OCR
RECOGNISING ACHIEVEMENT

Thursday 31 May 2012 – Morning

AS GCE MEI STATISTICS

G242 Statistics 2 (Z2)

QUESTION PAPER

Candidates answer on the Printed Answer Book.

OCR supplied materials:
- Printed Answer Book G242
- 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**.
- This 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 manufacturer produces capacitors for the computer industry. The manufacturer's quality control department has found that 0.8% of capacitors produced are faulty. A sample of 50 capacitors is tested and the number of faulty capacitors is noted.

(i) State the conditions required for the binomial distribution B(50, 0.008) to be a suitable model for the number of faulty capacitors in the sample. **[1]**

(ii) Explain why the probability distribution in part (i) may be approximated by a Poisson distribution. State the parameter of this distribution. **[2]**

(iii) Use the approximating distribution in part (ii) to find the probability that a sample of 50 capacitors contains at least 3 which are faulty. **[2]**

The manufacturer also produces diodes. As part of the quality control process, diodes are tested in batches of 50. The number of faulty diodes in each batch is recorded. The following table summarises the results from 200 batches.

Number of faulty diodes	0	1	2	3	4
Observed frequency	40	61	64	29	6

The manufacturer wishes to test the goodness of fit of a Poisson model. A mean of 1.5, calculated from the data, is used as an estimate for the mean of the underlying population. The following table shows the corresponding expected frequencies.

Number of faulty diodes	0	1	2	3	$\geqslant 4$
Expected frequency	44.63	66.94	50.20	25.10	13.13

(iv) Show how the expected frequency of 66.94 for 1 faulty diode is obtained. **[2]**

(v) Given that the value of the chi-squared test statistic is 9.279, carry out the test of the goodness of fit of the Poisson model at the 5% level of significance. **[6]**

2 The mass of a tomato harvested from a particular variety of tomato plant is M grams, where M is Normally distributed with mean μ and variance 20.

(i) Given that $P(M < 72) = P(M > 98)$, explain why $\mu = 85$. **[1]**

(ii) Calculate the probability that a randomly selected tomato has a mass less than 80 g. **[3]**

A randomly selected sample of 4 tomatoes is chosen and the sample mean is calculated.

(iii) Explain the difference between a population mean and a sample mean. **[2]**

(iv) Calculate the standard error of the sample mean for the 4 tomatoes. **[2]**

(v) Calculate the probability that the sample mean exceeds 90 g. **[3]**

(vi) Explain whether or not it was necessary to apply the Central Limit Theorem to calculate the probability in part (v). **[2]**

3 A motorcycle racer wishes to improve his lap times for a particular racing circuit. He makes an adjustment to his motorcycle's gears which he hopes will reduce his average lap time. Before the adjustment was made, his median lap time for this circuit was 258 seconds. Following the adjustment, he recorded his lap times for a sample of 10 laps. The results, in seconds, are as follows.

253 255 251 246 250 271 264 256 248 267

(i) Stating any necessary assumptions, use a Wilcoxon test to examine, at the 5% significance level, whether the motorcyclist's adjustment has been successful in reducing the average lap time. **[13]**

(ii) State the advantage of using a Wilcoxon test rather than using a hypothesis test based on the t distribution. **[1]**

(iii) Discuss whether a test based on the t distribution would be suitable in this case. **[2]**

4 A car manufacturer is developing a battery-powered car. The manufacturer requires a battery which, when fully charged, is capable of powering the car for an average distance of 240 km. A particular type of battery is chosen for testing. In each test, a car with a fully charged battery is driven around a test track until the battery fails; the distance travelled is measured. The distances, in kilometres, for a random sample of 12 tests are as follows.

239 241 238 239 237 242 238 242 238 240 239 235

(i) Use these data to show that the sample mean is 239 and calculate the sample standard deviation. **[3]**

(ii) Explain, with reference to the sample, whether or not you think that these data could have an underlying Normal distribution. **[2]**

(iii) Assuming that these distances do have an underlying Normal distribution, obtain a 95% confidence interval, based on the t distribution, for the mean distance travelled. **[5]**

(iv) Discuss the confidence interval found in part **(iii)** in relation to the manufacturer's requirement. **[3]**

(v) Explain why, in this case, a confidence interval based on the t distribution is more suitable than a confidence interval based on the Normal distribution. **[2]**

[Question 5 is printed overleaf.]

5 A beekeepers' organisation is concerned about the continuing reduction in the number of bees. It funds a variety of research projects to investigate the reasons for this reduction. One such project aims to discover if there is an association between the change in size of a bee colony over the course of a year and the intensity of pesticide use over the area in which the colony is located. 120 colonies, regarded as a random sample, are selected and the results are summarised in the table below.

