GCE

Further Mathematics B (MEI)

Y432/01: Statistics minor

Advanced GCE

Mark Scheme for Autumn 2021

OCR (Oxford Cambridge and RSA) is a leading UK awarding body, providing a wide range of qualifications to meet the needs of candidates of all ages and abilities. OCR qualifications include AS/A Levels, Diplomas, GCSEs, Cambridge Nationals, Cambridge Technicals, Functional Skills, Key Skills, Entry Level qualifications, NVQs and vocational qualifications in areas such as IT, business, languages, teaching/training, administration and secretarial skills.

It is also responsible for developing new specifications to meet national requirements and the needs of students and teachers. OCR is a not-for-profit organisation; any surplus made is invested back into the establishment to help towards the development of qualifications and support, which keep pace with the changing needs of today's society.

This mark scheme is published as an aid to teachers and students, to indicate the requirements of the examination. It shows the basis on which marks were awarded by examiners. It does not indicate the details of the discussions which took place at an examiners' meeting before marking commenced.

All examiners are instructed that alternative correct answers and unexpected approaches in candidates' scripts must be given marks that fairly reflect the relevant knowledge and skills demonstrated.

Mark schemes should be read in conjunction with the published question papers and the report on the examination.

Annotations and abbreviations

Annotation in scoris	Meaning
✓and ✗	
BOD	Benefit of doubt
FT	Follow through
ISW	Ignore subsequent working
M0, M1	Method mark awarded 0, 1
A0, A1	Accuracy mark awarded 0, 1
B0, B1	Independent mark awarded 0, 1
E	Explanation mark 1
SC	Special case
^	Omission sign
MR	Misread
BP	Blank page
Highlighting	
Other abbreviations in mark scheme	**Meaning**
E1	Mark for explaining a result or establishing a given result
dep*	Mark dependent on a previous mark, indicated by *. The * may be omitted if only previous M mark.
cao	Correct answer only
oe	Or equivalent
rot	Rounded or truncated
soi	Seen or implied
www	Without wrong working
AG	Answer given
awrt	Anything which rounds to
BC	By Calculator
DR	This indicates that the instruction **In this question you must show detailed reasoning** appears in the question.

Question			Answer	Marks	AOs	Guidance	
1	(a)		$k + 2k + 5k + 10k + 17k = 1$ $35k = 1$ so $k = \frac{1}{35}$	M1 A1 [2]	2.4 1.1	AG	
1	(b)		The distribution has (strong) **negative skew**	B1 [1]	1.1		
1	(c)		$E(X) = 4\frac{1}{7} = \frac{29}{7} = 4.143\ldots$ $Var(X) = 1\frac{23}{245} = \frac{268}{245} = 1.094\ldots$	B1 B1 [2]	1.1a 1.1	BC Accept any equivalent form. BC Accept any equivalent form.	Decimal answers should agree to at least 2 significant figures.
1	(d)		$E(Y) = 10\frac{5}{7} = \frac{75}{7} = 10.714$ $Var(Y) = 27\frac{17}{49} = \frac{1340}{49} = 27.347$	B1FT B1FT [2]	1.1 1.1	BC Accept any equivalent form. FT their E(X) from **(c)** BC Accept any equivalent form. FT their Var(X) from **(c)**	Decimal answers should agree to at least 2 significant figures.

Question			Answer	Marks	AOs	Guidance	
2	(a)		*a* is the independent variable since the values of *a* are not subject to random variation	**B1** **[1]**	**2.4**	**B1**: values of *a* are controlled **B0:** *d* is dependent on *a*	Explanation required
2	(b)		$\boldsymbol{d} = -1.104\boldsymbol{a} + 197.1$	**M1** **A1** **[2]**	**3.3** **1.1**	For either $-1.104(a)$ or 197.1 BC	$y = -1.104x + 197.1$ scores **M1 A0**
2	(c)		estimate = 130.9 (m)	**B1FT** **[1]**	**1.1**	FT from **(b)** if the value is plausible from the scatter diagram.	Accept 130 as rounded to 2 significant figures.
2	(d)		Because this would be extrapolation and it is possible that the relationship is different for young children	**B1** **B1** **[2]**	**2.2b** **2.4**	For 'extrapolation' **B1:** a 5-year-old child may not be able to read yet	**B0** for comment about child not being able to drive
2	(e)		Residual = $150 - (-1.104\ldots \times 40 + 197.1\ldots)$ $= -3.0$	**M1** **A1FT** **[2]**	**1.1** **1.1**	Subtraction other way around scores **M1** only Allow –2.9 (using 1.104 and 197.1) FT from **(b)**	
2	(f)		Because the values of *a* are non-random so it makes no sense to try to predict them.	**B1** **[1]**	**3.2b**	Should show understanding of a purpose of a regression line being to make predictions.	

