Thursday 12 June 2014 – Afternoon

A2 GCE MATHEMATICS (MEI)

4768/01 Statistics 3

QUESTION PAPER

Candidates answer on the Printed Answer Book.

OCR supplied materials:
- Printed Answer Book 4768/01
- MEI Examination Formulae and Tables (MF2)

Other materials required:
- Scientific or graphical calculator

Duration: 1 hour 30 minutes

3144250734

INSTRUCTIONS TO CANDIDATES

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

- The Question Paper will be found inside 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 **all** the 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 **12** 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.

1 **(i)** Let X be a random variable with variance σ^2. The independent random variables X_1 and X_2 are both distributed as X. Write down the variances of $X_1 + X_2$ and $2X$; explain why they are different. **[3]**

A large company has produced an aptitude test which consists of three parts. The parts are called mathematical ability, spatial awareness and communication. The scores obtained by candidates in the three parts are continuous random variables X, Y and W which have been found to have independent Normal distributions with means and standard deviations as shown in the table.

	Mean	Standard deviation
Mathematical ability, X	30.1	5.1
Spatial awareness, Y	25.4	4.2
Communication, W	28.2	3.9

(ii) Find the probability that a randomly selected candidate obtains a score of less than 22 in the mathematical ability part of the test. **[3]**

(iii) Find the probability that a randomly selected candidate obtains a total score of at least 100 in the whole test. **[4]**

(iv) For a particular role in the company, the score $2X + Y$ is calculated. Find the score that is exceeded by only 2% of candidates. **[4]**

(v) For a different role, a candidate must achieve a score in communication which is at least 60% of the score obtained in mathematical ability. What proportion of candidates do not achieve this? **[3]**

2 **(i)** Explain what is meant by a simple random sample. **[2]**

A manufacturer produces tins of paint which nominally contain 1 litre. The quantity of paint delivered by the machine that fills the tins can be assumed to be a Normally distributed random variable.

The machine is designed to deliver an average of 1.05 litres to each tin. However, over time paint builds up in the delivery nozzle of the machine, reducing the quantity of paint delivered. Random samples of 10 tins are taken regularly from the production process. If a significance test, carried out at the 5% level, suggests that the average quantity of paint delivered is less than 1.02 litres, the machine is cleaned.

(ii) By carrying out an appropriate test, determine whether or not the sample below leads to the machine being cleaned.

0.994 1.010 1.021 1.015 1.016 1.022 1.009 1.007 1.011 1.026 **[9]**

Each time the machine has been cleaned, a random sample of 10 tins is taken to determine whether or not the average quantity of paint delivered has returned to 1.05 litres.

(iii) On one occasion after the machine has been cleaned, the quality control manager thinks that the distribution of the quantity of paint is symmetrical but not necessarily Normal. The sample on this occasion is as follows.

1.055 1.064 1.063 1.043 1.062 1.070 1.059 1.044 1.054 1.053

By carrying out an appropriate test at the 5% level of significance, determine whether or not this sample supports the conclusion that the average quantity of paint delivered is 1.05 litres. **[8]**

3 **(a)** A personal trainer believes that drinking a glass of beetroot juice an hour before exercising enables endurance tests to be completed more quickly. To test his belief he takes a random sample of 12 of his trainees and, on two occasions, asks them to carry out 100 repetitions of a particular exercise as quickly as possible. Each trainee drinks a glass of water on one occasion and a glass of beetroot juice on the other occasion.

The times in seconds taken by the trainees are given in the table.

Trainee	Water	Beetroot juice
A	75.1	72.9
B	86.2	79.9
C	77.3	71.6
D	89.1	90.2
E	67.9	68.2
F	101.5	95.2
G	82.5	76.5
H	83.3	80.2
I	102.5	99.1
J	91.3	82.2
K	92.5	90.1
L	77.2	77.9

The trainer wishes to test his belief using a paired *t* test at the 1% level of significance. Assuming any necessary assumptions are valid, carry out a test of the hypotheses $H_0: \mu_D = 0$, $H_1: \mu_D < 0$, where μ_D is the population mean difference in times (time with beetroot juice minus time with water). **[8]**

(b) An ornithologist believes that the number of birds landing on the bird feeding station in her garden in a given interval of time during the morning should follow a Poisson distribution. In order to test her belief, she makes the following observations in 60 randomly chosen minutes one morning.