		Pesticide use		
		High	Medium	Low
Change in bee population	Minimal change	3	16	20
	Decrease of 10% to 20%	11	16	15
	Over 20% decrease	15	13	11

A test to examine whether these data provide any evidence of an association between these classification factors is to be carried out. The following tables show some of the expected frequencies and contributions to the test statistic.

Expected frequencies		Pesticide use		
		High	Medium	Low
Change in bee population	Minimal change	9.425	14.625	14.95
	Decrease of 10% to 20%	10.15		
	Over 20% decrease	9.425		

Contributions to the test statistic		Pesticide use		
		High	Medium	Low
Change in bee population	Minimal change	4.380	0.129	1.706
	Decrease of 10% to 20%	0.071		
	Over 20% decrease	3.298		

(i) Calculate the remaining expected frequencies and contributions. Carry out the test at the 5% level of significance. **[11]**

(ii) The cell corresponding to high pesticide use and minimal change in population provides the largest single contribution to the test statistic. Explain how this can be interpreted in relation to your hypotheses. **[2]**

(iii) Which of the three levels of pesticide use shows the least association with change in population? Explain your answer. **[2]**

OCR
RECOGNISING ACHIEVEMENT

Thursday 31 May 2012 – Morning

AS GCE MEI STATISTICS

G242 Statistics 2 (Z2)

PRINTED ANSWER BOOK

Candidates answer on this Printed Answer Book.

OCR supplied materials:
- Question Paper G242 (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.

Turn over

1 (i)	
1 (ii)	
1 (iii)	
1 (iv)	

1 (v)

2 (i)	
2 (ii)	
2 (iii)	

2 (iv)	
2 (v)	
2 (vi)	

3 (i)	

3 (i)	(continued)
3 (ii)	
3 (iii)	

4 (i)

4 (ii)

4 (iii)

4 (iii)	(continued)
4 (iv)	
4 (v)	

5 (i)

Expected frequencies		Pesticide use		
		High	Medium	Low
Change in bee population	Minimal change	9.425	14.625	14.95
	Decrease of 10% to 20%	10.15		
	Over 20% decrease	9.425		

Contributions to the test statistic		Pesticide use		
		High	Medium	Low
Change in bee population	Minimal change	4.380	0.129	1.706
	Decrease of 10% to 20%	0.071		
	Over 20% decrease	3.298		

5 (i)	(continued)
5 (ii)	
5 (iii)	

PLEASE DO NOT WRITE ON THIS PAGE

RECOGNISING ACHIEVEMENT

GCE

Statistics (MEI)

Advanced Subsidiary GCE

Unit **G242:** Statistics 2 (Z2)

Mark Scheme for June 2012

Any enquiries about publications should be addressed to:

OCR Publications
PO Box 5050
Annesley
NOTTINGHAM
NG15 0DL

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

Annotations 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 awarde$ 0(1
B0$ B1	Independent mark awarded 0, 1
SC	Cpecial 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 Mr 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 (e.g. 2 or 4 significant figures where 3 is expected) should not normally be penalised, while answers which are grossly over- or under-specified should normally result in the loss of a mark. The situation regarding any particular cases where the accuracy of the answer may be a marking issue should be detailed in the mark scheme rationale. If in doubt, contact your Team Leader.

g. Rules for replaced work

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

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

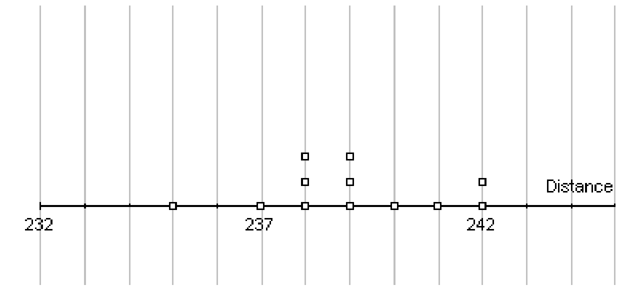
NB Follow these maths-specific instructions rather than those in the assessor handbook.

