OCR
RECOGNISING ACHIEVEMENT

Thursday 31 May 2012 – Morning

A2 GCE MATHEMATICS (MEI)

4767 Statistics 2

QUESTION PAPER

Candidates answer on the Printed Answer Book.

OCR supplied materials:
- Printed Answer Book 4767
- 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 **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 The times, in seconds, taken by ten randomly selected competitors for the first and last sections of an Olympic bobsleigh run are denoted by x and y respectively. Summary statistics for these data are as follows.

$$\Sigma x = 113.69 \quad \Sigma y = 52.81 \quad \Sigma x^2 = 1292.56 \quad \Sigma y^2 = 278.91 \quad \Sigma xy = 600.41 \quad n = 10$$

(i) Calculate the sample product moment correlation coefficient. **[5]**

(ii) Carry out a hypothesis test at the 10% significance level to investigate whether there is any correlation between times taken for the first and last sections of the bobsleigh run. **[6]**

(iii) State the distributional assumption which is necessary for this test to be valid. Explain briefly how a scatter diagram may be used to check whether this assumption is likely to be valid. **[2]**

(iv) A commentator says that in order to have a fast time on the last section, you must have a fast time on the first section. Comment briefly on this suggestion. **[2]**

(v) (*A*) Would your conclusion in part **(ii)** have been different if you had carried out the hypothesis test at the 1% level rather than the 10% level? Explain your answer. **[2]**

(*B*) State one advantage and one disadvantage of using a 1% significance level rather than a 10% significance level in a hypothesis test. **[2]**

2 A particular genetic mutation occurs in one in every 300 births on average. A random sample of 1200 births is selected.

(i) State the exact distribution of X, the number of births in the sample which have the mutation. **[2]**

(ii) Explain why X has, approximately, a Poisson distribution. **[2]**

(iii) Use a Poisson approximating distribution to find

(*A*) $P(X = 1)$,

(*B*) $P(X > 4)$. **[5]**

(iv) Twenty independent samples, each of 1200 births, are selected. State the mean and variance of a Normal approximating distribution suitable for modelling the total number of births with the mutation in the twenty samples. **[2]**

(v) Use this Normal approximating distribution to

(*A*) find the probability that there are at least 90 births which have the mutation, **[3]**

(*B*) find the least value of k such that the probability that there are at most k births with this mutation is greater than 5%. **[4]**

3 At a vineyard, the process used to fill bottles with wine is subject to variation. The contents of bottles are independently Normally distributed with mean $\mu = 751.4$ ml and standard deviation $\sigma = 2.5$ ml.

(i) Find the probability that a randomly selected bottle contains at least 750 ml. **[3]**

(ii) A case of wine consists of 6 bottles. Find the probability that all 6 bottles in a case contain at least 750 ml. **[2]**

(iii) Find the probability that, in a random sample of 25 cases, there are at least 2 cases in which all 6 bottles contain at least 750 ml. **[4]**

It is decided to increase the proportion of bottles which contain at least 750 ml to 98%.

(iv) This can be done by changing the value of μ, but retaining the original value of σ. Find the required value of μ. **[4]**

(v) An alternative is to change the value of σ, but retain the original value of μ. Find the required value of σ. **[3]**

(vi) Comment briefly on which method might be easier to implement and which might be preferable to the vineyard owners. **[2]**

[Question 4 is printed overleaf.]

4 **(a)** Mary is opening a cake shop. As part of her market research, she carries out a survey into which type of cake people like best. She offers people 4 types of cake to taste: chocolate, carrot, lemon and ginger. She selects a random sample of 150 people and she classifies the people as children and adults. The results are as follows.

		Classification of person		Row totals
		Child	Adult	
Type of cake	Chocolate	34	23	57
	Carrot	16	18	34
	Lemon	4	18	22
	Ginger	13	24	37
Column totals		67	83	150

The contributions to the test statistic for the usual χ^2 test are shown in the table below.

