OCR
RECOGNISING ACHIEVEMENT

OXFORD CAMBRIDGE AND RSA EXAMINATIONS

Advanced Subsidiary General Certificate of Education
Advanced General Certificate of Education

MEI STRUCTURED MATHEMATICS 2616

Statistics 4

Monday **21 JUNE 2004** 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 The random variables X and Y have independent Normal distributions as follows.

$$X \sim \mathrm{N}(\mu, \sigma^2) \qquad Y \sim \mathrm{N}(2\mu, 4\sigma^2).$$

It is proposed to estimate μ using an estimator

$$T = aX + bY$$

where a and b are constants.

(i) Show that T will be an unbiased estimator of μ if $b = \frac{1}{2}(1 - a)$. Find the variance of T in this case. [5]

(ii) Show that the variance found in part **(i)** is minimised when $a = \frac{1}{2}$, and state the distribution of T in this case. Explain why this indicates that, in this case, T would be better than either X or $\frac{1}{2}Y$ as an estimator of μ. [8]

(iii) A random sample of 10 values of T for the case $a = \frac{1}{2}$ in part **(ii)** is obtained. The values are as follows.

8.6 9.8 7.4 7.7 7.3 5.9 6.3 7.5 6.6 7.7

Taking the value of σ^2 to be 3, use this sample to obtain a one-sided 95% confidence interval giving a lower confidence bound for μ. [7]

2 Engineers are testing a new composite material for making aircraft wings. Some test pieces have been made by each of two experimental processes in the research laboratory. Their strengths are carefully measured on specially designed apparatus and are found to be as follows, in a convenient unit.

Process A	237	249	213	233	227	236
Process B	203	222	214	216	230	

It is not yet known whether the underlying distributions of these strengths may be taken as Normal, though it appears that they are of the same shape and symmetrical.

(i) Use a non-parametric procedure to test, at the 5% significance level, whether the median strengths from the two processes may be assumed equal. [8]

(ii) Stating carefully the required assumption concerning underlying Normality, use a t test to examine, at the 5% significance level, whether the mean strengths from the two processes may be assumed equal. [8]

(iii) Discuss briefly the advantages and disadvantages of the t procedure relative to the non-parametric procedure. [4]

3 A psychologist is studying the effect of a particular type of therapy on a phobic reaction. A random sample of 10 patients is available. For each patient, the intensity of the phobic reaction is measured, on a suitable scale, before and after the therapy. The results are as follows.

Patient	Intensity before therapy	Intensity after therapy
A	72	66
B	59	40
C	45	58
D	87	56
E	37	15
F	64	66
G	7	15
H	75	31
I	14	3
J	50	36

(i) Use an appropriate t procedure to test, at the 5% level of significance, whether the therapy on the whole reduces the intensity of the reaction, stating carefully your null and alternative hypotheses and the required distributional assumption. [14]

(ii) Provide a two-sided 95% confidence interval for the true mean difference in intensity of reaction before and after the therapy. [4]

(iii) If the distributional assumption in part **(i)** were not satisfied, name a procedure that might be used for carrying out the test. [2]

4 As part of a survey of interest in local elections, a random sample of 845 people was taken in towns which did not have directly-elected mayors. The people were classified according to age (< 30 or $\geqslant 30$) and their stated levels of interest (great or little) in local elections. The results were as follows.

		Level of interest	
		Great	Little
Age	< 30	49	216
	$\geqslant 30$	145	435

(i) Carry out the usual χ^2 test for independence, at the 5% significance level, stating carefully the null and alternative hypotheses and briefly discussing the conclusions. [12]

At a later stage in the survey, a random sample of 1327 people was taken in towns which had directly-elected mayors. These people were classified similarly, with the following results.

