OCR
RECOGNISING ACHIEVEMENT

OXFORD CAMBRIDGE AND RSA EXAMINATIONS

Advanced Subsidiary General Certificate of Education

MEI STATISTICS G242

Statistics 2 (Z2)

Monday **12 JUNE 2006** Afternoon 1 hour 30 minutes

Additional materials:
8 page answer booklet
Graph paper
MEI Examination Formulae and Tables (MF2)

TIME 1 hour 30 minutes

INSTRUCTIONS TO CANDIDATES

- Write your name, centre number and candidate number in the spaces provided on the answer booklet.
- Answer **all** the questions.
- You are permitted to use a graphical calculator in this paper.
- Final answers should be given to a degree of accuracy appropriate to the context.

INFORMATION FOR CANDIDATES

- The number of marks is given in brackets [ ] at the end of each question or part question.
- 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 question paper consists of 3 printed pages and 1 blank page.

© OCR 2006 [L/103/0483]
Registered Charity Number: 1066969
[Turn over

1 **(a)** The total playing time of a football match may be modelled using a Normal distribution with mean 93 minutes and standard deviation 0.9 minutes.

(i) Find the probability that the total playing time of a particular football match is less than 94 minutes. [3]

(ii) Find the probability that, in four football matches, there is at least one in which the total playing time is **more** than 94 minutes. [3]

(b) Footballers may be classified either as goalkeepers or as outfield players. The heights, x cm, of a random sample of 60 outfield players in a football competition are measured. The results are summarised as follows.

$$\Sigma x = 10\,920 \qquad \Sigma x^2 = 1\,989\,670$$

(i) Use these data to show that the sample variance is $37.80\,\text{cm}^2$, correct to 2 decimal places. [2]

(ii) Find a two-sided 95% confidence interval for the mean height of outfield players in this competition. [4]

(iii) The mean height for goalkeepers in the competition is 188 cm. Comment on this fact in relation to the confidence interval found in part **(ii)**. [4]

2 The manager of a busy regional railway station is investigating the arrival of customers at the ticket office. She counts the numbers of customers arriving during 100 two-minute intervals. Her observations are made randomly, at appropriate times, during a particular week. The results are shown in the following frequency table.

Number of customers arriving, x	0	1	2	3	4	5	6	7	$\geqslant 8$
Observed frequency, f	6	12	14	18	21	16	9	4	0

Summary statistics: $\Sigma f = 100$, $\Sigma fx = 340$, $\Sigma fx^2 = 1486$.

The manager believes that the number of customers, X, arriving in a two-minute interval may be modelled using a Poisson distribution.

(i) Verify that the mean of the manager's sample, $\bar{x}$, is 3.4 and find the sample variance.

Are the results of your calculations consistent with the manager's belief that the Poisson distribution is an appropriate model? Justify your answer. [4]

(ii) Taking 3.4 as the mean of the underlying population, use the appropriate cumulative probability tables to find the probabilities corresponding to the values of x in the table.

Hence obtain the expected frequencies corresponding to the observed frequencies. [5]

(iii) The expected and observed frequencies are used to carry out a test of the goodness of fit of the Poisson model. The cells for $x = 0$ and $x = 1$ are merged; the cells for $x = 7$ and $x \geqslant 8$ are also merged. The calculated statistic for the χ^2 test is 5.127.

(*A*) Explain why the cells were merged.

(*B*) What is the conclusion of the test when a 5% significance level is used? Justify your answer using an appropriate critical value. [5]

3 It is possible to monitor the amount of heavy-metal contamination in the atmosphere by measuring the pollution in outer tail feathers collected from a certain species of bird. The amount of pollution found in the birds' feathers may be assumed to follow a Normal distribution.

Regular monitoring of feathers showed that, at the time of closure of a copper smelting factory, the mean level of pollution was 4.9 units. Five years later a sample was collected, with pollution levels as follows.

4.6 3.5 2.1 4.8 6.4 4.6 2.2 3.9 1.6 1.8 2.1 3.6

(i) Use these data to estimate the population mean and standard deviation. [2]

(ii) Examine at the 5% significance level whether this sample provides evidence that there has been a reduction in the mean level of pollution. State your null and alternative hypotheses clearly. [10]

(iii) What assumption must be made about the sampling process for the above test to be valid? [1]

4 In a psychology experiment to determine whether personality and colour preference are related, a random sample of 200 people is taken. Their personalities are classified as either 'introvert' or 'extrovert' and they are asked their colour preferences. The results are summarised in the following table.

		Introvert	Extrovert
Preferred colour	Red	24	56
	Yellow	7	17
	Green	22	28
	Blue	28	18

(i) Examine, at the 5% level of significance, whether these data provide any evidence of an association between these classification factors. State clearly your null and alternative hypotheses. [12]

(ii) Discuss your findings. [3]

5 A pharmacologist is carrying out research on a possible treatment for asthma. She thinks that a particular drug may help to reduce the percentage bronchial restriction of asthma sufferers caused by prolonged exercise.

Following prolonged exercise, the median bronchial restriction of asthma sufferers is 12 per cent. The drug is given to a randomly chosen sample of 10 asthma sufferers and the percentage bronchial restriction, following prolonged exercise, is measured. The results for the sample are as follows.

20 11 9 17 19 1 6 10 3 2

(i) Use a Wilcoxon test to examine, at the 5% significance level, whether the drug is effective in reducing the median bronchial restriction. State your null and alternative hypotheses clearly. [12]

(ii) Suppose that it can be assumed that the underlying distribution of percentage bronchial restriction is Normal, but with unknown variance. Explain why a test for the population mean using the Normal distribution may be unsuitable. Suggest a more suitable test procedure. [2]

OCR
RECOGNISING ACHIEVEMENT

ADVANCED SUBSIDIARY GCE UNIT **G242/01**

MEI STATISTICS

Statistics 2 (Z2)

THURSDAY 14 JUNE 2007 Afternoon

Time: 1 hour 30 minutes

Additional Materials:
Answer booklet (8 pages)
Graph paper
MEI Examination Formulae and Tables (MF2)

INSTRUCTIONS TO CANDIDATES

- Write your name, centre number and candidate number in the spaces provided on the answer booklet.
- Answer **all** the questions.
- You are permitted to use a graphical calculator in this paper.
- Final answers should be given to a degree of accuracy appropriate to the context.

INFORMATION FOR CANDIDATES

- The number of marks is given in brackets [ ] at the end of each question or part question.
- The total number of marks for this paper is 72.

ADVICE TO CANDIDATES

- Read each question carefully and make sure you know what you have to do before starting your answer.
- 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.

This document consists of **4** printed pages.

© OCR 2007 [L/103/0483]

1 On work days, Rodney aims to catch the 0700 train at his local station. The 0700 train always departs on time. Each day, Rodney leaves home at 0630. The time for his journey to the station may be modelled using a Normal distribution with mean 26 minutes and standard deviation 2.4 minutes.

(i) Find the probability that Rodney arrives at the station in time to catch the train. [3]

Rodney would like to be at least 99% certain of arriving at the station in time to catch the 0700 train.

(ii) What is the latest time, to the nearest minute, that Rodney could leave home? [4]

The 0700 train is scheduled to arrive at Rodney's destination at 0805. Each day, over a period of 30 days, Rodney records the actual journey time. He finds that the sample mean is 67.4 minutes and the sample standard deviation is 2.45 minutes.