		Classification of person	
		Child	Adult
Type of cake	Chocolate	2.8646	2.3124
	Carrot	0.0436	0.0352
	Lemon	3.4549	2.7889
	Ginger	0.7526	0.6075

The sum of these contributions, correct to 2 decimal places, is 12.86.

(i) Calculate the expected frequency for children preferring chocolate cake. Verify the corresponding contribution, 2.8646, to the test statistic. **[3]**

(ii) Carry out the test at the 1% level of significance. **[5]**

(b) Mary buys flour in bags which are labelled as containing 5 kg. She suspects that the average contents of these bags may be less than 5 kg. In order to test this, she selects a random sample of 8 bags and weighs their contents. Assuming that weights are Normally distributed with standard deviation 0.0072 kg, carry out a test at the 5% level, given that the weights of the 8 bags in kg are as follows.

4.992 4.981 5.006 4.982 4.996 5.009 4.991 5.003 **[9]**

OCR
RECOGNISING ACHIEVEMENT

Thursday 31 May 2012 – Morning

A2 GCE MATHEMATICS (MEI)

4767 Statistics 2

PRINTED ANSWER BOOK

Candidates answer on this Printed Answer Book.

OCR supplied materials:
- Question Paper 4767 (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 **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.

OCR is an exempt Charity
DC (LEO) 43764/4

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

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

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

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

2 (v) (*A*)

2 (v) (*B*)

3 (i)

3 (ii)

3 (iii)

3 (iv)

3 (v)

3 (vi)

4 (a) (i)

4 (a) (ii)

(answer space continued on next page)

4 (a) (ii)	**(continued)**
4 (b)	
	(answer space continued overleaf)

4 (b)	(continued)

OCR
RECOGNISING ACHIEVEMENT

GCE

Mathematics (MEI)

Advanced GCE

Unit **4767:** Statistics 2

Mark Scheme for June 2012

OCR (Oxford Cambridge and RSA) is a leading UK awarding body, providing a wide range of qualifications to meet the needs of candidates of all ages and abilities. OCR qualifications include AS/A Levels, Diplomas, GCSEs, OCR Nationals, Functional Skills, Key Skills, Entry Level qualifications, NVQs and vocational qualifications in areas such as IT, business, languages, teaching/training, administration and secretarial skills.

It is also responsible for developing new specifications to meet national requirements and the needs of students and teachers. OCR is a not-for-profit organisation; any surplus made is invested back into the establishment to help towards the development of qualifications and support, which keep pace with the changing needs of today's society.

This mark scheme is published as an aid to teachers and students, to indicate the requirements of the examination. It shows the basis on which marks were awarded by examiners. It does not indicate the details of the discussions which took place at an examiners' meeting before marking commenced.

All examiners are instructed that alternative correct answers and unexpected approaches in candidates' scripts must be given marks that fairly reflect the relevant knowledge and skills demonstrated.

Mark schemes should be read in conjunction with the published question papers and the report on the examination.

OCR will not enter into any discussion or correspondence in connection with this mark scheme.

Any enquiries about publications should be addressed to:

OCR Publications
PO Box 5050
Annesley
NOTTINGHAM
NG15 0DL

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

Annotations 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

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, with 3 significant figures often being the norm. Small variations in the degree of accuracy to which an answer is given (eg 2 or 4 significant figures where 3 is expected) should not normally be penalised, while answers which are grossly over- or under-specified should normally result in the loss of a mark. The situation regarding any particular cases where the accuracy of the answer may be a marking issue should be detailed in the mark scheme rationale. If in doubt, contact your Team Leader.

g. Rules for replaced work

If a candidate attempts a question more than once, and indicates which attempt he / she wishes to be marked, then examiners should do as the candidate requests.

If there are two or more attempts at a question which have not been crossed out, examiners should mark what appears to be the last (complete) attempt and ignore the others.

NB Follow these maths-specific instructions rather than those in the assessor handbook.

h. For a *genuine* misreading (of numbers or symbols) which is such that the object and the difficulty of the question remain unaltered, mark according to the scheme but following through from the candidate's data. A penalty is then applied; 1 mark is generally appropriate, though this may differ for some units. This is achieved by withholding one A mark in the question.

