OXFORD CAMBRIDGE AND RSA EXAMINATIONS

Advanced Subsidiary General Certificate of Education
Advanced General Certificate of Education

MEI STRUCTURED MATHEMATICS 2616

Statistics 4

Friday **13 JUNE 2003** Morning 1 hour 20 minutes

Additional materials:
  Answer booklet
  Graph paper
  MEI Examination Formulae and Tables (MF12)

TIME 1 hour 20 minutes

INSTRUCTIONS TO CANDIDATES

- Write your Name, Centre Number and Candidate Number in the spaces provided on the answer booklet.
- Answer any **three** questions.
- You are permitted to use a graphical calculator in this paper.

INFORMATION FOR CANDIDATES

- The allocation 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.
- Final answers should be given to a degree of accuracy appropriate to the context.
- The total number of marks for this paper is 60.

This question paper consists of 4 printed pages.

1 **(i)** The random variable X has the exponential distribution with parameter λ ($\lambda > 0$), so that its probability density function is

$$f(x) = \lambda e^{-\lambda x}$$

for $x \geqslant 0$. Derive the mean and variance of X. [6]

(ii) Show that the cumulative distribution function of X is

$$F(x) = 1 - e^{-\lambda x}$$

for $x \geqslant 0$. Hence write down an expression for $P(X > y)$ where $y \geqslant 0$. [2]

(iii) The smallest of a random sample of n independent observations from X is denoted by Y. By first considering $P(Y > y)$ where $y \geqslant 0$, show that the cumulative distribution function of Y is

$$G(y) = 1 - e^{-n\lambda y}$$

(for $y \geqslant 0$). Deduce that the distribution of Y is exponential with parameter $n\lambda$. [4]

(iv) Show that nY is an unbiased estimator of μ, the mean of X.

Find the variance of nY.

Discuss briefly the use of nY as an estimator of μ. [8]

2 A pharmaceutical company is investigating whether a new drug for patients with heart disease successfully reduces blood cholesterol levels. Nine such patients receive the drug for a trial period of six weeks, while as a control eight other such patients receive a placebo (i.e. a treatment that should have no effect). Each of these groups of patients may be considered as a random sample from the relevant population. At the end of the trial period, the blood cholesterol levels of the patients are found to be as follows, measured in a suitable unit. For convenience, the observations in each group have been arranged in ascending order.

Drug	243	246	250	257	260	262	267	287	295
Placebo	249	266	273	280	284	285	288	293	

(i) Use an appropriate non-parametric test, at the 5% level of significance, to examine whether, on the whole, patients treated with the drug in this way may be assumed to have lower blood cholesterol levels than patients who receive the placebo. [6]

(ii) The drug regulatory authorities require that these data are also analysed using an appropriate parametric procedure. Carry out this analysis, again using a 5% level of significance. [8]

(iii) What distributional assumption is needed in part **(ii)** but not in part **(i)**? By considering the data, comment briefly and informally on whether this assumption appears to hold. [You may wish to use simple diagrams.] [3]

(iv) What further assumption is needed in part **(ii)** but not in part **(i)**? [1]

(v) Compare and comment on the results from parts **(i)** and **(ii)**. [2]

3 A taxi fleet manager thinks that fuel consumption might be improved by adopting a new design of tyre. An experiment is conducted to compare fuel consumption using this new design and using standard tyres. Ten taxis are selected at random from the fleet, fitted with the new tyres, and driven for a month in normal service, each keeping its own driver throughout the trial. The average fuel consumption over this period is measured for each taxi, and compared with the average fuel consumption for a previous similar period with the standard tyres (each taxi still having its same driver). The results for average fuel consumption, in litres per 100 kilometres, are as follows.

Taxi	1	2	3	4	5	6	7	8	9	10
New tyres	18.2	17.6	19.4	17.9	18.9	17.4	18.5	19.0	18.9	17.2
Standard tyres	19.0	17.1	19.6	19.0	18.8	18.9	18.8	19.7	18.3	18.4

It is desired to examine the null hypothesis that, on the whole, the fuel consumption is the same with new and standard tyres against the alternative that it is better (i.e. the result in litres per 100 kilometres is *smaller*) with the new tyres.