(iii) Find a two-sided 95% confidence interval for the mean journey time for this train. [4]

(iv) Comment on the scheduled 0805 arrival time in relation to the confidence interval found in part **(iii)**. [3]

(v) Comment on the appropriateness of Rodney's sample. [2]

2 The numbers of Manx Shearwaters flying past a bird observatory each day, for a sample of 15 days this year, are as follows.

191 224 278 300 281 217 412 490 410 433 360 376 365 290 178

(i) Records show that for several years the median number per day has been 280. Use a Wilcoxon test to examine, at the 5% significance level, whether there has been an increase in the numbers of Manx Shearwaters flying past the observatory this year. State your null and alternative hypotheses clearly. [12]

(ii) What assumptions about the variable being tested and the sample used are necessary when carrying out this Wilcoxon test? [2]

3 A brewing company wishes to test a large crop of hops for the presence of a particular fungicide. The agreed safe mean level of fungicide is 50 mg/kg. The brewing company measures the amount of fungicide in each of a random sample of ten specimens taken from the crop, with results in mg/kg as follows.

45 49 37 41 51 46 47 46 44 53

(i) Use these data to estimate the population mean and standard deviation. [2]

(ii) Use a *t* test to examine at the 5% significance level whether this sample provides evidence that the mean level of fungicide in this crop is less than the agreed safe level. State your null and alternative hypotheses clearly. [9]

(iii) What assumption is necessary for the above test to be valid? [2]

© OCR 2007

4 A researcher is monitoring the presence of bacteria of a certain kind in a river. He collects a random sample of water specimens from the river, each containing the same volume, and counts the number of these bacteria in each specimen. He believes that the number of bacteria in these specimens may be modelled using a Poisson distribution.

(i) Write down the conditions for a Poisson model to apply. [2]

The researcher's results are as follows.

Number of bacteria, x	0	1	2	3	4	5	6	7	$\geqslant 8$
Observed frequency, f	17	26	48	40	32	17	15	5	0

(ii) The sample standard deviation is 1.762, correct to 3 decimal places.

(*A*) Verify that the sample mean number of bacteria is 2.9. [2]

(*B*) Do these statistics give you any reason to doubt the appropriateness of the Poisson model? Justify your answer. [2]

(iii) Taking 2.9 as an estimate for the mean of the underlying population, use the appropriate cumulative probability tables to find the probabilities corresponding to $x = 0$, $x = 1$ and $x \geqslant 8$. Hence obtain the expected frequencies corresponding to the observed frequencies for these values of x. [5]

The expected and observed frequencies are used to carry out a test of the goodness of fit of the Poisson model. The cells for $x = 7$ and $x \geqslant 8$ are merged. The calculated statistic for the χ^2 test is 9.032.

(iv) What is the conclusion of the test when a 5% significance level is used? Justify your answer using an appropriate critical value. [4]

[Question 5 is printed overleaf.]

5 As part of a survey to determine whether television viewing preferences differ between females and males, a random sample of 250 people is taken. The television viewing preferences are classified as film, drama, news, sport, music and wildlife. The results are as follows.

	Female	Male
Film	26	20
Drama	24	12
News	29	10
Sport	24	39
Music	14	17
Wildlife	15	20
Total	132	118

(i) A test is to be carried out to examine whether these data provide any evidence of an association between these classification factors. State clearly the null and alternative hypotheses. The following tables show some of the expected frequencies and corresponding contributions to the test statistic. Calculate the remaining expected frequencies and contributions. Carry out the test, at the 5% level of significance. [11]

Expected frequencies

	Female	Male
Film	24.288	21.712
Drama	19.008	16.992
News	20.592	18.408
Sport	33.264	29.736
Music		
Wildlife		

Contributions to the test statistic

	Female	Male
Film	0.120 67	0.134 99
Drama	1.311 03	1.466 58
News	3.433 10	3.840 42
Sport	2.580 02	2.886 12
Music		
Wildlife		

(ii) Discuss briefly how television viewing preferences differ between females and males, as shown by the contributions to the test statistic. [3]

Permission to reproduce items where third-party owned material protected by copyright is included has been sought and cleared where possible. Every reasonable effort has been made by the publisher (OCR) to trace copyright holders, but if any items requiring clearance have unwittingly been included, the publisher will be pleased to make amends at the earliest possible opportunity.

OCR is part of the Cambridge Assessment Group. Cambridge Assessment is the brand name of University of Cambridge Local Examinations Syndicate (UCLES), which is itself a department of the University of Cambridge.

© OCR 2007

OCR
RECOGNISING ACHIEVEMENT

ADVANCED SUBSIDIARY GCE **G242/01**

MEI STATISTICS

Statistics 2 (Z2)

WEDNESDAY 21 MAY 2008 Afternoon

Time: 1 hour 30 minutes

Additional materials: Answer Booklet (8 pages)
Graph paper
MEI Examination Formulae and Tables (MF2)

INSTRUCTIONS TO CANDIDATES

- Write your name in capital letters, your Centre Number and Candidate Number in the spaces provided on the Answer Booklet.
- Read each question carefully and make sure you know what you have to do before starting your answer.
- Answer **all** the questions.
- You are permitted to use a graphical calculator in this paper.
- Final answers should be given to a degree of accuracy appropriate to the context.

INFORMATION FOR CANDIDATES

- The number of marks is given in brackets [ ] at the end of each question or part question.
- The total number of marks for this paper is 72.
- 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.

This document consists of **4** printed pages.

© OCR 2008 [L/103/0483]

1 An environmental group is monitoring the number of Scots pine seedlings on a large region of Scottish moorland. The moorland is divided into plots of equal area. The numbers of Scots pine seedlings in the plots are modelled using a Poisson distribution.

(i) Write down the conditions for a Poisson model to apply. [2]

Records for similar regions suggest the mean number of seedlings per plot is 8.

(ii) Find the probability that in a randomly chosen plot there will be

(*A*) exactly 7 seedlings, [2]

(*B*) more than 7 seedlings. [2]

(iii) Find the probability that, out of a random sample of 5 plots, at least one plot contains more than 7 seedlings. [3]

The heights, in cm, of seedlings are modelled using a Normal distribution with mean 56 and standard deviation 20.

(iv) Show that the probability that the height of a seedling exceeds 70 cm is 0.242. [2]

(v) Find the height exceeded by 95% of seedlings. [3]

2 A coffee company is investigating the cholesterol-lowering effects of removing a particular oil in the production of a new coffee product. Over a period of several weeks, a group of 12 volunteers drink this new coffee product. For each volunteer, the decrease in cholesterol level is measured. The results, in suitable units, are as follows.

182 194 202 209 187 165 168 171 178 201 190 173

(i) Use these data to estimate the mean and variance of the underlying population. [2]

The coffee company will promote the new coffee product if it believes that the population mean decrease in cholesterol level is more than 175.

(ii) Use a *t* test at the 5% level of significance to examine whether this sample provides evidence that the mean decrease in cholesterol level is more than 175. State your null and alternative hypotheses clearly. [9]

(iii) What assumptions about the underlying population and the sample are necessary for the above test to be valid? [2]

© OCR 2008

3 A medical research team is investigating whether there is an association between coronary artery disease and gum disease. They collect data from a random sample of 250 patients who have been tested for possible heart conditions. The results are as follows.

