OCR
Oxford Cambridge and RSA

Friday 5 June 2015 – Morning

A2 GCE MATHEMATICS (MEI)

4767/01 Statistics 2

QUESTION PAPER

Candidates answer on the Printed Answer Book.

OCR supplied materials:
- Printed Answer Book 4767/01
- MEI Examination Formulae and Tables (MF2)

Other materials required:
- Scientific or graphical calculator

Duration: 1 hour 30 minutes

1204423576

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 A random sample of wheat seedlings is planted and their growth is measured. The table shows their average growth, y mm, at half-day intervals.

Time t days	0	0.5	1	1.5	2	2.5	3
Average growth y mm	0	7	21	33	45	56	62

(i) Draw a scatter diagram to illustrate these data. **[3]**

(ii) Calculate the equation of the regression line of y on t. **[5]**

(iii) Calculate the value of the residual for the data point at which $t = 2$. **[3]**

(iv) Use the equation of the regression line to calculate an estimate of the average growth after 5 days for wheat seedlings. Comment on the reliability of this estimate. **[2]**

It is suggested that it would be better to replace the regression line by a line which passes through the origin.

You are given that the equation of such a line is $y = at$, where $a = \dfrac{\sum yt}{\sum t^2}$.

(v) Find the equation of this line and plot the line on your scatter diagram. **[4]**

2 It was stated in 2012 that 3% of £1 coins were fakes. Throughout this question, you should assume that this is still the case.

(i) Find the probability that, in a random selection of 25 £1 coins, there is exactly one fake coin. **[2]**

A random sample of 250 £1 coins is selected.

(ii) Explain why a Poisson distribution is an appropriate approximating distribution for the number of fake coins in the sample. **[2]**

(iii) Use a Poisson distribution to find the probability that, in this sample, there are

(A) exactly 10 fake coins, **[3]**

(B) at least 10 fake coins. **[2]**

(iv) Use a suitable approximating distribution to find the probability that there are at least 50 fake coins in a sample of 2000 coins. **[5]**

It is known that 0.2% of another type of coin are fakes.

(v) A random sample of size n of these coins is taken. Using a Poisson approximating distribution, show that the probability of at most one fake coin in the sample is equal to $e^{-\lambda} + \lambda e^{-\lambda}$, where $\lambda = 0.002n$. **[2]**

(vi) Use the approximation $e^{-\lambda} + \lambda e^{-\lambda} \approx 1 - \dfrac{\lambda^2}{2}$ for small values of λ to estimate the value of n for which the probability in part **(v)** is equal to 0.995. **[3]**

3 The random variable X represents the weight in kg of a randomly selected male dog of a particular breed. X is Normally distributed with mean 30.7 and standard deviation 3.5.

(i) Find

(A) $\mathrm{P}(X < 30)$, **[3]**

(B) $P(25 < X < 35)$. **[3]**

(ii) Five of these dogs are chosen at random. Find the probability that each of them weighs at least 30 kg. **[2]**

(iii) The weights of females of the same breed of dog are Normally distributed with mean 26.8 kg. Given that 5% of female dogs of this breed weigh more than 30 kg, find the standard deviation of their weights. **[4]**

(iv) Sketch the distributions of the weights of male and female dogs of this breed on a single diagram. **[4]**

4 **(a)** As part of an investigation into smoking, a random sample of 120 students was selected. The students were asked whether they were smokers, and also whether either of their parents were smokers. The results are summarised in the table below. Test, at the 5% significance level, whether there is any association between the smoking habits of the students and their parents.

	At least one parent smokes	Neither parent smokes
Student smokes	21	27
Student does not smoke	17	55

[10]

(b) The manufacturer of a particular brand of cigarette claims that the nicotine content of these cigarettes is Normally distributed with mean 0.87 mg. A researcher suspects that the mean nicotine content of this brand is higher than the value claimed by the manufacturer. The nicotine content, x mg, is measured for a random sample of 100 cigarettes. The data are summarised as follows.

$$\Sigma x = 88.20 \qquad \Sigma x^2 = 78.68$$

Carry out a test at the 1% significance level to investigate the researcher's belief. **[10]**

END OF QUESTION PAPER

OCR
Oxford Cambridge and RSA

Friday 5 June 2015 – Morning

A2 GCE MATHEMATICS (MEI)

4767/01 Statistics 2

PRINTED ANSWER BOOK

Candidates answer on this Printed Answer Book.

