\left(\frac{9.6^2}{104}+\frac{7.0^2}{95}\right)}}$ = 6.419..	M1	1.3	$10.7 - 3.1$
		M1	1.3	$\sqrt{\left(\frac{9.6^2}{104}+\frac{7.0^2}{95}\right)}$
	Test stat = 6.419..	A1	1.3	AWRT 6.42 Ignore sign
	Critical value = ± 1.96 OR *p*-value P(z > 6.419..) = 1.37 x 10^{-10}	B1	1.3	Z > 6.419 implied by correct *p*-value
	(6.419>1.96 or 1.37 x 10^{-10} <0.025) so reject H_0	A1dep	2.1b	cv correct and compared with ts OR *p*-value compared to 0.025 PI Dep on cv/ ts or *p*-value correct
	There is significant evidence of a **difference** in mean **weight loss** between patients assigned to **Diet A** and those assigned to **Diet B**	E1dep	2.1a	Conclusion correct and in context; test all correct Dep on cv/ ts or *p*-value correct

Question	Scheme	Marks	AO	Notes
	Alternative			
	H_0: $\mu_A = \mu_B$ H_1: $\mu_A \neq \mu_B$	(B1)		oe both
	Test stat = $\frac{10.7-3.1}{\sqrt{\left(\frac{8.46^2}{104}+\frac{8.46^2}{95}\right)}}$ = 6.33..	(M1)		$10.7-3.1$
	$\sqrt{\left(\frac{8.46^2}{104}+\frac{8.46^2}{95}\right)}$	(M1)		Use their s_p^2
	Test stat = 6.33..	(A1)		AWRT 6.33 Ignore sign
	Critical value = ± 1.972 OR *p*-value $P(z > 6.33..) = 1.62 \times 10^{-9}$	(B1)		$z > 6.33$ implied by correct *p*-value
	($6.33>1.97$ or $1.62 \times 10^{-9} <0.025$) so reject H_0	(A1dep)		cv correct and compared with ts OR *p*-value compared to 0.025 PI Dep on cv/ ts or *p*-value correct
	There is significant evidence of a **difference** in mean **weight loss** between patients assigned to **Diet A** and those assigned to **Diet B**	(E1dep)		Conclusion correct and in context; test all correct Dep on cv/ ts or *p*-value correct
6(c)	The sample means are approximately normally distributed large samples so CLT applies	E1	3.1a	Large samples, CLT

Question	Scheme	Marks	AO	Notes
6(d)	Patients did not deviate from diet			
	Same scales used			
	Sample variances can be used in place of unknown population variances (large samples)			
		E1	3.1a	Any sensible explanation
6(e)	H_0: $\mu_A - \mu_B = 4$ H_1: $\mu_A - \mu_B > 4$	B1	1.3	oe Both
	Test stat = $\frac{10.7-3.1-(4)}{\sqrt{\left(\frac{9.6^2}{104}+\frac{7.0^2}{95}\right)}}$ = 3.040..	M1	1.3	$7.6-(4-0)$
	Test stat = 3.04	A1	1.3	AWRT 3.04
	Critical value = 1.6449 OR *p*-value $P(z > 3.04) = 0.00118$	B1	1.3	$Z > 3.04$ implied by correct *p*-value
	($3.04 > 1.6449$ or $0.00118 < 0.05$) so reject H_0	A1dep	2.1b	PI Dep on cv/ ts or *p*-value correct
	There is significant evidence that the mean weight loss of patients assigned to Diet A is at least 4kg more than that of patients assigned to Diet B. OR There is significant evidence the mean weight loss of patients assigned to Diet A is medically worthwhile	E1dep	2.1a	Conclusion correct and in context; test all correct Dep on cv/ ts or *p*-value correct

Question	Scheme	Marks	AO	Notes
6(e)	**Alternative**			
	H_0: $\mu_A - \mu_B = 4$ H_1: $\mu_A - \mu_B > 4$	(B1)		oe Both
	Test stat = $\frac{10.7-3.1-(4)}{\sqrt{\left(\frac{8.46^2}{104}+\frac{8.46^2}{95}\right)}}$ = 2.99..	(M1)		$7.6-(4-0)$
	Test stat = 2.99…	(A1)		AWRT 3.0
	Critical value = 1.653 OR *p*-value (z > 2.99) = 0.001395	(B1)		z > 2.99 implied by correct *p*-value
	(2.99 > 1.653 or 0.001395 < 0.05) so reject H_0	(A1dep)		PI Dep on cv/ ts or *p*-value correct
	There is significant evidence that the mean weight loss of patients assigned to Diet A is at least 4kg more than that of patients assigned to Diet B. OR There is significant evidence the mean weight loss of patients assigned to Diet A is medically worthwhile	(E1dep)		Conclusion correct and in context; test all correct Dep on cv/ ts or *p*-value correct