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

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

Question			Answer	Marks	Guidance
1	(i)		Independence and constant probability	B1 **[1]**	
1	(ii)		Poisson approximation is appropriate as p is small and n is large The parameter of the Poisson distribution is 0.4	B1 B1 **[2]**	Allow 'np < 10' or 'mean ≈ variance' for 'p small' This mark may be recovered in part (iii)
1	(iii)		Using $X \sim$ Poisson(0.4) $P(X \geq 3) = 1 - P(X \leq 2)$ $= 1 - 0.9921$ $= 0.0079$	 M1 A1 **[2]**	 Correct structure with any λ. FT their λ from part (ii). 0.0076 from B(50, 0.008) gets 0/2
1	(iv)		Using $X \sim$ Poisson(1.5) $P(X = 1) = \dfrac{e^{-1.5} \times 1.5}{1!} = 0.33469\ldots$ $\ldots \times 200 = 66.9390\ldots = 66.94$ (2d.p.)	M1 M1 **[2]**	M1 for calculation as seen Or for $0.5578 - 0.2231$ from tables
1	(v)		H_0: The Poisson model is suitable H_1: The Poisson model is not suitable Number of degrees of freedom = 3 (5 – 1 – 1) At 5% significance level, critical value is 7.815 9.279 > 7.815 Result is significant. Evidence suggests that this Poisson model is not a good fit to these data.	 B1 B1 B1 B1 B1 E1 **[6]**	 Condone 'data is a good fit to the Poisson model' 3 degrees of freedom **seen** cao Sensible comparison **seen**. FT only c.v. = 9.488 [note that c.v. = 9.488 leads to 'not significant'] Conclusion in words must be non-assertive and consistent with their result and hypotheses.
2	(i)		By symmetry, μ lies halfway between 72 and 98	B1 **[1]**	Allow (72 + 98) ÷ 2 = 85. Allow μ is the mid-point of 72 and 98. Allow 98 = 85 + 13 & 72 = 85 – 13. Allow demonstration that the values standardise to give ± 2.907
2	(ii)		$P(M < 80) = P(Z < \dfrac{80-85}{\sqrt{20}})$ $= 1 - \Phi(1.118\ldots)$ $= 1 - 0.8682 = 0.132$ (3sf)	M1 M1 A1 **[3]**	M1 for standardising. Allow numerator reversed. M0 if continuity correction used. Correct structure FT their σ cao

Question			Answer	Marks	Guidance
2	(iii)		A population mean is the mean of all possible values and is a constant A sample mean is the mean of a sample and is variable.	E2,1,0 **[2]**	For E2 need recognition that population mean is constant and sample mean is variable. E1 for explaining that the sample mean represents the mean of only part of the population.
2	(iv)		Standard error = $\sqrt{20} \div \sqrt{4}$ $= \sqrt{5} = 2.24$ (3sf) cao	M1 A1 **[2]**	M1 for $\sqrt{20} \div \sqrt{4}$ even if seen as part of probability calculation. Allow M1 for $20 \div \sqrt{4}$ if use of $\sigma = 20$ has already been penalised in part (ii)
2	(v)		$\dfrac{90-85}{\sqrt{5}} = 2.236\ldots$ $1 - \Phi(2.236\ldots) = 0.0127$ (3sf)	M1 M1 A1 **[3]**	M1 for standardising. FT their standard error M1 for correct structure cao
2	(vi)		As the sample was taken from a Normal distribution, it is not necessary to use the CLT in part (v)	B1 E1 dep **[2]**	
3	(i)		Assume that the underlying population is distributed symmetrically and the sample is random. H_0: population median = 258 H_1: population median < 258 Actual differences –5 –3 –7 –12 –8 13 6 –2 –10 9 Associated ranks 3 2 5 9 6 10 4 1 8 7 $T^- = 3 + 2 + 5 + 9 + 6 + 1 + 8 = 34$ $T^+ = 10 + 4 + 7 = 21$ $\therefore T = 21$ From $n = 10$ tables – at the 5% level of significance in a one-tailed Wilcoxon single sample test, the critical value of T is 10 $21 > 10 \therefore$ the result is not significant The evidence does not suggest that there has been a reduction in the average lap time.	B1 B1 B1 B1 B1 M1*A1 B1 B1 M1* A1 M1dep* A1 **[13]**	 Condone 'the values are random' B1 for using 258 in hypotheses B1 for both correct and including 'population' Condone opposite signs M1 for ranking absolute values of their differences. A1 FT provided differences are used and the sum of the ranks is 55. B1 for $T^+ = 21$or $T^- = 34$. FT only if the sum of the ranks is 55 B1 FT for test statistic cao FT their test statistic only if both previous M1 marks earned A1 for non-assertive conclusion in context

