OCR
RECOGNISING ACHIEVEMENT

Thursday 24 May 2012 – Morning

A2 GCE MATHEMATICS (MEI)

4769 Statistics 4

QUESTION PAPER

Candidates answer on the Printed Answer Book.

OCR supplied materials:
- Printed Answer Book 4769
- MEI Examination Formulae and Tables (MF2)

Other materials required:
- Scientific or graphical calculator

Duration: 1 hour 30 minutes

4715880612

INSTRUCTIONS TO CANDIDATES

These instructions are the same on the Printed Answer Book and the Question Paper.

- The Question Paper will be found in the centre of the Printed Answer Book.
- Write your name, centre number and candidate number in the spaces provided on the Printed Answer Book. Please write clearly and in capital letters.
- **Write your answer to each question in the space provided in the Printed Answer Book**. Additional paper may be used if necessary but you must clearly show your candidate number, centre number and question number(s).
- Use black ink. HB pencil may be used for graphs and diagrams only.
- Read each question carefully. Make sure you know what you have to do before starting your answer.
- Answer any **three** questions.
- Do **not** write in the bar codes.
- You are permitted to use a scientific or graphical calculator in this paper.
- Final answers should be given to a degree of accuracy appropriate to the context.

INFORMATION FOR CANDIDATES

This information is the same on the Printed Answer Book and the Question Paper.

- The number of marks is given in brackets **[ ]** at the end of each question or part question on the Question Paper.
- You are advised that an answer may receive **no marks** unless you show sufficient detail of the working to indicate that a correct method is being used.
- The total number of marks for this paper is **72**.
- The Printed Answer Book consists of **16** pages. The Question Paper consists of **8** pages. Any blank pages are indicated.

INSTRUCTION TO EXAMS OFFICER / INVIGILATOR

- Do not send this Question Paper for marking; it should be retained in the centre or destroyed.

Option 1: Estimation

1 In a certain country, any baby born is equally likely to be a boy or a girl, independently for all births. The birthweight of a baby boy is given by the continuous random variable X_B with probability density function (pdf) $f_B(x)$ and cumulative distribution function (cdf) $F_B(x)$. The birthweight of a baby girl is given by the continuous random variable X_G with pdf $f_G(x)$ and cdf $F_G(x)$.

The continuous random variable X denotes the birthweight of a baby selected at random.

(i) By considering

$$P(X \leqslant x) = P(X \leqslant x \mid \text{boy})\, P(\text{boy}) + P(X \leqslant x \mid \text{girl})\, P(\text{girl}),$$

find the cdf of X in terms of $F_B(x)$ and $F_G(x)$, and deduce that the pdf of X is

$$f(x) = \tfrac{1}{2}\{f_B(x) + f_G(x)\}.$$ **[3]**

(ii) The birthweights of baby boys and girls have means μ_B and μ_G respectively. Deduce that

$$E(X) = \tfrac{1}{2}(\mu_B + \mu_G).$$ **[1]**

(iii) The birthweights of baby boys and girls have common variance σ^2. Find an expression for $E(X^2)$ in terms of μ_B, μ_G and σ^2, and deduce that

$$\text{Var}(X) = \sigma^2 + \tfrac{1}{4}(\mu_B - \mu_G)^2.$$ **[7]**

(iv) A random sample of size $2n$ is taken from all the babies born in a certain period. The mean birthweight of the babies in this sample is $\overline{X}$. Write down an approximation to the sampling distribution of $\overline{X}$ if n is large. **[4]**

(v) Suppose instead that a stratified sample of size $2n$ is taken by selecting n baby boys at random and, independently, n baby girls at random. The mean birthweight of the $2n$ babies in this sample is $\overline{X}_{st}$. Write down the expected value of $\overline{X}_{st}$ and find the variance of $\overline{X}_{st}$. **[4]**

(vi) Deduce that both $\overline{X}$ and $\overline{X}_{st}$ are unbiased estimators of the population mean birthweight. Find which is the more efficient. **[5]**