		Level of interest	
		Great	Little
Age	< 30	118	314
	$\geqslant 30$	260	635

(ii) The organisation that commissioned the survey then asked whether, for people under the age of 30, the level of interest in local elections is independent of whether or not there is a directly-elected mayor. Using the data in the tables above, write down the 2×2 table, including its margins, to be analysed. [6]

(iii) Explain why the usual χ^2 test might not be appropriate for the 2×2 table you have written down in part (ii). [2]

Mark Scheme

General 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 (‘explanation’) are for explanation and/or interpretation. These will frequently be subdividable 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 marking should have been resolved at the standardising 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*)
	Work of no value

Q1 $X \sim N(\mu, \sigma^2)$, $Y \sim N(2\mu, 4\sigma^2)$; $T = aX + bY$

(i) We want $\mu = E[aX + bY]$ **M1**

$= a.\mu + b.2\mu$ **1**

$\therefore 2b = 1 - a$ i.e. $b = \frac{1}{2}(1 - a)$ **1** Beware printed answer

The Var(T) $= a^2\sigma^2 + \left\{\frac{1}{2}(1\text{-}a)\right\}^2 (4\sigma^2)$ **M1** Substitution of $b = \frac{1}{2}(1\text{-}a)$ reqd

$= \sigma^2\{a^2 + (1 - a)^2\}$

$= \sigma^2\{2a^2 - 2a + 1\}$ **1** **5**

(ii) Consider $\frac{d}{da}(2a^2 - 2a + 1) = 0$ **M1**

i.e. $0 = 4a - 2$ **1**

$\therefore a = \frac{1}{2}$ **1** Beware printed answer

Verification that this is a <u>mini</u>mum (e.g. trivially by $\frac{d^2}{da^2}$) **1**

$T = \frac{1}{2}X + \frac{1}{4}Y \sim N\left(\mu, \frac{1}{2}\sigma^2\right)$

2 if all three items are correct;
award **1** if any two are correct

[Both X and $\frac{1}{2}$Y are u. b. for m and both are Normally distributed – all of which is also true for T; but] T has <u>smaller variance</u> $\left[\text{Var}(X) = \sigma^2, \text{Var}\left(\frac{1}{2}Y\right) = \sigma^2\right]$

E2 **8**

(iii) $\bar{t} = 7.48$ **B1** FT if wrong

One-sided CI is given by

$7.48 - 1.645 \frac{\sqrt{\frac{1}{2} \times 3}}{\sqrt{10}}$ **M1** (use of $\frac{1}{2}\sigma^2$ as Var(T))

M1 M1 B1 **M1** $= 7.48 - 0.63(71)$

$= 6.84(29)$ **A1** C.A.O. **7**

Q2

A	237	249	213	233	227	236
B	203	222	214	216	230	

(i) Wilcoxon rank sum test (or Mann-Whitney from thereof).

Ranks are

A	10	11	2	8	6	9	**M1** for attempt
B	1	5	3	4	7		**A1** if all correct

Rank sum is 20 (from B, otherwise the tables can't be used)
(Mann-Whitney is 5) **1**

Refer to tables of Wilcoxon rank sum (or Mann-Whitney) statistics. **1**
Lower $2\frac{1}{2}$ % tail is needed. **1**
Value for (5, 6) is 18 (or 3 for Mann-Whitney). **1**
Result is not significant. **1**
Seems medians are the same. **1** **8**

(ii) Normality of both underlying populations/distributions. **1**

$n_1 = 6 \quad \bar{x} = 232.5 \quad s_{n-1}^2 = 143.1 \left(s_{n-1} = 11.9624\right) \quad \left[s_n^2 = 119.25, s_n = 10.9202\right]$

$n_2 = 5 \quad \bar{y} = 217.0 \quad s_{n-1}^2 = 100.0 \left(s_{n-1} = 10.0\right) \quad \left[s_n^2 = 80.0, s_n = 8.9443\right]$

Pooled $s^2 = \frac{5\times 143.1 + 4\times 100.0}{9} = 123.9\dot{4}$ **M1** for any reasonable attempt at pooling (and FT into test)
A1 if correct

Test statistic is

M1 $\frac{232.5 - 217.0 (-0)}{\sqrt{123.9\dot{4}}\sqrt{\frac{1}{6}+\frac{1}{5}}} = \frac{15.5}{6.7414} = 2.29(92)$ **A1**

= 11.1330
FT reasonable attempt
Refer to t_9. **1** May be awarded even if test statistic is wrong. No FT if wrong.
Double-tailed 5% point is 2.262. **1** No FT if wrong
Significant, seems means differ. **1** **8**