Question			Answer	Marks	AOs	Guidance	
3	(a)		The **sample** must be random	**B1** **[1]**	**1.2**		
3	(b)		E8: $\frac{23\times 29}{120} = 5.5583$	**B1**	**1.1**		
			C13: $\frac{(28-33.1417)^2}{33.1417}$	**M1**	**1.1a**		
			$= 0.7977$	**A1** **[3]**	**1.1**		
3	(c)		H_0: No association between age and smoking (status) H_1: Some association between age & smoking (status)	**B1**	**3.4**	Both hypotheses needed Use of 'correlation' in place of 'association' is **B0**	
			Degrees of freedom = 2	**B1**	**3.3**		
			Critical value = 5.991	**B1**	**1.1**	or $\chi_2^2(6.4801) = 0.9608$ or p-value = 0.0392	
			Test statistic = 3.3642 + 0.6964 + … + 0.2792 = 6.4801	**B1FT**	**1.1**	FT their value of C13	
			6.4801 > 5.991	**M1**	**2.2b**	or 0.9608 > 0.95 or 0.0392 < 0.05	Comparing their test and critical values leading to a conclusion.
			There is sufficient evidence at the 5% level to suggest that there is association between age and smoking (status)	**A1** **[6]**	**3.5a**	Correct test and critical values required Use of 'correlation' in place of 'association' is **A0**	Conclusion in context
3	(d)		For 16-34 year olds the contribution of 3.3642 suggests that more are smokers than would be expected.	**E1**	**2.3**	Max of 2 marks out of 3 if no contributions are mentioned.	Should take each age group in turn and discuss status
			For 35-59 year olds things are (approximately) as expected if there were no association.	**E1**	**3.5a**	Allow equivalent statements about non-smokers	Max 2 marks if done differently
			For people aged 60 and over the contribution of 1.1775 suggests that fewer are smokers than would be expected.	**E1** **[3]**	**3.2a**		

Question			Answer	Marks	AOs	Guidance	
4	(a)		Because (the grouping of points on) the scatter diagram appears to be very roughly elliptical, the distribution may be bivariate Normal.	**E1** **E1** **[2]**	**3.5a** **2.4**	For elliptical For full answer (dep. on first mark) "the **data** is bivariate Normal" is **E0**	
4	(b)		$S_{xy} = 12317.2 - \frac{1}{10} \times 351.9 \times 350.0 = 0.7$ $S_{xx} = 12384.5 - \frac{1}{10} \times 351.9^2 = 1.139$ $S_{yy} = 12251.2 - \frac{1}{10} \times 350.0^2 = 1.2$ $r = \frac{S_{xy}}{\sqrt{S_{xx} S_{yy}}} = \frac{0.7}{\sqrt{1.139 \times 1.2}}$ $= 0.60$	**M1** **M1** **M1** **A1** **[4]**	**1.1a** **1.1** **3.3** **1.1**	 For either S_{xx} or S_{yy} For general form including square root **SC2** for correct value without any intermediate calculations	**Detailed reasoning** required. (0.598750…)
4	(c)		H_0: $\rho = 0$ H_1: $\rho > 0$ where ρ is the population product moment correlation coefficient between x and y For $n = 10$, 5% critical value (one tailed) = 0.5494 Since 0.60 > 0.5494 the result is significant. There is sufficient evidence (at the 5% level) to suggest that there is positive correlation between directly measured and satellite measured **salinity level**.	**B1** **B1** **B1** **M1** **A1FT** **[5]**	**3.3** **2.5** **3.4** **1.1** **2.2b**	For both hypotheses For defining ρ For critical value For comparison of test statistic and critical value leading to a conclusion FT for conclusion in context	H_0: no correlation in the **population** H_1: positive correlation in the **population** scores first **B1**
4	(d)		It means that one can be more confident that the correlation is genuine, rather than simply the result of random variation.	**E1** **[1]**	**2.4**		
4	(e)		The test shows that there is almost certainly some real correlation in the population. However, it is uninformative/of little use since the effect size is so small.	**E1** **E1** **[2]**	**3.5a** **2.2b**		

Question			Answer	Marks	AOs	Guidance	
5	(a)	(i)	$P(X \geq 10) = 1 - 0.9161$ $= 0.0839$	**M1** **A1** **[2]**	**3.1b** **1.1**	Or $P(X \geq 10) = 1 - P(X \leq 9)$ BC (0.083924…)	
5	(a)	(ii)	Poisson (60) $P(X < 50) = 0.0844$	**M1** **A1** **[2]**	**3.3** **1.1**	soi BC (0.084406…)	
5	(b)		0.9161^{20} $= 0.1733$	**M1** **A1FT** **[2]**	**3.1a** **1.1**	soi or 0.1732 from calculator value in **(a)**	
5	(c)		**Use** of $\text{Var}(X) = \text{E}(X)$ $\text{Var}(X) = \text{E}(X^2) - (\text{E}(X))^2$ $\mu = 12 - \mu^2$ $\mu = 3$ $P(X < 5) = 0.8153$	**M1** **M1** **A1** **A1** **[4]**	**3.1b** **2.2a** **1.1** **1.1**	 For equation BC (0.815263…)	

Question			Answer	Marks	AOs	Guidance	
6	(a)	(i)	Uniform distribution on {1, 2, … , n} $P\left(X \le \frac{1}{4}n\right) = \frac{1}{4}$	**B1** **[1]**	**3.3**		
6	(a)	(ii)	$P\left(X \le \frac{1}{4}n\right) = \frac{k}{4k+1}$ $= \frac{\frac{1}{4}(n-1)}{n}$ $= \frac{n-1}{4n}$	**M1** **M1** **A1** **[3]**	**2.2a** **1.1** **1.1**		Single fraction required for **A1**
6	(b)		E(X) = 51 Var(X) = 850 SD(X) = 29.1 So require P(21.9 < X < 80.1) $P(22 \le X \le 80) = \frac{59}{101}$	**M1** **M1** **A1** **[3]**	**3.1b** **1.1a** **1.1**	For either For required interval	

OCR (Oxford Cambridge and RSA Examinations)
The Triangle Building
Shaftesbury Road
Cambridge
CB2 8EA

OCR Customer Contact Centre

Education and Learning
Telephone: 01223 553998
Facsimile: 01223 552627
Email: general.qualifications@ocr.org.uk

www.ocr.org.uk