Number of birds	0	1	2	3	4	5	6	$\geqslant 7$
Frequency	2	5	10	17	14	7	4	1

Given that the data in the table have a mean value of 3.3, use a goodness of fit test, with a significance level of 5%, to investigate whether the ornithologist is justified in her belief. **[11]**

Question 4 begins on page 4

4 The probability density function of a random variable X is given by

$$f(x) = \begin{cases} kx & 0 \leqslant x \leqslant a, \\ k(2a-x) & a < x \leqslant 2a, \\ 0 & \text{otherwise,} \end{cases}$$

where a and k are positive constants.

(i) Sketch $f(x)$. Hence explain why $E(X) = a$. **[3]**

(ii) Show that $k = \frac{1}{a^2}$. **[3]**

(iii) Find $\text{Var}(X)$ in terms of a. **[4]**

In order to estimate the value of a, a random sample of size 50 is taken from the distribution. It is found that the sample mean and standard deviation are $\bar{x} = 1.92$ and $s = 0.8352$.

(iv) Construct a symmetrical 95% confidence interval for a. Give one reason why the answer is only approximate. **[5]**

(v) A non-statistician states that the probability that a lies in the interval found in part **(iv)** is 0.95. Comment on this statement. **[2]**

END OF QUESTION PAPER

Thursday 12 June 2014 – Afternoon

A2 GCE MATHEMATICS (MEI)

4768/01 Statistics 3

PRINTED ANSWER BOOK

Candidates answer on this Printed Answer Book.

OCR supplied materials:
- Question Paper 4768/01 (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 inside 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 **all** the 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 **12** pages. The Question Paper consists of **4** pages. Any blank pages are indicated.

Turn over

1 (i)	
1 (ii)	

1 (iii)	
1 (iv)	

1 (v)	
2 (i)	
2 (ii)	
	(answer space continued on next page)

2 (ii)	(continued)

2 (iii)	

3 (a)	

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

3 (b)	(continued)
4 (i)	

4 (ii)	
4 (iii)	

4 (iv)

4 (v)	

OCR
Oxford Cambridge and RSA

GCE

Mathematics (MEI)

Unit **4768**: Statistics 3

Advanced GCE

Mark Scheme for June 2014

1. Annotations and abbreviations

Annotation in scoris	Meaning
BP	Blank Page – this annotation **must** be used on all blank pages within an answer booklet (structured or unstructured) and on each page of an additional object where there is no candidate response.
✓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

2. 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)		$\text{Var}(X_1 + X_2) = 2\sigma^2$	B1	Allow 2Var(X) and 4Var(X)
			$\text{Var}(2X) = 4\sigma^2$	B1	
			$X_1 + X_2$ means two independent values from X are added together. $2X$ means that one value from X is multiplied by 2.	E1 **[3]**	Any comment explaining why $X_1 + X_2$ is different from $2X$
1	(ii)		$\text{P}(X < 22) = \text{P}\left(Z < \frac{22 - 30.1}{5.1}\right)$	M1	For standardising. Award once, here or elsewhere
			$= \text{P}(Z < -1.5882)$	A1	Correct z value
			= 0.0561	A1 **[3]**	cao
1	(iii)		$X + Y + W \sim \text{N}(83.7, 58.86)$	B1 B1	Mean Variance (or sd = 7.67)
			$\text{P}(X + Y + W > 100) = \text{P}\left(\text{Z} > \frac{100 - 83.7}{\sqrt{58.86}}\right)$	M1	Correct set up
			$\text{P}(Z > 2.1246) = 0.0168$	A1 **[4]**	cao
1	(iv)		$2X \sim \text{N}(60.2, 104.04)$		
			$\rightarrow 2X + Y \sim \text{N}(85.6, 121.68)$	B1	Variance
			$\text{P}(2X + Y > b) = 0.02$		
			$\rightarrow \frac{b - 85.6}{\sqrt{121.68}} = 2.054$	B1 M1	2.054 seen Correct set up
			$\rightarrow b = 108.26$ Score exceeded by 2% is 108.3	A1 **[4]**	cao

