OCR
Oxford Cambridge and RSA

GCE

Mathematics (MEI)

Unit **4769**: Statistics 4

Advanced GCE

Mark Scheme for June 2015

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.

OCR will not enter into any discussion or correspondence in connection with this mark scheme.

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
SC	Special case
^	Omission sign
MR	Misread
Highlighting	
Other abbreviations in mark scheme	**Meaning**
E1	Mark for explaining
U1	Mark for correct units
G1	Mark for a correct feature on a graph
M1 dep*	Method mark dependent on a previous mark, indicated by *
cao	Correct answer only
oe	Or equivalent
rot	Rounded or truncated
soi	Seen or implied
www	Without wrong working

Subject-specific Marking Instructions for GCE Mathematics (MEI) Statistics strand

a Annotations should be used whenever appropriate during your marking.

The A, M and B annotations must be used on your standardisation scripts for responses that are not awarded either 0 or full marks. It is vital that you annotate standardisation scripts fully to show how the marks have been awarded.

For subsequent marking you must make it clear how you have arrived at the mark you have awarded.

b An element of professional judgement is required in the marking of any written paper. Remember that the mark scheme is designed to assist in marking incorrect solutions. Correct *solutions* leading to correct answers are awarded full marks but work must not be judged on the answer alone, and answers that are given in the question, especially, must be validly obtained; key steps in the working must always be looked at and anything unfamiliar must be investigated thoroughly.

Correct but unfamiliar or unexpected methods are often signalled by a correct result following an *apparently* incorrect method. Such work must be carefully assessed. When a candidate adopts a method which does not correspond to the mark scheme, award marks according to the spirit of the basic scheme; if you are in any doubt whatsoever (especially if several marks or candidates are involved) you should contact your Team Leader.

c The following types of marks are available.

M
A suitable method has been selected and *applied* in a manner which shows that the method is essentially understood. Method marks are not usually lost for numerical errors, algebraic slips or errors in units. However, it is not usually sufficient for a candidate just to indicate an intention of using some method or just to quote a formula; the formula or idea must be applied to the specific problem in hand, eg by substituting the relevant quantities into the formula. In some cases the nature of the errors allowed for the award of an M mark may be specified.

A
Accuracy mark, awarded for a correct answer or intermediate step correctly obtained. Accuracy marks cannot be given unless the associated Method mark is earned (or implied). Therefore M0 A1 cannot ever be awarded.

B
Mark for a correct result or statement independent of Method marks.

E
A given result is to be established or a result has to be explained. This usually requires more working or explanation than the establishment of an unknown result.

Unless otherwise indicated, marks once gained cannot subsequently be lost, eg wrong working following a correct form of answer is ignored. Sometimes this is reinforced in the mark scheme by the abbreviation isw. However, this would not apply to a case where a candidate passes through the correct answer as part of a wrong argument.

d When a part of a question has two or more 'method' steps, the M marks are in principle independent unless the scheme specifically says otherwise; and similarly where there are several B marks allocated. (The notation 'dep *' is used to indicate that a particular mark is dependent on an earlier, asterisked, mark in the scheme.) Of course, in practice it may happen that when a candidate has once gone wrong in a part of a question, the work from there on is worthless so that no more marks can sensibly be given. On the other hand, when two or more steps are successfully run together by the candidate, the earlier marks are implied and full credit must be given.

e The abbreviation ft implies that the A or B mark indicated is allowed for work correctly following on from previously incorrect results. Otherwise, A and B marks are given for correct work only — differences in notation are of course permitted. A (accuracy) marks are not given for answers obtained from incorrect working. When A or B marks are awarded for work at an intermediate stage of a solution, there may be various alternatives that are equally acceptable. In such cases, exactly what is acceptable will be detailed in the mark scheme rationale. If this is not the case please consult your Team Leader.

Sometimes the answer to one part of a question is used in a later part of the same question. In this case, A marks will often be 'follow through'. In such cases you must ensure that you refer back to the answer of the previous part question even if this is not shown within the image zone. You may find it easier to mark follow through questions candidate-by-candidate rather than question-by-question.

f Wrong or missing units in an answer should not lead to the loss of a mark unless the scheme specifically indicates otherwise. Candidates are expected to give numerical answers to an appropriate degree of accuracy, with 3 significant figures often being the norm. Small variations in the degree of accuracy to which an answer is given (e.g. 2 or 4 significant figures where 3 is expected) should not normally be penalised, while answers which are grossly over- or under-specified should normally result in the loss of a mark. The situation regarding any particular cases where the accuracy of the answer may be a marking issue should be detailed in the mark scheme rationale. If in doubt, contact your Team Leader.

g Rules for replaced work

If a candidate attempts a question more than once, and indicates which attempt he/she wishes to be marked, then examiners should do as the candidate requests.