OCR supplied materials:
- Question Paper 4767/01 (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.

1 (i)

1 (ii)

(answer space continued on next page)

1 (ii)	**(continued)**
1 (iii)	
1 (iv)	

1 (v)	
2 (i)	
2 (ii)	
2 (iii) (*A*)	

2 (iii) (*B*)	
2 (iv)	

2 (v)	
2 (vi)	
3 (i) (*A*)	

3 (i) (*B*)	
3 (ii)	

3 (iii)	
3 (iv)	

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

4 (a)	(continued)

4 (b)

(answer space continued on next page)

4 (b)	**(continued)**

OCR
Oxford Cambridge and RSA

Annotations and abbreviations

Annotation in scoris	**Meaning**
✓and ✘	
BOD	Benefit of doubt
FT	Follow through
ISW	Ignore subsequent working
M0, M1	Method mark awarded 0, 1
A0, A1	Accuracy mark awarded 0, 1
B0, B1	Independent mark awarded 0, 1
SC	Special case
^	Omission sign
MR	Misread
Highlighting	
Other abbreviations in mark scheme	**Meaning**
E1	Mark for explaining
U1	Mark for correct units
G1	Mark for a correct feature on a graph
M1 dep*	Method mark dependent on a previous mark, indicated by *
cao	Correct answer only
oe	Or equivalent
rot	Rounded or truncated
soi	Seen or implied
www	Without wrong working
awrt	Answers which round to

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. Where over-specification is penalised, no more than two marks per question and no more than four marks in total per script should be lost.

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.

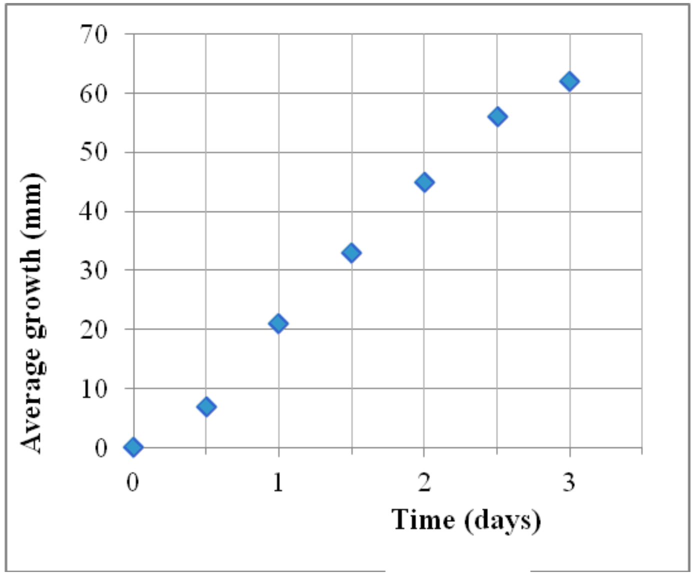
A misread should normally attract *some* penalty, though this would often be only 1 mark and should rarely if ever be more than 2. Commonly in sections of questions where there is a numerical answer either at the end of the section or to be obtained and commented on (eg the value of a test statistic), 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.

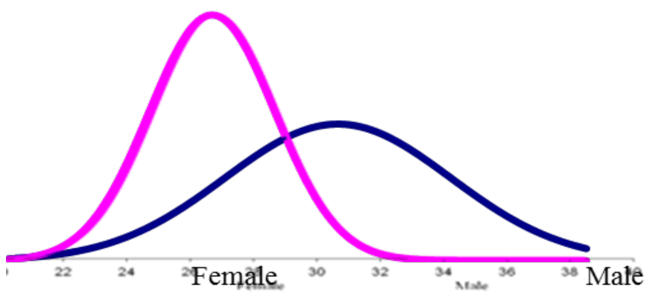
Note that a miscopy of the candidate’s own working is not a misread but an accuracy error.