	Patients with gum disease	Patients without gum disease
Patients with coronary artery disease	40	45
Patients without coronary artery disease	103	62

(i) Carry out a test, at the 5% level of significance, to examine whether these data provide evidence of an association between these classification factors. State clearly the null and alternative hypotheses. [10]

In a further investigation, patients with gum disease are classified by age and by occurrence of coronary artery disease. For a random sample of 160 such patients, the results are as follows.

	Age $\leqslant 45$	$45 <$ Age $\leqslant 60$	Age > 60
Patients with coronary artery disease	4	12	17
Patients without coronary artery disease	49	32	46

(ii) These results are used to carry out a test for association between age and occurrence of coronary artery disease among people with gum disease. The calculated statistic for the χ^2 test is 8.2808. Given that no cells needed to be merged, what is the conclusion of the test when a 5% significance level is used? Justify your answer with an appropriate critical value. [4]

(iii) In the light of the conclusions obtained in parts **(i)** and **(ii)**, discuss briefly what the results suggest about links between gum disease and coronary artery disease. [2]

4 It is suspected that, due to global warming, the population of a certain species of ant is decreasing. As part of an ongoing field survey regarding the distribution of this species of ant, a number of traps are located at different sites over a large area. One hour after each trap is set, the number of trapped ants is counted. The results for a sample of 12 traps this year are as follows.

23 43 16 15 22 33 27 11 18 24 10 20

(i) Records for this species of ant taken over a number of years show the median number of trapped ants to be 26. Use a Wilcoxon test to examine, at the 5% significance level, whether there has been a decrease in the median number of ants this year. State your null and alternative hypotheses clearly. [12]

(ii) What assumptions about the variable being tested and the sample used are necessary when carrying out this Wilcoxon test? [2]

[Question 5 is printed overleaf.]

© OCR 2008

5 A pet products researcher is investigating the bedding material preferences of hamsters. He wishes to find out if hamsters prefer to sleep on a new man-made bedding material rather than the more traditional bedding materials. He has eight hamsters on which to carry out his research. Each hamster is given a choice of three bedding compartments, only one of which contains the new man-made material. The researcher records the numbers of hamsters, x, that choose the new material over a period of 150 days. The results are as follows.

Number of hamsters choosing new material, x	0	1	2	3	4	5	$\geqslant 6$
Frequency, f	10	31	42	34	19	8	6

(i) Explain why, assuming the hamsters choose independently and randomly, the researcher should model the data using a binomial distribution with $p = \frac{1}{3}$. [1]

(ii) Use the cumulative binomial probability table with $n = 8$ and $p = \frac{1}{3}$ to find the probabilities corresponding to the values of x in the table. Hence obtain the expected frequencies corresponding to the observed frequencies. [5]

(iii) Carry out a test of the goodness of fit of the above binomial model. Use a 5% significance level. State your null and alternative hypotheses clearly. [9]

Permission to reproduce items where third-party owned material protected by copyright is included has been sought and cleared where possible. Every reasonable effort has been made by the publisher (OCR) to trace copyright holders, but if any items requiring clearance have unwittingly been included, the publisher will be pleased to make amends at the earliest possible opportunity.

OCR is part of the Cambridge Assessment Group. Cambridge Assessment is the brand name of University of Cambridge Local Examinations Syndicate (UCLES), which is itself a department of the University of Cambridge.

OCR
RECOGNISING ACHIEVEMENT

ADVANCED SUBSIDIARY GCE

MEI STATISTICS

Statistics 2 (Z2)

G242

Candidates answer on the Answer Booklet

OCR Supplied Materials:
- 8 page Answer Booklet
- Graph paper
- MEI Examination Formulae and Tables (MF2)

Other Materials Required:
None

Monday 1 June 2009
Morning

Duration: 1 hour 30 minutes

INSTRUCTIONS TO CANDIDATES

- Write your name clearly in capital letters, your Centre Number and Candidate Number in the spaces provided on the Answer Booklet.
- Use black ink. Pencil may be used for graphs and diagrams only.
- Read each question carefully and make sure that 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 graphical calculator in this paper.
- Final answers should be given to a degree of accuracy appropriate to the context.

INFORMATION FOR CANDIDATES

- The number of marks is given in brackets **[ ]** at the end of each question or part question.
- 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 document consists of **8** pages. Any blank pages are indicated.

© OCR 2009 [L/103/0483]
OCR is an exempt Charity

1 A wine producer wishes to determine whether the quality of the grapes harvested on his estate is associated with the part of the estate on which they are grown. To investigate this, the estate is divided into three different areas. A random sample of 300 bunches of grapes is taken from across the entire estate. The area from which each bunch is taken is noted and the quality of each bunch is classified as excellent, good or satisfactory. The results are as follows.

	Area A	Area B	Area C
Excellent	36	18	30
Good	36	50	42
Satisfactory	21	31	36

(i) State null and alternative hypotheses for a test to examine whether these data provide any evidence of an association between these classification factors. **[1]**

The following tables show some of the expected frequencies and contributions to the test statistic.

Expected frequencies

	Area A	Area B	Area C
Excellent	26.04	27.72	30.24
Good	39.68		
Satisfactory	27.28		

Contributions to the test statistic

	Area A	Area B	Area C
Excellent	3.8096	3.4083	0.0019
Good	0.3413		
Satisfactory	1.4457		

(ii) Calculate the remaining expected frequencies and contributions to the statistic for the χ^2 test, and carry out the test at the 5% significance level. **[10]**

(iii) With reference to the values in the above tables, discuss briefly how the quality of grapes harvested differs between areas. **[3]**

© OCR 2009

2 **(i)** A coal merchant supplies smokeless fuel in bags labelled as containing 25 kg. A machine is used to fill the bags. The quantity delivered to each bag, X kg, may be modelled using a Normal distribution with mean 25.2 kg and standard deviation 0.1 kg.

(*A*) Find the probability that the machine delivers less than 25 kg of smokeless fuel to a bag. **[3]**

(*B*) Find the probability that, in a customer's order of five bags, there is at least one bag containing less than 25 kg. **[3]**

(ii) The coal merchant also sells house coal in bags labelled as containing 25 kg. The coal merchant investigates whether or not the machine used to fill bags with house coal is delivering the nominal amount of 25 kg on average. The amount, y kg, in each of a random sample of 50 bags is measured. The results are summarised as follows.

$$\Sigma y = 1295 \qquad \Sigma y^2 = 33\,544$$

(*A*) Use these data to show that the sample variance is $0.071\,43\ \text{kg}^2$, correct to 4 significant figures. **[2]**

(*B*) Find a two-sided 95% confidence interval for the mean amount delivered by this machine. **[4]**

(*C*) With reference to the confidence interval found in part (*B*), describe what the coal merchant could conclude about the amount delivered by this machine. **[3]**

3 A practice manager is monitoring the length of time, during evening surgeries, that patients wait to see a doctor. She has found the median waiting time to be 23 minutes. She decides to change the system of allocating appointment times.

Following the introduction of a new system, she measures the waiting times, during evening surgeries, of a random sample of 12 patients. The results, in minutes, for the sample are as follows.

14 33 12 11 6 16 27 18 29 8 9 20

(i) Use a Wilcoxon test to examine, at the 5% significance level, whether the new appointment booking system has been successful in reducing the median waiting time. State your null and alternative hypotheses clearly. **[12]**

(ii) Suppose that it could be assumed that the underlying distribution of these waiting times is Normal. State, with a reason, what the most appropriate test procedure would be. **[2]**

© OCR 2009

4 A botanist is investigating how the seeds of the Black Nightshade plant spread. He believes that seeds spread randomly, and that X, the number of seedlings in plots of a given area, may be modelled using a Poisson distribution. To test his belief, the botanist divides a region into plots of equal area, and then counts the number of Black Nightshade seedlings in each of a random sample of 120 plots.

