OCR
RECOGNISING ACHIEVEMENT

GCE

Mathematics (MEI)

Advanced GCE

Unit **4769:** Statistics 4

Mark Scheme for June 2012

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, OCR Nationals, 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.

Any enquiries about publications should be addressed to:

OCR Publications
PO Box 5050
Annesley
NOTTINGHAM
NG15 0DL

Telephone: 0870 770 6622
Facsimile: 01223 552610
E-mail: publications@ocr.org.uk

Annotations

Annotation	**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

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)		$\mathrm{P}(X \le x) = \mathrm{F}_B(x).\frac{1}{2} + \mathrm{F}_G(x).\frac{1}{2}$ ie cdf of X is $\mathrm{F}(x) = \frac{1}{2}\{\mathrm{F}_B(x) + \mathrm{F}_G(x)\}$ ie (by differentiating) pdf of X is $\mathrm{f}(x) = \frac{1}{2}\{\mathrm{f}_B(x) + \mathrm{f}_G(x)\}$	M1 A1 A1 **[3]**	use of cdfs Answer given
1	(ii)		$\mathrm{E}(X)\left(= \frac{1}{2}\left\{\int x\mathrm{f}_B(x)\mathrm{d}x + \int x\mathrm{f}_G(x)\mathrm{d}x\right\}\right) = \frac{1}{2}\mu_B + \frac{1}{2}\mu_G$	M1 **[1]**	[answer given; needs *some* indication of method]
1	(iii)		$\mathrm{E}(X^2) = \int x^2\mathrm{f}(x)\mathrm{d}x$ $= \frac{1}{2}\left\{\int x^2\mathrm{f}_B(x)\mathrm{d}x + \int x^2\mathrm{f}_G(x)\mathrm{d}x\right\}$ Use of "$\mathrm{E}(X^2) = \sigma^2 + \mu^2$" $= \frac{1}{2}\left\{\sigma^2 + {\mu_B}^2 \quad + \quad \sigma^2 + {\mu_G}^2\right\}$ $\therefore \mathrm{Var}(X) = \mathrm{E}(X^2) - \{\mathrm{E}(X)\}^2$ $= \sigma^2 + \frac{1}{2}{\mu_B}^2 + \frac{1}{2}{\mu_G}^2 - \frac{1}{4}{\mu_B}^2 - \frac{1}{4}{\mu_G}^2 - \frac{1}{2}\mu_B\mu_G$ $= \sigma^2 + \frac{1}{4}(\mu_B - \mu_G)^2$	M1 M1 M1 A1 M1 A1 A1 **[7]**	 Answer given
1	(iv)		[Central Limit Theorem] Approx dist of $\overline{X}$ is $\mathrm{N}\left(\frac{1}{2}\mu_B + \frac{1}{2}\mu_G,\ \frac{1}{2n}\left(\sigma^2 + \frac{1}{4}(\mu_B - \mu_G)^2\right)\right)$ B1 B1 B1 B1	 B4 **[4]**	 4 marks as shown
1	(v)		$\overline{X}_{st} = \frac{1}{2}(\overline{X}_B + \overline{X}_G)$ $\mathrm{Var}(\overline{X}_{either}) = \frac{\sigma^2}{n}$ $\therefore \mathrm{E}(\overline{X}_{st}) = \frac{1}{2}(\mu_B + \mu_G)$ and $\mathrm{Var}(\overline{X}_{st}) = \frac{1}{4}\left(\frac{\sigma^2}{n} + \frac{\sigma^2}{n}\right) = \frac{\sigma^2}{2n}$	M1M1 B1 B1 **[4]**	