Option 2: Generating Functions

2 The random variable X ($X = 1, 2, 3, 4, 5, 6$) denotes the score when a fair six-sided die is rolled.

(i) Write down the mean of X and show that $\text{Var}(X) = \frac{35}{12}$. **[3]**

(ii) Show that G(t), the probability generating function (pgf) of X, is given by

$$G(t) = \frac{t(1 - t^6)}{6(1 - t)}. \quad \textbf{[2]}$$

The random variable N ($N = 0, 1, 2, \ldots$) denotes the number of heads obtained when an unbiased coin is tossed repeatedly until a tail is first obtained.

(iii) Show that $P(N = r) = \left(\frac{1}{2}\right)^{r+1}$ for $r = 0, 1, 2, \ldots$. **[1]**

(iv) Hence show that H(t), the pgf of N, is given by $H(t) = (2 - t)^{-1}$. **[2]**

(v) Use H(t) to find the mean and variance of N. **[4]**

A game consists of tossing an unbiased coin repeatedly until a tail is first obtained and, each time a head is obtained in this sequence of tosses, rolling a fair six-sided die. The die is not rolled on the first occasion that a tail is obtained and the game ends at that point. The random variable Q ($Q = 0, 1, 2, \ldots$) denotes the total score on all the rolls of the die. Thus, in the notation above, $Q = X_1 + X_2 + \ldots + X_N$ where the X_i are independent random variables each distributed as X, with $Q = 0$ if $N = 0$. The pgf of Q is denoted by K(t). The familiar result that the pgf of a sum of independent random variables is the product of their pgfs does **not** apply to K(t) because N is a random variable and not a fixed number; you should instead **use without proof** the result that $K(t) = H(G(t))$.

(vi) Show that $K(t) = 6(12 - t - t^2 - \ldots - t^6)^{-1}$. **[4]**

[Hint. $(1 - t^6) = (1 - t)(1 + t + t^2 + \ldots + t^5)$.]

(vii) Use K(t) to find the mean and variance of Q. **[6]**

(viii) Using your results from parts **(i)**, **(v)** and **(vii)**, verify the result that (in the usual notation for means and variances)

$$\sigma_Q{}^2 = \sigma_N{}^2 \mu_X{}^2 + \mu_N \sigma_X{}^2. \quad \textbf{[2]}$$

Option 3: Inference

3 At an agricultural research station, trials are being made of two fertilisers, A and B, to see whether they differ in their effects on the yield of a crop. Preliminary investigations have established that the underlying variances of the distributions of yields using the two fertilisers may be assumed equal. Scientific analysis of the fertilisers has suggested that fertiliser A may be inferior in that it leads, on the whole, to lower yield. A statistical analysis is being carried out to investigate this.

The crop is grown in carefully controlled conditions in 14 experimental plots, 6 with fertiliser A and 8 with fertiliser B. The yields, in kg per plot, are as follows, arranged in ascending order for each fertiliser.

Fertiliser A	9.8	10.2	10.9	11.5	12.7	13.3		
Fertiliser B	10.8	11.9	12.0	12.2	12.9	13.5	13.6	13.7

(i) Carry out a Wilcoxon rank sum test at the 5% significance level to examine appropriate hypotheses. **[9]**

(ii) Carry out a t test at the 5% significance level to examine appropriate hypotheses. **[11]**

(iii) Goodness of fit tests based on more extensive data sets from other trials with these fertilisers have failed to reject hypotheses of underlying Normal distributions. Discuss the relative merits of the analyses in parts **(i)** and **(ii)**. **[4]**

Option 4: Design and Analysis of Experiments

4 **(i)** In an engineering research laboratory, a study is being made of the strength of steel girders supplied by four different manufacturers. Four techniques for casting the girders are to be used, as are four slightly different chemical compositions of the steel. Sixteen girders are to be supplied for testing purposes, four by each manufacturer.