The botanist's results are as follows.

Number of seedlings, x	0	1	2	3	4	5	6	$\geqslant 7$
Observed frequency	16	25	24	18	12	13	12	0

The sample standard deviation is 1.880 correct to 3 decimal places.

(i) (*A*) Verify that the sample mean is 2.6. **[2]**

(*B*) Do the sample statistics provide any reason to doubt the appropriateness of the Poisson model? Justify your answer. **[2]**

(*C*) Does the pattern of observed frequencies provide any reason to doubt the appropriateness of the Poisson model? Justify your answer. **[2]**

The botanist wishes to carry out a test of the goodness of fit of the Poisson model. He uses 2.6 as an estimate for the mean of the underlying population. The following tables show some of the expected frequencies and corresponding contributions to the test statistic.

Expected frequencies

Number of seedlings, x	0	1	2	3	4	5	$\geqslant 6$
Expected frequency	8.916	23.172	30.120	26.112	16.968		

Contributions to the test statistic

Number of seedlings, x	0	1	2	3	4	5	$\geqslant 6$
Contribution	5.6284	0.1442	1.2435	2.5201			6.3698

(ii) Use the appropriate cumulative probability tables to find $P(X = 5)$ and $P(X \geqslant 6)$ and hence calculate the remaining expected frequencies and contributions. Carry out the test at the 5% level of significance. **[10]**

© OCR 2009

5 A golf club manufacturer is testing clubs made with a new alloy to find out if the average ball striking distance is greater than for clubs made with the standard alloy. A random sample of 10 observations from a test using the new alloy is as follows. All distances are in yards.

269 275 279 266 273 274 276 275 267 276

(i) Use these data to estimate the mean and variance of the underlying population. **[2]**

In previous tests it has been found that, when using clubs made with the standard alloy, the ball striking distance is Normally distributed with a mean of 268 yards.

(ii) Use a t test at the 5% level of significance to examine whether this sample provides evidence that the mean striking distance for clubs made with the new alloy is greater than 268 yards. State your null and alternative hypotheses clearly. **[9]**

(iii) When testing golf clubs, a robot arm is used to ensure that all balls are struck with a club head speed of 100 mph. Tests are carried out at an outdoor golf driving range. Describe briefly two other factors that the golf club manufacturer could consider to help ensure fairness in the testing procedure. **[2]**

© OCR 2009

OCR
RECOGNISING ACHIEVEMENT

ADVANCED SUBSIDIARY GCE

MEI STATISTICS

Statistics 2 (Z2)

G242

Candidates answer on the Answer Booklet

OCR Supplied Materials:
- 8 page Answer Booklet
- Graph paper
- MEI Examination Formulae and Tables (MF2)

Other Materials Required:
- Scientific or graphical calculator

Wednesday 9 June 2010
Afternoon

Duration: 1 hour 30 minutes

INSTRUCTIONS TO CANDIDATES

- Write your name clearly in capital letters, your Centre Number and Candidate Number in the spaces provided on the Answer Booklet.
- Use black ink. Pencil may be used for graphs and diagrams only.
- Read each question carefully and make sure that 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 graphical calculator in this paper.
- Final answers should be given to a degree of accuracy appropriate to the context.

INFORMATION FOR CANDIDATES

- The number of marks is given in brackets **[ ]** at the end of each question or part question.
- 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 document consists of **4** pages. Any blank pages are indicated.

© OCR 2010 [L/103/0483]
OCR is an exempt Charity

1 A birdwatcher has learned to recognise different species of birds by their song. He notices that three particular types of warbler regularly found in his 'patch' prefer to sing in trees. He decides to investigate whether there is any association between the type of warbler and the type of tree from which they are heard singing. 200 warblers, regarded as a random sample, are selected and the numbers of warblers in each category are summarised in the table below.

		Tree		
		Willow	Birch	Oak
Warbler	Chiffchaff	10	13	20
	Willow Warbler	39	43	12
	Whitethroat	24	20	19

(i) A test is to be carried out to examine whether these data provide any evidence of an association between these classification factors. State clearly the null and alternative hypotheses. The following tables show some of the expected frequencies and contributions to the test statistic. Calculate the remaining expected frequencies and contributions. Carry out the test at the 5% level of significance. **[11]**

Expected frequencies		Tree		
		Willow	Birch	Oak
Warbler	Chiffchaff	15.695	16.340	10.965
	Willow Warbler	34.310		
	Whitethroat	22.995		

Contributions to the test statistic		Tree		
		Willow	Birch	Oak
Warbler	Chiffchaff	2.0665	0.6827	7.4447
	Willow Warbler	0.6411		
	Whitethroat	0.0439		

(ii) For each type of warbler, comment briefly on how its distribution compares with what would be expected if there were no association. **[3]**

(iii) While out for a walk, the birdwatcher hears the song of a whitethroat. Use the given data to estimate the probability that it is singing from a birch tree. **[2]**

© OCR 2010

2 A doctor working in a hospital in a poor area of a large city is concerned about the low average birth weight of babies born in the hospital. For babies born in this hospital, the mean birth weight is 2800 grams, which is well below the ideal birth weight. The doctor introduces an extensive prenatal care programme in an attempt to increase the mean birth weight. Following the introduction of the programme, the doctor measures the birth weight of each of a random sample of 12 babies born in the hospital, with results in grams as follows.

2430 2720 2910 3000 3230 2840 2660 3350 3210 2870 2820 3540

(i) Explain why, in this situation, it would not be appropriate to carry out a hypothesis test for a population mean using the Normal distribution. State the assumption necessary for a test based on the t distribution to be valid. **[3]**

(ii) Use these data to estimate the population mean and the population standard deviation. **[3]**

(iii) Use a t test to examine at the 5% significance level whether this sample provides evidence that the prenatal care programme has been successful in increasing the mean birth weight of babies born in this hospital. State your null and alternative hypotheses clearly. **[10]**

3 A regional highway authority is concerned about the high numbers of accidents involving cyclists at roundabouts. A random sample of 150 roundabouts is selected, and the number of accidents involving cyclists at each of these roundabouts over a four-week period is recorded. The results are shown in the following frequency table.

Number of accidents, x	0	1	2	3	4	5	6	$\geqslant 7$
Observed frequency, f	21	36	26	24	23	12	8	0

(i) The sample standard deviation is 1.734, correct to 3 decimal places.

(*A*) Verify that the sample mean number of accidents is 2.4. **[2]**

(*B*) Do these statistics give you any reason to doubt the belief that the number of accidents may be modelled using a Poisson distribution? Justify your answer. **[2]**

(ii) The highway authority wishes to carry out a test of the goodness of fit of the Poisson model. The sample mean of 2.4 is used as an estimate of the mean of the underlying population. The following tables show some of the expected frequencies and corresponding contributions to the test statistic. Use the appropriate cumulative probability tables to find the remaining expected frequencies, and calculate the remaining contributions. Carry out the test at the 5% level of significance. **[10]**

Expected frequencies

Number of accidents, x	0	1	2	3	4	5	$\geqslant 6$
Expected frequency	13.605		39.195	31.350	18.810	9.030	

Contributions to the test statistic

Number of accidents, x	0	1	2	3	4	5	$\geqslant 6$
Contribution	4.0196	0.3426			0.9333	0.9768	1.3064

4 As part of a research project involving a particular colony of common seals, a biologist is investigating the length of time that seals spend under water each time they dive. The dive durations, in seconds, for a random sample of 10 adolescent seals are as follows.