Question			Answer	Marks	Guidance
1	(vi)		$E(\bar{X}) = E(\bar{X}_{st}) = \frac{1}{2}(\mu_B + \mu_G) = E(X)$	E1	
			ie they are unbiased.	E1	
			Clearly $\text{Var}(\bar{X}) > \text{Var}(\bar{X}_{st})$,	M1	for any attempt to compare variances Candidates are not required to note that the variances are equal in the case $\mu_B = \mu_G$.
				M1	for deduction that $\text{Var}(\bar{X}) > \text{Var}(\bar{X}_{st})$ [FT c's variances]
			$\therefore$ $\bar{X}_{st}$ is the more efficient.	E1 **[5]**	More efficient
2	(i)		Mean of $X = 3.5$ (immediate by symmetry)	B1	
			$\text{E}(X^2) = \frac{1}{6}(1 + 4 + ... + 36) = \frac{91}{6}$	M1	
			$\therefore \text{Var}(X) = \frac{91}{6} - \left(\frac{7}{2}\right)^2 = \frac{35}{12}$	A1 **[3]**	Answer given
2	(ii)		$\text{G}(t) = \text{E}(t^X) = (t^1.\frac{1}{6}) + (t^2.\frac{1}{6}) + ... + (t^6.\frac{1}{6})$	M1	
			$= \frac{1}{6}(t + t^2 + ... + t^6) = \frac{t(1-t^6)}{6(1-t)}$	A1 **[2]**	Answer given
2	(iii)		[P($N = 0$) = ½, P($N = 1$) = (½)(½),] P($N = r$) = $(½)^r.(½)$	B1 **[1]**	answer given; must be convincing
2	(iv)		$\text{H}(t) = \text{E}(t^N) = (t^0.\frac{1}{2}) + (t^1.\frac{1}{4}) + (t^2.\frac{1}{8}) + ...$	M1	
			$= \frac{\frac{1}{2}}{1-\frac{t}{2}} = \frac{1}{2-t} = (2-t)^{-1}$	A1	Answer given

Question			Answer	Marks	Guidance
2	(v)		Mean = H'(1), variance = H"(1) + mean – mean^2. $H'(t) = (-1)(2-t)^{-2}(-1) = (2-t)^{-2}$ $\therefore$ mean = 1 $H''(t) = (-2)(2-t)^{-3}(-1) = 2(2-t)^{-3}$ $\therefore$ variance = 2 + 1 – 1 = 2	M1 A1 M1 A1 **[4]**	for use of 1st derivative for use of 2nd derivative For variance
2	(vi)		$K(t) = H\{G(t)\} = \{2 - G(t)\}^{-1}$ $= \left(2 - \frac{t(1-t^6)}{6(1-t)}\right)^{-1} = \left(\frac{12(1-t) - t(1-t)(1+t+t^2+...+t^5)}{6(1-t)}\right)^{-1}$ $= \left(\frac{12 - t - t^2 - t^3 - ... - t^6}{6}\right)^{-1} = 6(12 - t - t^2 - ... - t^6)^{-1}$	M1 M1 M1 A1 **[4]**	inserting G(t) use of hint given Answer given
2	(vii)		$K'(t) = 6(12 - t - t^2 - ... - t^6)^{-2}(1 + 2t + 3t^2 + 4t^3 + 5t^4 + 6t^5)$ $K''(t) = 12(12 - t - t^2 - ... - t^6)^{-3}(1 + 2t + 3t^2 + 4t^3 + 5t^4 + 6t^5)^2$ $+ 6(12 - t - t^2 - ... - t^6)^{-2}(2 + 6t + 12t^2 + 20t^3 + 30t^4)$ $\therefore$ mean = K'(1) = $6(12 - 6)^{-2}(21) = 21/6 = 7/2$ $\therefore$ K"(1) = $(12 \times 6^{-3} \times 21^2) + (6 \times 6^{-2} \times 70) = (49/2) + (70/6)$ $\therefore$ variance = $\frac{49}{2} + \frac{70}{6} + \frac{7}{2} - \frac{49}{4} = \frac{294 + 140 + 42 - 147}{12} = \frac{329}{12}$	M1 M1 M1 A1 A1 A1 **[6]**	reasonable attempt to differentiate K(t) reasonable attempt at 2nd derivative for use of derivatives Substitution shown Ft c's K'(1) and/or K"(1) provided variance positive Exact.

Question			Answer	Marks	Guidance
2	(viii)		We have: $\mu_X = 7/2$ $\sigma_X{}^2 = 35/12$ $\mu_N = 1$ $\sigma_N{}^2 = 2$ $\sigma_Q{}^2 = 329/12$ Inserting in the quoted formula gives $\left[2\times\left(\frac{7}{2}\right)^2\right]+\left[1\times\frac{35}{12}\right] = \frac{294+35}{12} = \frac{329}{12}$ as required.	M1 A1 **[2]**	for correct use of candidate's values for means and variances answer honestly obtained (common denominator shown). A0 if different from (vii)
3	(i)		H_0: population medians are equal H_1: population median for A < population median for B Wilcoxon rank sum test (or Mann-Whitney form of test) Ranks are: A 1 2 4 5 9 11 B 3 6 7 8 10 12 13 14 $W = 1 + 2 + 4 + 5 + 9 + 11 = 32$ [or 0 + 0 + 1 + 1 + 4 + 5 = 11 if M-W used] Refer to $W_{6,8}$ [or $MW_{6,8}$] tables Lower 5% critical point is 31 [or 10 if M-W used] Result is not significant Seems median yields may be assumed equal	B1 B1 M1 A1 B1 M1 A1 A1 A1 **[9]**	[*Note: "population" must be explicit*] 1) Explicit statement re shapes of distributions. (eg that they are the same shape) is not required. 2) More formal statements of hypotheses gain both marks [eg cdfs are $F(x)$ and $F(x - \Delta)$, H_0 is $\Delta = 0$ etc).] Combined ranking Correct [allow up to 2 errors; FT provided M1 earned] No FT if wrong No FT if wrong