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

Question			Answer	Marks	Guidance	
1	(i)		**EITHER** $S_{xy} = \Sigma xy - \frac{1}{n}\Sigma x \Sigma y = 600.41 - \frac{1}{10} \times 113.69 \times 52.81 = 0.01311$ $S_{xx} = \Sigma x^2 - \frac{1}{n}(\Sigma x)^2 = 1292.56 - \frac{1}{10} \times 113.69^2 = 0.01839$ $S_{yy} = \Sigma y^2 - \frac{1}{n}(\Sigma y)^2 = 278.91 - \frac{1}{10} \times 52.81^2 = 0.02039$ $r = \frac{S_{xy}}{\sqrt{S_{xx}S_{yy}}} = \frac{0.01311}{\sqrt{0.01839 \times 0.02039}} = 0.677$ **OR** $\text{cov}(x,y) = \frac{\sum xy}{n} - \bar{x}\bar{y} = 600.41/10 - 11.369 \times 5.281 = 0.001311$ $\text{rmsd}(x) = \sqrt{\frac{S_{xx}}{n}} = \sqrt{(0.01839/10)} = \sqrt{0.001839} = 0.04288$ $\text{rmsd}(y) = \sqrt{\frac{S_{yy}}{n}} = \sqrt{(0.02039/10)} = \sqrt{0.002039} = 0.04516$ $r = \frac{\text{cov}(x,y)}{rmsd(x)\,rmsd(y)} = \frac{0.001311}{0.04288 \times 0.04516} = 0.677$	M1 M1 A1 M1 A1 M1 M1 A1 M1 A1 Methods mixed – max M0M1A1M0A0 **[5]**	For method for S_{xy} For method for at least one of S_{xx} or S_{yy} For at least one of S_{xy}, S_{xx} or S_{yy} correct For fully correct structure of r For answer rounding to 0.68 For method for cov (x,y) For method for at least one msd or rmsd For at least one of cov (x,y), msd or rmsd correct For fully correct structure of r For answer rounding to 0.68	
1	(ii)		H_0: $\rho = 0$ H_1: $\rho \neq 0$ (two-tailed test) where ρ is the population correlation coefficient For $n = 10$, 10% critical value = 0.5494	B1 B1 B1	For H_0, H_1 in symbols. Hypotheses in words must refer to population. Do not allow alternative symbols unless clearly defined as the population correlation coefficient. For defining ρ. Condone omission of "population" if correct notation ρ is used, but if ρ is defined as the **sample** correlation coefficient then award **B0**. CAO Note that critical values for a one-tailed test at the 10% level are not available in tables.	

			Since 0.677 > 0.5494 the result is significant. (Thus we have sufficient evidence to) reject H_0 There is sufficient evidence at the 10% level to suggest that there is correlation between times for the first and last sections.	M1 A1* E1dep* **[6]**	For sensible comparison leading to a conclusion provided that $\|r\| < 1$. The comparison can be in the form of a diagram as long as it is clear and unambiguous. Sensible comparison: e.g. 0.677 > 0.5494 is 'sensible' whereas 0.677 > - 0.5494 is 'not sensible'. Reversed inequality sign e.g. 0.677 < 0.5494 etc. gets max M1 A0. For reject H_0 o.e. FT their r and critical value from 10% 2-tail column. For correct, non-assertive conclusion in context. Allow 'x and y' for context. E0 if H_0 and H_1 not stated, reversed or mention a value other than zero for ρ in H_0. Do not allow 'positive correlation' or 'association'	
1	**(iii)**		The underlying population must have a bivariate Normal distribution. The points in the scatter diagram should have a roughly elliptical shape.	B1 E1 **[2]**	Condone "bivariate Normal distribution", "underlying bivariate Normal distribution", but **do not allow** "the **data** have a bivariate Normal distribution" Condone 'oval' or suitable diagram	
1	**(iv)**		The hypothesis test has shown that there appears to be correlation. However it could be that there is a third causal factor	E1 E1 **[2]**	For relevant comment relating to the test result or positive value of r in supporting (unless FT leads to not supporting) the commentator's suggestion. Or correlation does not imply causation. There may be a third factor. For questioning the use of the word 'must' Allow any two suitable, statistically based comments.	
1	**(v)**	(*A*)	Yes because the critical value at the 1% level is 0.7646 which is larger than the test statistic	B1* E1dep* **[2]**	B1 for 0.7646 seen E1 for comment consistent with their (ii) provided $r\| < 1$	

