Friday 24 May 2013 – Morning

A2 GCE MATHEMATICS (MEI)

4769/01 Statistics 4

QUESTION PAPER

Candidates answer on the Printed Answer Book.

OCR supplied materials:
- Printed Answer Book 4769/01
- MEI Examination Formulae and Tables (MF2)

Other materials required:
- Scientific or graphical calculator

Duration: 1 hour 30 minutes

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 **4** 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 recycled. Please contact OCR Copyright should you wish to re-use this document.

DC (CW) 66630/1

Option 1: Estimation

1 Traffic engineers are studying the flow of vehicles along a road. At an initial stage of the investigation, they assume that the average flow remains the same throughout the working day. An automatic counter records the number of vehicles passing a certain point per minute during the working day. A random sample of these records is selected; the sample values are denoted by $x_1, x_2, \ldots, x_n$.

(i) The engineers model the underlying random variable X by a Poisson distribution with unknown parameter θ. Obtain the likelihood of $x_1, x_2, \ldots, x_n$ and hence find the maximum likelihood estimate of θ. **[10]**

(ii) Write down the maximum likelihood estimate of the probability that no vehicles pass during a minute. **[3]**

(iii) The engineers note that, in a sample of size 1000 with sample mean $\bar{x} = 5$, there are no observations of zero. Suggest why this might cast some doubt on the investigation. **[3]**

(iv) On checking the automatic counter, the engineers find that, due to a fault, no record at all is made if no vehicle passes in a minute. They therefore model X as a Poisson random variable, again with an unknown parameter θ, except that the value $x = 0$ cannot occur. Show that, under this model,

$$\mathrm{P}(X = x) = \frac{\theta^x}{\left(\mathrm{e}^\theta - 1\right)x!}, \quad x = 1, 2, \ldots,$$

and hence show that the maximum likelihood estimate of θ satisfies the equation

$$\frac{\theta \mathrm{e}^\theta}{\mathrm{e}^\theta - 1} = \bar{x}.$$

[8]

Option 2: Generating Functions

2 The random variable X takes values -2, 0 and 2, each with probability $\frac{1}{3}$.

(i) Write down the values of

(A) μ, the mean of X,

(B) $\mathrm{E}(X^2)$,

(C) σ^2, the variance of X. **[3]**

(ii) Obtain the moment generating function (mgf) of X. **[2]**

A random sample of n independent observations on X has sample mean $\overline{X}$, and the standardised mean is denoted by Z where

$$Z = \frac{\overline{X} - \mu}{\frac{\sigma}{\sqrt{n}}}.$$

(iii) Stating carefully the required general results for mgfs of sums and of linear transformations, show that the mgf of Z is

$$\mathrm{M}_Z(\theta) = \left\{\frac{1}{3}\left(1 + \mathrm{e}^{\frac{\theta\sqrt{3}}{\sqrt{2n}}} + \mathrm{e}^{-\frac{\theta\sqrt{3}}{\sqrt{2n}}}\right)\right\}^n. \qquad \textbf{[8]}$$

(iv) By expanding the exponential functions in $\mathrm{M}_Z(\theta)$, show that, for large n,

$$\mathrm{M}_Z(\theta) \approx \left(1 + \frac{\theta^2}{2n}\right)^n. \qquad \textbf{[7]}$$

(v) Use the result $\mathrm{e}^y = \lim\limits_{n \to \infty}\left(1 + \frac{y}{n}\right)^n$ to find the limit of $\mathrm{M}_Z(\theta)$ as $n \to \infty$, and deduce the approximate distribution of Z for large n. **[4]**

Option 3: Inference

3 **(i)** Explain the meaning of the following terms in the context of hypothesis testing: Type I error, Type II error, operating characteristic, power. **[8]**

(ii) A test is to be carried out concerning a parameter θ. The null hypothesis is that θ has the particular value θ_0. The alternative hypothesis is $\theta \neq \theta_0$. Draw a sketch of the operating characteristic for a perfect test that never makes an error. **[3]**

(iii) The random variable X is distributed as $\mathrm{N}(\mu, 9)$. A random sample of size 25 is available. The null hypothesis $\mu = 0$ is to be tested against the alternative hypothesis $\mu \neq 0$. The null hypothesis will be accepted if $-1 < \bar{x} < 1$ where $\bar{x}$ is the value of the sample mean, otherwise it will be rejected. Calculate the probability of a Type I error. Calculate the probability of a Type II error if in fact $\mu = 0.5$; comment on the value of this probability. **[9]**