Question			Answer	Marks	Guidance
1	(v)		$P(W < 0.6X) = P(W - 0.6X < 0)$	M1	Either way round
			$W - 0.6X \sim N(0.14, 24.5736)$	B1	Mean and variance
			$P(W - 0.6X < 0) = P(Z < -2.0455)$		
			$=0.0204$	A1 **[3]**	Cao. Allow convincing recovery
2	(i)		A simple random sample is one where every sample of the same size has an equal probability of being selected.	E2,1,0 **[2]**	Allow E1 for every item has the same probability of being selected
2	(ii)		$H_0 : \mu = 1.02 \quad H_1 : \mu < 1.02$	B1	Both hypotheses. Hypotheses in words only must include "population". Do NOT allow "$\bar{X} = ...$" or similar unless $\bar{X}$ is clearly and explicitly stated to be a <u>population</u> mean.
			Where μ is the population mean volume	B1	For adequate verbal definition. Allow absence of "population" if correct notation μ is used.
			$\bar{x} = 1.0131 \quad s = 0.009146$	B1	Do not allow $s_n = 0.00868$ here or in construction of test statistic
			Test statistic is $\dfrac{1.0131 - 1.02}{0.009146/\sqrt{10}}$	M1*	For method, allow candidates $\bar{x}, s$. Allow confidence interval approach.
			=-2.3857 value between –2.38 and –2.39	A1	cao
			Refer to t_9	M1	For t_9 No FT from here
			5% point is ± 1.833	A1	For matching 1.833 seen. No FT from here
			$-2.3857 < -1.833$ reject H_0	M1 dep	For rejection. Must compare test statistic with <u>matching</u> 1.833 unless absolute values are being compared. No ft from here if wrong.
			Conclude mean appears to be below 1.02 and so machine will be cleaned.	E1 dep **[9]**	Needs context

Question		Answer	Marks	Guidance
2	(iii)	H_0 : median of population volumes = 1.05 H_1 : median of population volumes $\neq$ 1.05	B1	Both Allow location parameter or mean (if explanation is given)

Difference	0.005	0.014	0.013	-0.007	0.012	0.020	0.009	-0.006	0.004	0.003
Rank	3	9	8	5	7	10	6	4	2	1

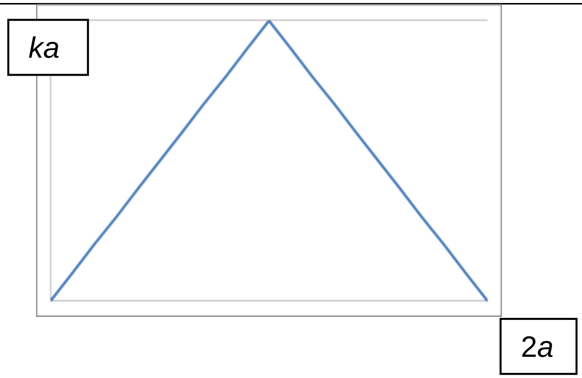
Answer	Marks	Guidance
	M1	For differences from 1.05. ZERO from this point if differences not used.
	M1	For ranks.
	A1	ft from here if ranks wrong.
$W_- = 4 + 5 = 9$	B1	Or $W_+ = 1 + 2 + 3 + 6 + 7 + 8 + 9 + 10 = 46$
Refer to tables of Wilcoxon paired (/single sample) statistic for *n* = 10	M1	No FT from here if wrong.
Critical value = 8. (47 for upper)	A1	No FT from here if wrong (i.e. 1 tail)
Cannot reject H_0 and conclude median volume appears to be 1.05	E1	Must have context.
	[8]	