Question			Answer	Marks	Guidance	
1	(i)			G1*	Both axes labeled (allow t and y) with indication of scale	Allow axes interchanged Condone x for t
				G1dep*	for values of time BOD if (0,0) not clearly visible	(evenly spaced)
				G1dep*	for values of average growth BOD if (0,0) not clearly visible. BOD if confusion arises from points plotted for part (v).	visually correct SC1 for points having the correct distribution and G0* awarded. Line through origin should appear but this is rewarded in part **(v)**
				[3]		
1	(ii)		$\bar{t} = 1.5$, $\bar{y} = 32$	B1	For $\bar{t}$ and $\bar{y}$ seen or implied by final answer.	Seen either in calculating b or in forming the equation of the line.
			$b = \frac{S_{yt}}{S_{tt}} = \frac{490-(224\times10.5/7)}{22.75-10.5^2/7} = \frac{154}{7} = 22$ OR $b = \frac{490/7-(32\times1.5)}{22.75/7-1.5^2} = \frac{22}{1} = 22$	M1	For attempt at gradient (b)	Correct structure needed. See additional notes. FT their $\bar{t}$ and $\bar{y}$ for M1
				A1	For 22 cao	
			hence least squares regression line is: $y - \bar{y} = b(t - \bar{t})$	M1	For equation of line	With their $b > 0$, $\bar{t}$ and $\bar{y}$

Question			Answer	Marks	Guidance	
			$\Rightarrow \quad y - 32 = 22(t - 1.5)$ $\Rightarrow \quad y = 22t - 1$	A1 **[5]**	CAO	A0 for $y = 22x - 1$
1	(iii)		$t = 2 \Rightarrow$ predicted average growth $= (22 \times 2) - 1 = 43$ Residual = 45 – 43 = 2	 B1 M1 A1 **[3]**	 for prediction for subtraction (either way) FT	 FT their equation 45 – their prediction
1	(iv)		$(22 \times 5) - 1 = 109$ Likely to be **unreliable as extrapolation** (oe)	B1 B1 **[2]**	Estimate calculated using equation	FT their equation
1	(v)		$a = \dfrac{490}{22.75} = 21.538\ldots = 21.5$ (3 s.f.) Equation is $y = 21.5t$ Line plotted on diagram	M1 A1 A1 A1 **[4]**	 Allow $y = 21.54t$ CAO For line correctly plotted CAO A0 if axes not scaled or $a \neq 21.5$ to 3 sf	 Allow $y = (280/13)t$ Through (0,0) and between (3, 64) and (3,65)
2	(i)		P(Exactly one) = $\binom{25}{1} \times 0.03^1 \times 0.97^{24}$ = 0.361	M1 A1 **[2]**	Binomial calculation with correct structure Allow 0.3611 and 0.36 www A0 for 0.3612	$25 \times p \times (1 - p)^{24}$
2	(ii)		n is large p is small.	B1 B1 **[2]**	n large or sample is large p is small, or $np \approx np(1 - p)$ B0 for the “probability” is small unless “probability” is correctly defined.	or $n > 30$ or $np < 10$

Question			Answer	Marks	Guidance	
2	(iii)	(*A*)	Mean = 250 × 0.03 = 7.5 P(exactly 10) = $e^{-7.5}\dfrac{7.5^{10}}{10!}$ Or from tables = 0.8622 – 0.7764 = 0.0858	B1 M1 A1 **[3]**	For mean (SOI) For Poisson probability calculation Allow 0.08583 or 0.086www	Or using $P(X \le 10)$ - $P(X \le 9)$ with Poisson tables
2	(iii)	(*B*)	P(At least 10) = 1 – $P(X \le 9)$ = 1 – 0.7764 = 0.2236	M1 A1 **[2]**	For **using** 1 - $P(X \le 9)$ CAO	Allow 0.224 www
2	(iv)		Mean 2000 × 0.03 = 60 Variance = 2000 × 0.03 × 0.97 = 58.2 Using Normal approx. to the binomial, $X \sim N(60, 58.2)$ $P(X \ge 50) = P\left(Z \ge \dfrac{49.5-60}{\sqrt{58.2}}\right)$ = $P(Z > -1.376)$ = $\Phi(1.376)$ = 0.9157 (allow 0.9156 and 0.916)	B1 B1 B1 M1 A1 **[5]**	Normal approximation **used** For parameters (soi) For correct continuity correction For probability using correct structure. CAO **(**Do not FT wrong or omitted CC)	Award full credit for use of Normal approximation to Poisson distribution N(60, 60) N(60, 60) leads to $P(Z > -1.356)$ = 0.9125 (or 0.913) Allow 0.9124 (or 0.912)
2	(v)		Using a Poisson approximation to the binomial the mean λ = ***np*** (= 0.002n) or λ = **n × 0.002** (= 0.002*n*). P(At most one fake coin) = P(zero or one fake coins) = $e^{-\lambda}\dfrac{\lambda^0}{0!}+e^{-\lambda}\dfrac{\lambda^1}{1!}=e^{-\lambda}+\lambda e^{-\lambda}$ **AG**	B1 B1 **[2]**	For evidence of using *np* from binomial distribution (to give λ = 0.002*n*) Evidence of using P(X = 0) + P(X = 1) with λ = 0.002*n* ***NB* ANSWER GIVEN**	Need to **see use of *n*×*p*** or obtaining 0.002*n* from B(*n*, 0.002)