(iii) If the assumptions for the t procedure are satisfied, it is 'better' (more sensitive/powerful), **E2**
but if not it might be seriously misleading and the non-par procedure safer. **E2** **4**

Q3 (i) $H_0 : \mu D = 0$ (or $\mu_{AFTER} = \mu_{BEFORE}$) **1**
$H_1 : \mu D < 0$ (or $\mu_{AFTER} < \mu_{BEFORE}$) **1**

Where μ_D is the population mean difference 'after – before' **1** for verbal def^n of μ
[**NOTE** – candidate might of course define D as 'before – after' – take core that H_1 agrees]

Requires Normality of population **1** of differences **1**.
must be clear, or clearly implied

The test procedure, and the CI in (ii), MUST be PAIRED COMPARISON t.
Differences are [as 'after – before', candidate might use 'before – after']

–6 –19 13 –31 –22 2 8 –44 –11 –14

$\bar{d} = -12.4$ $s_{n-1} = 17.621$ $\left(s_{n-1}^2 = 310.49\right)$ **A1** Accept $s_n = 16.716(5)$
$\left[s_n^2 = 279.44\right]$ ONLY if correctly used in sequel.

Test statistic is
$\frac{-12.4 - 0}{\frac{17.621}{\sqrt{10}}} = -2.22(535)$ **A1** **M1, M1, M1** (don't FT to 2nd M1)

Refer to t_9 **1** May be awarded even if test statistic is wrong. No FT if wrong

Lower s.t. 5% pt is –1.833 **1** Sign **must** agree with H_1/test statistic, unless a **clear** argument based on modulus is used. No FT if wrong.

Significant **1**. Seems mean afterwards is lower. **1** **14**

(ii) CI is given by
$-12.4 \pm 2.262 \frac{17.621}{\sqrt{10}} = -12.4 \pm 12.60(4) = (-25.00(4), 0.20(4))$
M1 **B1** **M1** **A1** c.a.o. **4**

Xero out of 4 if not same dist as for test. Some wrong dist can score max M1 B0 M1 A0. Recovery to t_9 is ok.

(iii) Any non-parametric procedure **1**
Paired Wilcoxon **1** [allow sign test] **2**

Q4 (i) H_0 : no association between age and level of interest. **B1**
H_1 : association between age and level of interest. **B1**

o_i

49	216	265
145	435	580
194	651	845

e_i

60.84	204.16	**A2**	Award **A1** if any one is correct. But **deduct 1** if not at least 2 dp
133.16	446.84		

$o_i - e_i = \pm 11.84$
or ± 11.34 with Yates' correction

$x^2 = 3.99(71)$ with Yates — **M1** for either, near-enough correct
$[4.35(73)$ without Yates$]$ — **A1** if Yates used

Refer to χ_1^2 **1** [FT if 2 or 3 df averred]
Upper 5% point is 3.84 **1**
Significant **1**
Seems there is association **1***
Seems under-30s have less interest than would be expected, and over-30s more, then if there were no association. **2***

* These 3 marks are not available if $H_0 \leftrightarrow H_1$

(ii)

		Level of interest: Great	Level of interest: Little	Total
Directly-elected mayor	Yes	118	314	432
	No	49	216	265
	Total	167	530	697

M1 for table with correctly labelled rows and columns.
M1 if **all** margins correctly add up from the individual values.
A1, **A1**, **A1**, **A1** for each individual cell (118, 314, 49, 216). **6**

(iii) We do not [at least prima facie] have a random sampler of 697 people who were classified over the 4 cells. The usual sample χ^2 approach requires such an assumption. **E2** **2**

Examiner’s Report

2616 Statistics 4

General Comments

Most candidates appeared to be well prepared for this examination and there was no evidence that candidates had insufficient time to complete the paper. In fact, some candidates gave full answers to all four questions.

As in previous years candidates performed much more strongly when carrying out the numerical parts of questions than they did when discussing assumptions or analysing results. The two most common examples of this weakness were firstly the assumptions required for the various t-tests to be valid – many candidates were not clear about whether parent populations, samples, means or data had to be normally distributed or whether they were looking at one distribution, two distributions or the difference between two distributions.