Question			Answer	Marks	Guidance
3	(ii)		The Wilcoxon test does not require the underlying population to be Normally distributed whereas the t test does.	E1 **[1]**	Allow 'lap times do not need to be Normally distributed' Condone 'data do not need to be Normally distributed'
3	(iii)		Sensible comment with justification. e.g. Lap times are unlikely to be distributed Normally so a t test is probably not appropriate. e.g. consideration of dot plot	E2,1,0 **[2]**	E2 for 'lap times are unlikely to be distributed Normally so a t test is (probably) not appropriate'. i.e. not overly assertive. E1 for 'lap times are not Normally distributed so a t test is not appropriate'. Allow sensible alternatives
4	(i)		$\Sigma x \div n = 2868 \div 12$ (= 239 A.G.) $\sqrt{\dfrac{685498 - \dfrac{2868^2}{12}}{11}} = 2.045$ (4s.f.)	B1 M1 A1 **[3]**	Full method shown or 2868 ÷ 12 seen Allow M1A0 for σ^2
4	(ii)		Sensible comment, with justification.  eg These (continuous) data appear to be symmetrically distributed and so it is possible that they have an underlying Normal distribution.	E1 E1 **[2]**	E1 for suitable comment regarding similarity between mean and median or for suitable comment relating to central tendency or comment about continuous variable. E1 for suitable comment regarding symmetry
4	(iii)		$239 \pm 2.201 \times \dfrac{2.0449..}{\sqrt{12}}$ (237.70, 240.30)	B1 B1 M1 A1A1 **[5]**	centred on 239 for 2.201 structure Allow A1A0 for (237.700…, 240.299…)

<table>
<tr><th colspan="3">Question</th><th>Answer</th><th>Marks</th><th>Guidance</th></tr>
<tr><td>4</td><td>(iv)</td><td></td><td>The interval in (iii) contains 240.
This supports the view that this type of battery is suitable.
The mean could be at least 240.</td><td>E1*
E1dep*
E1
[3]</td><td>Can be implied. Allow ‘manufacturer’s requirement’ for ‘240’.
FT their CI
Allow other sensible comments.</td></tr>
<tr><td>4</td><td>(v)</td><td></td><td>Variance is unknown.
Sample is small.</td><td>E1
E1
[2]</td><td>
Allow ‘n is small’ or ‘n < 30’</td></tr>
<tr><td>5</td><td>(i)</td><td></td><td>H_0: No association between pesticide use and change in bee population.
H_1: There is an association between pesticide use and change in bee population.
Expected frequencies
<table>
<tr><th colspan="2" rowspan="2">Expected frequencies</th><th colspan="3">Pesticide use</th></tr>
<tr><th>High</th><th>Medium</th><th>Low</th></tr>
<tr><td rowspan="3">Change in bee population</td><td>Minimal change</td><td>9.425</td><td>14.625</td><td>14.95</td></tr>
<tr><td>Decrease of 10% to 20%</td><td>10.15</td><td>15.75</td><td>16.1</td></tr>
<tr><td>Over 20% decrease</td><td>9.425</td><td>14.625</td><td>14.95</td></tr>
</table>
Contributions to X^2
<table>
<tr><th colspan="2" rowspan="2">Contributions to the test statistic</th><th colspan="3">Pesticide use</th></tr>
<tr><th>High</th><th>Medium</th><th>Low</th></tr>
<tr><td rowspan="3">Change in bee population</td><td>Minimal change</td><td>4.380</td><td>0.129</td><td>1.706</td></tr>
<tr><td>Decrease of 10% to 20%</td><td>0.071</td><td>0.004</td><td>0.075</td></tr>
<tr><td>Over 20% decrease</td><td>3.298</td><td>0.181</td><td>1.044</td></tr>
</table>
$X^2 = 10.887$
4 degrees of freedom
Critical value for 5% significance level is 9.488
As 10.887 > 9.488 the result is significant
There is evidence to suggest an association between pesticide use and change in bee population</td><td>B1

B2,1,0

M1
A1

A1
B1
B1
M1, A1
B1

[11]</td><td>Hypotheses in context

B1 if at least two expected frequencies correct
B2 if all correct

Attempt at $(\text{obs} - \text{exp})^2 \div \text{exp}$
All correct

cao AWRT 10.89

M1 for sensible comparison, A0 if wrong cv used or if first M mark not awarded
B1 for non-assertive conclusion in context. If hypotheses reversed, do not award first B1 or final A1 B1</td></tr>
</table>