Question			Answer	Marks	Guidance	
2	(vi)		$1-\frac{\lambda^2}{2} = 0.995$ $\lambda^2 = 0.01$ so $\lambda = 0.1$ $n = 50$	M1 A1 A1 **[3]**	For equation in λ or **equivalent** equation in n For λ SOI or for $n^2 = 2500$ CAO	
3	(i)	(*A*)	P(X < 30) $= \text{P}\left(Z < \frac{30-30.7}{3.5}\right)$ = P(Z < -0.20) = Φ(-0.20) = 1 – Φ(0.20) = (1 – 0.5793) = 0.4207	M1 M1 A1 **[3]**	For standardising For correct structure CAO	Penalise erroneous continuity corrections and wrong sd. Condone numerator reversed. **1 –** Φ(positive *z*) Allow 0.421 www
3	(i)	(*B*)	P(25 < X < 35) $= \text{P}\left(\frac{25-30.7}{3.5} < Z < \frac{35-30.7}{3.5}\right)$ = P(-1.629 < X < 1.229) = Φ(1.229) – Φ(-1.629) = 0.8904 – (1 – 0.9483) = 0.8904 – 0.0517 = 0.8387	M1 M1 A1 **[3]**	Correctly standardising both. For correct structure Use of differences column required	Penalise erroneous continuity corrections and wrong sd. Condone both numerators reversed. Φ(1.23) – Φ(-1.63) leads to 0.8907 – 0.0516 =0.8391 Only allow 0.839 if 0.8387 is seen.

Question			Answer	Marks	Guidance	
3	(ii)		P(all 5 weigh at least 30kg) $= 0.5793^5$	M1	Allow FT $(1 - \text{their (i)}(A))^5$ or [their $P(X \geq 30)]^5$	
			$= 0.0652$	A1	FT only $(1 - \text{their (i)}(A))^5$	Allow 0.06524, allow 0.065 www
				[2]		
3	(iii)		P(weight > 30) = 0.05 $P(Z > \frac{30-26.8}{\sigma}) = 0.05$			
			$\Phi^{-1}(0.95) = 1.645$	B1	For 1.645. B0 for 1 – 1.645 or 0.1645	NOTE use of -1.645 allowed only if numerator reversed. Condone use of spurious c.c. if already penalised in parts (i)(*A*) or (i)(*B*). See additional guidance notes.
			$\frac{30-26.8}{\sigma} = 1.645$	M1*	For equation as seen or equivalent, with their $z > 1$.	
			$\sigma = \frac{30-26.8}{1.645} = 1.945$ kg	M1dep*	Rearranging for σ	Allow $\sigma = 1.95$ www
				A1	CAO	
				[4]		
3	(iv)			G1	For two Normal shapes including attempt at asymptotic behaviour with horizontal axis at each of the four ends	Penalise clear asymmetry
				G1	For means, shown explicitly or by scale **on a single diagram**	If shown explicitly, the positions must be consistent with horizontal scale if present.
				G1	For lower max height for Male	If not labelled, assume the larger mean represents Male
				G1	For visibly greater width for	If not labelled, assume the

