OXFORD CAMBRIDGE AND RSA EXAMINATIONS

Advanced Subsidiary General Certificate of Education
Advanced General Certificate of Education

MEI STRUCTURED MATHEMATICS 2616

Statistics 4

Monday **20 JUNE 2005** 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 $X_1, X_2, \ldots, X_n$ $(n > 1)$ are independent random variables all having the same Normal distribution. The common variance is denoted by σ^2.

The random variable Y is defined by

$$Y = \sum_{i=1}^{n} (X_i - \bar{X})^2 \qquad \text{where } \bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i .$$

You may in this question *use* the results

$$\mathrm{E}(Y) = (n-1)\sigma^2,$$
$$\mathrm{Var}(Y) = 2(n-1)\sigma^4 .$$

(i) It is proposed to estimate σ^2 by an estimator of the form $T = kY$ where k is a constant. Write down the mean and variance of T. [2]

(ii) Hence write down an expression for the bias of T as an estimator of σ^2. [2]

(iii) Show that the mean square error of T as an estimator of σ^2 is

$$\mathrm{MSE}(T) = \sigma^4\left[\left\{2(n-1) + (n-1)^2\right\}k^2 - 2(n-1)k + 1\right].$$ [4]

(iv) Hence, using calculus or otherwise, show that $k = \dfrac{1}{n+1}$ gives the minimum value of MSE(T). [6]

(v) Find the minimum value of MSE(T). [2]

(vi) Deduce from the answer to part **(ii)** that T is an unbiased estimator of σ^2 only if $k = \dfrac{1}{n-1}$. Find the value of MSE(T) in this case. [4]

2 Research and development engineers are studying the strengths of steel rods made by two processes, A and B. The strengths are measured, in a conventional unit, for a random sample of 9 rods made by process A and a random sample of 8 rods made, independently, by process B. These strengths are found to be as follows.

Process A	111	117	99	113	132	107	116	123	114
Process B	134	122	106	123	114	125	139	127	

(i) State the null and alternative hypotheses and the required assumptions for the usual t test that would be applied to these data. [4]

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

(iii) Obtain a two-sided 99% confidence interval for the mean difference between the strengths. [4]

(iv) Name the non-parametric test that could be used as an alternative method of analysing these data to investigate the strengths. [2]

3 At an agricultural research station, a trial is conducted to see whether use of an experimental fertiliser has any effect on the yield of carrots compared with using a standard fertiliser. 10 plots are available for the trial. They are of the same size but are known to differ from each other in terms of natural soil fertility. Each plot is divided into two halves, and in one half, chosen at random, carrots are grown using the standard fertiliser; in the other half, carrots are grown using the experimental fertiliser. All other conditions in the trial are controlled carefully. The yields, in kg per half-plot, are as follows.

Plot	1	2	3	4	5	6	7	8	9	10
Standard fertiliser	20.2	22.0	17.8	20.6	26.8	20.9	21.2	16.5	20.8	12.9
Experimental fertiliser	20.8	24.3	17.0	21.2	27.7	19.4	22.6	17.3	20.9	13.1

(a) Use an appropriate t test to examine, at the 5% level of significance, whether the mean yields of carrots with the experimental and standard fertilisers may be assumed equal, stating carefully your null and alternative hypotheses and the required distributional assumption. [11]

(b) Using the data for the experimental fertiliser in the table above, provide a one-sided 95% confidence interval giving a lower confidence bound for the mean yield, stating the required distributional assumption. Interpret the confidence interval carefully. [9]

4 Managers at a busy international airport are studying the times taken by arriving passengers to collect their luggage and pass through passport control and customs.

(a) For a random sample of 12 arriving passengers, the times in minutes are as follows, arranged in ascending order.

29 32 34 38 40 46 51 52 59 63 71 95

Use a Wilcoxon test to examine, at the 5% level of significance, whether it is reasonable to assume that the population median time is 60 minutes. [9]

(b) An extensive survey is made, using a random sample of 200 arriving passengers. Their times, t minutes, are recorded in a frequency table as shown below. The mean and standard deviation of the data are 62.0 and 27.3 minutes, respectively. The corresponding expected frequencies given by the Normal distribution with $\mu = 62.0$ and $\sigma = 27.3$ are also shown in the table.