(iv) Without carrying out any further calculations, draw a sketch of the operating characteristic for the test in part **(iii)**. **[4]**

Option 4: Design and Analysis of Experiments

4 **(i)** Explain the advantages of randomisation and replication in a statistically designed experiment. **[6]**

(ii) The usual statistical model underlying the one-way analysis of variance is given, in the usual notation, by

$$x_{ij} = \mu + \alpha_i + e_{ij}$$

where x_{ij} denotes the jth observation on the ith treatment. Define carefully all the terms in this model and state the properties of the term that represents experimental error. **[7]**

(iii) A trial of five fertilisers is carried out at an agricultural research station according to a completely randomised design in which each fertiliser is applied to four experimental plots of a crop (so that there are 20 experimental units altogether). The sums of squares in a one-way analysis of variance of the resulting data on yields of the crop are as follows.

Source of variation	Sum of squares
Between fertilisers	219.2
Residual	304.5
Total	523.7

State the customary null and alternative hypotheses that are tested. Provide the degrees of freedom for each sum of squares. Hence copy and complete the analysis of variance table and carry out the test at the 5% level. **[11]**

RECOGNISING ACHIEVEMENT

Friday 24 May 2013 – Morning

A2 GCE MATHEMATICS (MEI)

4769/01 Statistics 4

PRINTED ANSWER BOOK

Candidates answer on this Printed Answer Book.

OCR supplied materials:
- Question Paper 4769/01 (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 **4** pages. Any blank pages are indicated.

DC (CW) 66631/1

Turn over

1 (i)

(answer space continued on next page)

1 (i)	**(continued)**
1 (ii)	

1 (iii)

1 (iv)

(answer space continued on next page)

1 (iv)	**(continued)**

Turn over

2 (i) (*A*)

2 (i) (*B*)

2 (i) (*C*)

2 (ii)

2 (iii)	

2 (iv)

2 (v)

3 (i)

3 (ii)	
3 (iii)	

(answer space continued on next page)

Turn over

3 (iii)	**(continued)**
3 (iv)	

4 (i)

4 (ii)

4 (iii)

(answer space continued on next page)

4 (iii)	**(continued)**

OCR
RECOGNISING ACHIEVEMENT

GCE

Mathematics (MEI)

Advanced GCE

Unit **4769:** Statistics 4

Mark Scheme for June 2013

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.

© OCR 2013

Annotations

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. 3 significant figures may often be the norm for this, but this always needs to be considered in the context of the problem in hand. For example, in quoting probabilities from Normal tables, we generally expect *some* evidence of interpolation and so quotation to 4 decimal places will often be appropriate. But even this does not always apply – quotations of the standard critical points for significance tests such as 1.96, 1.645, 2.576 (maybe even 2.58 – but not 2.57) will commonly suffice, especially if the calculated value of a test statistic is nowhere near any of these values. Sensible discretion *must* be exercised in such cases.

Discretion must also be exercised in the case of small variations in the degree of accuracy to which an answer is given. For example, if 3 significant figures are expected (either because of an explicit instruction or because the general context of a problem demands it) but only 2 are given, loss of an accuracy ("A") mark is likely to be appropriate; but if 4 significant figures are given, this should not normally be penalised. Likewise, answers which are slightly deviant from what is expected in a very minor manner (for example a Normal probability given, after an attempt at interpolation, as 0.6418 whereas 0.6417 was expected) should not be penalised. However, answers which are *grossly* over- or under-specified should normally result in the loss of a mark. This includes cases such as, for example, insistence that the value of a test statistic is (say) 2.128888446667 merely because that is the value that happened to come off the candidate's calculator. Note that this applies to answers that are given as final stages of calculations; intermediate working should usually be carried out, and quoted, to a greater degree of accuracy to avoid the danger of premature approximation.

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. Genuine misreading (of numbers or symbols, occasionally even of text) occurs. If this results in the object and/or difficulty of the question being considerably changed, it is likely that all the marks for that question, or section of the question, will be lost. However, misreads are often such that the object and/or difficulty remain substantially unaltered; these cases are considered below.