Question			Answer	Marks	Guidance	
				[4]	Male	larger mean represents Male
4	(a)		H_0: no association between student smoking and parent smoking H_1: some association between student smoking and parent smoking	B1	Correct hypotheses in context NB if H_0 H_1 reversed do not award first B1 or final B1dep*	Allow hypotheses in terms of independence, in context. Do not allow "relationship" or "correlation" for "association"
			Expected frequency table: Student smokes: Parent smokes 15.2, Parent does not smoke 32.8 Student does not smoke: Parent smokes 22.8, Parent does not smoke 49.2	B1 B1	For at least one row/column of expected values correct May be implied by correct contributions or correct X^2 All correct	
			Contribution table: Student smokes: Parent smokes 2.213, Parent does not smoke 1.026 Student does not smoke: Parent smokes 1.475, Parent does not smoke 0.684	M1 A1	For valid attempt at (O-E)^2/E All correct (to 3 d.p.)	NB These M1A1 marks cannot be implied by a correct final value of X^2
			$X^2 = 5.398$	B1	Allow awrt 5.40	Do not penalise use of Yates correction, giving $X^2 = 4.51$
			Refer to χ_1^2 Critical value at 5% level = 3.841	B1 B1	For 1 degree of freedom CAO for cv.	 p value = 0.02016

Expected frequency	Parent smokes	Parent does not smoke
Student smokes	15.2	32.8
Student does not smoke	22.8	49.2

Contribution	Parent smokes	Parent does not smoke
Student smokes	2.213	1.026
Student does not smoke	1.475	0.684

Question			Answer	Marks	Guidance	
					No further marks from here if wrong or omitted, unless *p*-value used instead.	
			Result is significant	B1*		For significant oe FT their test statistic
			There is sufficient evidence to **suggest** that there is **association between student smoking and parent smoking.**	B1dep*	NB if H_0 H_1 reverse do not award first B1 or final B1dep*	For **non-assertive** conclusion in context Allow conclusion in terms of independence FT their test statistic. Do not allow "relationship" or "correlation" for "association".
				[10]		
4	**(b)**		$\bar{x} = 88.2/100 = 0.882$	B1	For 0.882 seen.	
			$s = \sqrt{\frac{78.68-(88.2)^2/100}{99}} = \sqrt{\frac{0.8876}{99}}$ $= \sqrt{0.0089657}$	M1	For correctly structured calculation for the sample standard deviation or variance	
			= 0.0947 (allow 0.095 www)	A1	Allow A1 for $s^2 = 0.0089657$	or 0.00897 (allow 0.009**0**)
			H_0: $\mu = 0.87$; H_1 : $\mu > 0.87$	B1	For both correct	Hypotheses in words only must refer to population.
			Where μ denotes the **mean nicotine** content (of cigarettes of this brand in the population)	B1	For definition of μ in context.	Do not allow other symbols unless clearly defined as population mean.

Question			Answer	Marks	Guidance	
			Test statistic = $\frac{0.882-0.87}{0.0947/\sqrt{100}}=\frac{0.012}{0.00947}$	M1*	including correct use of √100.	FT their s (not s = 0.87 or √0.87 or 88.2 or √88.2 or 78.68 or √78.68)
			=1.267	A1	CAO	
			Upper 1% level 1 tailed critical value of z = 2.326	B1	For 2.326 **No further marks from here if wrong**	
			1.267 < 2.326 (Not significant.)	M1dep*	For sensible comparison leading to a conclusion (even if incorrect)	
			There is **insufficient evidence to suggest** that the **mean nicotine content** of this brand is **greater** than 0.87mg.	A1 **[10]**	For non-assertive conclusion in words and in context. FT only candidate's test statistic	

Additional Notes on Correct Structure in Q1(ii)

Equivalent calculations for finding b are allowed. For example use of $7S_{yt}/7S_{tt}$ is allowed. However, where these are mixed we award M0. e.g. use of $7S_{yt}/S_{tt}$ would earn M0. For M1 to be awarded, the calculation must be structurally equivalent to the one provided – NOTE if it is believed that the candidate has made an error in transcription of a number (for example using 244 instead of 224) we can allow M1 BOD if the structure is otherwise correct.

