

Mark Scheme (Results)

Summer 2018

Pearson Edexcel GCE In Statistics (8ST0) Paper 02

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 <u>www.edexcel.com</u> or <u>www.btec.co.uk</u>. Alternatively, you can get in touch with us using the details on our contact us page at <u>www.edexcel.com/contactus</u>.

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 2018 Publications Code 8ST0_02_1806_MS All the material in this publication is copyright © Pearson Education Ltd 2018

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.

Question	Sch	Marks	AO	Notes	
1(a)	Disproportional		E1	1.1	Seen anywhere
	stratifie	d (sampling)	E1	1.1	Seen anywhere
					Subtract second E1 for any additional terms except:
					randomunbiased
1(b)	Unrestricted	E1	1.1	Seen anywhere	
	random (E1	1.1	Seen anywhere	
					Accept 'simple random sampling with replacement' for E2
					Subtract second E1 for any additional terms except:
					• unbiased
		Total	4		

Question	Scheme	Marks	AO	Notes
2	Possible criticisms			
	• The bar length is inconsistent with the numbering (for prison officers and school teachers).			
	• All (or most) pay has gone down in the bar chart between 2005 and 2015 whereas pay has continued to rise since 2010 in the (line) graph.			Clear explanation that pay has decreased in the bar chart. Clear explanation that pay has increased in the line graph.
	• We don't know whether inflation has been included.			
	• 'CPI' has not been defined.			or 'index not defined'
	• Units on vertical axis unclear.			
	• The horizontal scale is unclear.			Do not accept 'vertical axis does not start at 0'
	• Not all public-sector workers are included (e.g. MPs)			
	• It is difficult to read values from the line graph.			Accept 'cannot read off exact values'.
	• The data in one table starts at 2005, and the other starts at 2010.			or 'one table ends in 2015 and the other ends in 2017'
	• It is not clear what date in 2005 and 2015 the bar data is from			
		E1, E1, E1	3.1a, 3.1a, 3.1a	Each bullet point scores E1 (max E3)
		E1	2.1a	Explanations in context

Special case			
Insufficient inform regarding number 2.	nation given s used in Figure	(E1)	This solution scores max E1
	Total	4	

Question		Sche	me			Mark	ĸs	AO		Ν	otes	
3(a)	H ₀ : $\eta_M = \eta_F$ H ₁ : $\eta_M < \eta_F$							1.3	oe [Ha me po] [Ha fro po] [Ha not fro	b: male dian = pulation b: samp m ident pulation b: femal longer gs, on a	popula female 1 medi les tak ical 1s] e frog than r	ation e an] en s are nale e]
	М	60	58	55	62	67	53	55	67	54	55	56
	F	57	72	77	86							
	M (rank)	9	8	4	10	11.5	1	4	11.:	5 2	4	6
	F (rank)	7	13	14	15							
						M1		1.3	Sej me two	oarating asurem o group	g ents ir s.	ito
						A1		1.3	Alleit	ranks oner grou	correct 1p.	t for
	$T_M = 9 + 8 +$ or $T_F = 7 + 1$	+6= 3+14	= 71 +15 =	49		M1		1.3	Att	empt a either	t rank group.	total
	$U_M = 71 - \frac{1}{2}(11 \times 12) = 5$ or $U_F = 49 - \frac{1}{2}(4 \times 5) = 39$							1.3	Att Mu or	empt a st inclu 4×5'	t either ide '11	r <i>U</i> ∣×12,
	ts = 5 (or 39)				A1		1.3	cae)		
	(1-tail test, a	$\alpha = 0.0$	05)									
	cv = 9 (or 35	5)				B1		1.3	cae)		