Question			Answer	Marks	Guidance
5	(ii)		This cell provides the strongest evidence of association between the classification factors. There were significantly fewer observed values than would be expected if there were no association.	E1 E1 **[2]**	Do not allow 'positive association'. Allow 'this supports the alternative hypothesis' Allow other sensible comments related to the hypothesis test
5	(iii)		Medium pesticide use shows the least association. The column for medium use has the lowest total contribution to the test statistic.	E1 E1 **[2]**	

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

OCR Customer Contact Centre

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

www.ocr.org.uk

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

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

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

PART OF THE CAMBRIDGE ASSESSMENT GROUP

LLOYD'S REGISTER QUALITY ASSURANCE ISO9001

UKAS MANAGEMENT SYSTEMS 001

GCE

Statistics (MEI)

Advanced Subsidiary GCE **AS H132**

OCR Report to Centres

June 2012

G242 Statistics 2

General Comments

Many candidates showed good ability to apply their statistical knowledge and performed particularly well on the hypothesis tests, though the goodness of fit test was not completed as successfully as the Wilcoxon test or the chi-squared test for association. Many candidates lacked the knowledge of different modelling assumptions and properties of probability distributions needed to be able to tackle the questions requiring explanation. Some candidates did not read questions carefully enough to take advantage of the instructions given.

Comments on Individual Questions

1) (i) Few candidates supplied adequate justification for the use of a binomial model.

1) (ii) Many candidates identified the correct parameter but did not provide a correct reason for the appropriateness of the Poisson model.

1) (iii) Many followed the instructions to use the Poisson approximating distribution. Others opted to attempt to use a binomial distribution. Of those using the correct distribution, many scored 2/2 but some evaluated $P(X = 3)$ or $1 - P(X \leq 3)$ for $P(X \geq 3)$.

1) (iv) Many fully correct answers were seen. Most candidates knew what was required.

1) (v) Many candidates tackled this question well. For those familiar with the chi-squared goodness of fit test, the main source of lost marks was through providing incorrect or overly assertive conclusions.

2) (i) This was well answered by most candidates.

2) (ii) This question was, on the whole, well answered with many gaining full marks. Some candidates did not standardise correctly, using variance in place of standard deviation.

2) (iii) Few correct answers were seen. Many candidates realised that the sample, being a subset of the population, would likely produce a different mean to the population mean.

2) (iv) Few candidates were familiar with the term ‘standard error’.

2) (v) Many candidates attempted a Normal probability calculation here. Few fully correct responses were seen. Many of those attempting the calculation standardised incorrectly. Some did not use the Normal table with sufficient accuracy.

2) (vi) Very few candidates could give the correct reason as to why it was not necessary to apply the Central Limit Theorem.

3) (i) This question proved to be a rich source of marks for many candidates. Many candidates managed to identify at least one of the necessary assumptions. Most provided appropriate hypotheses involving a median, but few referred to 'population median'. The majority of those who managed to rank their differences between sample values and population median went on to secure most of the remaining marks, though mistakes were seen in some scripts and some final conclusions were either incorrect, overly assertive or did not refer back to the context of the question.

3) (ii) Few candidates provided a correct response.

3) (iii) Few correct responses were seen. Many provided general comments relating to sample size rather than focussing on necessary assumptions.

4) (i) Most candidates answered this well. Some lost a mark through providing the standard deviation correct to 5 or more significant figures. Some missed the instruction to 'show that the sample mean is 239' and simply quoted the figure given.

4) (ii) Some candidates were able to explain, with reference to the data, whether or not they thought the sample could have an underlying Normal distribution.

4) (iii) Many candidates were familiar with the structure of a confidence interval. Of these, many based their interval on the Normal distribution rather than the *t* distribution despite this being indicated in the question. Some of those using the *t* distribution over-specified their answers, giving 6 or more significant figures. Some candidates interpreted the instruction 'obtain a 95% confidence interval' to mean 'carry out a hypothesis test'.

4) (iv) Few candidates scored 3/3 in this question. Many of those who obtained a confidence interval in part (iii) used it appropriately to gain credit.

4) (v) Most recognised that the small sample was reason to support the use of a *t* distribution. Many also recognised that the population variance being unknown provided further support.

5) (i) This question was well answered and a rich source of marks for most candidates. Many scored at least 10/11. Other than making mistakes in calculations, reasons for losing marks commonly included use of incorrect hypotheses, over-assertive conclusions and stating the wrong value for the number of degrees of freedom.

5) (ii) Some candidates correctly explained the relevance of the indicated contribution to the outcome of the hypothesis. Many provided comments based only on the context of the question. In such questions, candidates should aim to provide statistically based comments.

5) (iii) Most candidates connected medium pesticide use as showing least association with change in population. Many provided a correct reason to support their decision.