Mark Scheme (Results)

Summer 2023

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

Summer 2023
Publications Code 9ST0_02_2306_MS

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.

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>Question</th><th>Scheme</th><th>Marks</th><th>AO</th><th>Notes</th></tr>
<tr><td>1(a)</td><td>Wilcoxon Rank Sum</td><td>M1</td><td>2.1a</td><td>Used or clearly stated
or
Mann-Whitney (U test) stated</td></tr>
<tr><td></td><td>$H_0: \eta_A = \eta_B$
$H_1: \eta_A > \eta_B$</td><td>B1</td><td>1.3</td><td>H_0: Samples from identical populations
H_1: Samples from different populations
oe</td></tr>
<tr><td></td><td><table><tr><th>Asia (A)</th><th>Not Asia (B)</th></tr><tr><td>1 20
2 19
3 18
5 16
7 14
9 12
16 5
17 4</td><td>4 17
6 15
8 13
10 11
11 10
12 9
13 8
14 7
15 6
18 3
19 2
20 1</td></tr></table></td><td>M1</td><td>1.3</td><td>Attempt at ranking as one group</td></tr>
<tr><td></td><td></td><td>A1</td><td>1.3</td><td>All ranking correct
Or reversed ranks</td></tr>
<tr><td></td><td>T_A=60 or T_B=150</td><td>M1dep</td><td>1.3</td><td>Effort to obtain T (reverse totals 108,102)</td></tr>
<tr><td></td><td>$U_A = 60 - \frac{1}{2}(8)(9) = 24$
$U_B = 150 - \frac{1}{2}(12)(13) = 72$</td><td>A1</td><td>1.3</td><td>Either U correct</td></tr>
</table>

Question	Scheme	Marks	AO	Notes
	cv=26 (or 70) "24" < "26" (or 72>70) Reject H_0 There is significant evidence at the 5% level that players from Asian countries do better than those from non-Asian countries	B1 M1 E1dep	1.3 2.1b 2.1a	Either cv correct Correct conclusion for their ts and their cv Dep on correct ts & cv
1(b)	**Possible improvements (not exhaustive)** Use a larger sample size Ensure sample is randomly selected Select sample from all professional players not just those taking part in the world cup Use more than one video game Use equal sample sizes Use sample from more than one event	E1	3.1a	Larger than 30 so CLT can be used Any sensible suggestion
1(c)	Numbers of points may not be normally distributed. Positions are discrete rather than continuous. It is not known whether the variances are equal.	E1	3.1a	Any of these

Question	Scheme	Marks	AO	Notes
1(d)	Can't necessarily extend to a category that it is a subset of/similar to (all video games) World Cup was the best players from each country – no relevance to a specific player Nothing suggesting it's related to the living situations Keito's statement is too definite	E1, E1	3.1b 3.1b	Any two comments
	Total	**13**		

Question	Scheme	Marks	AO	Notes

	Light pink	Pink	Red	Purple	Blue	Light Blue
Shipibo	6.77	12.31	10.88	9.03	10.26	5.75
Kastom	7.88	14.33	12.66	10.51	11.94	6.69
BaYaka	6.28	11.42	10.09	8.37	9.51	5.33
Brisbane	12.07	21.94	19.38	16.09	18.28	10.24

2(a)	See table above	M1	1.3	Any correct value in bold box to at least 1dp
		A1	1.3	All calculated expected values correct and shown

0.223	0.140	0.324	1.017	1.362	0.274
0.449	1.981	0.143	1.918	0.074	0.800
0.828	1.023	0.842	1.357	2.114	0.524
0.713	4.612	0.292	0.227	2.902	2.680

2(b)				PI
	$\frac{(6-9.03)^2}{9.03}+\cdots+\frac{(18-16.09)^2}{16.0009}$	M1	1.3	Method for contributions correct Give mark for one correct value seen Condone small slip
	= 26.819	A1	1.3	awfw 25~28 oe
	H_0: Colour preference and home location **are independent** H_1: Colour preference and home location **are not independent**	B1	1.3	Both hypotheses allow association
	cv = 24.996	B1	1.3	awrt 25
	"26.824" > "24.996" Reject H_0	M1	2.1b	Stated comparison of their ts and their cv or 0.0302 < 0.05

