

**ADVANCED SUBSIDIARY GCE UNIT
MEI STATISTICS**

G242/01

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.

- 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]

- 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	≥ 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 $x \geq 8$. 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 $x \geq 8$ 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.

**Mark Scheme G242
June 2007**

1(i)	$P(X < 30) = P\left(Z < \frac{30 - 26}{2.4}\right)$ $= P(Z < 1.666\dots)$ $= 0.952 \text{ (3sf)}$	M1 A1 A1	3
1(ii)	$\Phi^{-1}(0.99) = 2.326$ $\frac{x - 26}{2.4} = 2.326 \quad (\text{Equation in positive } z)$ $x = 31.58 \text{ (CAO)}$ <p>must leave at 06 28 (dependent on M1)</p>	B1 M1 A1 A1 FT	4
1(iii)	<p>C.I. centred on 67.4</p> $67.4 \pm 1.96 \times 2.45 \div \sqrt{30} \quad (\text{B1 for } 1.96)$ <p>(66.5, 68.3)</p>	M1 B1 M1 A1 (FT sensible z)	4
1(iv)	<p>The CI does not contain the scheduled journey time of 65 minutes.</p> <p>This suggests that the scheduled time is not accurate.</p> <p>The mean journey time could be greater than 65 minutes.</p>	E1(65 not in) E1(suggests) E1(mean greater)	3
1(v)	<p>Valid comment on sample size.</p> <p>Valid comment on randomness.</p>	E1 E1	2
2(i)	<p>H_0: population median = 280</p> <p>H_1: population median > 280</p> <p>Actual differences -89 -56 -2 20 1 -63 132 210 130 153 80 96 85 10 -102</p> <p>Associated ranks 9 5 2 4 1 6 13 15 12 14 7 10 8 3 11</p> <p>$T = 9 + 5 + 2 + 6 + 11 = 33$</p> <p>$T^+ = 4 + 1 + 13 + 15 + 12 + 14 + 7 + 10 + 8 + 3 = 87$</p> <p>$\therefore T = 33$</p> <p>From tables – at the 5% level of significance in a one-tailed Wilcoxon signed rank test, the critical value of T is 30</p> <p>$33 > 30 \therefore$ the result is not significant</p> <p>The evidence does not suggest an increase in the numbers of Manx Shearwaters flying past the observatory at this time of year.</p>	B1 B1 M1 M1 A1 B1 B1 B1FT B1 M1 A1 E1	12
2(ii)	<p>Variable - symmetry about median</p> <p>Sample - random</p>	E1 E1	2

3(i)	mean = 45.9 standard deviation = 4.65	B1 B1	2
3(ii)	$H_0 : \mu = 50$ & $H_1 : \mu < 50$ Where μ represents the mean fungicide level in the underlying population. $t = \frac{45.9 - 50}{\frac{SD}{\sqrt{10}}} = -2.79 \text{ (3s.f.)}$ 9 degrees of freedom At 5% level, critical value of $t = 1.833$ $-2.79 < -1.833$ so the result is significant Evidence suggests the mean level of fungicide in the crop is below the agreed safe level.	B1 E1 M1 A1 B1 B1 M1 A1 A1	9
3(iii)	The amount of fungicide in the underlying population follows a Normal distribution	E1(Normal) E1dep(pop ^d)	2
4(i)	Bacteria occur randomly with a uniform mean rate of occurrence	E1(random) E1(uniform mean rate)	2
4(ii)A	$\Sigma fx \div \Sigma f$ $= 580 \div 200 = 2.9$ (A.G.)	M1 A1	2
4(ii)B	Variance = $1.762^2 = 3.1046\dots$ mean \approx variance so does not give reason to doubt suitability of the Poisson model	B1 E1dep	2
4(iii)	$P(X = 0) = 0.055$ $P(X = 1) = 0.1596$ $P(X \geq 8) = 0.0099$ 11 ($X = 0$), 31.92 ($X = 1$), 1.98 ($X \geq 8$)	B1 B1 B1 M1 ($\times 200$) A1	5
4(iv)	6 degrees of freedom ($8 - 1 - 1$) Critical value at 5% level is $\chi^2 = 12.59$ $9.032 < 12.59$ so not significant The Poisson model seems a good fit.	M1 (for 8-1-1) A1 M1 A1 (F.T. sensible c.v.)	4