5<9 (or 39 so re) > 35) ject Ho).			B1ft	-		2.1b	Com sensi sensi tail	parisc ble ts ble cv	on of ti and in sa	heir me	
Conclude that there is evidence (at the 5% significance level) that the population median body length for <u>female</u> frogs is <u>longer</u> than that of <u>male</u> frogs. or Conclude that there is evidence (at the 5% significance level) to support Ramesh's belief.					E1de	р		2.1a	Musi Shou defir conc Dep and o	t be in ild not iite in lusion on bot cv.	conte t be th corr	ext. rect ts	
Alternative (rank descending) H_0: $\eta_M = \eta_F$ H_1: $\eta_M < \eta_F$								(1.3)	oe [H ₀ : medi popu [H ₀ : from popu	male j an = f ilatio ilation ident ilation	popul emale n med es tak ical s]	ation ian] en	
М	60	58	55	62	67	53	3	55	67	54	55	56	
F	57	72	77	86									
M (rank)	7	8	12	6	4.5	15	5	12	4.5	14	12	10	
F (rank)	9	3	2	1									
					(M1))		(1.3)	Sepa meas two g atten	rating sureme groups npt at	ents in s, and ranks.	ito	
					(A1))		(1.3)	All r eithe	anks c r grou	orrect p	for	

$T_M = 7 + 8 + + 10 = 105$ or $T_F = 9 + 3 + 2 + 1 = 15$	(M1)	(1.3)	Attempt at rank total for either group
$U_M = 105 - \frac{1}{2}(11 \times 12) = 39$ or $U_F = 15 - \frac{1}{2}(4 \times 5) = 5$	(M1)	(1.3)	Attempt at either U Must include '11×12' or '4×5'
ts = 5 (or 39)	(A1)	(1.3)	cao
(1-tail test, $\alpha = 0.05$)			
cv = 9 (or 35)	(B1)	(1.3)	cao
5 < 9 (or $39 > 35$) so reject H ₀ .	(B1dep)	(2.1b)	Comparison of their sensible ts and sensible cv in same tail
Conclude that there is evidence (at the 5% significance level) that the population median body <u>length</u> for <u>female</u> frogs is longer than that of <u>male</u> frogs. or Conclude that there is evidence (at the 5% significance level) to support Ramesh's belief.	(E1dep)	(2.1a)	Must be in context. Should not be definite in conclusion. Dep on both correct ts and cv.

3(b)	[Let F = number of sample, or M = nu the sample]	of females in the umber of males in			
	F or $M \sim B(15, 1)$	0.5)	B1	2.1a	PI
	$P(F \le 4)$				
	or		M1	1.2	
	$P(M \ge 11)$				
	= 0.0592		A1	1.2	awfw 0.059~0.0593
					Accept 2-tail prob: awfw 0.117~0.12
3(c)				2.1b	oe
					Accept any reasonable significance level.
	0.0592 > 0.05 (or there is insufficien support Gill's beli	0.118 > 0.05) so at evidence to ef.	E1ft		Accept 'no, as the probability is still quite high'.
	11				Condone any sensible comment on potential frog behaviour.
					ft from (b)
		Total	13		

Question	Scheme	Marks	AO	Notes				
4 (a)	[Equation of form $y = a + bx$]							
	a = -0.241	M1	1.2	Actual: -0.24143212				
	<i>b</i> = 8.175	M1	1.2	Actual: 8.17459341				
	Regression line: y = -0.241 + 8.175x	A1	1.2	cao				
				–1 mark for coefficients switched				
	SC							
	Outlier not removed: $y = 3.756 + 5.617x$ scores M1 M1							
	Missing value = $1.85 \sim 1.86$: $y = -0.296 + 8.27x$ scores M1 M1							
	Outlier not removed AND missing value = $1.85 \sim 1.86$: $y = 3.685 + 5.77x$ scores M1							
	x on y: $y = 0.719 + 0.054x$ scores M1 M1							
	AbsMag on Distance: $y = 1.284 + 0.061x$ scores M1							
	Wrong outlier (clearly) and correct ed	quation for	m scores	M1				
4(b)	$y = 8.175 \times 1.2 - 0.241$	M1	2.1b	Attempt to sub x = 1.2 into their linear regression equation. PI				
	=9.57	A1dep	2.1b	awfw 9.56~9.6 Accept: • awrt 10.5 • awrt 9.63 • awrt 10.6 • awfw 0.783~0.784 • awrt 2.01 Dep on acceptable equation seen in (a)				