The simple rule is that *all* method ("M") marks [and of course all independent ("B") marks] remain accessible but at least some accuracy ("A") marks do not. It is difficult to legislate in an overall sense beyond this global statement because misreads, even when the object and/or difficulty remains unchanged, can vary greatly in their effects. For example, a misread of 1.02 as 10.2 (perhaps as a quoted value of a sample mean) may well be catastrophic; whereas a misread of 1.6748 as 1.6746 may have so slight an effect as to be almost unnoticeable in the candidate's work.

A misread should normally attract *some* penalty, though this would often be only 1 mark and should rarely if ever be more than 2. Commonly in sections of questions where there is a numerical answer either at the end of the section or to be obtained and commented on (eg the value of a test statistic), this answer will have an "A" mark that may actually be designated as "cao" [correct answer only]. This should be interpreted *strictly* – if the misread has led to failure to obtain this value, then this "A" mark must be withheld even if all method marks have been earned. It will also often be the case that such a mark is implicitly "cao" even if not explicitly designated as such.

On the other hand, we commonly allow "fresh starts" within a question or part of question. For example, a follow-through of the candidate's value of a test statistic is generally allowed (and often explicitly stated as such within the marking scheme), so that the candidate may exhibit knowledge of how to compare it with a critical value and draw conclusions. Such "fresh starts" are not affected by any earlier misreads.

A misread may be of a symbol rather than a number – for example, an algebraic symbol in a mathematical expression. Such misreads are more likely to bring about a considerable change in the object and/or difficulty of the question; but, if they do not, they should be treated as far as possible in the same way as numerical misreads, *mutatis mutandis*. This also applied to misreads of text, which are fairly rare but can cause major problems in fair marking.

The situation regarding any particular cases that arise while you are marking for which you feel you need detailed guidance should be discussed with your Team Leader.

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

Question			Answer	Marks	Guidance
1	(i)		$\mathrm{P}(X=x)=\dfrac{\mathrm{e}^{-\theta}\theta^{x}}{x!}$		
			$\mathrm{L}=\dfrac{\mathrm{e}^{-\theta}\theta^{x_1}}{x_1!}\cdots\cdots\dfrac{\mathrm{e}^{-\theta}\theta^{x_n}}{x_n!}=\dfrac{\mathrm{e}^{-n\theta}\theta^{\Sigma x_i}}{x_1!x_2!\ldots x_n!}$	M1 A1	M1 for general product form. A1 (a.e.f.) for answer.
			$\ln \mathrm{L} = -n\theta + \Sigma x_i \ln\theta - \Sigma \ln x_i!$	M1 A1	M1 is for taking logs (base e). Allow (±) constant instead of last term.
			$\dfrac{\mathrm{d}\ln \mathrm{L}}{\mathrm{d}\theta}=-n+\dfrac{\Sigma x_i}{\theta}$	M1 A1	M1 for differentiating, A1 for answer.
			$= 0$ for ML Est $\hat{\theta}$.	M1	Any or all of the four M1 marks down to here can be awarded in part (iv) if not awarded here.
			$\therefore n\hat{\theta}=\Sigma x_i$, i.e. $\hat{\theta}=\overline{x}$.	A1	
			Confirmation that this is a maximum:	M1	
			$\dfrac{\mathrm{d}^2\ln \mathrm{L}}{\mathrm{d}\theta^2}=-\dfrac{\Sigma x_i}{\theta^2}<0$.	A1 **[10]**	
1	(ii)		$\mathrm{P}(X=0)=\mathrm{e}^{-\theta}$.	M1	
			ML Est of $\mathrm{e}^{-\theta}=\mathrm{e}^{-\hat{\theta}}$, i.e. the estimate is $\mathrm{e}^{-\overline{x}}$.	M1 A1 **[3]**	M1 for "invariance property", not necessarily named.
1	(iii)		We have estimate of $\mathrm{P}(X=0)=\mathrm{e}^{-5}=0.0067$,	M1	
			so we might reasonably expect around $1000e^{-5}\approx 6.7$ cases of zero in a sample of size 1000	E1	Sensible use of $n=1000$ and $\overline{x}=5$.
			– finding no such cases seems suspicious.	E1 **[3]**	