243 251 218 227 205 232 198 224 187 264

Over a period of time, the biologist has found that, for this particular seal colony, the median length of dive is 210 seconds. Use a Wilcoxon test to examine, at the 5% significance level, whether the sample provides evidence of a difference between the median dive duration of these adolescent seals and that of the seal colony as a whole. State your null and alternative hypotheses clearly. **[13]**

5 A large brewery supplies beer in bottles labelled as containing 500 ml. The bottles are filled by machine. The random variable X represents the volume of beer, in ml, delivered to each bottle. X is Normally distributed with mean μ and standard deviation 1.29. The value of μ can be adjusted by a machine operator.

(i) Given that $\mu = 502$, find $P(X < 500)$. **[3]**

(ii) Find the value of μ needed to ensure that 1% of bottles filled by the machine contain less than 500 ml. **[3]**

The brewery also sells beer in casks labelled as containing 9 gallons. During one month, a random sample of 40 casks is selected. The sample mean volume of beer is 9.05 gallons and the sample standard deviation is 0.06 gallons.

(iii) Find a two-sided 95% confidence interval for the mean volume of beer per cask. **[5]**

(iv) The brewery aims to avoid the mean volume being less than the advertised 9 gallons. Comment on this, using the confidence interval found in part **(iii)** to support your answer. **[2]**

OCR
RECOGNISING ACHIEVEMENT

Copyright Information

OCR is committed to seeking permission to reproduce all third-party content that it uses in its assessment materials. OCR has attempted to identify and contact all copyright holders whose work is used in this paper. To avoid the issue of disclosure of answer-related information to candidates, all copyright acknowledgements are reproduced in the OCR Copyright Acknowledgements Booklet. This is produced for each series of examinations, is given to all schools that receive assessment material and is freely available to download from our public website (www.ocr.org.uk) after the live examination series.

If OCR has unwittingly failed to correctly acknowledge or clear any third-party content in this assessment material, OCR will be happy to correct its mistake at the earliest possible opportunity.

For queries or further information please contact the Copyright Team, First Floor, 9 Hills Road, Cambridge CB2 1GE.

OCR is part of the Cambridge Assessment Group; Cambridge Assessment is the brand name of University of Cambridge Local Examinations Syndicate (UCLES), which is itself a department of the University of Cambridge.

© OCR 2010

OCR
RECOGNISING ACHIEVEMENT

ADVANCED SUBSIDIARY 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

Monday 13 June 2011
Morning

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. 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 destroyed.

© OCR 2011 [L/103/0483]

OCR is an exempt Charity

1 When making bread, Nancy measures out the flour using a cup. The weight, in grams, of flour in the cup is Normally distributed with mean 144 and standard deviation 2.6.

(i) Find the probability that the weight of flour in the cup is less than 146 grams. **[3]**

(ii) 12% of the time the weight of flour in the cup exceeds k grams. Find the value of k. **[3]**

A recipe for a large loaf of bread requires 4 cups of flour.

(iii) Assuming that the weights of flour per cup are independent, find the probability that the total weight of flour in 4 cups is less than 568 grams. **[4]**

Nancy uses this recipe to make 5 large loaves of bread.

(iv) Find the probability that at least one of the loaves contains less than 568 grams of flour. **[3]**

2 The owner of a large vineyard regularly carries out tests to monitor his grape vines for signs of disease. His vineyard is divided into plots of size $10\,m^2$. One test, carried out in June each year, involves counting the number of 'scorched' leaves per plot, as scorched leaves can be a sign of a potentially destructive disease. From tests carried out in previous years, he has established that the median number of scorched leaves per plot is 25.

He takes steps intended to reduce the number of scorched leaves. The following June he counts the numbers of scorched leaves per plot for a sample of 12 plots. The results are as follows.

24 21 18 23 28 20 31 9 36 13 15 17

(i) Use a Wilcoxon test to examine, at the 5% significance level, whether the vineyard owner has achieved his aim. **[12]**

(ii) What assumption about the sample used is necessary when carrying out this Wilcoxon test? **[1]**

3 Marine scientists have been studying populations of the great scallop, *Pecten maximus*, at a number of sites around the coast of Britain. After over-fishing in the 20th century, the mean size of scallop caught fell below the legal minimum, 110 mm, in so many sites that conservation measures were introduced. At one particular site where the mean size had fallen to 85 mm, scallop fishing was banned for a period of 3 years. Following this 3-year period, a random sample of 15 scallops was obtained at this site and used to produce a 95% confidence interval for the population mean size, in mm, of scallop. The resulting confidence interval, based on the t distribution, was (96.6, 99.4).

(i) Use the confidence interval to show that, at this site, an estimate for the mean of the underlying population is 98.0 mm and obtain an estimate for the population variance. **[5]**

(ii) Explain how the limits of this confidence interval may be used to support the view that the conservation measures are working but the site is not yet ready for the fishing ban to be lifted. **[2]**

Following a further 3-year period, a random sample of 12 scallops is obtained at this site and the size of each scallop recorded. The results, in mm, are as follows.

112 103 115 118 109 107 115 116 111 112 114 113

(iii) Use a t test to examine at the 5% level of significance whether this sample provides evidence that the population mean scallop size now exceeds the legal minimum of 110 mm. **[9]**

© OCR 2011

4 A student ornithologist is investigating the feeding habits of Eurasian Oystercatchers. She knows that the bill shape of oystercatchers varies between individual birds and suspects that there is an association between bill shape and the location in which birds feed. The bill shapes of 140 oystercatchers, regarded as a random sample, are observed in different feeding locations. The results are summarised in the table below.

		Bill shape	
		Blunt tip	Pointed tip
Feeding location	Shoreline	27	11
	Mudflats	15	37
	Inland field	26	24

(i) A test is to be carried out to examine whether these data provide any evidence of an association between bill shape and feeding location. State the null and alternative hypotheses. The following tables show some of the expected frequencies and contributions to the test statistic. Calculate the remaining expected frequencies and contributions. Complete the test at the 5% level of significance. **[11]**

Expected frequencies		Bill shape	
		Blunt tip	Pointed tip
Feeding location	Shoreline	18.457	19.543
	Mudflats		
	Inland field		

Contributions to the test statistic		Bill shape	
		Blunt tip	Pointed tip
Feeding location	Shoreline	3.9540	3.7344
	Mudflats		
	Inland field		

(ii) With reference to the contributions to the test statistic, comment briefly on how bill shape in each feeding location compares with what would be expected if there were no association. **[3]**

[Question 5 is printed overleaf.]

5 In a bid to increase its number of subscribers, a satellite television company is targeting individual customers who have failed to renew their subscription packages. Individual customers are contacted, by telephone, and offered the chance to purchase a new subscription package at a favourable rate. For every 5 customers called, a record is kept of the number of acceptances of the new package. The results for a random sample of 60 groups of 5 customers are as follows.

Number of acceptances	0	1	2	3	4	5
Observed frequency	17	22	13	6	1	1

(i) Use these data to find the sample mean. **[2]**

A manager at the company proposes to use these data to carry out a test of the goodness of fit of the binomial model B(5, p).