1	**(v)**	(*B*)	One advantage of a 1% level is that one is less likely to reject the null hypothesis when it is true. One disadvantage of a 1% level is that one is more likely to accept the null hypothesis when it is false.	E1 E1 **[2]**	o.e. Wording must be clear. o.e.	
2	**(i)**		Binomial(1200,1/300)	B1 B1dep **[2]**	For binomial. For parameters Allow B(1200, 1/300) and B(1200, 0.00333)	
2	**(ii)**		Because n is large and p is small	E1, E1 **[2]**	Allow n is large and $np < 10$. Allow "sample is large" for n is large and "mean ≈ variance" for "p is small"	
2	**(iii)**		$\lambda = 1200 \times 1/300 = 4$ (*A*) $P(X = 1) = e^{-4}\frac{4^1}{1!} = 0.0733$ (3 s.f.) or from tables $= 0.0916 - 0.0183 = 0.0733$ (*B*) Using tables: $P(X > 4) = 1 - P(X \leq 4)$ $= 1 - 0.6288 = 0.3712$	B1 M1 A1 M1 A1 **[5]**	For λ FT their p For attempt to find $P(X = 1)$ using Poisson p.d.f. or tables Allow answers which round to 0.073 www. FT their λ (= np). No FT for $\lambda = 1/300$. For finding $1 - P(X \leq 4)$ CAO For answers rounding to 0.371 www	
2	**(iv)**		$\mu = 80$ $\sigma^2 = 80$	B1 B1 **[2]**	If symbols/words used then they must be correct. Allow σ^2 rounding to 79.7 from original binomial. FT their λ (= np)	
2	**(v)**	(*A*)	$P(Y \geq 90) = P\left(Z \geq \frac{89.5 - 80}{\sqrt{80}}\right)$ $= P(Z > 1.062) = 1 - \Phi(1.062)$ $= 1 - 0.8559 = 0.1441$	B1 M1 A1cao **[3]**	For correct continuity correction. For probability using correct tail and structure (condone omission of c.c.) $\sigma^2 = 79.73$ leads to $P(Z > 1.064)$ $\sigma^2 = 79.73$ leads to $1 - 0.8563 = 0.1437$. Allow 0.144 www. NOTE 0.1441 from B(24000, 1/300) gets 0/3	

2	**(v)**	(*B*)	P($Y \leq k$) > 0.05 From tables Φ^{-1} (0.05) = -1.645 $\frac{(k+0.5)-80}{\sqrt{80}}$ = -1.645 $k + 0.5 = 80 - (1.645 \times \sqrt{80}) = 65.29$ $k > 64.79$ So least value of $k = 65$	B1 M1 A1 A1 **[4]**	For ±1.645 seen For correct equation in k seen or equivalent – e.g. allow +1.645 used if numerator reversed. FT their μ, σ^2 and z-value. Condone omission of, or incorrect, continuity correction. A1 for 65.29 or 64.79 or 65.79 (σ^2 = 79.73 leads to 65.31 or 64.81 or 65.81) Allow 3s.f. For rounding **64.79 or 64.81 up to give k = 65**. See additional notes for alternative method	
3	**(i)**		P($X \geq 750$) = $P\left(Z \geq \frac{750-751.4}{2.5}\right)$ = P(Z > -0.56) = Φ(0.56) = 0.7123	M1 M1 A1 **[3]**	For standardizing For correct structure (M0 if continuity correction used) CAO Allow 0.712 www	
3	**(ii)**		P(all 6 at least 750ml) = 0.7123^6 = 0.1306	M1 A1 **[2]**	For (their answer to part (i))^6 FT 3s.f.	
3	**(iii)**		P(Y=0) = $\binom{25}{0} \times 0.8694^{25}$ (= 0.0302) P(Y=1) = $\binom{25}{1} \times 0.8694^{24} \times 0.1306$ (= 0.1135) P(Y=0) + P(Y=1) = 0.144 P($Y \geq 2$) = 1 – 0.144 =0.856	M1 M1 M1dep A1 **[4]**	For using Binomial(25, p) with their p from part (ii) For correct structure of either P(Y = 0) or P(Y = 1) with their p from part (ii) M0 if p and q reversed For 1 – sum of both probabilities CAO	