(a) Making an appropriate assumption about underlying Normality, which should be carefully stated, use a t test to examine the above hypotheses at the 5% level of significance. [9]

(b) Provide an alternative analysis using an appropriate Wilcoxon test, again at the 5% level of significance. [6]

(c) Suppose the *only* information reported by the ten drivers taking part in the experiment is that seven found fuel consumption had improved with the new tyres and three found it had become worse. What could be concluded from a test based on the $\mathrm{B}(10, \frac{1}{2})$ distribution? [5]

4 The managers of a waste disposal site are carrying out a survey of usage. Each member of a random sample of people using the site is asked where they have come from. The site is meant to serve two local authorities, A and B, but it is known that users also come from elsewhere. Each member of the sample is also asked the main purpose of the visit, categorised as disposal of household rubbish, of garden waste, of recyclable materials and of old appliances such as worn-out refrigerators. The results are as follows.

		Main purpose of visit			
		Household rubbish	Garden waste	Recyclable materials	Old appliances
Origin	A	30	27	16	8
	B	15	28	13	7
	Other areas	8	8	9	11

(i) State the null and alternative hypotheses under examination in the usual χ^2 test of whether or not there is association between 'origin' and 'main purpose of visit'. [2]

(ii) Carry out the usual test, at the 5% significance level. [10]

(iii) Discuss your conclusions. [4]

(iv) Explain briefly *why*, in a contingency table of this type, the expected frequencies are calculated in the way that they are. [4]

Mark Scheme

Marking Instructions

Some marks in the mark scheme are explicitly designated as 'M', 'A', 'B' or 'E'.

'M' marks ('method') are for an *attempt to use* a correct method (not merely for stating the method).

'A' marks ('accuracy') are for accurate answers and can only be earned if corresponding 'M' mark(s) have been earned. Candidates are expected to give answers to a sensible level of accuracy in the context of the problem in hand. The level of accuracy quoted in the mark scheme will sometimes deliberately be greater than is required, when this facilitates marking.

'B' marks are independent of all others. Typically they are available for correct quotation of points such as 1.96 from tables.

'E' marks ('explanation') are for explanation and/or interpretation. These will frequently be sub-dividable depending on the thoroughness of the candidate's answer.

Follow-through marking should normally be used wherever possible – there will however be an occasional designation of 'c.a.o.' for 'correct answer only'.

Full credit **MUST** be given when correct alternative methods of solution are used. If errors occur in such methods, the marks awarded should correspond as nearly as possible to equivalent work using the method in the mark scheme.

All queries about the mark scheme should have been resolved at the standardisation meeting. Assistant Examiners should telephone the Principal Examiner (or Team Leader if appropriate) if further queries arise during the marking.

Assistant Examiners may find it helpful to use shorthand symbols as follows:

FT	Follow-through marking
✓^	Correct work after error
✗^	Incorrect work after error
C	Condonation of a minor slip
BOD	Benefit of doubt
NOS	Not on scheme (to be used *sparingly*)
I	Work of no value