Name an experimental design that should be used for allocating the work to the manufacturers in such a way that any differences in strength of girders between the different manufacturers can be studied, whether or not there are consistent differences resulting from the casting techniques or from the chemical compositions. Give an example of a suitable layout of the experiment. **[5]**

(ii) After initial investigation, it is decided that differences in strength resulting from the casting techniques or the chemical compositions can be ignored. A one-way analysis of variance is therefore carried out on the results, which are as follows, measured in a convenient unit.

Strength of girder

Manufacturer			
A	B	C	D
109.4	114.4	114.8	115.1
110.0	113.1	113.7	114.0
110.9	113.5	115.4	114.7
110.3	112.5	114.3	115.6

[The sum of these data items is 1811.7 and the sum of their squares is 205 202.57.]

Construct the usual one-way analysis of variance table. Carry out the appropriate test and report your conclusion. **[12]**

(iii) Using the customary notation, write down the usual model underlying the one-way analysis of variance. Carefully interpret the terms in this model. State the assumptions that are usually made for the error term in the model. **[7]**

BLANK PAGE

BLANK PAGE

OCR
RECOGNISING ACHIEVEMENT

Thursday 24 May 2012 – Morning

A2 GCE MATHEMATICS (MEI)

4769 Statistics 4

PRINTED ANSWER BOOK

Candidates answer on this Printed Answer Book.

OCR supplied materials:
- Question Paper 4769 (inserted)
- MEI Examination Formulae and Tables (MF2)

Other materials required:
- Scientific or graphical calculator

Duration: 1 hour 30 minutes

Candidate forename		Candidate surname	

Centre number						Candidate number				

INSTRUCTIONS TO CANDIDATES

These instructions are the same on the Printed Answer Book and the Question Paper.

- The Question Paper will be found in the centre of the Printed Answer Book.
- Write your name, centre number and candidate number in the spaces provided on the Printed Answer Book. Please write clearly and in capital letters.
- **Write your answer to each question in the space provided in the Printed Answer Book**. Additional paper may be used if necessary but you must clearly show your candidate number, centre number and question number(s).
- Use black ink. HB pencil may be used for graphs and diagrams only.
- Read each question carefully. Make sure you know what you have to do before starting your answer.
- Answer any **three** questions.
- Do **not** write in the bar codes.
- You are permitted to use a scientific or graphical calculator in this paper.
- Final answers should be given to a degree of accuracy appropriate to the context.

INFORMATION FOR CANDIDATES

This information is the same on the Printed Answer Book and the Question Paper.

- The number of marks is given in brackets **[ ]** at the end of each question or part question on the Question Paper.
- You are advised that an answer may receive **no marks** unless you show sufficient detail of the working to indicate that a correct method is being used.
- The total number of marks for this paper is **72**.
- The Printed Answer Book consists of **16** pages. The Question Paper consists of **8** pages. Any blank pages are indicated.

1 (i)	
1 (ii)	

1 (iii)

1 (iv)	
1 (v)	

1 (vi)	
2 (i)	

2 (ii)	
2 (iii)	

2 (iv)	
2 (v)	

2 (vi)	
2 (vii)	
	(answer space continued on next page)

(answer space continued on next page)

2 (vii)	**(continued)**
2 (viii)	

3 (i)	
	(answer space continued on next page)

3 (i)	(continued)
3 (ii)	
	(answer space continued on next page)

3 (ii)	**(continued)**

3 (iii)	
4 (i)	

4 (ii)	
	(answer space continued on next page)

4 (ii)	(continued)

4 (iii)	

© OCR 2012

RECOGNISING ACHIEVEMENT

GCE

Mathematics (MEI)

