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.

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 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]

BLANK PAGE

Q1			
(a)(i)	$P(\text{time} < 94) = P(Z < \frac{94-93}{0.9})$ $P(Z < 1.111)$ 0.8667	M1 A1 A1	**3**
(a)(ii)	For 4 games, the probability that all four last less than 94 minutes is $(0.8667)^4 = 0.5642\ldots$ $1 - (0.8667)^4$ 0.4357…	M1 M1 A1	**3**
(b)(i)	$\dfrac{1989670 - \frac{10920^2}{60}}{59}$ = 37.7966…. = 37.80 (2d.p.)	M1 A1	**2**
(b)(ii)	Sample mean = 182 $182 \pm 1.96\times\sqrt{37.80}\div\sqrt{60}$ (B1 for 1.96) (180.4,183.6)	B1 B1M1 A1	**4**
(b)(iii)	The C.I. does not contain 188 suggesting the mean height of goalkeepers could be greater than that of outfield players. Relevant comment based on probability – in context - e.g. A wider CI based on the outfield players data – 99%, say - would still not contain 188 cm.	E1(188 not in) E1(suggesting) E1(mean greater) E1(prob^y link)	**4**

Q2			
(i)	$\Sigma fx \div \Sigma f = 340 \div 100$ (=3.4) $(1486 - 340^2/100) \div 99$ Sample variance = $3\frac{1}{3}$ No reason to doubt manager as mean $\approx$ variance	B1 M1 A1 E1	**4**
(ii)	0.0334, 0.1134, 0.1929, 0.2187, 0.1858, 0.1263, 0.0716, 0.0348, 0.0231 3.34, 11.34, 19.29, 21.87, 18.58, 12.63, 7.16, 3.48, 2.31	M1 A2 M1 A1	**5**
(iii)	(*A*) Low expected frequencies have a disproportionate influence on the value of X^2 and may make the procedure a poor approx$^{\underline{n}}$ (*B*) 5 degrees of freedom (7 – 1 – 1) Critical value at 5% level is $\chi^2 = 11.07$ 5.127 < 11.07 so not significant The Poisson model seems a good fit.	E1 B1 B1 M1 E1	**5**

Q3			
(i)	Mean = $3\frac{13}{30}$ & SD = 1.49989…	B1, B1	**2**
(ii)	$H_0 : \mu = 4.9$ & $H_1 : \mu < 4.9$ Where μ represents the population mean pollution level *t* distribution needed $t = \frac{3\frac{13}{30} - 4.9}{SD/\sqrt{12}}$ = -3.39 (3s.f.) 11 degrees of freedom At 5% level, critical value of *t* = 1.796 -3.394 < -1.796 so the result is significant Evidence suggests there is a reduction in mean pollution level	B1 B1 B1 M1 A1 (FT) B1 B1 M1 A1 E1	**10**
(iii)	Sample is random	B1	**1**

<table>
<tr><td>Q4</td><td></td><td></td><td></td></tr>
<tr><td>(i)</td><td>
H₀: there is no association between personality and colour preference
H₁: there is an association between personality and colour preference

Expected frequencies
<table>
<tr><td></td><td></td><td>Introvert</td><td>Extrovert</td></tr>
<tr><td>Preferred colour</td><td>Red</td><td>32.4</td><td>47.6</td></tr>
<tr><td></td><td>Yellow</td><td>9.72</td><td>14.28</td></tr>
<tr><td></td><td>Green</td><td>20.25</td><td>29.75</td></tr>
<tr><td></td><td>Blue</td><td>18.63</td><td>27.37</td></tr>
</table>
Contribution to X^2
<table>
<tr><td></td><td></td><td>Introvert</td><td>Extrovert</td></tr>
<tr><td>Preferred colour</td><td>Red</td><td>2.18</td><td>1.48</td></tr>
<tr><td></td><td>Yellow</td><td>0.76</td><td>0.52</td></tr>
<tr><td></td><td>Green</td><td>0.15</td><td>0.10</td></tr>
<tr><td></td><td>Blue</td><td>4.71</td><td>3.21</td></tr>
</table>
X^2= 13.11 (13.11399… without rounding)

3 degrees of freedom

Critical value for 5% significance level is 7.815
As 13.11 > 7.815 the result is significant

There is evidence of an association between personality and colour preference.
</td><td>B1
B1
M1A1
M1A1
A1
B1
B1
M1A1
E1</td><td>12</td></tr>
<tr><td>(ii)</td><td>People classed as extrovert tend to prefer red.
People classed as introvert tend to prefer blue.
Third relevant comment e.g. referring to specific contribution to X^2</td><td>E1
E1
E1</td><td>3</td></tr>
</table>

Q5			
(i)	H_0: population median = 12 H_1: population median < 12 Actual differences +8 -1 -3 +5 +7 -11 -6 -2 -9 -10 Associated ranks 7 1 3 4 6 10 5 2 8 9 $T^- = 1 + 3 + 10 + 5 + 2 + 8 + 9 = 38$ $T^+ = 7 + 4 + 6 = 17$ $\therefore\ T = 17$ From tables – at the 5% level of significance in a one-tailed Wilcoxon signed rank test, the critical value of T is 10 $17 > 10$ $\therefore$ the result is not significant The evidence does not suggest the drug is effective.	B1 B1 B1 M1A1 B1 B1 B1 B1 M1A1 E1	**12**
(ii)	Sample too small *t* distribution	B1 B1	**2**