Q1	(i)	$f(x) = \lambda e^{-\lambda x} \quad x \geqslant 0 \; (\lambda > 0)$		
		$E(X) = \int_0^\infty \lambda x e^{-\lambda x} dx$		
		$= \lambda\left\{\left[x\frac{e^{-\lambda x}}{-\lambda}\right]_0^\infty + \int_0^\infty \frac{e^{-\lambda x}}{\lambda} dx\right\}$ (attempt to integrate correct expression by parts)	M1	
		$= 0 + \left[\frac{e^{-\lambda x}}{-\lambda}\right]_0^\infty = \frac{1}{\lambda}$	1	
		$E(X^2) = \int_0^\infty \lambda x^2 e^{-\lambda x} dx$		
		$= \lambda\left\{\left[x^2\frac{e^{-\lambda x}}{-\lambda}\right]_0^\infty + \int_0^\infty \frac{2xe^{-\lambda x}}{\lambda} dx\right\}$ (attempt to integrate correct expression by parts)	M1	
		$0 + 2 \,.\, \frac{1}{\lambda}E(X)$	M1	
		(or by parts again) $= \frac{2}{\lambda^2}$	1	
		$\therefore \text{Var}(X) = E(X^2) - \{E(X)\}^2 = \frac{2}{\lambda^2} - \frac{1}{\lambda^2} = \frac{1}{\lambda^2}$ cao, and only from correct E(X) and E(X^2)	1	6
	(ii)	$\text{cdf } F(x) = \int_0^x \lambda e^{-\lambda t} dt = \left[e^{-\lambda t}\right]_0^x = 1 - e^{-\lambda x}$	1	
		Must be CONVINCING – beware printed answer		
		$\therefore P(X > y) = 1 - (1 - e^{-\lambda y}) = e^{-\lambda y}$	1	2
	(iii)	$P(Y > y) = P(\text{all } X\text{s are} > y)$	M1	
		$= \{e^{-\lambda y}\}^n$	1	
		$\therefore$ cdf of Y, $G(y) = 1 - \{e^{-\lambda y})^n = 1 - e^{-n\lambda y}$	1	
		This is same functional form as the cdf F(x) with λ replaced by $n\lambda$, $\therefore$ exponential distribution with parameter $n\lambda$ (or, differentiate and find g(y), etc)	1	4
	(iv)	Consider E[nY]	M1	
		$= n$. (mean of exponential ($n\lambda$))		
		$= n \times \frac{1}{n\lambda}$	1	
		$= \mu$	1	
		$\therefore$ unbiased	1	
		Var(nY) = n^2 (variance of exponential ($n\lambda$))	M1	
		$= n^2\left(\frac{1}{n\lambda}\right)^2 = \frac{1}{\lambda^2}\left[= \mu^2\right]$	1	
		Variance does not depend (inversely!) on sample size	E1	
		- no increase in precision as samples get larger (Two E1 marks are available for any two intelligent comments. Allow E1 for 'unbiased and this is a good thing'.)	E1	8

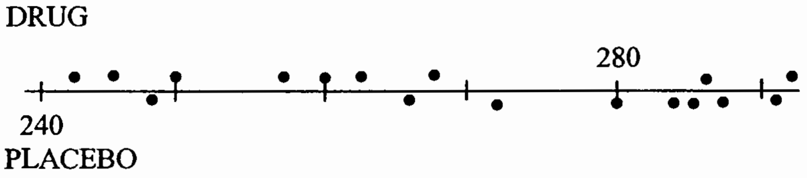
Q2

Drug	243	246	250	257	260	262	267	287	295
Placebo	249	266	273	280	284	285	288	293	

(i) Ranks are:

Drug	1	2	4	5	6	7	9	14	17
Placebo	3	8	10	11	12	13	15	16	

(i)	Rank sum is 88 (using placebo; 65 using drug. Mann-Whitney, if directly calculated, is 52 or 20)	1	
	If drug is effective, placebo should have *higher* ranks		
	A: (may be implicit) fo use of an upper tail	M1	
	so for one-sided 5% test we need *upper* 5% point of $W_{8,9}$		
	B: for using $W_{8,9}$	M1	
	Which is 90 [lower 5% point is 54, mean is 72]	1 cao	
	(upper 5% pt of M-$W_{8,9}$ is 54; lower 5% pt 18, mean 36)		
	Result is not significant		
	C: (Depends on getting the 90 right)	1	
	Drug does not appear successful	1	6
(ii)	$m = 9 \quad \bar{x} = 263.0 \quad s_{m-1} = 17.75 \quad s_{m-1}^2 = 315.0$		
	$n = 8 \quad \bar{y} = 277.25 \quad s_{n-1} = 14.24 \quad s_{n-1}^2 = 202.786$		
	Pooled $s^2 = \frac{8\times315.0+7\times202.786}{15} = 262.6\dot{3}$	M1 A1	
	For any reasonable attempt (but not unweighted average) at pooled estimate, and FT (M1)		
	If correct (A1)		
	Test statistic is $\frac{263.0-277.25}{\sqrt{262.6\dot{3}}\sqrt{\frac{1}{9}+\frac{1}{8}}} = -1.81$	M1 A1	
	Refer to t_{15} — No FT if wrong	1	
	Lower 5% point is –1.753 — No FT if wrong	1	
	Significant	1	
	Drug appears successful	1	8
(iii)	Normality of both populations	1	
	Consider e.g. dotplots:-		
	or for any other relevant display/discussion of the data	M1	3
	Neither looks very Normal	E1	
(iv)	Equal population variances	1	1
(v)	If assumptions (particularly Normality) are satisfied, *t* test is 'better' ('more powerful') and thus better at detecting a real difference. But if assumptions not satisfied, *t* test may mislead ('not robust') – Wilcoxon safer.	E2	2



Q3	(a)	Normality of *differences*	1	
		MUST be PAIRED COMPARISON t test Differences are −0.8 0.5 −0.2 −1.1 0.1 −1.5 −0.3 −0.7 0.6 −1.2	M1	
		$\bar{d} = -0.46$ $s_{n-1} = 0.7199$, $s_{n-1}^2 = 0.5182$	A1	
		Accept $s_n = 0.6829$, $s_n^2 = 0.4664$, but ONLY if correctly used in sequel Test statistics (for test of $\mu_D = 0$ against $\mu_D < 0$) is		
		$\frac{-0.46-0}{\frac{0.7199}{\sqrt{10}}} = -2.02(07)$	M1 A1	
		Refer to t_9 May be awarded even if test statistic is wrong. No FT if wrong	1	
		(lower) st 5% point is −1.833 Candidate must (somehow) make clear that correct tail is being used. No FT if wrong	1	
		Significant	1	
		Seems consumption is better with new design of tyre	1	9
	(b)	MUST be PAIRED WILCOXON test Ranks of $\lvert d \rvert$ are $7 \quad \underset{+}{\underset{\uparrow}{4}} \quad 2 \quad 8 \quad \underset{+}{\underset{\uparrow}{1}} \quad 10 \quad 3 \quad 6 \quad \underset{+}{\underset{\uparrow}{5}} \quad 9$	M1	
		Test statistic = 1 + 4 + 5 = 10 [or 45]	A1	
		Refer to paired Wilcoxon table with $n = 10$	M1	
		Lower 5% point is 10 [upper is 45]	1	
		∴ the observed 10 [or 45] *is* significant	1	
		Seems consumption is better with new design of tyre	1	6
	(c)	[sign test] 7 negative differences out of 10 If positive and negative differences are equally likely (ie if the null hypothesis is true), we would have B (10, ½) here	E1	
		P(B(10, ½) ≥ 7	M1	
		= 1 − 0.8281 = 0.1719	A1	
		This is quite a high probability, 'not significant' at any sensible significance level	E1	
		So no evidence of any change in fuel consumption	1	5