Question			Answer	Marks	Guidance
1	(iv)		X has Poisson distribution "scaled up" so that , therefore ,	M1	M1 for this idea, however expressed.
			where $1 = k\sum_{x=1}^{\infty}\frac{e^{-\theta}\theta^{x}}{x!} = k\left\{\sum_{x=0}^{\infty}\frac{e^{-\theta}\theta^{x}}{x!} - e^{-\theta}\right\} = k\left\{1-e^{-\theta}\right\}$.	M1	M1 for sum from 0 to ∞ minus value for $x = 0$.
			$\therefore k = \frac{1}{1-e^{-\theta}}$.	M1	
			$\therefore P(X = x) = \frac{1}{1-e^{-\theta}}.\frac{e^{-\theta}\theta^{x}}{x!} = \frac{\theta^{x}}{(e^{\theta}-1)x!}$ [for $x = 1, 2, \ldots$].	A1	**Beware printed answer.**
			$\therefore$ L = $\frac{\theta^{\Sigma x_i}}{(e^{\theta}-1)^{n} x_1!x_2!\ldots x_n!}$	A1	
			$\therefore \ln L = \Sigma x_i \ln\theta - n\ln(e^{\theta}-1) - \Sigma \ln x_i!$	A1	
			$\therefore \frac{d\ln L}{d\theta} = \frac{\Sigma x_i}{\theta} - \frac{ne^{\theta}}{e^{\theta}-1}$,	A1	
			and on setting this equal to zero we get that $\hat{\theta}$ satisfies $\frac{\theta e^{\theta}}{e^{\theta}-1} = \frac{\Sigma x_i}{n} = \bar{x}$.	A1	**Beware printed answer.**
				[8]	
2	(i)	(*A*)	$\mu = 0$.	B1	
		(*B*)	$E(X^2) = 8/3$.	B1	
		(*C*)	$Var(X) = 8/3$.	B1	
				[3]	
2	(ii)		Mgf of X is $M_X(\theta) = E(e^{\theta X})$ $= \left(e^{-2\theta}.\frac{1}{3}\right) + \left(e^{0}.\frac{1}{3}\right) + \left(e^{2\theta}.\frac{1}{3}\right)$	M1	
			$= \frac{1}{3}\left(1 + e^{2\theta} + e^{-2\theta}\right)$.	A1	Any equivalent form.
				[2]	

Question			Answer	Marks	Guidance
2	**(iii)**		General results:		
			[Convolution theorem] Mgf of sum of independent random variables = product of their mgfs.	B1 M1	B1 for explicit mention of "independent". Note M1 mark required for f.t. to A marks below.
			[Linear transformation result] $\mathrm{M}_{aX+b}(\theta) = \mathrm{e}^{b\theta}\mathrm{M}_X(a\theta)$.	M1	Allow implicit b = 0. Note M1 mark required for f.t. to A marks below.
			$\mathrm{M}_{\Sigma X}(\theta) = \left\{\frac{1}{3}\left(1 + \mathrm{e}^{2\theta} + \mathrm{e}^{-2\theta}\right)\right\}^n$	A1	
			$\mathrm{M}_{\bar{X}}(\theta) = \left\{\frac{1}{3}\left(1 + \mathrm{e}^{2\frac{\theta}{n}} + \mathrm{e}^{-2\frac{\theta}{n}}\right)\right\}^n$	A1	
			$Z = \frac{\bar{X}\sqrt{n}\ n}{\sigma} - \frac{\sqrt{\ }\ n}{\sigma} = \frac{n\bar{X}\sqrt{3}}{2\sqrt{2}}$	B1	Might be implicit in what follows.
			$\mathrm{M}_Z(\theta) = \left\{\frac{1}{3}\left(1 + \mathrm{e}^{\frac{2}{n}\frac{\sqrt{3n}}{2\sqrt{2}}\theta} + \mathrm{e}^{-\frac{2}{n}\frac{\sqrt{3n}}{2\sqrt{2}}\theta}\right)\right\}^n$	A1	
			$= \left\{\frac{1}{3}\left(1 + \mathrm{e}^{\frac{\theta\sqrt{3}}{\sqrt{2n}}} + \mathrm{e}^{-\frac{\theta\sqrt{3}}{\sqrt{2n}}}\right)\right\}^n$.	A1	**Beware printed answer.**
				[8]	

