Mark Scheme (Results)

Summer 2019

Pearson Edexcel GCE

In Mathematics (8ST0) 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 8ST0_01_1906_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.

Paper 2 mark scheme

Question	Scheme	Marks	AO	Notes
1	**Possible stratifying factors (not exhaustive)** Notes on strata **Number of employees** Must not have any gaps between strata. First stratum must start at 1 or 0. Final stratum must be of form e.g. '100+'. **Business sector** At least 3 specific sensible sectors included. Must include 'other' category or similar. **Age of company** Strata must have units (e.g. years) clearly labelled. Must not have any gaps between strata. First stratum must start at 0. Final stratum must be of form e.g. '100 years +'. **Company turnover** Strata must have units (£) clearly labelled. Must not have any gaps between strata. First stratum must start at £0. Final stratum must be of form e.g. '£10 million +'.			

	Company profits Strata must have units (£) clearly labelled. Must not have any gaps between strata. Negative profits must be considered. First stratum most be of form e.g. 'less than £0'. Final stratum must be of form e.g. '£10 million +'.			
		E1	1.1	E1 for first **sensible** stratifying factor
		E1	1.1	E1 for second **sensible** stratifying factor
		E1dep	1.1	E1 for **sensible** and **complete** set of strata Dep on at least one of previous E1 marks
	Total	**3**		

Question	Scheme	Marks	AO	Notes
2(a)	It is **dependent**…	E1	2.1b	
	...on whether there are the same number of responses for each question.	E1	2.1a	
	Special case Binomial is **suitable**, $n=10$, $p=\frac{1}{\text{Number of choices}}$	(E1)		All three pieces of information seen. NOTE: This solution can score 1 mark max.
2(b)	Binomial is **unsuitable**…	E1	2.1b	
	…as the trials are not independent. **or** …as the probability of success is not fixed.	E1	2.1a	
	Total	**4**		

<table>
<tr><th>Question</th><th>Scheme</th><th>Marks</th><th>AO</th><th>Notes</th></tr>
<tr><td>3(a)</td><td>$\mathrm{P}(X > 7) = 0.097$</td><td>B1</td><td>1.2</td><td>awfw 0.09~0.102</td></tr>
<tr><td>3(b)</td><td>[Let X = number of first digits > 7]

$X \sim B(5, 0.097)$

$\mathrm{P}(X \geq 1) = 0.399$</td><td>

M1

A1</td><td>

2.1a

1.2</td><td>
PI
Evidence of binomial with n = 5

awfw 0.375~0.417</td></tr>
<tr><td>3(c)</td><td><table><tr><th>x</th><th>p</th><th>xp</th></tr><tr><td>1</td><td>0.3~0.302</td><td>0.3~0.302</td></tr><tr><td>2</td><td>0.17~0.18</td><td>0.34~0.36</td></tr><tr><td>3</td><td>0.12~0.13</td><td>0.36~0.39</td></tr><tr><td>4</td><td>0.095~0.1</td><td>0.38~0.4</td></tr><tr><td>5</td><td>0.078~0.08</td><td>0.39~0.4</td></tr><tr><td>6</td><td>0.065~0.07</td><td>0.39~0.42</td></tr><tr><td>7</td><td>0.055~0.06</td><td>0.385~0.42</td></tr><tr><td>8</td><td>0.05~0.052</td><td>0.4~0.416</td></tr><tr><td>9</td><td>0.04~0.05</td><td>0.36~0.45</td></tr></table>
$\mathrm{E}(X) = 3.44$</td><td>M1

A1

A1</td><td>1.1

1.1

1.2</td><td>PI
Attempt at finding xp

PI
At least one value correct

awfw 3.3~3.56</td></tr>
<tr><td>3(d)</td><td>If you took the first digits of the areas of the population of British lakes, you would expect the mean of these to equal $\mathrm{E}(X)$.</td><td>E1

E1dep</td><td>1.2

2.1a</td><td>Mention of population (or all, or large sample, or lots of different samples) and mean.

Fully correct explanation in context.
Must include ‘first digit’
Dep on previous E1</td></tr>
<tr><td></td><td>Total</td><td>8</td><td></td><td></td></tr>
</table>

Question	Scheme	Marks	AO	Notes
4	$H_0 : \mu = 8.5$ $H_1 : \mu \neq 8.5$	B1	1.3	Both correct If in words, must see '**population mean**'
	Test statistic, $z = \dfrac{8.0848 - 8.5}{\dfrac{1.763537}{\sqrt{50}}}$	M1	1.2	PI Correct method for z-score
	$= -1.665$	A1	1.2	awfw $-1.66 \sim -1.67$ Actual: -1.6647835
	(Lower) cv $= -1.960$	B1	1.3	awrt **Alt**: p-value = 0.048 (0.04~0.05) or 0.096 (0.08~0.1)
	$-1.665 > -1.960$ so accept H_0	B1dep	2.1b	Comparison of ts and cv and correct conclusion.(or *p*-value and sig level) **Alt**: 0.048 > 0.025 or 0.096 > 0.05 Dep on both ts and cv correct.(or *p*-value and sig level)
	Conclude that there is insufficient evidence (at the 5% significance level) to suggest that the (population) mean is not equal to 8.5.	E1dep	2.1b	Must not be definite in conclusion (e.g. do not accept 'the mean is not 8.5'). dep on both correct ts and cv .(or *p*-value and sig level)
	Total	**6**		