<table>
<tr><td>Q4</td><td>(i)</td><td>H_0 : no association
H_1 : association</td><td>1
1</td><td>2</td></tr>
<tr><td></td><td>(ii)</td><td>
<table>
<tr><td>o_i</td><td>H</td><td>G</td><td>Re-C</td><td>App</td><td></td></tr>
<tr><td>A</td><td>30</td><td>27</td><td>16</td><td>8</td><td>81</td></tr>
<tr><td>B</td><td>15</td><td>28</td><td>13</td><td>7</td><td>63</td></tr>
<tr><td>Others</td><td>8</td><td>8</td><td>9</td><td>11</td><td>36</td></tr>
<tr><td></td><td>53</td><td>63</td><td>38</td><td>26</td><td>180</td></tr>
</table>
<table>
<tr><td>e_i</td><td></td><td></td><td></td><td></td></tr>
<tr><td></td><td>23.85</td><td>28.35</td><td>17.1</td><td>11.7</td></tr>
<tr><td></td><td>18.55</td><td>22.05</td><td>13.3</td><td>9.1</td></tr>
<tr><td></td><td>10.6</td><td>12.6</td><td>7.6</td><td>5.2</td></tr>
</table>
Deduct 1 per error. Must be to this level of accuracy</td><td>A3</td><td></td></tr>
<tr><td></td><td></td><td>Contributions to X^2
<table>
<tr><td>1.5858</td><td>0.0643</td><td>0.0708</td><td>1.1701</td></tr>
<tr><td>0.6794</td><td>1.6056</td><td>0.0068</td><td>0.4846</td></tr>
<tr><td>0.6377</td><td>1.6794</td><td>0.2579</td><td>6.4692</td></tr>
</table>
$X^2 = 14.71$</td><td>M1
A1</td><td></td></tr>
<tr><td></td><td></td><td>Refer to χ^2_6 FT if df wrong, unless ≈ 180</td><td>2</td><td></td></tr>
<tr><td></td><td></td><td>Upper 5% point is 12.59</td><td>1</td><td></td></tr>
<tr><td></td><td></td><td>Significant</td><td>1</td><td></td></tr>
<tr><td></td><td></td><td>Seems there is association ZERO if $H_0 \leftrightarrow H_1$</td><td>1</td><td>10</td></tr>
<tr><td></td><td>(iii)</td><td>The key feature is that people from 'other areas' bring far more old appliances than would be expected if there were no association. Any other associations suggested by the data are comparatively slight ZERO if $H_0 \leftrightarrow H_1$</td><td>E4</td><td>4</td></tr>
<tr><td></td><td>(iv)</td><td>Absence of association</td><td></td><td></td></tr>
<tr><td></td><td></td><td>⇒ rows and columns are independent</td><td>E1</td><td></td></tr>
<tr><td></td><td></td><td>⇒ P (in (i, j) cell) = P (in row i) × P (in column j)</td><td>E1</td><td></td></tr>
<tr><td></td><td></td><td>which we can estimate by $\frac{n_{i\cdot}}{n} \times \frac{n_{\cdot j}}{n}$</td><td>E1</td><td></td></tr>
<tr><td></td><td></td><td>⇒ estimated expected frequency in (i, j) cell is n × this = $\frac{n_{i\cdot} n_{\cdot j}}{n}$</td><td>E1</td><td>4</td></tr>
</table>

Examiner’s Report

2616 Statistics 4

General Comments

The quality of the candidates' work was pleasingly high, with the majority managing to give substantial answers to 3 questions. There were not many very weak scripts and a large number of high quality scripts. Although the question on estimation was on very familiar territory, this question was by far the least popular of the four questions. In all questions, calculations and procedures were generally carried out accurately and confidently, but candidates were weaker where discussion or analysis were required. The candidates appeared to have sufficient time to finish the paper comfortably, indeed a significant number answered all 4 questions.

Comments on Individual Questions

Q.1 Estimation

This was the least popular question and was answered by about a quarter of all candidates. The derivation of the mean and variance posed few problems for most candidates, but a number of candidates were unable to integrate by parts accurately, and faked the final answers.

The derivations of the cumulative distribution functions of X and Y were well understood by candidates. Few candidates were able to note that the functional form of G(y) was the same as that of F(x) and hence deduce the distribution of Y. Instead they differentiated G(y) and considered the functional form of g(y). Of course, this is perfectly acceptable.