Advanced GCE

Unit **4769:** Statistics 4

Mark Scheme for June 2012

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)		$\text{P}\left(X \le x\right) = \text{F}_B(x).\tfrac{1}{2} + \text{F}_G(x).\tfrac{1}{2}$ ie cdf of X is $\text{F}\left(x\right) = \tfrac{1}{2}\left\{\text{F}_B(x) + \text{F}_G(x)\right\}$ ie (by differentiating) pdf of X is $\text{f}\left(x\right) = \tfrac{1}{2}\left\{\text{f}_B(x) + \text{f}_G(x)\right\}$	M1 A1 A1 **[3]**	use of cdfs Answer given
1	(ii)		$\text{E}\left(X\right)\left(= \tfrac{1}{2}\left\{\int x\text{f}_B(x)\text{d}x + \int x\text{f}_G(x)\text{d}x\right\}\right) = \tfrac{1}{2}\mu_B + \tfrac{1}{2}\mu_G$	M1 **[1]**	[answer given; needs *some* indication of method]
1	(iii)		$\text{E}\left(X^2\right) = \int x^2\text{f}(x)\text{d}x$ $= \tfrac{1}{2}\left\{\int x^2\text{f}_B(x)\text{d}x + \int x^2\text{f}_G(x)\text{d}x\right\}$ Use of "$\text{E}\left(X^2\right) = \sigma^2 + \mu^2$" $= \tfrac{1}{2}\left\{\sigma^2 + {\mu_B}^2 \quad + \quad \sigma^2 + {\mu_G}^2\right\}$ $\therefore \text{Var}(X) = \text{E}(X^2) - \left\{\text{E}(X)\right\}^2$ $= \sigma^2 + \tfrac{1}{2}{\mu_B}^2 + \tfrac{1}{2}{\mu_G}^2 - \tfrac{1}{4}{\mu_B}^2 - \tfrac{1}{4}{\mu_G}^2 - \tfrac{1}{2}\mu_B\mu_G$ $= \sigma^2 + \tfrac{1}{4}\left(\mu_B - \mu_G\right)^2$	M1 M1 M1 A1 M1 A1 A1 **[7]**	 Answer given
1	(iv)		[Central Limit Theorem] Approx dist of $\overline{X}$ is $\text{N}\left(\tfrac{1}{2}\mu_B + \tfrac{1}{2}\mu_G,\ \tfrac{1}{2n}\left(\sigma^2 + \tfrac{1}{4}\left(\mu_B - \mu_G\right)^2\right)\right)$ B1 B1 B1 B1	B4 **[4]**	4 marks as shown
1	(v)		$\overline{X}_{st} = \tfrac{1}{2}\left(\overline{X}_B + \overline{X}_G\right) \qquad \text{Var}\left(\overline{X}_{either}\right) = \tfrac{\sigma^2}{n}$ $\therefore \text{E}\left(\overline{X}_{st}\right) = \tfrac{1}{2}\left(\mu_B + \mu_G\right)$ and $\text{Var}\left(\overline{X}_{st}\right) = \tfrac{1}{4}\left(\tfrac{\sigma^2}{n} + \tfrac{\sigma^2}{n}\right) = \dfrac{\sigma^2}{2n}$	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	More efficient
				[5]	
2	(i)		Mean of $X = 3.5$ (immediate by symmetry)	B1	
			$\text{E}(X^2) = \frac{1}{6}(1+4+\ldots+36) = \frac{91}{6}$	M1	
			$\therefore \text{Var}(X) = \frac{91}{6} - \left(\frac{7}{2}\right)^2 = \frac{35}{12}$	A1	Answer given
				[3]	
2	(ii)		$\text{G}(t) = \text{E}(t^X) = (t^1.\frac{1}{6}) + (t^2.\frac{1}{6}) + \ldots + (t^6.\frac{1}{6})$	M1	
			$= \frac{1}{6}(t + t^2 + \ldots + t^6) = \frac{t(1-t^6)}{6(1-t)}$	A1	Answer given
				[2]	
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)		$\text{H}(t) = \text{E}(t^N) = (t^0.\frac{1}{2}) + (t^1.\frac{1}{4}) + (t^2.\frac{1}{8}) + \ldots$	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.	M1	for <u>use</u> of 1st derivative
			$\text{H}'(t)=(-1)(2-t)^{-2}(-1)=(2-t)^{-2}$ $\quad\therefore$ mean = 1	A1	
			$\text{H}''(t)=(-2)(2-t)^{-3}(-1)=2(2-t)^{-3}$	M1	for <u>use</u> of 2nd derivative
			$\therefore$ variance = 2 + 1 – 1 = 2	A1	For variance
				[4]	
2	(vi)		K(t) = H{G(t)} = {2 – G(t)}$^{-1}$	M1	
			$=\left(2-\frac{t(1-t^6)}{6(1-t)}\right)^{-1}=\left(\frac{12(1-t)-t(1-t)(1+t+t^2+\ldots+t^5)}{6(1-t)}\right)^{-1}$	M1 M1	inserting G(t) use of hint given
			$=\left(\frac{12-t-t^2-t^3-\ldots-t^6}{6}\right)^{-1}=6(12-t-t^2-\ldots-t^6)^{-1}$	A1	Answer given
				[4]	
2	(vii)		$\text{K}'(t)=6(12-t-t^2-\ldots-t^6)^{-2}(1+2t+3t^2+4t^3+5t^4+6t^5)$	M1	reasonable attempt to differentiate K(t)
			$\text{K}''(t)=12(12-t-t^2-\ldots-t^6)^{-3}(1+2t+3t^2+4t^3+5t^4+6t^5)^2$	M1	reasonable attempt at 2nd derivative
			$+\ 6(12-t-t^2-\ldots-t^6)^{-2}(2+6t+12t^2+20t^3+30t^4)$	M1	for <u>use</u> of derivatives
			$\therefore$ mean = K'(1) = 6(12 – 6)$^{-2}$(21) = 21/6 = 7/2	A1	Substitution shown
			$\therefore$ K"(1) = (12 × 6^{-3} × 21^2) + (6 × 6^{-2} × 70) = (49/2) + (70/6)	A1	
			$\therefore$ variance = $\frac{49}{2}+\frac{70}{6}+\frac{7}{2}-\frac{49}{4}=\frac{294+140+42-147}{12}=\frac{329}{12}$	A1	Ft c's K'(1) and/or K"(1) provided variance positive Exact.
				[6]	