Question			**Answer**	**Marks**	**Guidance**
2	**(iv)**		$\mathrm{M}_Z(\theta) = \{ \ \frac{1}{3}(1$	M1	M1 for reasonable attempt to expand exponentials.
			$+1+\frac{\theta\sqrt{3}}{\sqrt{2n}}+\frac{3\theta^2}{\sqrt{2}n.2!}+\text{terms in } n^{-3/2}, n^{-2}, \ldots$	A1	A1 for this line.
			$+1-\frac{\theta\sqrt{3}}{\sqrt{2n}}\sqrt{\ }+\frac{3\theta^2}{\sqrt{2}n.2!}+\text{terms in } n^{-3/2}, n^{-2}, \ldots)\}^n$	A1	A1 for this line.
			cancel $\sqrt{\ }$	M1	M1 for cancelling first order terms, may be implicit.
			neglect	M1	M1 for neglecting higher order terms in n^{-1} , MUST be explicit.
			$\approx \left\{\frac{1}{3}\left(3+\frac{3\theta^2}{2n}\right)\right\}^n$	A1	
			$= \left(1+\frac{\theta^2}{2n}\right)^n$.	A1	**Beware printed answer.**
				[7]	
2	**(v)**		Limit is $\mathrm{e}^{\theta^2/2}$. This is mgf of N(0, 1).	M1 M1	M1 for limit, M1 for recognising mgf.
			Mgfs are unique (even in this limiting process).	B1	Bracketed phrase is not needed to earn the mark.
			So (approximately) distribution of Z is N(0, 1).	A1	Bracketed word is not needed to earn the mark.
				[4]	
3	**(i)**		Type I error: rejecting null hypothesis	B1	Allow B1 out of 2 for P(...).
			when it is true	B1	
			Type II error: accepting null hypothesis	B1	Allow B1 out of 2 for P(...).
			when it is false	B1	
			OC: P(accepting null hypothesis	B1	P(Type II error | the true value of the parameter) scores B1+B1.
			as a function of the parameter under investigation)	B1	
			Power: P(rejecting null hypothesis	B1	P(Type I error | the true value of the parameter) scores B1+B1.
			as a function of the parameter under investigation)	B1	
				[8]	

Question			Answer	Marks	Guidance
3	**(ii)**		Correct "spike" shape, or point shown.	G1	
			Location of spike or point at θ_0 correct and correctly labelled.	G1	
			Height of spike correct and correctly labelled as 1.	G1	
				[3]	
3	**(iii)**		$X \sim \mathrm{N}(\mu, 9)$. $n = 25$. H_0: $\mu = 0$. H_1: $\mu \neq 0$. Accept H_0 if $-1 < \bar{x} < 1$.		
			P(Type I error) = $\mathrm{P}\left(\lvert\bar{X}\rvert > 1 \;\middle\vert\; \bar{X} \sim \mathrm{N}\left(0, \frac{9}{25}\right)\right)$	M1 M1	M1 for use of $\lvert\bar{X}\rvert > 1$, M1 for distribution of $\bar{X}$. Either or both marks might be implicit in what follows.
			$= \mathrm{P}\left(\lvert\mathrm{N}(0,1)\rvert > \frac{1-0}{3/5}\right) = 2 \times 0.0478 = 0.0956.$	A1	Accept a.w.r.t. 0.096.
			P(Type II error when $\mu = 0.5$) $= \mathrm{P}\left(-1 < \bar{X} < 1 \;\middle\vert\; \bar{X} \sim \mathrm{N}\left(0.5, \frac{9}{25}\right)\right)$	M1 M1	As for the two M1 marks for the Type I error.
			$= \mathrm{P}\left(\frac{-1.5}{3/5} < \mathrm{N}(0,1) < \frac{0.5}{3/5}\right) = \mathrm{P}\left(-2.5 < \mathrm{N}(0,1) < 0.8333\right)$	M1	Standardising with correct end-points.
			$= 0.7976 - 0.0062 = 0.7914.$	A1	
			This is high.	E1	
			We are trying to detect only a small departure from H_0, the "error" (σ^2) being comparatively large.	E1	
				[9]	
3	**(iv)**		Correct shape – must be symmetrical, and reasonable approximation to Normal pdf.	G1	
			"Centre" at 0 and clearly labelled.	G1	
			Height at 0 distinctly less than 1 by about 10% (ft candidate's P(Type I error)).	G1	
			Some indication that height at 0.5 is about 0.8 (ft candidate's P(Type II error)).	G1	
				[4]	