Part (iv) identified the best candidates, who simply noted that $\mathrm{E}[nY] = n\mathrm{E}[Y]$ and that $\mathrm{E}[Y] = \frac{1}{n\lambda}$ to deduce that nY was an unbiased estimator of μ, the mean of X. Weaker candidates, however, resorted to integration, and often got lost in the detail. A number of candidates were able to show that $\mathrm{Var}(nY) = \frac{1}{\lambda^2}$, but did not always realise the key point that this is independent of n and hence will not reduce as the sample size increases. Some candidates thought that this level of variance was "good", as long as λ was large.

(i) $\frac{1}{\lambda}$ and $\frac{1}{\lambda^2}$, (ii) $\mathrm{P}(X > y) = \mathrm{e}^{-\lambda y}$.

Q.2 Wilcoxon rank-sum and two sample *t* test

This was a very popular question and most candidates were able to score well. The Wilcoxon rank sum test in the first part was best answered by those candidates who used the Mann-Whitney approach as they were able to overcome the upper-tail problem by comparing the Mann-Whitney test statistic value of 20 with the tabulated (lower-tail) critical value of 18. Candidates using the Wilcoxon approach should have been careful to compare a calculated value of 88 (not 65 – this is the calculated value from the other sample) with an upper-tail critical value of 90. Few, however, did this directly. Many incorrectly compared 65 with the tabulated lower-tail critical value of 54, not realising that 65 would refer to an (m=9, n=8) case whereas 54 is the critical value for (m=8, n=9). An approach which was successful when using the Wilcoxon procedure was to compare min(R, $m(m+n+1) - R$) with the tabulated value, where R is the rank sum of the sample of size m.

Candidates also reacted very positively to part (ii) and most were clearly comfortable with the two sample t test. When calculating the pooled estimate of variance, the usual confusions between n and $(n - 1)$ and variance and standard deviation were rare. The calculations were generally correct and so were the degrees of freedom, critical value and conclusion.

Candidates were less successful in the final 3 parts of the question. In part (iii), the additional assumption required was the Normality of both populations. Many candidates talked about Normality of the population and many about the Normality of the samples. Many candidates were able to comment on the Normality, or otherwise, of the two populations, particularly those who took the hint and drew simple diagrams. The best candidates distinguished themselves by commenting that there were insufficient data to be sure one way or the other. In part (iv) answers were again marred by the lack of the word population and by the inclusion of the word sample.

Only a few candidates understood the importance of the final part of this question. Ideally, the *t* test should be used as it is a more powerful test, however, if the necessary assumptions are not true, then it is safer to use a non parametric test.

(i) Drug does not appear successful; (ii) Drug appears successful.

Q.3 Paired *t* test and paired Wilcoxon test

Virtually all candidates realised that a paired test was appropriate here. As with the previous question, many candidates were not precise enough in their descriptions of the required assumption. Apart from this lapse, most candidates were able to show that this was part of the syllabus that they understand well. It is particularly pleasing to report that virtually all candidates now give their conclusion to a hypothesis in context.

In part (b) candidates were again able to demonstrate their knowledge and also coped very well with the potential problem of having the test statistic equal to the critical value. A small minority of candidates tried to carry out an unpaired test.

The response of candidates to part (c) was disappointing with a large proportion of candidates simply evaluating $P(X = 7)$, others evaluating $P(X > 7)$ or the wrong tail or even using a Normal approximation.

(a) seems consumption is better with new tyres;
(b) seems consumption is better with new tyres;
(c) no evidence of any change in fuel consumption.

Q.4 Contingency table

A very popular question. Virtually all candidates were able to give the hypotheses the right way round. The test was carried out efficiently and with a high level of accuracy by virtually all candidates. This is clearly well understood material. The response to part (iii) was much weaker. Any discussion of the conclusions of this test without taking note of the contributions to the χ^2 statistic can only be very limited. This analysis makes it very clear that the dominant feature is that people from other areas bring far more old appliances than would be expected if there were no association. Some candidates identified that this was the key feature, but failed to mention whether there were more, or less, visits than expected.

Many candidates showed that they had a good understanding of the origin of the calculation of the expected frequencies, although the more subtle points concerning independence and the expected values only being estimates eluded all but the very best candidates.