4(c)	Possible explanat	tions			
	There is another g stars which were n (the first scatter di The distribution is normal. The small sample representative of t An outlier was rer not an outlier in th	roup of (brighter) not in Figure 3 iagram). s not bivariate is clearly not he population. moved, but this is ne larger sample.	E1	3.1b	oe Accept 'the relationship is nonlinear'
4(d)	Separate into diffe	erent <u>star types</u>	E1	3.1a	oe
	and produce <u>dif</u> <u>lines</u> (or curves) f	E1	3.1a		
		Total	8		

Question	Scheme	Marks	AO	Notes
5(a)(i)	$P(S) = \frac{8387}{12308}$ or $\frac{16774}{24616}$ or 0.681	B1	1.2	oe awfw 0.68~0.682 Actual: 0.681426714
5(a)(ii)	$P(L') = \frac{12308 - 933}{12308}$	M1	1.2	12308-933 or 10749+626 or 11375 seen
	$=\frac{11375}{12308} \text{ or } \frac{22750}{24616} \text{ or } 0.924$	A1	1.2	oe awfw 0.92~0.925 Actual: 0.9241956
	Alternative			
	$P(L) = \frac{933}{12308} = 0.076$	(M1)	(1.2)	PI
	P(L') = 1 - 0.076 = $\frac{11375}{12308}$ or $\frac{22750}{24616}$ or 0.924	(A1)	(1.2)	oe awfw 0.92~0.925 Actual: 0.9241956
	SC $\frac{10749}{12308}$ or awfw 0.87~0.874 scores N	I1	·	

5(b)				Outer box drawn
		B1	1.1	Two overlapping circles
				Circles labelled S & L
		B 1	1.1	At least two figures correct and in correct region
	C 7625 (762) 171	B1	1.1	All four figures correct.
	3750			SC: Venn diagram with probabilities scores max B2
				S: 0.6195 $S \cap L$: 0.0619 L: 0.0139 Ext: 0.3047
				SC: -B1 for S/L interchanged
5(c)				PI
	$P(S \mid L') = \frac{P(S \cap L')}{P(L')} = \frac{n(S \cap L')}{n(L')}$	M1	1.2	Evidence of multiplication rule used (prob or freq)
				Denominator = 11375, 22750, or 91 seen scores M1
	$=\frac{7625}{15250}=\frac{15250}{100}=\frac{61}{100}=0.670$	A1	1.2	oe awfw 0.67~0.671
	11375 22750 91			Actual: 0.67032967
	SC:			
	$P(L' S)$ calculated correctly $=\frac{7623}{8387}$	$\frac{5}{7} = 0.909$	scores N	11

5(d)	$P(S L') \neq P(S)$ or $P(S L') \neq P(S)$ or $P(S) \times P(L') \neq P(S)$	$S \mid L$) $\neq P(S \cap L')$	B1	2.1b	Dep on previous B1
	So S and L are r Special case $\frac{7625}{11375} \neq \frac{8387}{12308}$ (o or $\frac{7625}{11375} \neq \frac{762}{933}$ (or or $\frac{8387}{12308} \times \frac{11375}{12308} \neq \frac{11375}{12308} \neq \frac{11375}{12308}$ (or 0.681× so S and L' are r	or $0.670 \neq 0.681$) $0.670 \neq 0.817$) $\frac{7625}{2308}$ $(0.924 \neq 0.620)$ not independent .	(B1)	2.10	Correct comparison of numerical probability and correct conclusion scores single B1
		Total	10		