Total	**17**

<table>
<tr><th>Question</th><th>Scheme</th><th>Marks</th><th>AO</th><th>Notes</th></tr>
<tr><td>7(a)</td><td>$H_0: \mu_A = \mu_B = \mu_C = \mu_D$
H_1: at least two of μ_A, μ_B, μ_C, μ_D are different</td><td>B1</td><td>2.1a</td><td>or 1-factor ANOVA stated</td></tr>
<tr><td></td><td>T = 654.528
$\Sigma\Sigma\, x_{ij}^2 = 15508.274$</td><td>B1</td><td>1.3</td><td>Either</td></tr>
<tr><td></td><td>$SS_T = \Sigma\Sigma\, x_{ij}^2 - \frac{T^2}{n}$
$= "15508.274" - \frac{"654.528"^2}{29}$
$= 735.622$</td><td>M1</td><td>1.3</td><td>SS Total</td></tr>
<tr><td></td><td>$SS_B = \Sigma \frac{T_i^2}{n_i} - \frac{T^2}{n}$
$SS_B = \frac{190.078^2}{7} + \frac{184.401^2}{8} + \frac{133.191^2}{6} + \frac{146.858^2}{8} - \frac{"654.528"^2}{29}$
$= 291.742$</td><td>M1</td><td>1.3</td><td>SS between tyre brands</td></tr>
<tr><td></td><td>SS_E = 735.622-291.742 = 443.880</td><td>M1ft</td><td>1.3</td><td></td></tr>
<tr><td></td><td colspan="4"><table>
<tr><th>Source of variation</th><th>df</th><th>SS</th><th>MSS</th><th>Ratio</th></tr>
<tr><td>Between brands</td><td>3</td><td>291.742</td><td>97.247</td><td>5.477</td></tr>
<tr><td>Error</td><td>25</td><td>443.879</td><td>17.755</td><td></td></tr>
<tr><td>Total</td><td>28</td><td>735.621</td><td></td><td></td></tr>
</table></td></tr>
<tr><td></td><td></td><td>B1</td><td>1.3</td><td>df correct</td></tr>
<tr><td></td><td></td><td>M1</td><td>1.3</td><td>MS=SS/df for between brands and error</td></tr>
<tr><td></td><td>$F = \frac{97.247}{17.755} = 5.477$</td><td>A1</td><td>1.3</td><td>AWRT 5.5
OR p = 0.0049</td></tr>
</table>

Question	Scheme	Marks	AO	Notes
7(a) (cont)	**cv:** $F^3_{25}(0.05) = 2.991$ **or** $F^3_{25}(0.01) = 4.675$	B1	1.3	Either cv
	5.477 > cv so reject H_0	M1	2.1b	**or** p=0.0049<0.05 **or** 0.01 **PLUS** correct conclusion
	There is significant evidence to suggest that at least two of the four mean lives of the tyre brands are different.	A1	2.1a	Conclusion in context condone statement that brands A and D differ

Question	Scheme	Marks	AO	Notes
7(b)	**Possible comments**			
	Scatter diagram			
	The scatter diagram is consistent with normality within tyre brands so there is no reason to doubt the validity of the test on this basis.			
	The scatter diagram appears to show similar spread (variances) for all tyre brands so there is no reason to doubt the validity of the test on this basis.			Condone brand B has a greater spread so not valid
		E1	3.1a	Either of these
	Data collection			
	The tyre sample should be a completely randomised design (CRD). There may be bias due to the opportunistic nature of Daniel's sampling method.			
	Larger samples would have enabled tests for normality and comparisons of variances within 'treatments'.			
		E1	3.1a	Either of these
		E1	3.1a	One extra comment from either category

<table>
<tr><th>Question</th><th>Scheme</th><th>Marks</th><th>AO</th><th>Notes</th></tr>
<tr><td rowspan="2">7(c)(i)</td><td>Treatment factor/variable – tyre brand A, B, C, D
Blocking factor/variable – front or rear wheels
or
Table which could have rows/columns reversed
<table>
<tr><td></td><td colspan="4">Brand</td></tr>
<tr><td></td><td>A</td><td>B</td><td>C</td><td>D</td></tr>
<tr><td>Front Wheels</td><td></td><td></td><td></td><td></td></tr>
<tr><td>Rear Wheels</td><td></td><td></td><td></td><td></td></tr>
</table></td><td>B1</td><td>2.1a</td><td>Consideration of brands of tyre with front and rear wheels</td></tr>
<tr><td>Measure the Tyre life for each factor combination</td><td>B1</td><td>2.1a</td><td>oe
combine both</td></tr>
<tr><td>7(c)(ii)</td><td>2-factor ANOVA</td><td>B1</td><td>2.1a</td><td>Condone 2-way ANOVA</td></tr>
<tr><td></td><td>Total</td><td>17</td><td></td><td></td></tr>
</table>