Additional Notes for Q3 (iii)

M1* is for forming a suitable equation using their z-value but it must be reasonably clear that the value used is a z-value – for example we do not allow 0.05 or 0.95 to be treated as z-values here. The M1dep* can be awarded if the candidate correctly rearranges their equation to find σ. Hence, use of an incorrect z-value could earn max B0M1*M1dep*A0. However, if it is clear that the z-value is from the wrong tail (e.g. -1.645 used in place of +1.645) then award 0/4. In cases where -1.645 is used and the numerator of the equation is reversed allow full credit and annotate with BOD.

Additional Notes on Sensible Comparisons

In Q4 (b) Neither 1.267 > 0.05 nor 0.1026 < 2.326 are considered sensible as each compares a z-value with a probability.
For 1.267 > 2.326 leading to a conclusion, allow M1A0.

Additional Notes on Conclusions to Hypothesis Tests

The following are examples of conclusions which are considered too assertive.

There is sufficient evidence to reject H_0 and **conclude** that…
"there is a positive association between…" or
"there seems to be evidence that there is a positive association between…" or
"the mean nicotine content is greater …."
"there doesn't appear to be association between…"

Also note that final conclusions **must refer to H_1 in context** for the final mark to be given.
e.g. In Q4 (a), a conclusion just stating that "the evidence suggests that there is association" gets A0 as this does not refer to the context.

Additional Notes on Alternative Methods in Q4 (b)

Critical value method	cv = 0.87 + 2.326 × 0.0946÷ √100	gets M1* B1 (for 2.326)
	= 0.8920	gets A1 cao (replacing the A1 for 1.267)
	0.882 < 0.8920	gets M1dep* if a conclusion is made. The final A1 available as before if 2.326 used.
Probability Method	P(sample mean > 1.267) = 0.1026	gets M1*A1 B1 (the B1 for 0.1026 (allow 0.1025), from tables, replaces the B1 for 2.326).
	0.1026 > 0.01	gets M1dep* if a conclusion is made. The final A1 available as before provided that B1 for 0.1026 awarded NOTE Condone B1 0.8974 (0.8975) if compared with 0.99 at which point the final M1dep*A1 are available. B0M0A0A0 if 0.8974 obtained from P(sample mean > -1.267).

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

4767 Statistics 2

General Comments

The overall performance of candidates taking this paper was very good. It was again pleasing to see candidates taking care over the wording used when carrying out hypothesis tests, with the majority providing non-assertive conclusions referring to the alternative hypothesis. Most candidates demonstrated good understanding of the processes required. In general, statistical calculations were completed with success, though a minority of candidates seemed unsure where inequalities were involved. A minority of candidates used spurious continuity corrections when calculating Normal probabilities. Once again, fewer over-specified answers were seen than in previous years. There appeared to be more candidates making use of graphical display calculators this year, often with good success; however, where incorrect answers were provided and no method described, examiners could give no credit. Candidates seemed to have sufficient time to complete the paper and very few failed to complete all 4 questions.

Comments on Individual Questions

Question 1

(i) The scatter diagram allowed candidates to make a successful start to the paper, the only common errors being either omitting scales or labels.

(ii) The many successful attempts at this usually found the means, $\bar{t}$ and $\bar{y}$, first and then calculated the other required totals: $\sum t^2$, $\sum y^2$and $\sum yt$ before finding the regression line gradient and using this in the equation of the line, referenced to ($\bar{t}, \bar{y}$).

(iii) Many full marks here also with the errors being either finding the prediction and not finding a residual at all, or reversing the required order of $y_{observed} - y_{predicted}$.

(iv) Once candidates had found the predicted value, the majority were aware of the dangers of extrapolation and thus gained maximum marks on this part of the question. Those who opted to describe the extrapolation process tended to succeed but a small minority didn't. Comments such as "unreliable, as the point is not in the data set" were not given credit. . Some candidates chose to write about biological, rather than mathematical, reasons for unreliability.

(v) This part, as intended, tested more candidates but there were still many who gained full marks. The most frequent errors in the final equation were either related to excessive accuracy being used or to not stating the equation at all.