3	**(iv)**		$P(Z < \frac{750-\mu}{2.5}) = 0.02$ $\Phi^{-1}(0.02) = -2.054$ $\frac{750-\mu}{2.5} = -2.054$ $\mu = 750 + 2.054 \times 2.5$ $= 755.1$	 B1 M1 M1 A1 **[4]**	 For ±2.054 seen. Allow ±2.055 For correct equation as seen or equivalent. FT σ = √2.5. M0 if c.c. used. For correctly rearranging their equation (if 750 used in numerator) for μ, FT their *z* cao Condone 755 or 5 s.f. rounding to 755.1 www	
3	**(v)**		$P(Z < \frac{750-751.4}{\sigma}) = 0.02$ $\frac{750-751.4}{\sigma} = -2.054$ $\sigma = \frac{-1.4}{-2.054}$ $= 0.682$	 M1 M1 A1 **[3]**	 For correct equation as seen or equivalent For correctly rearranging their equation (if 750 used in numerator) for σ unless this leads to σ < 0 cao Allow answers rounding to 0.68 www	
3	**(vi)**		Probably easier to change the mean (as reducing the standard. deviation would require a much more accurate filling process). However increasing the mean would result in fewer bottles being filled overall and so less profit for the owners, so **reducing** the **standard deviation** would be **preferable** to the vineyard owners.	E1 E1 **[2]**	 For "preferable to reduce the standard deviation" with valid reason.	
4	**(a)**	**(i)**	Expected frequency = 67/150 × 57 = 25.46 Contribution = $(34 - 25.46)^2 / 25.46$ = 2.8646	B1 M1 A1 **[3]**	For 25.46 For valid attempt at $(O-E)^2/E$ Correct values used to give answer which rounds to 2.8646 ***NB Answer given***	

4	**(a)**	**(ii)**	H_0: no association between type of cake and classification of person. H_1: some association between type of cake and classification of person. Test statistic $X^2 = 12.86$ Refer to X_3^2 Critical value at 1% level = 11.34 Result is significant There is evidence to suggest association between type of cake and classification of person. NB if H_0 H_1 reversed, omitted or 'correlation' mentioned, do not award first B1 or final E1	B1 B1 B1 B1 E1 **[5]**	For both hypotheses in context For 3 degrees of freedom CAO For cv. No FT from here if wrong/omitted For significant For correct, non-assertive conclusion, in context.	
4	**(b)**		$\bar{x} = 4.995$ H_0: $\mu = 5$ H_1 : $\mu < 5$ Where μ denotes the **mean** content of the bags of flour (in the population) Test statistic $= \dfrac{4.995 - 5.0}{0.0072/\sqrt{8}} = \dfrac{-0.005}{0.002546} = -1.964$ Lower 5% level 1 tailed critical value of $z = -1.645$ $-1.964 < -1.645$ so significant. There is sufficient evidence to reject H_0 There is sufficient evidence to suggest that the average contents of bags is less than 5kg.	B1 B1 B1 B1 M1* A1 B1* M1 dep* A1 **[9]**	For 4.995 seen For use of 5 in hypotheses. For both correct. Hypotheses in words must refer to population. Do not allow alternative symbols unless clearly defined as the population mean. For definition of μ in context. Condone omission of "population" if correct notation μ is used, but if μ is defined as the **sample** mean then award **B0**. must include $\sqrt{8}$ FT their $\bar{x}$. Allow +1.964 only if later compared with +1.645 For –1.645 No FT from here if wrong. Must be –1.645 unless it is clear that absolute values are being used. For sensible comparison with correct c.v. leading to a conclusion. For non-assertive conclusion in words and in context. No FT here. See additional notes.	