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 H_1: population mean for A < population mean for B	B1 B1	 *"population" must be explicit, either in words or by notation*
			For A: $\bar{x} = 11.4$, $s_{n-1}^2 = 1.912$ [$s_{n-1} = 1.38275$] For B: $\bar{y} = 12.575$, $s_{n-1}^2 = 1.051$ [$s_{n-1} = 1.025$]	B1	For all. Use of s_n scores B0
			Pooled $s^2 = \frac{(5\times 1.912)+(7\times 1.051)}{12} = \frac{16.915}{12} = 1.4096$	M1 A1	for any reasonable attempt at pooling (but *not* if s_n^2 used) If correct
			Test statistic $= \frac{11.4 - 12.575}{\sqrt{1.4096}\sqrt{\frac{1}{6}+\frac{1}{8}}} = \frac{-1.175}{0.6412} = -1.83(25)$	M1 A1	Ft if incorrect
			Refer to t_{12} Lower single-tailed 5% critical point is –1.782	M1 A1	No FT if wrong No FT if wrong *must compare –1.83 with –1.782 unless it is clear and explicit that absolute values are being used*
			Significant Seems mean yield for A is less than that for B	A1 A1 **[11]**	
3	(iii)		t test is "more sensitive" if Normality is correct. Non-rejection of Normality supports t. But Wilcoxon is more reliable if not Normal – and we do not have *proof* of Normality.	E1 E1 E1 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 **[5]**	within the square