Question	Scheme	Marks	AO	Notes
5(a)(i)	[Let X = change in share price] $P(-5 < X < 5) = 0.433$	B1	1.2	awrt 0.43 Actual: 0.433175…
5(a)(ii)	$P(X > 35) = 0.00003$ **or** 3×10^{-5}	B1	1.2	awrt Actual: 3.091348×10^{-5} Do **not** accept 3E-5 oe
5(b)	**Possible strengths** The model is reliable close to the mean (or close to 0). The data is roughly symmetric/bell-shaped. Reference to large sample	E1	3.1a	Any sensible point
	Possible weakness The model underestimates probabilities in the tail(s). Figure 2 has evidence of outliers. Historical data (back to 1988) may not still be relevant today.	E1	3.1a	oe Accept 'is unreliable for high/low values' Any sensible point
	Alternative Consideration of $x \pm s$, $x \pm 2s$, or $x \pm 3s$ (without the clear conclusions of the above answers)	(E1)		E1 only
	Total	**4**		

Question	Scheme	Marks	AO	Notes
6(a)	Total (by end of) 28th Sep = 350 Total (by end of) 29th Sep = 525	M1	1.1	awfw 310 ~ 390 awfw 500 ~ 550 PI
	Total received on 29th Sep = 525 – 350 = 175	A1	1.1	awfw 110 ~ 240
6(b)	4 days	B1	1.1	awfw 3 ~ 4
6(c)	Predicting into the future may be inaccurate.	E1		Condone extrapolation
	The pattern of entry in 2017 may be different to that in 2016.	E1		or mention of spikes oe
	The prediction is likely to be too high.	E1		
	Alternative 1			
	The 2017 application figures have not been consistently higher than 2016…	(E1)	3.1b	
	…and there has been a recent surge in applications in 2017…	(E1)	3.1b	
	…so the difference (or ratio) between the two figures (on 7th Oct) may be abnormally high (leading to an overestimate).	(E1dep)	3.1b	Dep on at least one previous E1
	Alternative 2			
	There is a weekly surge of applications (on Fridays)…	(E1)		
	…(This) surge happened on the 6th of October in 2017, but not until the 7th of October in 2016…	(E1)		Clear explanation that the surges happen on different dates
	…so the difference (or ratio) between the two figures (on 7th Oct)	(E1dep)		Dep on at least one previous E1

	is likely to be abnormally high (leading to an overestimate).			
	Total	**6**		

<table>
<tr><th>Question</th><th>Scheme</th><th>Marks</th><th>AO</th><th>Notes</th></tr>
<tr><td>7(a)</td><td>(approx.) £33.4 million</td><td>B1</td><td>1.1</td><td>awrt
Actual: £33 361 000
Condone lack of units</td></tr>
<tr><td>7(b)</td><td>Assumption:
The distribution of (the population of) percentage changes is symmetrical.

H_0: (Population) median (percentage) change $= -6\%$
H_1: (Population) median (percentage) change $> -6\%$

Differences from -6%:
<table><tr><th>District</th><th>Differences (%)</th></tr><tr><td>A</td><td>3.08</td></tr><tr><td>B</td><td>-6.90</td></tr><tr><td>C</td><td>10.39</td></tr><tr><td>D</td><td>5.93</td></tr><tr><td>E</td><td>8.00</td></tr><tr><td>F</td><td>1.09</td></tr><tr><td>G</td><td>6.92</td></tr><tr><td>H</td><td>-3.80</td></tr><tr><td>I</td><td>3.65</td></tr></table>
Rank table:
<table><tr><th>District</th><th>+</th><th>-</th></tr><tr><td>A</td><td>2</td><td></td></tr><tr><td>B</td><td></td><td>6</td></tr><tr><td>C</td><td>9</td><td></td></tr><tr><td>D</td><td>5</td><td></td></tr><tr><td>E</td><td>8</td><td></td></tr><tr><td>F</td><td>1</td><td></td></tr><tr><td>G</td><td>7</td><td></td></tr><tr><td>H</td><td></td><td>4</td></tr><tr><td>I</td><td>3</td><td></td></tr></table></td><td>E1

B1

M1

M1ft</td><td>1.3

1.3

1.2

1.3</td><td>Accept
‘reduction = 6%’ etc

Accept
mean in place of median (since distribution assumed to be symmetrical)