Question 3

This proved to be a very straightforward question with most candidates scoring high marks.

(i) Generally well-answered - a minority of candidates not recognising the need to use the binomial distribution, opting to use Poisson instead.

(ii) Generally well-answered, though many candidates included extra information that wasn't relevant.

(iii) Both parts were answered well.

(iv) Generally answered well with most candidates correctly identifying an acceptable approximation using the Normal distribution. The main issues were with not using the appropriate continuity correction or getting the wrong structure.

(v) Not well-answered, with many candidates failing to show sufficient detail in arriving at the given answer. Many seemed to reiterate the given result, merely changing the order of the terms or adding some multiplication signs or just replacing lambda with 0.002n.. Candidates who used lambda rather than 0.002n usually fared better.

(vi) This question was answered well. It was very pleasing to see that the vast majority of candidates used an algebraic approach. Some were let down by poor rearranging skills. A very small minority of candidates used a trial and improvement strategy, with mixed success.

Question 3

There were many good responses to this question, especially on the early parts. Spurious continuity corrections were wrongly applied, throughout parts (i) to (iii), in a minority of scripts - most involving +/- 0.5 but others involving +/- 1. These continuity corrections led to a loss of both method and accuracy marks.

(iA) The majority of candidates correctly standardised the given value then found the correct probability. Some candidates made good use of diagrams to indicate their intent.

(iB) For a question that had potential for many errors, either in standardising, approximating too early, or looking up values incorrectly, this question was answered well. This part was in general done rather better than part (iA).

(ii) The binomial situation, within a question fundamentally on the Normal distribution, was dealt with well by many candidates. The errors that did occur were often about forgetting that (iA) found $p(X < 30)$ whereas this part asked for all five dogs to weigh *more than* 30 kg.

(iii) Few candidates used the wrong tail for this calculation, most correctly identifying that +1.645 was the appropriate value to use as 5% weighed *more than* 30 kg.

(iv) This part of the question tested most candidates. Thus many found dealing with the correct shape and the various relative constraints on the curves difficult. The mark scheme will help with interpretation of the requirements.

Question 4

Most candidates seemed to have been well-coached in successfully finding chi-squared test statistics and applying the test. Hypotheses and final conclusions were generally well-worded and contributions given to an appropriate degree of accuracy. Part b caused the most problems with many candidates being unfamiliar with the techniques for testing for the mean when the population variance was unknown.

(a) Many candidates did very well on this question. The hypotheses were generally correct, though a few candidates referred to correlation. The expected values were generally calculated correctly. The contributions were often calculated correctly but not always shown to three significant figures and sometimes not shown at all. Not all candidates stated the number of degrees of freedom. Most students phrased conclusions correctly but some reached the wrong conclusion from correct values and others were too assertive in their conclusion or failed to put the conclusion in context.

(b) This wasn't as well done as the first part but a lot of good work was seen.
The sample mean was generally found but it not always recorded explicitly. Finding the sample standard deviation caused problems for many candidates; spurious formulae were prevalent and the n-1 divisor often missing. Hypotheses were generally correct though some candidates neglected to define μ; when the definition was provided it was usually in context as the population mean. Few candidates defined μ as the sample mean. The test statistic was usually correctly structured, though accuracy was sometimes lost through inappropriate rounding of the sample standard deviation (often to one significant figure).

Most candidates provided the correct critical value, made a sensible comparison which lead to the correct conclusion. Marks were sometimes lost from a failure to refer to the context in the final conclusion. Overly assertive conclusions and conclusions focusing on the null hypothesis were sometimes seen, but not as frequently as in previous years. Alternative approaches were seen (critical value and probability method) though these tended not to be as well done.