If there are two or more attempts at a question which have not been crossed out, examiners should mark what appears to be the last (complete) attempt and ignore the others.

NB Follow these maths-specific instructions rather than those in the assessor handbook.

h For a *genuine* misreading (of numbers or symbols) which is such that the object and the difficulty of the question remain unaltered, mark according to the scheme but following through from the candidate's data. A penalty is then applied; 1 mark is generally appropriate, though this may differ for some units. This is achieved by withholding one A mark in the question.

Note that a miscopy of the candidate's own working is not a misread but an accuracy error.

Question			Answer	Marks	Guidance	
1	(i)		The integral is $-\exp(-x^2/a)$, limits 0 and infinity. (Which evaluates to 0 – (–1) =1.)	B1 [1]	Answer given	
	(ii)		$\mathrm{E}(X^2) = \int_0^\infty \frac{2}{a} x^3 \exp\left(-\frac{x^2}{a}\right)\mathrm{d}x$ *or* $\int_0^\infty x^2 \frac{2}{a} x \exp\left(-\frac{x^2}{a}\right)\mathrm{d}x$	M1		
			$= \left[x^2\left(-\exp\left(-\frac{x^2}{a}\right)\right)\right]_0^\infty + \int_0^\infty 2x\exp\left(-\frac{x^2}{a}\right)\mathrm{d}x$	M1A1	M1 parts	
			$= 0 + a = a$	A1	Cao	
			$\mathrm{E}(X^4) = \int_0^\infty \frac{2}{a} x^5 \exp\left(-\frac{x^2}{a}\right)\mathrm{d}x$ *or* $\int_0^\infty x^4 \frac{2}{a} x \exp\left(-\frac{x^2}{a}\right)\mathrm{d}x$	M1		
			$= \left[x^4\left(-\exp\left(-\frac{x^2}{a}\right)\right)\right]_0^\infty + \int_0^\infty 4x^3\exp\left(-\frac{x^2}{a}\right)\mathrm{d}x$	A1		
			$= 0 + 2a \times a = 2a^2$	A1 [7]	Answer given so working must be convincing	
	(iii)		Likelihood is $\left(\frac{2}{a}\right)^n \prod X_i \exp\left(-\frac{1}{a}\sum X_i^2\right)$	M1A1A1	M1 for recognisable attempt to obtain the likelihood. A1 for $\prod X_i$. A1 for the exp fn.	
			(log likelihood as given) Differentiate log likelihood	M1		
			to obtain $-\frac{n}{a} + \frac{1}{a^2}\sum X_i^2$	A1		
			Set this to zero to obtain $\mathrm{MLE}(a) = \frac{1}{n}\sum X_i^2$	M1A1 [7]	Justification of a maximum not required	

Question			Answer	Marks	Guidance	
	(iv)		E(MLE(a)) = $\frac{1}{n}\sum \mathrm{E}\left(X_i^2\right) = \frac{1}{n} n a = a$	B1		
			That is, the MLE is unbiased	E1	Seen or very clearly implied	
			Var(MLE(a)) = $\frac{1}{n^2}\sum \mathrm{Var}\left(X_i^2\right) = \frac{1}{n^2}\sum\left(\mathrm{E}\left(X_i^4\right) - \left(\mathrm{E}\left(X_i^2\right)\right)^2\right)$	M1M1		
			$= \frac{1}{n^2} n\left(2a^2 - a^2\right) = \frac{a^2}{n}$	A1		
				[5]		
	(v)		Maximum likelihood estimate (also unbiased) of a is 1.471	B1	Explanation not required	
			with estimated standard error (a / √ n) 0.1471.	B1	if calculations are correct	
			95% CI is 1.471 ± 1.960 × 0.1471 = 1.471 ± 0.288	M1A1	Accept 2 for 1.960	
				[4]		
				[24]		