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 = $\frac{440.6^2}{4}+\frac{453.5^2}{4}+\frac{458.2^2}{4}+\frac{459.4^2}{4}-\text{CF}$ = 205196.55 – CF = 55.4969 Residual SS (by subtraction) = 61.5144 – 55.4969 = 6.0175	 M1 M1 A1	 for attempt to form three sums of squares. for correct method for any two if each calculated SS is correct.
			Source of variation / SS / df / MS / MS ratio Between treatments / 55.4969 / 3 / 18.4989 / 36.89 Residual / 6.0175 / 12 / 0.5014(6) Total / 61.5144 / 15	B1 B1 M1 M1 A1	Between treatments df Residual df Method for calculating either mean square MS ratio cao at least 3sf, condone up to 6 sf
			Refer MS ratio to $F_{3,12}$. quotation of an extreme point (not just the 5%point); eg 1% point, 5.95, or 0.1% point, 10.80. Must be some reference to *very highly* significant and *very strong/overwhelming* evidence that "the manufacturers are not all the same".	M1 A1 A1 A1 **[12]**	No FT if wrong
4	(iii)		$y_{ij}=\mu+\alpha_i+e_{ij}$ μ = population … …grand mean for whole exp't α_i = population mean amount by which the ith treatment differs from μ	B2 B1 B1 B1	B1 if any two RHS terms correct "Population"; award here or for α_i 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 **[7]**	if all three * components correct, B1 if any two correct.

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

UKAS MANAGEMENT SYSTEMS 001

4769 Statistics 4

General Comments

There were nearly twice as many entries this year compared with last. The standard was on the whole very good, with a small number of outstanding scripts. Very few candidates showed signs of being out of their depth.

No candidates answered more than the required three questions from the four options. Question 2 was most popular despite being somewhat unfamiliar, but with the guidance given there were many strong answers. Question 4 was least often chosen, but often well done.

Comments on Individual Questions

1) This question on estimation was not as popular as either Q2 or Q3.

The information given allowed nearly all candidates to gain full marks in (i). In (ii) the best solutions showed explicitly the use of the density function, likewise in (iii) when $\mathrm{E}(X^2)$ was needed. Credit was given to those who argued from the conditional expectations. Here most candidates realised that $\sigma^2 + \mu^2$ for each sex was needed but not all could find the correct expression to use. There was some evidence of forcing the required result, where accuracy marks were forfeit. In (iv) most candidates were able to deduce the correct distribution, errors in the denominator of the variance being most common. Explicit derivation of $\bar{X}_{st}$ was not always seen in (iv); those candidates who defined the expression through the observed random variables were the most successful in finding the expectation and the variance. In (vi) candidates with the correct variances for $\bar{X}$ and $\bar{X}_{st}$ were easily able to discern the more efficient estimator, but where there were mistakes in one or other expression this was not always possible.

2) Every candidate except one attempted this question on generating functions.

It is important that full working is shown where the required result is given, as in parts (i) to (iv). In the remaining sections nearly all candidates successfully used the information presented to obtain $\mathrm{K}(t)$. In (vii) most candidates successfully negotiated the differentiation required for $\mathrm{K}'(t)$ and found the expectation of Q. With more difficulty, many obtained $\mathrm{K}''(t)$, and the required variance. The final part (viii) was again requiring candidates to show full working in order to earn all the marks, including a common denominator and the exact form of σ_Q^2 .

3) This 'Inference' option question revealed a variable set of responses. The statements of appropriate hypotheses in both parts (i) and (ii) were mostly carefully given with explicit mention of 'population' values. The alternative hypothesis did not always take account of the suspected inferiority of fertiliser A, leading to a two-tailed test which was inappropriate and lost marks. In a few cases in (i) the wrong conclusion was reached when the size of the Wilcoxon test statistic was compared with the critical value. In (ii) most candidates successfully found the pooled estimate of variance and calculated the correct test statistic. Any error in the degrees of freedom inevitably lost the following marks. In both parts (i) and (ii) most candidates were careful to give a non-assertive interpretation of the test results, in context.