(ii) Show that the manager should use $p = 0.25$. **[2]**

The following table shows the expected frequencies obtained using B(5, 0.25).

Number of acceptances	0	1	2	3	4	5
Expected frequency	14.238	23.730	15.822	5.274	0.876	0.060

(iii) Explain why, in this case, it is appropriate to use 2 degrees of freedom when carrying out the goodness of fit test. **[3]**

(iv) Given that the resulting test statistic is 1.6812, carry out the test at the 10% level of significance. **[3]**

The satellite television company has a customer support service which receives telephone calls at a uniform average rate of 3 calls every 10 seconds.

(v) Use a Poisson model to calculate the probability that

(*A*) exactly 3 calls are received in a 10-second period, **[2]**

(*B*) at least 3 calls are received in a 20-second period. **[3]**

(vi) What additional assumption is needed to justify the use of the Poisson model in part **(v)**? **[1]**

OCR
RECOGNISING ACHIEVEMENT

Copyright Information

OCR is committed to seeking permission to reproduce all third-party content that it uses in its assessment materials. OCR has attempted to identify and contact all copyright holders whose work is used in this paper. To avoid the issue of disclosure of answer-related information to candidates, all copyright acknowledgements are reproduced in the OCR Copyright Acknowledgements Booklet. This is produced for each series of examinations and is freely available to download from our public website (www.ocr.org.uk) after the live examination series.

If OCR has unwittingly failed to correctly acknowledge or clear any third-party content in this assessment material, OCR will be happy to correct its mistake at the earliest possible opportunity.

For queries or further information please contact the Copyright Team, First Floor, 9 Hills Road, Cambridge CB2 1GE.

OCR is part of the Cambridge Assessment Group; Cambridge Assessment is the brand name of University of Cambridge Local Examinations Syndicate (UCLES), which is itself a department of the University of Cambridge.

© OCR 2011

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

G242590612

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.

© OCR 2012 [L/103/0483]

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]**

© OCR 2012

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

Copyright Information

OCR is committed to seeking permission to reproduce all third-party content that it uses in its assessment materials. OCR has attempted to identify and contact all copyright holders whose work is used in this paper. To avoid the issue of disclosure of answer-related information to candidates, all copyright acknowledgements are reproduced in the OCR Copyright Acknowledgements Booklet. This is produced for each series of examinations and is freely available to download from our public website (www.ocr.org.uk) after the live examination series.

If OCR has unwittingly failed to correctly acknowledge or clear any third-party content in this assessment material, OCR will be happy to correct its mistake at the earliest possible opportunity.

For queries or further information please contact the Copyright Team, First Floor, 9 Hills Road, Cambridge CB2 1GE.

OCR is part of the Cambridge Assessment Group; Cambridge Assessment is the brand name of University of Cambridge Local Examinations Syndicate (UCLES), which is itself a department of the University of Cambridge.

© OCR 2012

OCR
RECOGNISING ACHIEVEMENT

Thursday 6 June 2013 – Morning

AS GCE MEI STATISTICS

G242/01 Statistics 2 (Z2)

QUESTION PAPER

G21599061 3

Candidates answer on the Printed Answer Book.

OCR supplied materials:
- Printed Answer Book G242/01
- MEI Examination Formulae and Tables (MF2)

Other materials required:
- Scientific or graphical calculator

Duration: 1 hour 30 minutes

INSTRUCTIONS TO CANDIDATES

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

- The Question Paper will be found in the centre of the Printed Answer Book.
- Write your name, centre number and candidate number in the spaces provided on the Printed Answer Book. Please write clearly and in capital letters.
- **Write your answer to each question in the space provided in the Printed Answer Book.** Additional paper may be used if necessary but you must clearly show your candidate number, centre number and question number(s).
- Use black ink. HB pencil may be used for graphs and diagrams only.
- Read each question carefully. Make sure you know what you have to do before starting your answer.
- Answer **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.

© OCR 2013 [L/103/0483]
DC (LEG) 66321/1
OCR is an exempt Charity

1 A health care inspector is monitoring the occurrence of the 'superbug' MRSA in a large hospital. Each day, he checks for the presence of MRSA in different areas of the hospital and records the number of areas where MRSA is detected. The following table summarises the results for a random sample of 300 days.

Number of areas where MRSA is detected	0	1	2	3	4
Observed frequency	86	108	74	24	8

(i) Show that the sample mean is 1.2. Given also that the sample standard deviation is 1.03, explain why the Poisson distribution may provide a suitable model for these data. **[3]**

The inspector wishes to test the goodness of fit of a Poisson model. The mean calculated from the data is used as an estimate for the mean of the underlying population to produce the following expected frequencies.

Number of areas where MRSA is detected	0	1	2	3	$\geqslant 4$
Expected frequency	90.36	108.43	65.06	26.02	10.13

(ii) Show how the expected frequency of 90.36 for 0 areas containing MRSA is calculated. **[3]**

(iii) Carry out the test of the goodness of fit of the Poisson model at the 5% level of significance. **[9]**

2 A gas boiler manufacturer is working to improve its domestic hot water boilers. It wishes to ensure that the boilers heat water as quickly as possible. New boilers are tested in controlled conditions by measuring the temperature of the water after 15 seconds of heating. The resulting temperatures, in degrees Celsius, for a random sample of 8 observations are as follows.

51.0 50.7 49.8 50.4 50.6 50.8 49.3 50.6

It is required to produce a confidence interval for the population mean temperature using these data.

(i) Stating a necessary assumption, explain why a confidence interval based on the *t* distribution is more appropriate, in this case, than a confidence interval based on the Normal distribution. **[4]**

(ii) Obtain a 95% confidence interval, based on the *t* distribution, for the population mean temperature. **[7]**

The manufacturer aims to produce a boiler that, in 15 seconds, heats water to 50 degrees Celsius on average.

(iii) With reference to the confidence interval found in part **(ii)**, comment on whether the manufacturer has been successful in its aim. **[2]**

© OCR 2013

3 A musical instrument manufacturer is developing a new range of violin strings made from a synthetic material. Part of the development process involves testing breaking strength. The breaking strengths of a random sample of 12 strings are measured. The results, in suitable units, are as follows.

347 345 349 348 350 349 349 344 354 347 351 350

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

It is required that the mean breaking strength of synthetic strings should not be below 350.

(ii) Given that the sample standard deviation is 2.678, use a t test to examine, at the 5% significance level, whether this sample provides evidence that the mean breaking strength of these synthetic strings is below the minimum requirement. **[11]**

4 'Cool Milk Dairy' offers a delivery service for its milk. Milk is delivered each day in glass bottles. Empty bottles are collected so that they can be washed and used again. The dairy manager is monitoring the return of bottles to ensure there are enough bottles in stock to cover future demand. Over a long period of time she has established that, on average, 13 fewer bottles are collected each day than are delivered. For the past six months the dairy has been replacing lost stock with bottles from a different supplier and the dairy manager suspects that the average of 13 might have changed. A daily record of the difference between the number of bottles delivered, x, and the number of empty bottles collected, y, is kept. The differences, $x - y$, for ten consecutive days are as follows.

25 3 4 17 18 11 19 14 20 24

(i) Stating any necessary assumptions, use a Wilcoxon test to examine, at the 5% significance level, whether these data support the dairy manager's suspicions. **[13]**

(ii) Comment on the validity of any assumptions made in part **(i)**. **[2]**

[Question 5 is printed overleaf.]