Question			Answer	Marks	Guidance
3	(ii)		H_0: population means are equal	B1	
			H_1: population mean for A < population mean for B	B1	*"population" must be explicit, either in words or by notation*
			For A: $\bar{x} = 11.4,\ s_{n-1}^2 = 1.912\ [\ s_{n-1} = 1.38275]$	B1	For all. Use of s_n scores B0
			For B: $\bar{y} = 12.575,\ s_{n-1}^2 = 1.051\ [\ s_{n-1} = 1.025]$		
			Pooled $s^2 = \dfrac{(5\times1.912)+(7\times1.051)}{12} = \dfrac{16.915}{12} = 1.4096$	M1	for any reasonable attempt at pooling (but *not* if s_n^2 used)
				A1	If correct
			Test statistic $= \dfrac{11.4-12.575}{\sqrt{1.4096}\sqrt{\frac{1}{6}+\frac{1}{8}}} = \dfrac{-1.175}{0.6412} = -1.83(25)$	M1 A1	Ft if incorrect
			Refer to t_{12}	M1	No FT if wrong
			Lower single-tailed 5% critical point is –1.782	A1	No FT if wrong *must compare –1.83 with –1.782 unless it is clear and explicit that absolute values are being used*
			Significant	A1	
			Seems mean yield for A is less than that for B	A1	
				[11]	
3	(iii)		*t* test is "more sensitive" if Normality is correct.	E1	
			Non-rejection of Normality supports *t*.	E1	
			But Wilcoxon is more reliable if not Normal –	E1	
			and we do not have *proof* of Normality.	E1	
				[4]	
4	(i)		Latin square	B1	
			4 × 4 layout,		
			with rows clearly representing casting techniques		
			and columns clearly representing chemical compositions [*or vice versa*]	B2,1,0	B2 if completely correct, B1 if one point omitted
			Labels each appearing exactly once in each row and in each column	B1	for correct structure
			representing manufacturers.	B1	within the square
				[5]	

Question			Answer	Marks	Guidance
4	(ii)		Totals are 440.6 453.5 458.2 459.4 all from samples of size 4 Grand total 1811.7 "Correction factor" CF = $1811.7^2/16$ = 205141.06 Total SS = 205202.57 – CF = 61.5144		
			Between manufacturers SS = $\frac{440.6^2}{4}+\frac{453.5^2}{4}+\frac{458.2^2}{4}+\frac{459.4^2}{4}-\text{CF}$	M1	for attempt to form three sums of squares.
			= 205196.55 – CF = 55.4969	M1	for correct method for any two
			Residual SS (by subtraction) = 61.5144 – 55.4969 = 6.0175	A1	if each calculated SS is correct.
			Source of variation SS df MS MS ratio	B1	Between treatments df
			Between treatments 55.4969 3 18.4989 36.89	B1	Residual df
			Residual 6.0175 12 0.5014(6)	M1	Method for calculating either mean square
			Total 61.5144 15	M1	MS ratio
				A1	cao at least 3sf, condone up to 6 sf
			Refer MS ratio to $F_{3,12}$.	M1	No FT if wrong
			quotation of an extreme point (not just the 5%point); eg 1% point, 5.95, or 0.1% point, 10.80.	A1	
			Must be some reference to *very highly* significant	A1	
			and *very strong/overwhelming* evidence that "the manufacturers are not all the same".	A1	
				[12]	
4	(iii)		$y_{ij} = \mu + \alpha_i + e_{ij}$	B2	B1 if any two RHS terms correct
			μ = population …	B1	"Population"; award here <u>or</u> for α_i
			…grand mean for whole exp't	B1	
			α_i = population mean amount by which the *i*th treatment differs from μ	B1	
					e_{ij} is experimental error – does not need to be stated explicitly here, is subsumed in error assumptions below.
			e_{ij} ~ ind N **[*] (** accept "uncorrelated") (0 **[*]** , σ^2 **[*]**)	B2	if all three * components correct, B1 if any two correct.
				[7]	

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