Question			Answer	Marks	Guidance
4	(i)		Randomisation is mainly to guard against possible sources of bias, which may be due to subjective allocation of treatments to units or unsuspected. Replication enables an estimate of experimental error to be made.	B1 B1 B1 B1 B1 B1 **[6]**	Award up to B3 marks. These include B1 for idea of "sources of bias" and another B1 for either idea of " unsuspected" or “subjective”. Award up to B3 marks. These include B1 for *some* concept of replication addressing experimental error and another B1 for the explicit idea that it enables this error to be estimated. The third B1 should be awarded for the general quality of the response, especially the lack of extraneous material in it Alternative suggestions offered by candidates may be rewarded *provided they are statistically sound.* Suggestions that are limited to particular designs (eg for the advantages of randomisation within a randomised blocks design) may be rewarded if statistically sound, but for at most 2 of the B3 marks in each set.
4	(ii)		μ is the population mean for the entire experiment. α_i is the population amount by which the mean for the ith treatment differs from μ. e_{ij} ~ ind N (0, σ^2)	B1 B1 B1 B1 B1 B1 B1 **[7]**	B1 for explicit mention of "population", B1 for idea of mean for whole experiment. B1 for explicit mention of "population", B1 for idea of difference of means. For "ind N"; allow "uncorrelated". For mean 0. For variance σ^2 [i.e. that the variance is constant].

<table>
<tr><th colspan="3">Question</th><th>Answer</th><th>Marks</th><th>Guidance</th></tr>
<tr><td>4</td><td>(iii)</td><td></td><td>Null hypothesis: $\alpha_1 = \alpha_2 = ... = \alpha_k$ (= 0)</td><td>B1</td><td>No need for definition of <i>k</i> as the number of treatments, and accept simply $\alpha_1 = \alpha_2 =...$ with no explicit upper end to the sequence.</td></tr>
<tr><td></td><td></td><td></td><td>Alternative hypothesis: not all α_i are equal
Note that alternative hypothesis is NOT that all the α_i are different – B0 for alternative hypothesis if this is stated.</td><td>B1</td><td>Accept hypotheses stated verbally provided it is clear that <i>population</i> parameters (means) are being referred to.</td></tr>
<tr><td></td><td></td><td></td><td>
<table>
<tr><th>Source of variation</th><th>Sum of squares</th><th>d.f.</th><th>Mean squares</th><th>MS ratio</th></tr>
<tr><td>Between fertilisers</td><td>219.2</td><td>4</td><td>54.8</td><td>2.7</td></tr>
<tr><td>Residual</td><td>304.5</td><td>15</td><td>20.3</td><td></td></tr>
<tr><td>Total</td><td>523.7</td><td>19</td><td></td><td></td></tr>
</table>
</td><td>B1 B1
M1
M1
A1</td><td>B1 for <i>each</i> d.f. (4 and 15).
For <i>method</i> of mean squares.
For <i>method</i> of mean square ratio.
A1 c.a.o. for 2.7 (2.6695).
f.t. from here provided all M marks earned.</td></tr>
<tr><td></td><td></td><td></td><td>Refer 2.7 to $F_{4,15}$.</td><td>M1</td><td>No f.t. if wrong but allow M1 for F with candidate’s df if both positive and totalling 19.</td></tr>
<tr><td></td><td></td><td></td><td>Upper 5% point is 3.06.</td><td>A1</td><td>cao. No f.t. if wrong (or if not quoted).</td></tr>
<tr><td></td><td></td><td></td><td>Not significant.</td><td>A1</td><td></td></tr>
<tr><td></td><td></td><td></td><td>Seems mean effects of fertilisers are all the same.</td><td>E1
[11]</td><td>Verbal conclusion in context, and not "too assertive".</td></tr>
</table>

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

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

GCE

Mathematics (MEI)

Advanced GCE **A2 7895-8**

Advanced Subsidiary GCE **AS 3895-8**

OCR Report to Centres

June 2013

4769 Statistics 4

General Comments

The majority of candidates for this paper were very well prepared. There were instances where candidates attempted more than the three questions demanded by the rubric, but on the whole this was not to the detriment of the questions that would count. As in previous years, question 4 on Design and Analysis of Experiments was the least popular, attempted by only half the candidates, with Question 3 on Inference marginally more popular than the other two.

Comments on Individual Questions

1) *Estimation*

The first part was, on the whole, very well done with many scripts scoring full marks. Ideally candidates should show explicitly that the derivative of the log-likelihood function is set to zero to obtain the estimator, but the lack of this was not penalised. The final step of demonstrating that the Likelihood function achieved a maximum was not always fully explained.