5 A sea ferry operator is working to improve punctuality on one of its sailing routes. It decides to test for an association between the sea condition at the time of sailing and punctuality. The punctuality and sea condition of a random sample of 200 ferry crossings are recorded. The results are as follows.

		Punctuality	
		Not late	Late
Sea condition	Rough	36	12
	Moderate	91	12
	Slight	44	5

The following tables show some of the expected frequencies and contributions to the test statistic.

Expected frequencies		Punctuality	
		Not late	Late
Sea condition	Rough	41.040	6.960
	Moderate	88.065	
	Slight	41.895	

Contributions to the test statistic		Punctuality	
		Not late	Late
Sea condition	Rough	0.6189	3.6497
	Moderate	0.0978	
	Slight	0.1058	

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

The journey times for this particular route may be modelled using a Normal distribution with mean 160 minutes and standard deviation 5.5 minutes.

(ii) Given that the ferry departs at 08:30 and is scheduled to arrive at 11:15, find the probability that it is not late. **[5]**

OCR
RECOGNISING ACHIEVEMENT

Copyright Information

OCR is committed to seeking permission to reproduce all third-party content that it uses in its assessment materials. OCR has attempted to identify and contact all copyright holders whose work is used in this paper. To avoid the issue of disclosure of answer-related information to candidates, all copyright acknowledgements are reproduced in the OCR Copyright Acknowledgements Booklet. This is produced for each series of examinations and is freely available to download from our public website (www.ocr.org.uk) after the live examination series.

If OCR has unwittingly failed to correctly acknowledge or clear any third-party content in this assessment material, OCR will be happy to correct its mistake at the earliest possible opportunity.

For queries or further information please contact the Copyright Team, First Floor, 9 Hills Road, Cambridge CB2 1GE.

OCR is part of the Cambridge Assessment Group; Cambridge Assessment is the brand name of University of Cambridge Local Examinations Syndicate (UCLES), which is itself a department of the University of Cambridge.

© OCR 2013

OCR

Friday 6 June 2014 – Afternoon

AS GCE MEI STATISTICS

G242/01 Statistics 2 (Z2)

QUESTION PAPER

12408670 49

Candidates answer on the Printed Answer Book.

OCR supplied materials:
- Printed Answer Book G242/01
- MEI Examination Formulae and Tables (MF2)

Other materials required:
- Scientific or graphical calculator

Duration: 1 hour 30 minutes

INSTRUCTIONS TO CANDIDATES

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

- The Question Paper will be found 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.

© OCR 2014 [L/103/0483]

1 A dietician is investigating a claim that the dietary supplement 'red-yeast rice' can reduce cholesterol levels by 15 units on average. The dietician believes that the reduction will be greater than this. A group of volunteers is given the supplement for a period of eight weeks. The reduction in cholesterol levels of a sample of 10 of these volunteers is measured. The results are as follows.

16 24 27 22 11 13 26 21 18 20

(i) Stating any necessary assumptions, use a Wilcoxon test to examine, at the 5% significance level, whether these data support the dietician's belief. **[13]**

(ii) What further assumption is necessary for a test based on the *t* distribution to be appropriate? **[1]**

2 A medical researcher is investigating alternative methods for determining glucose levels in diabetes patients' blood. The most common method involves taking a small sample of blood; this is accurate but may be stressful for some patients. One alternative method involves measuring the glucose level in tear fluid. To assess the accuracy of this alternative method, the researcher records the difference in values given by the two methods for each of a random sample of 15 patients. The results, in suitable units, are as follows.

0.3 0.5 –0.2 0.8 0.3 –0.1 0.1 –0.2 0.4 –0.1 0.5 0.3 0.4 0.1 –0.7

(i) Given that the sample standard deviation is 0.374, use the *t* distribution to test, at the 5% significance level, the null hypothesis $H_0: \mu = 0$ against the alternative hypothesis $H_1: \mu \neq 0$, where μ represents the mean of the underlying population of differences. **[8]**

Another alternative method for determining blood glucose levels involves a test using ultrasound. The researcher records the difference between the value given by an ultrasound test and the value given by a blood test for a random sample of 12 patients.

(ii) Given that the sample mean is 0.19 and the sample standard deviation is 0.281, calculate a 95% confidence interval based on the *t* distribution for the mean difference in level given by these two methods. **[5]**

(iii) Given that underlying Normality holds, why is the *t* distribution required in parts **(i)** and **(ii)**? **[1]**

© OCR 2014

3 A sports equipment manufacturer produces 'yellow dot' and 'red dot' squash balls. The weights of each type of ball may be assumed to be Normally distributed. The manufacturer carries out regular checks of the production process. The weights, in grams, for a random sample of 10 yellow dot balls are as follows.

23.7 24.2 24.6 24.1 24.3 24.4 23.9 23.8 24.0 24.2

(i) Calculate the sample mean. **[1]**

The weight of a yellow dot ball should be 24 grams.

(ii) Given that the population standard deviation is 0.3 grams, use a test based on the Normal distribution, at the 5% significance level, to examine whether this sample provides evidence that the mean weight of yellow dot balls is not 24 grams. **[9]**

The weights of red dot balls are Normally distributed with mean 24.7 grams and standard deviation 0.4 grams. Red dot balls are sold in packs of 6. Balls in a pack may be regarded as a random sample.

(iii) Find the probability that the total weight of the 6 balls in a pack exceeds 150 grams. **[4]**

4 A seismologist is monitoring the global occurrence of earthquakes. Each week, she records the number of 'moderate strength' earthquakes that occur around the world. The following table summarises the results for a random sample of 200 weeks.

Number of earthquakes	0	1	2	3	4	5	6	$\geqslant 7$
Observed frequency	7	47	49	46	25	19	7	0

(i) The sample mean is 2.6. Calculate the sample standard deviation. Hence comment briefly on whether or not the Poisson distribution may provide a suitable model for these data. **[3]**

The seismologist decides to test the goodness of fit of a Poisson model. She uses the sample mean as an estimate for the mean of the underlying population to produce the following expected frequencies.

Number of earthquakes	0	1	2	3	4	5	6	$\geqslant 7$
Expected frequency	14.86	38.62	50.21	43.52	28.28	14.71	6.37	3.43

(ii) Show how the expected frequency of 28.28 for the number of weeks in which 4 earthquakes occur is calculated. **[3]**

(iii) Carry out the test of the goodness of fit of the Poisson model at the 5% level of significance. **[10]**

Question 5 begins on page 4

© OCR 2014

5 A car manufacturer decides to investigate whether there is an association between the age of customers and their 'brand loyalty' (ie whether they are returning customers or first time customers). A random sample of 150 customers is selected and classified as follows.

		Brand loyalty	
		Returning customer	First time customer
Age (in years)	Under 35	25	12
	35 to 50	38	35
	Over 50	31	9

The following tables show some of the expected frequencies and contributions to the test statistic for use in part **(i)**.

Expected frequencies		Brand loyalty	
		Returning customer	First time customer
Age (in years)	Under 35	23.187	13.813
	35 to 50	45.747	
	Over 50		14.933

Contributions to the test statistic		Brand loyalty	
		Returning customer	First time customer
Age (in years)	Under 35	0.142	0.238
	35 to 50	1.312	
	Over 50		2.357

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

(ii) With reference to the contributions to the test statistic, comment briefly on brand loyalty in the different age groups. **[3]**

END OF QUESTION PAPER

OCR

Copyright Information

OCR is committed to seeking permission to reproduce all third-party content that it uses in its assessment materials. OCR has attempted to identify and contact all copyright holders whose work is used in this paper. To avoid the issue of disclosure of answer-related information to candidates, all copyright acknowledgements are reproduced in the OCR Copyright Acknowledgements Booklet. This is produced for each series of examinations and is freely available to download from our public website (www.ocr.org.uk) after the live examination series.