<p>5(i)</p>	<p>H_0: there is no association between sex and viewing preference H_1: there is an association between sex and viewing preference</p> <p>Expected frequencies</p> <table border="1" data-bbox="400 353 1209 600"> <thead> <tr> <th colspan="2"></th> <th>Female</th> <th>Male</th> </tr> </thead> <tbody> <tr> <th rowspan="6">Programme Type</th> <th>Film</th> <td>24.288</td> <td>21.712</td> </tr> <tr> <th>Drama</th> <td>19.008</td> <td>16.992</td> </tr> <tr> <th>News</th> <td>20.592</td> <td>18.408</td> </tr> <tr> <th>Sport</th> <td>33.264</td> <td>29.736</td> </tr> <tr> <th>Music</th> <td>16.368</td> <td>14.632</td> </tr> <tr> <th>Wildlife</th> <td>18.48</td> <td>16.52</td> </tr> </tbody> </table> <p>Contribution to X^2</p> <table border="1" data-bbox="400 667 1209 913"> <thead> <tr> <th colspan="2"></th> <th>Female</th> <th>Male</th> </tr> </thead> <tbody> <tr> <th rowspan="6">Programme Type</th> <th>Film</th> <td>0.12067</td> <td>0.13499</td> </tr> <tr> <th>Drama</th> <td>1.31103</td> <td>1.46658</td> </tr> <tr> <th>News</th> <td>3.43310</td> <td>3.84042</td> </tr> <tr> <th>Sport</th> <td>2.58002</td> <td>2.88612</td> </tr> <tr> <th>Music</th> <td>0.34258</td> <td>0.38323</td> </tr> <tr> <th>Wildlife</th> <td>0.65532</td> <td>0.73308</td> </tr> </tbody> </table> <p>$X^2 = 17.88715$ 5 degrees of freedom Critical value at 5% level is 11.07 As $17.88715 > 11.07$ the result is significant</p> <p>There is evidence of an association between sex and viewing preference.</p>			Female	Male	Programme Type	Film	24.288	21.712	Drama	19.008	16.992	News	20.592	18.408	Sport	33.264	29.736	Music	16.368	14.632	Wildlife	18.48	16.52			Female	Male	Programme Type	Film	0.12067	0.13499	Drama	1.31103	1.46658	News	3.43310	3.84042	Sport	2.58002	2.88612	Music	0.34258	0.38323	Wildlife	0.65532	0.73308	<p>B1</p> <p>M1 A1</p> <p>M1 M1</p> <p>A1 B1 B1 M1A1</p> <p>A1</p>	<p>11</p>
		Female	Male																																														
Programme Type	Film	24.288	21.712																																														
	Drama	19.008	16.992																																														
	News	20.592	18.408																																														
	Sport	33.264	29.736																																														
	Music	16.368	14.632																																														
	Wildlife	18.48	16.52																																														
		Female	Male																																														
Programme Type	Film	0.12067	0.13499																																														
	Drama	1.31103	1.46658																																														
	News	3.43310	3.84042																																														
	Sport	2.58002	2.88612																																														
	Music	0.34258	0.38323																																														
	Wildlife	0.65532	0.73308																																														
<p>5(ii)</p>	<p>The largest contributions to the test statistic come from news and sport programmes. The value 3.43310 shows a positive association between females and news programmes whereas the value 3.84042 shows a negative association between males and news programmes. With sport, the preferences are reversed. The smallest contributions to the test statistic were for film programmes showing similarities between male and female.</p>	<p>E1</p> <p>E1</p> <p>E1</p>	<p>3</p>																																														