Question			Answer	Marks	Guidance
3	(a)		$H_0 : \mu_D = 0 \quad H_1 : \mu_D < 0$ Where μ_D is the population mean of the differences. MUST BE PAIRED COMPARISON *t* test. Differences are –2.2, –6.3, –5.7, 1.1, 0.3, –6.3, –6.0, –3.1, –3.4, –9.1, –2.4, 0.7	M1	
			$\overline{D} = -3.533 \quad s_D = 3.225$	A1	Do not allow $s_n = 3.088$
			Test statistic is $\dfrac{-3.533 - 0}{3.225/\sqrt{12}}$	M1*	For method, allow their $\overline{D}, s$. Allow confidence interval approach.
			$= -3.79$	A1	cao
			Refer to t_{11}	M1	No FT from here if wrong
			Single-tailed 1% point is –2.718	A1	No FT from here if wrong
			–3.79 < –2.718 so reject H_0	A1dep	FT their –3.79 if relevant M1 earned
			Conclude mean time appears reduced	A1dep **[8]**	In context. FT their –3.79 if relevant M1 earned

Question		Answer	Marks	Guidance
3	(b)	H_0 : The Poisson model fits the data H_1 : The Poisson model does not fit the data	B1	Both hypotheses. Must be the right way round. Do not accept "data fits model" or equivalent.
			M1	At least 3 probabilities to 3dp or better or 3 expected values to 3sf or better
			M1	Multiply by 60 to obtain expected values
			A1	All correct to 3sf or better

r	$P(X=r)$	Expected value	Combined
0	0.03688	2.213	9.516
1	0.12171	7.303	
2	0.20083	12.050	
3	0.22091	13.255	
4	0.18225	10.935	
5	0.12029	7.217	
6	0.06616	3.969	7.027
≥ 7	0.05097	3.058	

Answer	Marks	Guidance
	M1	Merge first 2 and last 2 cells
$X^2 = \frac{2.516^2}{9.516} + \frac{2.050^2}{12.050} + \frac{3.745^2}{13.255} + \frac{3.065^2}{10.935} + \frac{0.217^2}{7.217} + \frac{2.027^2}{7.027}$	M1*	Calculation of X^2
= 0.6652+0.3488+1.0581+0.8591+0.0065+0.5847		
= 3.522 awrt 3.52	A1	cao
Refer to χ_4^2	M1	Allow correct df from wrongly grouped table.
Upper 5% point is 9.49	B1	No FT from here if wrong.
3.522 < 9.49 cannot reject H_0	A1dep	FT candidates 3.522 if relevant M1 earned.
Poisson model appears to fit data.	A1dep	FT candidates 3.522 if relevant M1 earned. Do not accept "data fits Poisson model" or equivalent.
	[11]	

Question			Answer	Marks	Guidance
4	(i)			G1	Shape
				G1	Scales on axes
			$\mathrm{E}(X) = a$ because the distribution is symmetrical about $x = a$	B1	Do not allow integration method
				[3]	
4	(ii)		Total area = $\frac{1}{2}2a.ka$ Or $\int_0^a kx\,\mathrm{d}x + \int_a^{2a} k(2a - x)\,\mathrm{d}x$ $k\left[\frac{x^2}{2}\right]_0^a + k\left[2ax - \frac{x^2}{2}\right]_a^{2a}$ $k\left(\frac{a^2}{2} - 0\right) + k\left(2a^2 - \frac{3a^2}{2}\right)$	M1	Attempting to find area of triangle or setting up the correct integrals including limits (which may appear later).
			$= ka^2$	A1	oe Correctly finding area in terms of k, a
			$ka^2 = 1$ $k = \frac{1}{a^2}$	A1	Equating area to 1 and convincingly obtaining result. Answer given
				[3]	

Question			Answer	Marks	Guidance
4	(iii)		$\text{Var}(X) = k\int_0^a x^3\,\mathrm{d}x + k\int_a^{2a}(2ax^2 - x^3)\,\mathrm{d}x - a^2$	M1	Correct integral for $\mathrm{E}(X^2)$ including limits (which may appear later).
			$k\left[\frac{x^4}{4}\right]_0^a + k\left[\frac{2ax^3}{3} - \frac{x^4}{4}\right]_a^{2a} - a^2$	M1	Correctly integrated (dependent on M1 above)
			$\frac{a^2}{4} + \frac{16a^2}{3} - 4a^2 - \frac{2a^2}{3} + \frac{a^2}{4} - a^2$	M1	Using $\mathrm{E}(X^2) - (\mathrm{E}(X))^2$
			$\text{Var}(X) = \frac{a^2}{6}$	A1	cao
				[4]	
4	(iv)		$\bar{x} = 1.92 \qquad s = 0.8352$		Given
				B1	1.96 seen
			Interval is $1.92 \pm 1.96\frac{0.8352}{\sqrt{50}}$	M1	Correct SE
				M1	Centred on 1.92
			$= (1.69, 2.15)$	A1	cao
			The distribution of $\bar{X}$ is *approximately* Normal (CLT) or s is only an estimate.	E1	
				[5]	
4	(v)		This statement is incorrect	E1	
			The value of a either lies within this particular interval or it does not – so the probability is either 0 or 1. If a large number of such intervals were constructed, then *a* would lie in 95% of them.	E1	A comment either about p=0 or 1, or about a large number of intervals
				[2]	