Part (ii) was mostly well done but precision in the expressions employed is expected. The expression for $\mathrm{P}(X=0)$ should employ the theoretical parameter θ , it is an estimator of this probability that uses $\hat{\theta}$.

In part (iii), most candidates made sensible use of the figures given, 1000 trials and sample mean 5, commenting that the observed number of minutes reported, in which no cars were seen, was well short of the expected number they found, predicted by the model. Some candidates attempted to calculate the probability of observing no minutes without cars but did not always turn the event into "$X>0$" in each minute observed. Some numerical justification was expected. The use of the Normal approximation was of doubtful validity in this borderline situation.

In part (iv), many candidates failed to give a convincing argument deriving the probability distribution, with several not attempting this at all. Most of the successful answers used the idea of scaling the Poisson probabilities so that the sum from 1 to infinity was one, and most achieved the right scale factor. With undue haste to get to the required result, some algebraic mistakes did occur. There were some odd notations used seen. Candidates who used the idea of conditioning on $X \neq 0$ did not always express the argument coherently, but this was a quick and effective method to use. The final section of the question was mostly well done, where candidates could proceed with the quoted distribution, having earned the method points in part (i).

2) *Generating Functions*

In part (i), three marks were almost always earned, as was the case for moment generating function in part (ii).

Part (iii) required some clear thinking and careful algebra. Not all candidates started by stating the results requested. When these were given, initially or later on, the need for the independence of the random variables in the sum was mostly ignored or forgotten.

The complicating factor of $e^{\frac{-\mu\sqrt{n}}{\sigma}}$ could have been removed by using $\mu = 0$ early on. The most successful and convincing arguments defined Z in terms of $\sum_{i=1}^{n} X_i$ rather than $\bar{X}$ which resolved the problem of where $\sqrt{n}$ appeared in the exponents. It was not always clear that candidates knew how to deal with all the steps required. Some of this was due to confused notation, between $\mathrm{M}_X(\theta)$, $\mathrm{M}_{\bar{X}}(\theta)$, and $\mathrm{M}_{\sum X}(\theta)$.

Part (iv) was mostly well done. The most common mistake was to omit explicit reference to the effect of large n on the expansion as a power series, or mentioning powers of θ instead of n^{-1} .

Part (v) was also well done, with most candidates recognising the mgf of the standard Normal distribution and quoting the uniqueness property.

3) *Inference*

Part (i) well answered but for the few candidates who wanted to define the types of error as probabilities. The Operating Characteristic and the Power were sometimes not always described clearly as being functions of the parameter in question.
Power = 1 – Operating Characteristic is not acceptable as a definition of the power function.
Apart from some instances of strange labelling in (ii), the required extreme operating characteristic of the perfect test was well drawn.

In part (iii), most candidates were able to find the Type I error probability correctly. The Type II error probability was occasionally erroneous. Symmetrically placed bounds on the acceptance region were sometimes used; incorrect bounds and incorrect probability calculations were also seen. Most candidates said that their Type II error probability was high, not many made sensible comments on why this was so.

Part (iv) was usually well done, but for a few incomplete curves and instances of inadequate or incorrect labelling.

4) *Design and analysis of experiments*

Part (i) produced responses which tended to lack concisely expressed relevant points. Where examples of designs were given the descriptions could be helpful, but could also be vague. Randomisation, to counter sources of bias, was on the whole more successfully described than the merits of replication. Not many candidates explicitly stated that the latter made possible the estimation of experimental error variance.

Part (ii) was mostly well answered. The definitions should be well known and likewise the emphasis on the parameters being those of the *population* under consideration.

In part (iv,) the hypotheses were carefully stated and analysis was nearly always successful. There were some instances of over-assertive conclusions and some where the context, about the mean yields from different fertilizers, forgotten.

Unit level raw mark and UMS grade boundaries June 2013 series

AS GCE / Advanced GCE / AS GCE Double Award / Advanced GCE Double Award

GCE Mathematics (MEI)