Question	Scheme	Marks	AO	Notes
	There is sufficient evidence that the colour preference of females aged 4-11 years is not independent of their home location.	E1dep	2.1a	oe Conclusion in context Dep on correct cv and ts
2(c)	Females aged 4-11 years in Brisbane were more likely to like pink than expected 4.612 is the largest contribution to the test statistic	E1 B1	2.1b 2.1b	 PI by 32 and 21.94
2(d)	The association for colour preference may be based on age rather than home location Select ages of females in equal ratios across 4-11 years.	E1 E1	3.1b 2.1b	
	Total	**12**		

Question	Scheme	Marks	AO	Notes
3(a)	H_0: $\mu = 2.82$ H_1: $\mu > 2.82$	B1	1.3	Both hypotheses
	$\bar{x} = 2.917$ $s = 2.193$	B1	1.2	PI Both correct $\bar{x}$ awrt 2.92 s awrt 2.2
	ts = $\frac{"2.917"-2.82}{\frac{"2.193"}{\sqrt{12}}}$	M1	1.3	PI Method for ts
	= 0.153	A1ft	1.3	PI awfw 0.15~0.16 **or** p-value = 0.441
	"0.153" < 1.796, Do not reject H_0	M1	2.1b	Comparison of their ts and correct cv **or** 0.441 > 0.05
	There is insufficient evidence to suggest that the average number of goals scored per match is greater for Women's Super League than men's Premier League	E1dep	2.1a	**or** use of population mean Conclusion in context dep correct cv and ts

Question	Scheme	Marks	AO	Notes
Alternative 1				
3(a)	H_0: $\mu = 2.82$ H_1: $\mu > 2.82$	(B1)		Both hypotheses
	$\bar{x} = 2.917$ $s = 2.193$	(B1)		PI Both correct $\bar{x}$ awrt 2.92 s awrt 2.2
	cv = $2.82 + 1.796 \times \frac{"2.193"}{\sqrt{12}}$	(M1)		PI Method for cv
	3.957	(A1ft)		PI awrt 3.96 **or** p-value = 0.441 ft only for incorrect rounding of sample mean and sd not *t*
	"2.917" < 3.957, Do not reject H_0	(M1)		Comparison of their ts and correct cv **or** 0.441 > 0.05
	There is insufficient evidence to suggest that the average number of goals scored per match is greater for women's Super League than men's Premier League	(E1dep)		**or** use of population mean Conclusion in context dep correct cv and ts

Question	Scheme	Marks	AO	Notes
Alternative 2				
3(a)	H_0: $\eta_W = 2.82$ H_1: $\eta_W > 2.82$	(B1)		oe H_0: (Population) median for women = 2.82 H_1: (Population) median for women > 2.82

Diff	-1.82	-0.82	+0.18	+2.18	+4.18	+1.18	-2.82	+1.18	+2.18	+0.18	-1.82	-2.82
Ranks	6.5	3	1.5	8.5	12	4.5	10.5	4.5	8.5	1.5	6.5	10.5

	Differences (see table above)	(M1)		Attempt at calculating differences
	Ranks (see table above)	(B1)		Attempt at ranking the differences
	T_+=41 or T-=37	(A1)		Either T correct
	"37">17 Do not reject H_0	(M1)		Comparison their smaller T with 17 oe 61
	There is insufficient evidence to suggest that the average number of goals scored per match is greater for Women's Super League than Men's Premier League	(E1dep)		Dep on correct ts & cv

Question	Scheme	Marks	AO	Notes
3(b)	H_0: $\mu = 3.37$ H_1: $\mu < 3.37$	B1	1.3	Both hypotheses
	ts = $\frac{1.93-3.37}{\sqrt{\frac{2.76}{100}}}$	M1	1.3	PI Method for ts
	= -8.67	A1	1.3	PI **or** p-value ≈ 0
	"-8.67" < -1.6449, Reject H_0	M1	2.1b	Comparison of their ts and correct cv **or** 0 < 0.05
	There is sufficient evidence to suggest that the average number of yellow cards per player is fewer for women's Super League than men's Premier League	E1dep	2.1a	**or** use of population mean Conclusion in context dep correct cv and ts