The second weakness was in the contextualisation of the results of a hypothesis test. Many candidates did not make any statement beyond "reject" H_0, whilst at the other end of the scale, candidates were too definitive, making statements such as "reject H_0, hence the median strength using process A is greater than the median strength using process B.

Once again, Question 1 on estimation was by far the least popular question. However most candidates who attempted question 1 scored well.

Comments on Individual Questions

Q.1 This question was only attempted by about 20% of candidates.

Virtually all candidates knew what they had to do in part (i) and were able to verify the value of *b*. Most were also able to calculate the variance of *T,* although poor algebra let down some candidates.

In part (ii) most candidates used calculus to show that the variance was minimised when *a* = 0.5, although some showed only that the variance had a stationary value. A few candidates used a method involving completing the square.

Candidates who got this far were almost all able to state the distribution of *T* and explain why it was a better estimator of *µ* than either *X* or ½*Y.*

Most candidates who attempted part (iii) knew what they were doing but a number failed to realise that Var(*T*) = $\frac{1}{2}\sigma^2$ and a number also did not realise that because the value of σ^2 was known, the normal distribution should be used – indeed one candidate used specifically because the sample was small.

Q.2 This was the most popular question on the paper, being attempted by all but 2 candidates.

Part (i) was obviously familiar ground for most candidates and most scored very well here. The method of choice for most candidates was to calculate the

Wilcoxon rank sum statistic, covert to the Mann-Whitney statistic and then use the Mann-Whitney tables. Only a small minority of candidates calculated a statistic (Wilcoxon or Mann-Whitney) and then moved directly to the relevant statistical table. However, this part of the question was answered better than any other part of the paper.

Part (ii) was not answered as well with many candidates not realising that Normality of both underlying populations was required. The pooled variance also caused some confusion with some candidates trying to pool standard deviations, some adding variances and others being confused about the use of s_n^2 and/or s_{n-1}^2.

Once a variance had been obtained, most candidates were then able to calculate the test statistic correctly and compared it with the two-tailed value of t_9.

In both parts (i) and (ii) a significant number of candidates were too definitive in their interpretation of the rejection, or otherwise, of the null hypothesis.

Answers to part (iii) tended to be too vague with very few candidates mentioning the fact that the t-test is a more powerful, or sensitive, test than the non-parametric alternatives, as long as the assumptions are satisfied. However, if the assumptions are not satisfied, results can be seriously misleading.

Q.3 In part (i) many candidates lost a significant number of marks because they did not carefully state their hypotheses or take sufficient care with the distributional assumption. Hypotheses such as “ the intensity remains the same” and “the intensity reduces” were common. What is required are explicit statements about either the mean of the population of differences, or about the means of the populations before and after. In addition all terms used should be defined. The required distributional assumption was the Normality of the population of differences.

As with other questions, most candidates were able to carry out the calculations competently and most used the correct value of *t.*

Part (ii) was very well done by the majority of candidates, although a few did use the Normal distribution.

Virtually all candidates correctly named the paired Wilcoxon test in part (iii)

Q.4 Most candidates were obviously on comfortable ground here and tended to score well.

In part (i) most candidates were able to state the hypotheses correctly, although some got the hypotheses the wrong way round and some talked about correlation.

Calculations were inevitably done correctly, but a few candidates only gave the expected values to 1 decimal place or even to the nearest integer.

Many candidates obviously realised that it would be appropriate to use Yates' correction, but few actually did. Of those that did, some were unsure whether to add or subtract 0.5.

Most candidates correctly used 1 degree of freedom for the χ^2 test and were able to give the correct critical value. A small minority used 2 or 3 degrees of freedom.

There was a definite improvement on previous years in the discussion of the results of the hypothesis test, with many candidates considering the contributions to the χ^2 statistic, or at the very least considering the differences between observed and expected values.

Most candidates scored full marks in part (ii)

Candidates struggled with part (iii), with the most common suggestion being about different sample sizes. The actual reason was that we do not have a random sample of people who were classified over the 4 cells.