Question			Answer	Marks	Guidance	
2	(i)		$\mathrm{M}_Y(\theta) = \mathrm{E}(\mathrm{e}^{\theta Y}) = \int_0^\infty \mathrm{e}^{\theta y} \frac{1}{\sqrt{2\pi y}} \mathrm{e}^{-\frac{1}{2}y} \mathrm{d}y$	M1, A1		
			$= \int_0^\infty \frac{1}{\sqrt{2\pi y}} \mathrm{e}^{-\frac{1}{2}y(1-2\theta)} \mathrm{d}y$	A1		
			Substitute $u = y(1-2\theta)$, $\mathrm{d}u = \mathrm{d}y(1-2\theta)$ to obtain	M1		
			$\int_0^\infty \frac{\sqrt{1-2\theta}}{\sqrt{2\pi u}} \mathrm{e}^{-\frac{1}{2}u} \frac{1}{1-2\theta} \mathrm{d}u$	A1		
			$= (1-2\theta)^{-1/2} \int_0^\infty \frac{1}{\sqrt{2\pi u}} \mathrm{e}^{-\frac{1}{2}u} \mathrm{d}u$	B1		
			$= (1-2\theta)^{-1/2}$		Answer given	
				[6]		
	(ii)		Either expand the mgf as a power series:			
			$(1-2\theta)^{-1/2} = 1 + \theta + 3\frac{\theta^2}{2!} + \ldots$	M1A1		
			$\mathrm{E}(Y) = 1$ (coefficient of θ)	B1		
			$\mathrm{E}(Y^2) = 3$ (coefficient of $\theta^2/2!$)	B1		
			Hence $\mathrm{Var}(Y) = 2$	B1		
			– –	**[5]**		
			Or differentiate the mgf:	M1		
			1st derivative simplifies to $(1-2\theta)^{-3/2}$	A1		
			Putting $\theta = 0$ gives $\mathrm{E}(Y) = 1$	B1		
			2nd derivative simplifies to $3(1-2\theta)^{-5/2}$	A1		
			Putting $\theta = 0$ gives $\mathrm{E}(Y^2) = 3$, hence $\mathrm{Var}(Y) = 2$	B1		
				[5]		
	(iii)		For independent rvs X and Y,			
			$\mathrm{M}_{X+Y}(\theta) = \mathrm{M}_X(\theta)\,\mathrm{M}_Y(\theta)$	B1		

Question			Answer	Marks	Guidance	
			Hence $\mathrm{M}_U(\theta) = (1-2\theta)^{-n/2}$	B1		
			$\mathrm{E}(U) = n$	B1		
			$\mathrm{Var}(U) = 2n$	B1		
				[4]		
	(iv)		$\mathrm{M}_W(\theta) = \mathrm{E}\left(\exp\left(\theta\left(\frac{U-n}{\sqrt{2n}}\right)\right)\right)$	B1		
			$= \exp\left(-\frac{\theta n}{\sqrt{2n}}\right)\mathrm{E}\left(\exp\left(\frac{\theta}{\sqrt{2n}}U\right)\right)$	B1	Or by use of general linear transformation result	
			$= \exp\left(-\frac{\theta n}{\sqrt{2n}}\right)\mathrm{M}_U\left(\frac{\theta}{\sqrt{2n}}\right)$	B1		
			$= \exp\left(-\frac{\theta n}{\sqrt{2n}}\right)\left(1-\frac{2\theta}{\sqrt{2n}}\right)^{-n/2}$	B1		
			Hence $\ln(\mathrm{M}_W(\theta))$ as given Expanding $\ln(\mathrm{M}_W(\theta))$ gives			
			$-\sqrt{\frac{n}{2}}\,\theta+\frac{n}{2}\left(\sqrt{\frac{2}{n}}\,\theta+\frac{1}{2}\left(\sqrt{\frac{2}{n}}\,\theta\right)^2+\text{terms of order } n^{-3/2}\right)$	M1A1	Award a maximum of 3 marks from here on if no account is taken of terms beyond θ^2.	
			Convincing simplification to $\frac{1}{2}\theta^2$ + terms of order $n^{-1/2}$	B1	(answer given)	
			Hence tends to $\frac{1}{2}\theta^2$.			
			The mgf therefore tends to $\exp(\frac{1}{2}\theta^2)$.	B1	Uniqueness of mgfs may be implied	
			That is, the distribution of W tends to the standard Normal.	B1		
				[9]		
				[24]		

Question			Answer	Marks	Guidance
3	(a)	(i)	Sample means:		
			standard 107.2766, GM 113.6192	B1	
			Sample variances:		
			standard 138.2243, GM 121.9372	B1	
			H_0: $\mu_1 = \mu_2$ (the means of the underlying distributions)		Must be clear that the hypotheses refer to underlying means
			H_0: $\mu_1 \neq \mu_2$	B1	
			Test statistic: $\dfrac{113.6192 - 107.2766}{\sqrt{\frac{138.2243}{30} + \frac{121.93721}{26}}} = 2.08(0103)$	M1A1	
			Compare with *z* distribution	B1	
			Critical value, 2 tails, 5%, 1.960	B1	
			So 2.08(0103) is just in the critical region		
			Hence reject H_0 and conclude that there is a difference in mean weight between standard and GM tomatoes.	B1B1	
			Assumption:		
			the tomatoes may be regarded, in some sense, as random samples of their respective varieties.	B1	
				[10]	
		(ii)	Additional assumptions:		
			underlying Normality of the tomatoes' weights	B1	
			common variance in the underlying distributions	B1	
			Given the sample sizes it seems safe to use the Normal distribution.	E1	
			The Normal test is better in that it makes fewer assumptions.	E1	
				[4]	
	(b)	(i)	The tomatoes should be chosen at random.	B1	
			The panel should not know which tomatoes are GM / standard.	B1	
			Making fine judgements on the appearances of tomatoes is unlikely to be reliable.	B1	Accept other sensible comments
				[3]	