ADDITIONAL NOTES REGARDING QUESTION 2 (v) *B*

M1 for using a trial and improvement method with N(80,80) or N(80, 79.73) to find P($Y \leq k$) for any k. The distribution being used needs to be made clear.

A1 for P($Y \leq 66$) = 0.0587… (0.0584… from $\sigma^2 = 79.73$) or P($Y \leq 65$) = 0.0467… (0.0464… from $\sigma^2 = 79.73$)
A1 for both
Final A1 not available if 66 and 65 used

Or
A1 for P($Y \leq 65.5$) = 0.0524… (0.0521… from $\sigma^2 = 79.73$) or P($Y \leq 64.5$) = 0.0415… (0.0412… from $\sigma^2 = 79.73$)
A1 for both
A1 for least value of $k = 65$, dependent on previous two A marks earned.

ADDITIONAL NOTES REGARDING QUESTION 4 (b)

Critical Value Method
$5 - 1.645 \times 0.0072 \div \sqrt{8}$ gets M1*B1*
= 4.9958… gets A1
4.995 < 4.99581.. gets M1dep* for sensible comparison
A1 still available for correct conclusion in words & context

"Confidence Interval" Method
$4.995 + 1.645 \times 0.0072 \div \sqrt{8}$ gets M1* B1*
= 4.9991.. gets A1
NOTE that the final M1dep* A1 available only if 1.645 used.
5 > 4.9991… gets M1
A1 still available for correct conclusion in words & context

Probability Method
Finding P(sample mean < 4.995) = 0.0248 gets M1* A1 B1
0.0248 < **0.05*** gets M1dep* for a sensible comparison if a conclusion is made.
A1 available for a correct conclusion in words & context.
Condone P(sample mean > 4.995) = 0.9752 for M1 but only allow A1 B1 if later compared with 0.95, at which point the final M1and A1 are still available

ADDITIONAL NOTE REGARDING OVER-SPECIFICATION OF ANSWERS

Over-specification by providing final answers correct to 5 or more significant figures will be penalised. When this applies, candidates may lose no more than 2 marks per question and no more than 4 marks in total. The only exception to this rule is in Question 3 part (iv) – see guidance note.

OCR (Oxford Cambridge and RSA Examinations)
1 Hills Road
Cambridge
CB1 2EU

OCR Customer Contact Centre

Education and Learning
Telephone: 01223 553998
Facsimile: 01223 552627
Email: general.qualifications@ocr.org.uk

www.ocr.org.uk

For staff training purposes and as part of our quality assurance programme your call may be recorded or monitored

Oxford Cambridge and RSA Examinations
is a Company Limited by Guarantee
Registered in England
Registered Office; 1 Hills Road, Cambridge, CB1 2EU
Registered Company Number: 3484466
OCR is an exempt Charity

OCR (Oxford Cambridge and RSA Examinations)
Head office
Telephone: 01223 552552
Facsimile: 01223 552553

4767 Statistics 2

General Comments

Most candidates demonstrated good knowledge in all questions on this paper. The parts which proved to be most accessible were those involving hypothesis tests and also basic calculations such as evaluating the product moment correlation coefficient. Not so convincingly well answered were parts requiring knowledge of statistical distributions. Knowledge of modelling assumptions and application of approximating distributions was not as secure as that shown in hypothesis tests. The manipulation of expressions associated with Normal distribution calculations could have been handled better by many candidates; the provision of a diagram to indicate intention would have helped many candidates identify the correct tail. It was pleasing to see sensible answers provided to the parts requiring interpretation, with most candidates providing statistical justification as well as referring to the context of the question. The issue of over-specification in final answers was noticed, but the vast majority of candidates provided answers rounded to a suitable degree of accuracy for the context provided.