If OCR has unwittingly failed to correctly acknowledge or clear any third-party content in this assessment material, OCR will be happy to correct its mistake at the earliest possible opportunity.

For queries or further information please contact the Copyright Team, First Floor, 9 Hills Road, Cambridge CB2 1GE.

OCR is part of the Cambridge Assessment Group; Cambridge Assessment is the brand name of University of Cambridge Local Examinations Syndicate (UCLES), which is itself a department of the University of Cambridge.

© OCR 2014

OCR
Oxford Cambridge and RSA

Wednesday 3 June 2015 – Morning

AS GCE MEI STATISTICS

G242/01 Statistics 2 (Z2)

QUESTION PAPER

4422783791

Candidates answer on the Printed Answer Book.

OCR supplied materials:
- Printed Answer Book G242/01
- MEI Examination Formulae and Tables (MF2)

Other materials required:
- Scientific or graphical calculator

Duration: 1 hour 30 minutes

INSTRUCTIONS TO CANDIDATES

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

- The Question Paper will be found in the centre of the Printed Answer Book.
- Write your name, centre number and candidate number in the spaces provided on the Printed Answer Book. Please write clearly and in capital letters.
- **Write your answer to each question in the space provided in the Printed Answer Book.** Additional paper may be used if necessary but you must clearly show your candidate number, centre number and question number(s).
- Use black ink. HB pencil may be used for graphs and diagrams only.
- Read each question carefully. Make sure you know what you have to do before starting your answer.
- Answer **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 **8** 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.

© OCR 2015 [L/103/0483]

OCR is an exempt Charity

1 An engineering company is developing a device to convert tidal energy into electricity. To be viable, the device should produce, on average, 8000 units of electricity per day. The device is tested over a period of several months. The number of units of electricity generated per day on 10 particular days is as follows.

8970 7830 8303 8045 8641 7942 8701 8670 7840 8532

(i) Stating any necessary assumptions, use a Wilcoxon test to examine, at the 5% significance level, whether these data suggest that this device produces more than the required 8000 units of electricity per day on average. **[13]**

Having established Normality of the underlying population, the engineering company decides to carry out a hypothesis test for the population mean.

(ii) State, giving reasons, the test that should be used. **[2]**

2 A food standards officer is investigating the level of salt contained in a particular brand of biscuit. The officer is concerned that, on average, these biscuits contain more than the claimed amount of 0.16 g per biscuit. The officer selects a random sample of 200 biscuits and measures the amount of salt, x g, contained in each biscuit. The results are summarised as follows.

$$\sum x = 34.70 \qquad \sum x^2 = 6.737$$

(i) Calculate the sample mean and show that an estimate for the population variance is 0.00360 correct to 3 significant figures. **[3]**

(ii) Use a test based on the Normal distribution to examine, at the 10% significance level, whether the officer's concern is justified. **[8]**

(iii) Explain why a test based on the Normal distribution is appropriate in this case. **[1]**

(iv) For a test, also carried out at the 10% level, on a different brand of biscuit, the officer summarised a significant result as follows.

This proves that the mean amount of salt exceeds the stated value.

Comment briefly on the wording used in the officer's summary. **[2]**

© OCR 2015

3 An astronomer is investigating the arrival of photons from a distant astronomical source. The number of photons detected each second during a period of 10 minutes is recorded. The results are shown in the following table.

Number of photons detected each second	0	1	2	3	4	5
Observed frequency	157	220	152	54	15	2

(i) Describe the conditions required for a Poisson distribution to be an appropriate model for the number of photons detected each second. **[2]**

The astronomer decides to use these data to carry out a test of the goodness of fit of the Poisson model, using a mean of 1.26 calculated from the data. Some of the expected frequencies are shown in the table below.

Number of photons detected each second	0	1	2	3	4	$\geqslant 5$
Expected frequency	170.19	214.44	135.10	56.74		

(ii) Calculate the remaining expected frequencies. **[5]**

(iii) Carry out the test using a 5% level of significance. **[8]**

[Questions 4 and 5 are printed overleaf.]

© OCR 2015

4 A cycle racing team coach is investigating the performance of bicycles with tyres which are inflated to higher than usual pressure. She believes that the performance will be improved on some types of road surface when compared with the average result for tyres inflated to the usual pressure. She decides to carry out a χ^2 test to investigate whether there is an association between bicycle performance and type of road surface. The performance and type of road surface for a random sample of 125 test runs are recorded. The results are as follows.

		Bicycle performance		
		Worse	Same	Improved
Type of road surface	Smooth	3	14	25
	Rough	11	14	10
	Very Rough	19	18	11

The following tables show some of the expected frequencies and contributions to the test statistic.

Expected frequencies		Bicycle performance		
		Worse	Same	Improved
Type of road surface	Smooth	11.088	15.456	15.456
	Rough	9.240		
	Very Rough	12.672		

Contributions to the test statistic		Bicycle performance		
		Worse	Same	Improved
Type of road surface	Smooth	5.8997	0.1372	5.8934
	Rough	0.3352	0.0974	0.6440
	Very Rough	3.1600		

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

(ii) With reference to the contributions to the test statistic, comment briefly on the bicycle performance for each type of road surface. **[3]**

5 Eleanor works in a sugar processing factory. She is responsible for ensuring that the mean weight of sugar per bag does not fall below 1 kg, as labelled. From past experience, she assumes that the weights are Normally distributed.

Eleanor measures the weight, in kg, of sugar in each of a random sample of 8 bags to enable her to calculate a confidence interval for the mean weight of sugar per bag. The results are as follows.

0.998 0.999 0.991 1.005 0.999 0.989 0.998 1.003

(i) Explain why, in this case, it is not appropriate to calculate a confidence interval based on the Normal distribution. **[2]**

(ii) Obtain an appropriate 95% confidence interval for the mean weight of sugar per bag. **[9]**

(iii) Comment briefly on whether Eleanor should be concerned about the weight of sugar per bag. **[3]**

END OF QUESTION PAPER

© OCR 2015

BLANK PAGE

© OCR 2015

BLANK PAGE

© OCR 2015

OCR
Oxford Cambridge and RSA

Copyright Information

OCR is committed to seeking permission to reproduce all third-party content that it uses in its assessment materials. OCR has attempted to identify and contact all copyright holders whose work is used in this paper. To avoid the issue of disclosure of answer-related information to candidates, all copyright acknowledgements are reproduced in the OCR Copyright Acknowledgements Booklet. This is produced for each series of examinations and is freely available to download from our public website (www.ocr.org.uk) after the live examination series.

If OCR has unwittingly failed to correctly acknowledge or clear any third-party content in this assessment material, OCR will be happy to correct its mistake at the earliest possible opportunity.

For queries or further information please contact the Copyright Team, First Floor, 9 Hills Road, Cambridge CB2 1GE.

OCR is part of the Cambridge Assessment Group; Cambridge Assessment is the brand name of University of Cambridge Local Examinations Syndicate (UCLES), which is itself a department of the University of Cambridge.

© OCR 2015