			Max Mark	a	b	c	d	e	u
4751/01	(C1) MEI Introduction to Advanced Mathematics	Raw	72	62	56	51	46	41	0
		UMS	100	80	70	60	50	40	0
4752/01	(C2) MEI Concepts for Advanced Mathematics	Raw	72	54	48	43	38	33	0
		UMS	100	80	70	60	50	40	0
4753/01	(C3) MEI Methods for Advanced Mathematics with Coursework: Written Paper	Raw	72	58	52	46	40	33	0
4753/02	(C3) MEI Methods for Advanced Mathematics with Coursework: Coursework	Raw	18	15	13	11	9	8	0
4753/82	(C3) MEI Methods for Advanced Mathematics with Coursework: Carried Forward Coursework Mark	Raw	18	15	13	11	9	8	0
4753	(C3) MEI Methods for Advanced Mathematics with Coursework	UMS	100	80	70	60	50	40	0
4754/01	(C4) MEI Applications of Advanced Mathematics	Raw	90	66	59	53	47	41	0
		UMS	100	80	70	60	50	40	0
4755/01	(FP1) MEI Further Concepts for Advanced Mathematics	Raw	72	63	57	51	45	40	0
		UMS	100	80	70	60	50	40	0
4756/01	(FP2) MEI Further Methods for Advanced Mathematics	Raw	72	61	54	48	42	36	0
		UMS	100	80	70	60	50	40	0
4757/01	(FP3) MEI Further Applications of Advanced Mathematics	Raw	72	60	52	44	36	28	0
		UMS	100	80	70	60	50	40	0
4758/01	(DE) MEI Differential Equations with Coursework: Written Paper	Raw	72	62	56	51	46	40	0
4758/02	(DE) MEI Differential Equations with Coursework: Coursework	Raw	18	15	13	11	9	8	0
4758/82	(DE) MEI Differential Equations with Coursework: Carried Forward Coursework Mark	Raw	18	15	13	11	9	8	0
4758	(DE) MEI Differential Equations with Coursework	UMS	100	80	70	60	50	40	0
4761/01	(M1) MEI Mechanics 1	Raw	72	57	49	41	33	25	0
		UMS	100	80	70	60	50	40	0
4762/01	(M2) MEI Mechanics 2	Raw	72	50	43	36	29	22	0
		UMS	100	80	70	60	50	40	0
4763/01	(M3) MEI Mechanics 3	Raw	72	64	56	48	41	34	0
		UMS	100	80	70	60	50	40	0
4764/01	(M4) MEI Mechanics 4	Raw	72	56	49	42	35	29	0
		UMS	100	80	70	60	50	40	0
4766/01	(S1) MEI Statistics 1	Raw	72	55	48	41	35	29	0
		UMS	100	80	70	60	50	40	0
4767/01	(S2) MEI Statistics 2	Raw	72	58	52	46	41	36	0
		UMS	100	80	70	60	50	40	0
4768/01	(S3) MEI Statistics 3	Raw	72	61	55	49	44	39	0
		UMS	100	80	70	60	50	40	0
4769/01	(S4) MEI Statistics 4	Raw	72	56	49	42	35	28	0
		UMS	100	80	70	60	50	40	0
4771/01	(D1) MEI Decision Mathematics 1	Raw	72	58	52	46	40	35	0
		UMS	100	80	70	60	50	40	0
4772/01	(D2) MEI Decision Mathematics 2	Raw	72	58	52	46	41	36	0
		UMS	100	80	70	60	50	40	0
4773/01	(DC) MEI Decision Mathematics Computation	Raw	72	46	40	34	29	24	0
		UMS	100	80	70	60	50	40	0
4776/01	(NM) MEI Numerical Methods with Coursework: Written Paper	Raw	72	56	50	44	38	31	0
4776/02	(NM) MEI Numerical Methods with Coursework: Coursework	Raw	18	14	12	10	8	7	0
4776/82	(NM) MEI Numerical Methods with Coursework: Carried Forward Coursework Mark	Raw	18	14	12	10	8	7	0
4776	(NM) MEI Numerical Methods with Coursework	UMS	100	80	70	60	50	40	0
4777/01	(NC) MEI Numerical Computation	Raw	72	55	47	39	32	25	0
		UMS	100	80	70	60	50	40	0
4798/01	(FPT) Further Pure Mathematics with Technology	Raw	72	57	49	41	33	26	0
		UMS	100	80	70	60	50	40	0

GCE Statistics (MEI)

			Max Mark	a	b	c	d	e	u
G241/01	(Z1) Statistics 1	Raw	72	55	48	41	35	29	0
		UMS	100	80	70	60	50	40	0
G242/01	(Z2) Statistics 2	Raw	72	55	48	41	34	27	0
		UMS	100	80	70	60	50	40	0
G243/01	(Z3) Statistics 3	Raw	72	56	48	41	34	27	0
		UMS	100	80	70	60	50	40	0