Question	Scheme	Marks	AO	Notes
6(a)(i)	[Let <i>M</i> represent number of claims made by motorcycle riders]			
	E(M) = 0.223	B1	1.2	awfw 0.22~0.224 Actual: 0.2226
6(a)(ii)	$E\left(M^2\right) = 0.304$	M1	1.2	Clear attempt at finding $E(M^2)$ PI
	$\operatorname{Var}(M) = E(M^{2}) - (E(M))^{2}$ $= 0.254$	A1	1.2	awfw 0.25~0.255 Actual: 0.2544492
	SC sd=0.504 scores M1			
6(b)	[Let <i>C</i> represent number of claims made by car drivers]			
	$\operatorname{Var}(C) = 0.486^2 = 0.236$			awrt
	or $\sigma_{_{M}} = \sqrt{0.254} = 0.504$	B1ft	2.1b	awrt Actual: 0.5044296
	$ \mu_{M} = 0.223 $ or $ E(C) = 0.146 $	B1ft	2.1b	Recognition that mean = expected value Must be convinced
	The motorcycle riders have more claims, on average, than car drivers and the spread is roughly the same.	E1dep	2.1a	Both statements needed Must see <u>average</u> and <u>spread</u> oe Dep on both B1 marks

6(c)	Possible statemer exhaustive)	nts (not			
	The text just says may not necessari	'average'. This ly be the mean.			
	Responses in the 2 questionnaire may 2015.	2016 v include claims in			
	People may not w they've had lots of	ant to admit if f claims.			or people may have lied
	The (car) sample is representative, as people would be h questionnaire with questions.	may not be only certain happy to answer a h (at least) 18			
	The (car) sample may not be representative as it only includes people who have visited the website.				
	The motorcycles are only from a single company, so there may be bias (e.g. low risk insurance companies).		E1, E1, E1	3.1b, 3.1b, 3.1b	E1 for each sensible statement (Max E3)
		Total	9		

Question	Scheme	Marks	AO	Notes
7(a)	[Let X represent shares at close on 23^{rd} , and Y represent shares at close on 24^{th}]			
	H ₀ : $\eta_X = \eta_Y$ H ₁ : $\eta_X > \eta_Y$	B1	1.3	oe [H ₀ : population median share price at close on 23/6 = population median share price at close on 24/6] [H ₀ : Population median difference=0] [H ₀ : Share prices at close on 24/6 were not lower than close on 23/6, on average]
	Vodafone+Worldpay-Sky-TUI AG-St. James's Place-Direct Line-	M1	1.3	Attempt to calculate signs of differences (or signed differences) between the correct two columns.
	SSE-Schroders-Smiths Group-Royal Dutch Shell+	A1	1.3	2+, 8–
	(Under H ₀),			
	$S \sim B(10, 0.5)$	M1	1.3	PI
	$P(S \ge 8) = 0.0547$ or $P(S \le 2) = 0.0547$	A1	1.2	awfw 0.0546~0.055 Actual: 0.0546875 Alt: Critical region {9,10} or {0,1} with justification

0.0547 > 0.05	B1	2.1b	Comparing with 0.05 Alt: Comparing 2 or 8 with critical region in correct tail
so accept H ₀ The average share price was not any lower at close on 24/06/2016 than it was at close on 23/06/2016. or Brexit had no significant effect on the average share price.	E1ft dep	2.1a	Dep on previous B1 and attempt at sign test.
 (There is significant evidence that) the average share price fell between close on 23/06/2016 and open on 24/06/2016. or (There is significant evidence that) the average share price fell straight after the Brexit vote. 	E1	2.1a	
	E1	2.1a	Explanations in context. Must see 'share price', 'close/open', and dates.
	E1	2.1a	Style appropriate to broadsheet. No difficult statistical vocabulary used. No oversimplification.

7(b)	Possible suggestie exhaustive)	ons (not			
	Use a much bigge companies (e.g. F LSE)	r sample of TSE 100, whole			Accept 'companies outside LSE'
	Extend to further of June.	dates after the 24 th			
	Extend to further a 23 rd of June.	dates before the			
	Use open data for dates)	23/6 (or both			
	Consider hour by hour stock prices.				
	Group companies analyse which ind affected.	npanies by industry to hich industries have been		3.1a, 3.1a	E1 for each sensible comment (max E2)
		Total	12		