Effort to find differences

Ranking their differences, smallest mod value with rank 1</td></tr>
</table>

	ts: $T = 10$ (or $T = 35$) cv = 8 (or 37) $10 > 8$ (or ts > cv) **or** $35 < 37$ (or ts < cv) so accept H_0 I found no evidence to suggest that the figure of 6% is too high	A1 B1 B1dep E1dep E1	1.2 1.3 2.1b 2.1a 2.1a	cao for either correct Comparison of ts and cv in **same tail** and correct conclusion Dep on both correct ts and cv Must be in context. Must not be definite in conclusion. Dep on ts and cv **both** correct Style appropriate to non-specialist. No difficult vocabulary or concepts used.
7(c)	The sample is **not random** (which is a requirement for validity of the test). **or** The sample may not represent the population as it is an opportunity sample.	E1	3.1a	
7(d)	Make a numbered list of all district councils (in England)… ...and use the random number function (RAND or RANDBETWEEN) to produce a sample of numbers… …and look up (or use LOOKUP/VLOOKUP function) the numbers on the list to find the corresponding districts for your sample.	E1 E1 E1	3.1a 3.1a 3.1a	Must see **numbered** and **list** oe. Must see **random** oe
	Total	**14**		

Question	Scheme	Marks	AO	Notes
8(a)(i)	The beep test's data is split into two variables (or fields/columns)… **or** There are three data variables… …and Pearson's PMCC requires a single variable for each factor (or they need to be combined). **or** …and Pearson's PMCC requires precisely two variables. **Alternatives** • The data is not continuous • The data is not bivariate normal • The beep test data is not normally distributed	E1 E1 (E1) (E1) (E1)	3.1a 3.1a	**NOTE**: These are the **only** responses worth E2. Any one of these seen scores E1 (**E1 max**)
8(a)(ii)	(Rank the beep test data and) use Spearman's rank correlation coefficient. Calculate a new variable for beep test data equal to (e.g.): • total number of laps • level as a decimal with completed lap = 0.1 • distance run Only use the highest level reached in the beep test.	E1, E1	3.1a, 3.1a	Cannot gain E2 from this point only E1 for each sensible suggestion
8(b)	**Spearman's rank solution** H_0: no association (between laps completed and maximum leg press)	B1	1.3	**or** '…no correlation between the rank orders…'

	H_1: some association (between laps completed and maximum leg press) Ranked data (ascending): **Member / Beep / Leg** A 8 7 B 6 8 C 11 9 D 1 4 E 5 11 F 7 6 G 10 2 H 2 3 I 4 1 J 3 5 K 9 10 Alternative ranked data (descending): **Member / Beep / Leg** A 4 5 B 6 4 C 1 3 D 11 8 E 7 1 F 5 6 G 2 10 H 10 9 I 8 11 J 9 7 K 3 2 ts: $r_s = 0.391$ cv $= 0.609$	M1 A1 B1 B1 B1	1.3 1.3 1.2 1.3	**or** '…variables are independent…' Condone $r_s=0$ etc Condone one-tailed test Attempt at ranking PI Both columns correctly ranked PI r_s stated or seen **or** SRCC oe clearly stated awrt [Accept 0.5273 for one-tail test]

	$0.391 < 0.609$ (or ts < cv) so accept H_0 There is insufficient evidence to suggest that there is an association between the beep test and the leg press results (in the population).	E1dep	2.1a	oe Must be in context. Must not be definite in conclusion. Dep on ts and cv both correct
	Pearson's PMCC solution H_0: $\rho = 0$ H_1: $\rho \neq 0$	(B1)		**or** 'no correlation/ some correlation' **or** 'independent/ not independent' Condone one-tailed test
		(M1)		Attempt to calculate total beep laps completed or other variable
		(A1)		All beep laps calculated correctly
	Accept All beep laps out by 10 **Accept** All beep laps divided by 10 **Accept** All beep laps multiplied by 20 **Accept** Combinations of the above			
		(B1)		r stated or seen

Member	Beep laps
A	60
B	36
C	85
D	8
E	26
F	51
G	81
H	16
I	20
J	19
K	79

				or
	ts: $r = 0.310$			PMCC oe clearly stated
		(B1)		awrt
	cv $= 0.602$	(B1)		0.5214 for one-tailed test
	$0.310 < 0.602$ (or ts < cv) so accept H_0			
	There is insufficient evidence to suggest that there is correlation between the beep test and the leg press results.	(E1dep)		oe Must be in context. Must not be definite in conclusion. Dep on ts and cv both correct.
	SC **Completed laps column removed**			
	H_0: $\rho = 0$ H_1: $\rho \neq 0$	(B1)		**or** 'no correlation/ some correlation' **or** 'independent/ not independent' Condone one-tailed test
	Clear use of only 'highest level reached' column used.	(M1) (A0)		
	ts: $r = 0.252$	(B1)		r stated or seen **or** PMCC oe clearly stated
	cv $= 0.602$	(B0) (B1)		0.5214 for one-tailed test

		(E0)		
8(c)(i)	(Sprint cyclist is) E… …as she would need very strong legs (but not necessarily need too much stamina).	B1 E1	2.1b 2.1a	 oe **or** above average on both.
8(c)(ii)	(Marathon runner is) G… …as she would need good running skills and stamina (but not necessarily need to be able to push heavy weights with her legs).	B1 E1	2.1b 2.1a	 oe
	Total	**15**		

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