Question	Scheme	Marks	AO	Notes
Alternative				
3(b)	H_0: $\mu = 3.37$ H_1: $\mu < 3.37$	(B1)		Both hypotheses
	Standard error = $\sqrt{\frac{2.76}{100}}$	(M1)		PI
	cv = 3.10	(A1)		PI **or** p-value ≈ 0
	"1.93" < 3.10, Reject H_0	(M1ft)		Comparison of their ts and correct cv **or** 0 < 0.05
	There is sufficient evidence to suggest that the average number of yellow cards per player is fewer for Women's Super League than men's Premier League	(E1dep)		**or** use of population mean Conclusion in context dep correct cv and ts

Question	Scheme	Marks	AO	Notes
3(c)	Since sample is large and random therefore CLT applies	E1	3.1a	ref to $n > 30$ or CLT
	Therefore, population distribution is not required to be known (normal)	E1dep	3.1a	
3(d)	There is no evidence to suggest that women score more goals per game than men (based on the Super and Premier League)	E1ft	2.1b	Comment with regards to (a)
	However, in 7 out of the 12 womens' matches the number of goals per match was higher than the men's average.	E1	2.1b	Any relevant comment with regards to original sample data provided. e.g. The sample is only for one year with justification
	There is evidence to suggest that women do receive fewer yellow cards than men (based on the Super and Premier League)	E1ft	2.1b	Comment with regards (b)
		E1	2.1b	Style appropriate to general public. No difficult statistical vocabulary. No oversimplification.
	Total	**17**		

Question	Scheme	Marks	AO	Notes
4	**Similarities**			
	Both can be carried out on a single sample or a paired sample			
	Both are tests of the null hypothesis that the population median takes some value (or two population medians are equal in paired case)			Condone both test for a median (or difference in medians)
	Both require a random sample			
	Ignore differences of 0			
	Differences			
	Sign test can be used for non-numerical or summarised data, Wilcoxon signed-rank cannot			oe
	Wilcoxon signed-rank test takes into account size of differences, rather than just signs			Condone more powerful
	The critical value for the Wilcoxon signed-rank test is given in a table, but the critical value for a sign test is found using the binomial distribution			
	Wilcoxon signed-rank can be used to test for the mean			
		B1, B1, B1	1.3 1.3 1.3	Any 3 of the above reasons
	Assumptions			
	Sign test – no assumptions regarding distribution needed	B1	1.3	
	Wilcoxon Signed-Rank test – Need symmetry of distribution for single sample (or symmetry of differences for paired)	B1	1.3	
	Total	**5**		

<table>
<tr><th>Question</th><th>Scheme</th><th>Marks</th><th>AO</th><th>Notes</th></tr>
<tr><td>5(a)</td><td>H_0: $\mu_1 = \mu_2 = \mu_3$
H_1: At least two of the means differ</td><td>B1</td><td>1.3</td><td>oe
Or $\mu_i = \mu$
for $i = 1$ to 3</td></tr>
<tr><td></td><td>T = 28.81
$SS_T = 39.5831 - \frac{28.81^2}{21}$
$= 0.05852$</td><td>M1</td><td>1.3</td><td>SS_T method
PI</td></tr>
<tr><td></td><td>$SS_B = \frac{10.01^2}{7} + \frac{9.45^2}{7} + \frac{9.35^2}{7} - \frac{28.81^2}{21}$
$= 0.03615$</td><td>M1</td><td>1.3</td><td>SS_B method
PI</td></tr>
<tr><td></td><td></td><td>M1dep</td><td>1.3</td><td>SS_E method
No negative SS values.
Dep one previous M1</td></tr>
<tr><td></td><td><table><tr><td></td><td>SS</td><td>df</td><td>MS</td></tr><tr><td>Between generations</td><td>0.03615</td><td>2</td><td>0.018076</td></tr><tr><td>Error</td><td>0.02237</td><td>18</td><td>0.001243</td></tr><tr><td>Total</td><td>0.05852</td><td>20</td><td></td></tr></table></td><td>B1</td><td>1.3</td><td>PI
df 2 and 18</td></tr>
<tr><td></td><td></td><td>M1</td><td>1.3</td><td>PI
MS divide SS by df
ft on df values</td></tr>
<tr><td></td><td>$F = \frac{0.018076}{0.001243} = 14.5$</td><td>M1</td><td>1.3</td><td>PI
F method
ft on df values
or
p–value = 0.00017</td></tr>
<tr><td></td><td>Consideration of cvs
cv F^2_{15} = 3.682 and F^2_{20} = 3.493</td><td>B1</td><td>1.3</td><td>Testing both/either cvs
or
$F^2_{18} = 3.55$</td></tr>
</table>