4768 Statistics 3

General Comments:

As might be expected on a paper at this level, the scripts indicated that most candidates knew what they were doing most of the time. In addition, there were few scripts which showed evidence of candidates running out of time. Candidates seemed to be far more comfortable carrying out calculations than with the other requirements of the paper such as producing hypotheses and conclusions, interpreting results and providing definitions. In addition, as in previous years, many scripts suffered from a lack of precision. This manifested itself in many ways, inadequate hypotheses, over-assertive conclusions, over-specified final answers yet too little accuracy carried forward in calculations, inaccurate reading of tables, and finally a large number of scripts which were very difficult to read.

Comments on Individual Questions:

Question No. 1

Linear combinations of random variables

This question was done well by most candidates and many scored full marks for parts ii, iii, and iv.

Q1i nearly all candidates gained the first 2 marks correctly, showing that they could recognise the difference between adding two different values of X together, and multiplying one value of X by 2. Most candidates could not put this distinction into words, and few gained the third mark.

Q1ii This part was well done. The only errors of note were calculating z to only one or two decimal places, or the introduction of a spurious continuity correction.

Q1ii Again this part was done well and most candidates were able to sum the variables correctly and find the correct tail probability. The most common error was the lack of accuracy in calculating z, and then in reading the statistical tables.

Q1iv In this part, more candidates were unable to find the variance of $2X + Y$, but this was still rare. Most candidates were able to identify 2.054 as the relevant 2% point and thereby gave a final value in the correct tail. Some candidates used the value -2.054 and so ended up in the wrong tail. In this part many candidates gave their answers to 5 or more significant figures.

Q1v Many candidates in this part incorrectly started with the statement $W – 0.6X > 0$, but then a large proportion recovered at the end, often using a statement like “and so the proportion who did not achieve this =”. Candidates who took this approach often calculated the mean and variance correctly. Some candidates tried to deal with W and $0.6X$ separately.

Answers (ii) 0.0561 (iii) 0.0168 (iv) 108.3 (v) 0.0204

Question 2 Single sample *t* test and Wilcoxon test

Most candidates know how to carry out these tests, and most carried out the appropriate test in each case. Many candidates lost marks through taking insufficient care with hypotheses and conclusions. Symbols used in hypotheses need to be defined and in context. Conclusions also need to be in context and must not be over assertive.

Q2i Many candidates gave accurate definitions stating that every sample of a given size has the same chance of being selected. A significant number stated that every member of the population has the same chance of being selected, and a few simply described how a random sample might be carried out.

Q2ii Most candidates were able to calculate $\bar{x}$ and *s* correctly, although a significant number used truncated values in what followed. The test statistic was usually correctly carried out, although on occasion $\sqrt{10}$ was missing. The majority of candidates then chose the correct critical value of *t*, although occasionally two tailed values were seen, as were values of *z*.
The great majority of candidates made the correct decision in terms of rejecting the null hypothesis.

Q2iii This part was very well done, the most common error was not defining the population median in the hypotheses. The vast majority correctly found the differences from 1.05, although a few found the differences from 1.057, the median of the data. Most candidates identified 8 as the critical value, although a significant number gave 10 as the value. A few candidates carried out another t test here.

Question 3 Paired t test and goodness of fit test

Most candidates again knew how to carry out these tests. The comments in question 2 about hypotheses and conclusions are equally appropriate here.