Time (t minutes)	Observed frequency	Frequency as given by Normal model
$t \leqslant 20$	4	12.4
$20 < t \leqslant 40$	32	29.6
$40 < t \leqslant 60$	83	52.1
$60 < t \leqslant 80$	38	54.9
$80 < t \leqslant 100$	17	34.6
$100 < t \leqslant 120$	14	13.0
$120 < t$	12	3.4

(i) Verify that the expected frequency for the interval 80 – 100 minutes is 34.6. [2]

(ii) Test the fit of the Normal model to these data and comment on your findings. [6]

(iii) Comment on whether a t test might have been used for the analysis in part **(a)**. [3]

Mark Scheme 2616
June 2005

2616 Statistics 4

Q1	$X_1, \ldots, X_n \sim$ ind $N(\mu, \sigma^2)$ $\quad Y = \sum (X_i - \bar{X})^2$ $E(Y) = (n-1)\sigma^2 \quad Var(Y) = 2(n-1)\sigma^4 \quad T = kY$			
(i)	$E(T) = k(n-1)\sigma^2$ $Var(T) = 2k^2(n-1)\sigma^4$	B1 B1		2
(ii)	Bias = $E(T) - \sigma^2$ $= k(n-1)\sigma^2 - \sigma^2$	M1 A1	Allow M1A0 if $\sigma^2 - E(T)$.	2
(iii)	MSE(T) = Variance + bias^2 $= 2k^2(n-1)\sigma^4 + \{k(n-1)\sigma^2 - \sigma^2\}^2$ $= 2k^2(n-1)\sigma^4 + \{k^2(n-1)^2 - 2k(n-1) + 1\}\sigma^4$ $= \sigma^4[2(n-1) + (n-1)^2]k^2 - 2\sigma^4(n-1)k + \sigma^4$	M1 A1 A2	If both terms present, even if wrong. If both correct. Divisible for algebra. BEWARE printed answer.	4
(iv)	Consider $\frac{d\,MSE(T)}{dk} = 0$ $\frac{d\,MSE(T)}{dk} = \sigma^4\left[2(n-1) + (n-1)^2\right]2k - 2\sigma^4(n-1)$ $= 0 \quad \rightarrow \quad k = \frac{n-1}{2(n-1) + (n-1)^2}$ $= \frac{1}{n+1}$ Check minimum by considering $\frac{d^2\,MSE(T)}{dk^2} = \sigma^4\left[2(n-1) + (n-1)^2\right]2$ $> 0 \quad \therefore$ min	M1 A1 A1 A1 M1 A1	To include "=0", possibly implied. Correct derivative. Isolate k. BEWARE printed answer. Or other methods. (Since $n > 1$).	6
(v)	With $k = \frac{1}{n+1}$, $MSE(T) = \sigma^4\left\{\frac{2(n-1) + (n-1)^2}{(n+1)^2} - \frac{2(n-1)}{n+1} + 1\right\}$ $= \frac{\sigma^4}{(n+1)^2}\left\{2n - 2 + n^2 - 2n + 1 - 2n^2 + 2 + n^2 + 2n + 1\right\}$ $= \frac{\sigma^4}{(n+1)^2}\{2n+2\} = \frac{2\sigma^4}{n+1}$	B2	Divisible for algebra. Answer <u>not</u> printed.	2
(vi)	From (ii), we want $k(n-1)\sigma^2 - \sigma^2 = 0$ $\Rightarrow k = \frac{1}{n-1}$ In this case, MSE(T) = Var(T) $= \frac{2\sigma^4}{n-1}$	M1 A1 M1 A1	For the converse argument, with no support of "only if", award SC B1. Or substitute in expression for MSE in (iii) – this is not difficult.	4
				20