Question	Scheme	Mark	AO	Notes
	"14.5" > "3.682" and/or "3.493" Reject H_0	M1	2.1b	Comparison of their ts and their cv or "14.5" > 3.55 or 0.000174 < 0.05 Correct conclusion for their ts and their cv
	There is significant evidence to suggest that **mean** wing size for male butterflies for at least 2 of the generations differ	E1dep	2.1a	Correct conclusion in context Dep on test fully correct
5(b)	Use a 2 factor ANOVA using location as a blocking factor Randomised Block Design	E1 E1	3.1a 2.1b	
5(c)	The experimental error will **reduce**	E1	3.1a	oe
5(d)	r = -0.924	B1	1.2	
5(e)	H_0: No association H_1: Negative association	B1	1.3	oe Both Accept H_0: $\rho = 0$ H_1: $\rho < 0$
	cv = -0.6694 1-tail	B1	1.3	Ignore sign for cv
	"-0.924" < "-0.6694" Reject H_0	M1	2.1b	Comparison of their ts and their cv, consistent sign
	There is significant evidence to suggest that there is a negative association between the distance, in km, from the Fukushima Nuclear Power Plant and the average abnormality rate, %, for the male butterflies	A1	2.1a	Conclusion in context dep ts, cv correct oe
	Total	**18**		

Question	Scheme	Marks	AO	Notes
6	***p*-value** (0.04<0.05) evidence of significant difference (so reject H_0).	B1	1.3	Knowledge of *p*-value
	….there is evidence to suggest that adults playing games with blue light have decreased average sleepiness levels compared to those playing games without blue light.	E1	2.1a	Explained fully in context
	Cohen's *d* Medium effect …..	B1	1.3	Knowledge of Cohen's *d*
	….so suggests that there is a medium size magnitude in the difference between the average sleepiness levels for the two groups.	E1	2.1a	Explained fully in context
	Total	**4**		

Question	Scheme	Marks	AO	Notes
7(a)	The differences between the total amounts of food consumed by **each child** on each of the two Wednesdays	E1 E1	1.3 1.3	Clear indication that differences must be obtained Explained fully in context indication **both** samples are required
7(b)	The difference of two normal distributions is a normal distribution (with the means subtracted and variances added)	E1	3.1a	
7(c)	It is not necessary that **both samples** are selected from normally distributed populations provided that the population of **differences** are normally distributed	E1 E1	3.1a 3.1a	Must have an attempt at a correct justification
7(d)	H_0: $\mu_d = 0$ H_1: $\mu_d \neq 0$ $ts = \frac{0.03-0}{\frac{0.89}{\sqrt{20}}} = 0.151$ (3sf) $cv = \pm 2.093$ $"-2.093" < "0.151" < "2.093"$ Do not reject H_0. No significant evidence to suggest any differences in average hunger levels so it was reasonable for Justin to ignore it	B1 B1 B1 M1 E1dep	1.3 1.3 1.3 2.1b 2.1a	Both Hypotheses **or** p-value = 0.882 Comparison their ts with their cv **or** 0.882 > 0.025 Conclusion in context dep ts, cv correct

Question	Scheme	Marks	AO	Notes
Alternative				
7(d)	$H_0: \mu_d = 0$ $H_1: \mu_d \neq 0$	(B1)		Both Hypotheses
	cv = $0 \pm 2.093 \times \frac{0.89}{\sqrt{20}}$	(B1)		PI Method for cv
	$= \pm 0.4165$	(B1)		PI **or** p-value = 0.882
	$"-0.417" < "0.03" < "0.417"$ Do not reject H_0.	(M1)		Comparison their ts with their cv **or** 0.882 > 0.025
	No significant evidence to suggest any differences in average hunger levels so it was reasonable for Justin to ignore it	(E1dep)		Conclusion in context dep ts, cv correct
7(e)	A test which does not rely on the sample being selected from a specific probability distribution.	E1	1.3	
	Total	**11**		