Question			Answer	Marks	Guidance	
		(ii)	Rank sums: Standard 134, GM 76	M1A1		Or Mann-Whitney 79, 21
			H_0: GM and std tomatoes have, on the whole, the same appearance H_1: GM tomatoes have, on the whole, a better appearance than std tomatoes	B1	Hope for, but don't expect, a formulation in terms of a shift in location parameter for underlying distributions of appearances	
			(lower rank sum for GM indicates that the evidence is in the correct tail)	B1		
			Wilcoxon rank sum test	B1	May be implied	
			Critical value for $m = n = 10$, 1 tail, 1% level is 74	B1		Critical value 19
			The observed value of 76 is not in the critical region so accept H_0. That is, conclude that there is insufficient evidence to suppose that, on the whole, GM tomatoes have a better appearance than standard tomatoes.	B1		
				[7] **[24]**		

Question			Answer	Marks	Guidance	
4	(a)	(i)	Y_{ij} denotes the jth observation (or measurement) in the ith group	B1	Accept just Y_{ij} denotes the observations (measurements)	
			μ is the underlying mean for whole population	B1	Withhold these marks if there is no reference to populations	In marking section (a), evidence of understanding is more important than precise wording.
			α_i denotes the population mean difference for the ith group	B1		
			ε_{ij} denotes the random error in the ijth observation (or measurement)	B1		
				[4]		
		(ii)	ε_{ij} ~ ind N(0, σ^2),	B1B1	B1 ind N, B1 zero mean	
			where σ^2 is constant across groups.	B1	May be implied by later comment	
			The variance ratio test (F test) requires underlying Normality (and common variance).	B1	Common variance may be implied by earlier statement	
				[4]		
		(iii)	H_0: all the α_i are zero	B1	Accept clear wording in terms of all group effects being zero, or all groups having same underlying distribution	
			H_1: not all the α_i are zero	B1		H_1: all the α_i are non-zero scores zero
			If H_0 is accepted then we proceed on the basis that the group means are all equal. I.e. there are no ‘treatment effects’.	E1		
			If H_0 is rejected the we proceed on the basis that the group means are not all equal. I.e. there are some ‘treatment effects’.	E1		
				[4]		
	(b)	(i)	k = 4 (number of routes is 4)	B1		
			N = 19 (number of journeys is 19)	B1		
				[2]		

Question			Answer	Marks	Guidance	
		(ii)		M1	Evidence of understanding how to complete the table	

Source of var'n	SS	df	MS	F
Between groups	333.77	**3**	**111.2567**	**3.98(113)**
Within groups	**419.19**	15	**27.9460**	
Total	752.96	18		

Answer	Marks	Guidance
(table above)	A4	for the figures shown in bold -1 each error
H_0: all routes have the same underlying mean time H_1: routes do not all have the same underlying mean time	B1	Hypotheses may be implied by later comments
The observed variance ratio is compared with $F_{3,\,15}$	M1	May be implied by use of a correct value
The 5% critical value is 3.29 (tables) or 3.287382 (calculator)	A1	
Observed value is in critical region, so reject H_0.	B1	Or equivalent wording
That is, proceed on the assumption that the routes do differ in underlying mean time.	B1 **[10]**	
	[24]	

OCR (Oxford Cambridge and RSA Examinations)
1 Hills Road
Cambridge
CB1 2EU

OCR Customer Contact Centre

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

www.ocr.org.uk

For staff training purposes and as part of our quality assurance programme your call may be recorded or monitored

Oxford Cambridge and RSA Examinations
is a Company Limited by Guarantee
Registered in England
Registered Office; 1 Hills Road, Cambridge, CB1 2EU
Registered Company Number: 3484466
OCR is an exempt Charity

OCR (Oxford Cambridge and RSA Examinations)
Head office
Telephone: 01223 552552
Facsimile: 01223 552553

PART OF THE CAMBRIDGE ASSESSMENT GROUP