OCR
Oxford Cambridge and RSA

GCE Mathematics (MEI)				Max Mark	a	b	c	d	e	u
4751	01	C1 – MEI Introduction to advanced mathematics (AS)	Raw	72	63	58	53	48	43	0
			UMS	100	80	70	60	50	40	0
4752	01	C2 – MEI Concepts for advanced mathematics (AS)	Raw	72	56	50	44	39	34	0
			UMS	100	80	70	60	50	40	0
4753	01	(C3) MEI Methods for Advanced Mathematics with Coursework: Written Paper	Raw	72	56	51	46	41	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
			UMS	100	80	70	60	50	40	0
4754	01	C4 – MEI Applications of advanced mathematics (A2)	Raw	90	74	67	60	54	48	0
			UMS	100	80	70	60	50	40	0
4755	01	FP1 – MEI Further concepts for advanced mathematics (AS)	Raw	72	62	57	53	49	45	0
			UMS	100	80	70	60	50	40	0
4756	01	FP2 – MEI Further methods for advanced mathematics (A2)	Raw	72	65	58	52	46	40	0
			UMS	100	80	70	60	50	40	0
4757	01	FP3 – MEI Further applications of advanced mathematics (A2)	Raw	72	59	52	46	40	34	0
			UMS	100	80	70	60	50	40	0
4758	01	(DE) MEI Differential Equations with Coursework: Written Paper	Raw	72	63	57	51	45	38	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
			UMS	100	80	70	60	50	40	0
4761	01	M1 – MEI Mechanics 1 (AS)	Raw	72	62	54	46	39	32	0
			UMS	100	80	70	60	50	40	0
4762	01	M2 – MEI Mechanics 2 (A2)	Raw	72	54	47	40	33	27	0
			UMS	100	80	70	60	50	40	0
4763	01	M3 – MEI Mechanics 3 (A2)	Raw	72	64	56	48	41	34	0
			UMS	100	80	70	60	50	40	0
4764	01	M4 – MEI Mechanics 4 (A2)	Raw	72	53	45	38	31	24	0
			UMS	100	80	70	60	50	40	0
4766	01	S1 – MEI Statistics 1 (AS)	Raw	72	61	54	47	41	35	0
			UMS	100	80	70	60	50	40	0
4767	01	S2 – MEI Statistics 2 (A2)	Raw	72	65	60	55	50	46	0
			UMS	100	80	70	60	50	40	0
4768	01	S3 – MEI Statistics 3 (A2)	Raw	72	64	58	52	47	42	0
			UMS	100	80	70	60	50	40	0
4769	01	S4 – MEI Statistics 4 (A2)	Raw	72	56	49	42	35	28	0
			UMS	100	80	70	60	50	40	0
4771	01	D1 – MEI Decision mathematics 1 (AS)	Raw	72	56	51	46	41	37	0
			UMS	100	80	70	60	50	40	0
4772	01	D2 – MEI Decision mathematics 2 (A2)	Raw	72	54	49	44	39	34	0
			UMS	100	80	70	60	50	40	0
4773	01	DC – MEI Decision mathematics computation (A2)	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	45	40	34	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
			UMS	100	80	70	60	50	40	0
4777	01	NC – MEI Numerical computation (A2)	Raw	72	55	47	39	32	25	0
			UMS	100	80	70	60	50	40	0
4798	01	FPT - Further pure mathematics with technology (A2)	Raw	72	57	49	41	33	26	0
			UMS	100	80	70	60	50	40	0

OCR
Oxford Cambridge and RSA

GCE Statistics (MEI)											
				Max Mark	a	b	c	d	e	u
G241	01	Statistics 1 MEI (Z1)	Raw	72	61	54	47	41	35	0
			UMS	100	80	70	60	50	40	0
G242	01	Statistics 2 MEI (Z2)	Raw	72	55	48	41	34	27	0
			UMS	100	80	70	60	50	40	0
G243	01	Statistics 3 MEI (Z3)	Raw	72	56	48	41	34	27	0
			UMS	100	80	70	60	50	40	0

GCE Quantitative Methods (MEI)										
				Max Mark	a	b	c	d	e	u
G244	01	Introduction to Quantitative Methods MEI	Raw	72	58	50	43	36	28	0
G244	02	Introduction to Quantitative Methods MEI	Raw	18	14	12	10	8	7	0
			UMS	100	80	70	60	50	40	0
G245	01	Statistics 1 MEI	Raw	72	61	54	47	41	35	0
			UMS	100	80	70	60	50	40	0
G246	01	Decision 1 MEI	Raw	72	56	51	46	41	37	0
			UMS	100	80	70	60	50	40	0