Q3a Most candidates were able to calculate the mean and standard deviation correctly, although a few calculated “water – beetroot” rather than the other way around as specified in the question. Candidates usually calculated the mean and standard deviation correctly, but again many used truncated values. The test statistic was usually correctly carried out, although on occasion $\sqrt{12}$ was missing. The majority of candidates then chose the correct critical value of t, although occasionally two tailed values were seen, as were values of *z*.
The great majority of candidates made the correct decision in terms of rejecting the null hypothesis. A few candidates carried out a Wilcoxon test here

Q3b In goodness of fit tests it is particularly important to work to a high level of accuracy, and a significant minority did not do this. Most candidates were able to work out the Poisson probabilities correctly, but some calculated Prob($X = 7$) instead of Prob($X \geq 7$). Most candidates the multiplied by 60 to obtain the expected values, but a few multiplied by 100. Most candidates merged the first two and last two cells, although a few failed to merge any cells and a few merged only the last two cells. Those who had worked out Prob($X = 7$) often merged the last three cells. Most candidates were able to calculate the statistic correctly. The majority of candidates used the correct number of degrees of freedom, but a few failed to subtract 1 because the mean had been calculated from the data. A few candidates looked at the wrong tail in the chi-squared tables.

Question 4 Two part pdf and confidence interval

This question proved to be the most difficult for the candidates.

Q4i The great majority of candidates were able to sketch the correct shape. The horizontal axis was virtually always correct, but many either gave no value on the vertical axis or an incorrect value. Most candidates realised the key feature of the distribution was its symmetry, although a few stated that it was the fact that the distribution achieved its maximum value at a.

Q4ii The better candidates used the fact that the area of a triangle could be used at this point, or that the first integral should have a value of 0.5. These candidates invariably gained full marks. Those who stated that the sum of two integrals should equal 1 often made errors. These errors involved dealing with k and with signs. As the value of k was given, these candidates invariably ended with the correct value.

Q4iii Only the best candidates were able to obtain the correct value of $a^2/6$. Many candidates did not know how to calculate $E(X^2)$ and many were unable to deal with k, inserting the limits, and the signs correctly. In addition some omitted to subtract a^2 at the end.

Q4iv Most candidates knew how to calculate a confidence interval. Here the main errors seen were the use of 1.645 or 2.009 instead of 1.96 and giving the final interval to 5 or more significant figures.

Q4v Most candidates were able to give a correct definition of a 95% confidence interval, but a significant number still felt that the non-statistician was correct.

Unit level raw mark and UMS grade boundaries June 2014 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	61	56	51	46	42	0
	UMS	100	80	70	60	50	40	0
4752/01 (C2) MEI Concepts for Advanced Mathematics	Raw	72	57	51	45	39	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	47	42	36	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	68	61	54	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	60	54	48	42	36	0
	UMS	100	80	70	60	50	40	0
4757/01 (FP3) MEI Further Applications of Advanced Mathematics	Raw	72	57	51	45	39	34	0
	UMS	100	80	70	60	50	40	0
4758/01 (DE) MEI Differential Equations with Coursework: Written Paper	Raw	72	63	56	50	44	37	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	34	27	0
	UMS	100	80	70	60	50	40	0
4762/01 (M2) MEI Mechanics 2	Raw	72	57	49	41	34	27	0
	UMS	100	80	70	60	50	40	0
4763/01 (M3) MEI Mechanics 3	Raw	72	55	48	42	36	30	0
	UMS	100	80	70	60	50	40	0
4764/01 (M4) MEI Mechanics 4	Raw	72	48	41	34	28	22	0
	UMS	100	80	70	60	50	40	0
4766/01 (S1) MEI Statistics 1	Raw	72	61	53	46	39	32	0
	UMS	100	80	70	60	50	40	0
4767/01 (S2) MEI Statistics 2	Raw	72	60	53	46	40	34	0
	UMS	100	80	70	60	50	40	0
4768/01 (S3) MEI Statistics 3	Raw	72	61	54	47	41	35	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	51	46	41	36	31	0
	UMS	100	80	70	60	50	40	0
4772/01 (D2) MEI Decision Mathematics 2	Raw	72	46	41	36	31	26	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	54	48	43	38	32	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	61	53	46	39	32	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