   Part (iii) was on the whole found difficult to answer. Candidates were not always able to say why the t-test was better than the Wilcoxon test if the underlying population distribution was Normal, and not always willing to assert that the Normality was in any way questionable.

4) Design and Analysis of Experiments. This question was answered by half the candidates. The need for a Latin Square design was nearly always recognised, and the accompanying layout was usually correct. Some answers did not recognise the manufacturers as the 'treatments' under investigation and did not place them inside the square. In (ii) candidates were able to find the appropriate sums of squares, mean squares and degrees of freedom, but these were sometimes not laid out in the conventional Analysis of Variance table. Most candidates found the correct mean square ratio. The interpretation of the test result was disappointing. Candidates were not prepared to abandon the usual 5% significance level and point out how extreme the result was, even by comparison with the 0.1% point. The question was hoping for insight which was not forthcoming.

GCE Mathematics (MEI)

			Max Mark	90% cp	a	b	c	d	e	u
4753/01	(C3) MEI Methods for Advanced Mathematics with Coursework: Written Paper	Raw	72	66	60	53	47	41	34	0
4753/02	(C3) MEI Methods for Advanced Mathematics with Coursework: Coursework	Raw	18	16	15	13	11	9	8	0
4753/82	(C3) MEI Methods for Advanced Mathematics with Coursework: Carried Forward Coursework Mark	Raw	18	16	15	13	11	9	8	0
4753	(C3) MEI Methods for Advanced Mathematics with Coursework	UMS	100	90	80	70	60	50	40	0
4754/01	(C4) MEI Applications of Advanced Mathematics	Raw	90	73	65	57	50	43	36	0
		UMS	100	90	80	70	60	50	40	0
4756/01	(FP2) MEI Further Methods for Advanced Mathematics	Raw	72	66	61	53	46	39	32	0
		UMS	100	90	80	70	60	50	40	0
4757/01	(FP3) MEI Further Applications of Advanced Mathematics	Raw	72	61	54	47	40	34	28	0
		UMS	100	90	80	70	60	50	40	0

4758/01	(DE) MEI Differential Equations with Coursework: Written Paper	Raw	72	68	63	57	51	45	39	0
4758/02	(DE) MEI Differential Equations with Coursework: Coursework	Raw	18	16	15	13	11	9	8	0
4758/82	(DE) MEI Differential Equations with Coursework: Carried Forward Coursework Mark	Raw	18	16	15	13	11	9	8	0
4758	(DE) MEI Differential Equations with Coursework	UMS	100	90	80	70	60	50	40	0
4762/01	(M2) MEI Mechanics 2	Raw	72	65	58	51	44	38	32	0
		UMS	100	90	80	70	60	50	40	0
4763/01	(M3) MEI Mechanics 3	Raw	72	67	63	56	50	44	38	0
		UMS	100	90	80	70	60	50	40	0
4764/01	(M4) MEI Mechanics 4	Raw	72	63	56	49	42	35	29	0
		UMS	100	90	80	70	60	50	40	0
4767/01	(S2) MEI Statistics 2	Raw	72	66	61	55	49	43	38	0
		UMS	100	90	80	70	60	50	40	0
4768/01	(S3) MEI Statistics 3	Raw	72	65	58	51	44	38	32	0
		UMS	100	90	80	70	60	50	40	0
4769/01	(S4) MEI Statistics 4	Raw	72	63	56	49	42	35	28	0
		UMS	100	90	80	70	60	50	40	0
4772/01	(D2) MEI Decision Mathematics 2	Raw	72	62	56	50	44	39	34	0
		UMS	100	90	80	70	60	50	40	0
4773/01	(DC) MEI Decision Mathematics Computation	Raw	72	52	46	40	34	29	24	0
		UMS	100	90	80	70	60	50	40	0
4777/01	(NC) MEI Numerical Computation	Raw	72	63	55	47	39	32	25	0
		UMS	100	90	80	70	60	50	40	0