Comments on Individual Questions

1) (i) Most candidates found the PMCC value correctly and to a suitable level of accuracy. A few candidates gave their answer to 5dp, and a few rounded their answers for S_{xy} etc. and hence gave an inaccurate final answer. Very few candidates quoted the formula for r incorrectly.

(ii) Most candidates performed well on this question. To gain full marks, candidates needed to show awareness that their hypotheses concerned the population value of the pmcc, either by using correct notation or by including the word 'population' in their worded hypotheses. Most candidates reached the correct conclusion, based on their value of r. Fully correct, non-assertive conclusions were provided by many candidates.

(iii) Many candidates provided the correct distributional assumption. Others replaced 'population' with 'sample'. Most candidates provided a correct explanation of how the distributional assumption could be checked using a scatter diagram. Others thought that visual evidence of 'linear correlation' was sufficient.

(iv) This question was well answered. Responses attracting the least credit referred mainly to technical discussions about bobsleighing. Comments relating to the result of the hypothesis test or the value of *r* were usually well rewarded.

(v)(A) This question was well answered. Some candidates did not provide the critical value for the test at the 1% level

(v)(B) Some excellent responses were seen from some candidates. Many interpreted the smaller significance level as meaning the test was 'more accurate', rather than considering the implications that altering the level of the test can change the conclusion – as seen if parts (ii) and (vA) had been completed correctly.

2) (i) Most candidates gave the correct answer. Some gave the value of *p* as 0.003. Others provided the Poisson distribution as the 'exact' distribution at this stage.

(ii) This was well answered by most candidates. Some simply described the modelling assumptions for a Poisson distribution rather than answering the given question.

(iii)(A) This was correctly answered by most candidates.

(iii)(B) Correctly answered by most candidates. Many thought that $P(X > 4) = 1 - P(X \leq 3)$

(iv) Most candidates gained both available marks. Some candidates provided a correct value for the mean but an incorrect value for the variance.

(v)(A) Many candidates realised the need to apply a continuity correction and successfully reached a correct answer. Many did not realise the need for a continuity correction, though once a standardised value had been found, most knew how to use the Normal distribution table to produce an answer.

(v)(B) Most candidates identified the correct *z*-value for a 5%/95% tail and were able to de-standardise this correctly. Many candidates opted to use +1.645 (incorrectly) rather than -1.645 in their calculations. Few candidates showed appreciation that *k* needed to be an integer. Use of continuity corrections was rare and poorly handled in the few cases where seen.

3) (i) This question was answered very well by many candidates, with almost all realising that this continuous measure of volume did not require any continuity correction. Of the others who did apply erroneous continuity corrections, several opted for $\pm$ 1 unit instead of the usual $\pm$ 0.5.

(ii) Most candidates attained both of the available marks here. Some candidates mixed up their 'p and q' from part (i). Many who lost marks in part (i) earned both marks in part (ii)

(iii) Many candidates found this to be a challenging question and did not recognise the 'binomial' situation. Those candidates using the correct distribution usually applied $1 – [P(X=0) + P(X=1)]$ correctly.

(iv) Most candidates obtained the correct *z*-value of -2.054. Many used +2.054 appropriately and earned full marks. Others provided incorrect equations but managed to correctly rearrange them, though not recognise the absurdity of the subsequent answer.

(v) Many provided fully correct solutions to this question. Others did not correct their working despite it leading to a negative value for σ.

(vi) Most candidates understood the reasoning behind this question and correctly identified which method would be 'easier to implement' and which would be 'preferable'. To earn full marks, candidates needed to show that they understood that reduced variability was preferable.

4) (a)(i) Nearly all candidates scored 3/3 here.

(a)(ii) This question was well answered. Commonly, marks were lost through over-assertive final conclusions and poorly worded hypotheses, though the numerical aspect of the hypothesis test was well done.

(b) Many excellent answers were seen in this question. Often, lost marks were due to inappropriate hypotheses; those not using μ could still achieve full credit if they defined their replacement symbol as the population mean. Others provided over-assertive conclusions and some seemed unsure how to proceed once they had obtained the correct test statistic – inappropriate comparisons were seen quite frequently.

GCE Mathematics (MEI)

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

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