Q2				
(i)	$H_0 : \mu_A = \mu_B \qquad H_1 : \mu_A \neq \mu_B$	B1	Both hypotheses. Do **not** allow any other symbols, including, e.g., $\overline{X}_A = \overline{X}_B$ or similar, unless they are clearly and explicitly stated to be population means. Allow statements in words (see below).	
	Where μ_A, μ_B are the population mean strengths for processes A and B.	B1	For adequate verbal definitions of μ_A, μ_B. Must indicate "mean"; condone "average". Allow absence of "population" if correct notation μ is used, otherwise insist on "population".	
	Normality of both populations. Same variance.	B1 B1		4
(ii)	$n_1 = 9, \bar{x} = 114 \cdot 6667, s_{n-1}{}^2 = 87 \cdot 25, (s_{n-1} = 9 \cdot 3408)$ $n_2 = 8, \bar{y} = 123 \cdot 75, s_{n-1}{}^2 = 109 \cdot 07, (s_{n-1} = 10 \cdot 4437)$	B1	If all means and variances correct. Accept s_n's ONLY if correctly used in sequel. $s_n{}^2 = 77 \cdot \dot{5}, \quad s_n = 8 \cdot 8066$ $s_n{}^2 = 95 \cdot 4375, \quad s_n = 9 \cdot 7692$	
	Pooled $s^2 = \dfrac{698 + 763 \cdot 5}{15} = 97 \cdot 4\dot{3}$	M1 A1	For any reasonable attempt at pooling (and ft into test and CI). If correct.	
	Test statistic is $\dfrac{114 \cdot 6667 - 123 \cdot 75}{\sqrt{97 \cdot 4\dot{3}}\sqrt{\frac{1}{9} + \frac{1}{8}}} = \dfrac{-9 \cdot 0833}{\sqrt{23 \cdot 0051} = 4 \cdot 7964}$ $= -1 \cdot 89(38)$	M1 M1 A1	Overall structure. Allow c's pooled s. $\sqrt{\frac{1}{9} + \frac{1}{8}}$ ft c's pooled s^2.	
	Refer to t_{15}. Double tail 5% point is 2·131. Not significant. Seems mean strengths are the same for both processes.	M1 A1 E1 E1	No ft from here if wrong. No ft from here if wrong. ft only c's test statistic. ft only c's test statistic. Expect reference to means and context.	10
(iii)	CI is given by –9·0833 ± 2·947 × 4·7964 = –9·0833 ± 14·1349 = (–23·21(8), 5·05(2))	M1 B1 M1 A1	Must be c's $(\bar{x} - \bar{y}) \pm \ldots$ From t_{15}. Allow c's pooled s. c.a.o. Must be written as an interval.	4
(iv)	Wilcoxon Rank sum test	B1 B1	Or Mann-Whitney scores B2.	2
				20

Q3 (a)	$H_0 : \mu_D = 0$ or $\mu_E = \mu_S$ $H_1 : \mu_D \neq 0$ or $\mu_E \neq \mu_S$	B1	Both hypotheses. Do **not** allow any other symbols, including, e.g., $\overline{X}_E = \overline{X}_S$ or similar, unless they are clearly and explicitly stated to be population means. Allow statements in words (see below).	
	Where μ_D is "population mean for Experimental fertilizer – population mean for Standard fertilizer".	B1	For adequate verbal definition of μ. Must indicate "mean"; condone "average". Allow absence of "population" if correct notation μ is used, otherwise insist on "population".	
	Normality of differences is required.	B1	Must be explicit about the population.	
	MUST be PAIRED COMPARISON t test. Differences are 0·6 2·3 –0·8 0·6 0·9 –1·5 1·4 0·8 0·1 0·2	M1		
	$\bar{d} = 0 \cdot 46,\ s_{n-1} = 1 \cdot 0668(75),\ s_{n-1}{}^2 = 1 \cdot 1382$	B1	Accept $s_n = 1 \cdot 0121,\ s_n{}^2 = 1.0244$ ONLY if correctly used in sequel.	
	Test statistic is $\dfrac{0 \cdot 46 - 0}{\left(\dfrac{1 \cdot 0668(75)}{\sqrt{10}}\right)}$	M1	Allow c's $\bar{d}$ and/or s_{n-1}. Allow alternative: 0 (c's 2·262) × $\dfrac{1 \cdot 0668(75)}{\sqrt{10}}$ (= ±0·7631) for subsequent comparison with $\bar{d}$. (Or $\bar{d}$ ± (c's 2·262) × $\dfrac{1 \cdot 0668(75)}{\sqrt{10}}$ (= –0·303, 1·2231) for comparison with 0.)	
	= 1·36(35)	A1	c.a.o. (but ft from here if this is wrong.) Use of $\mu_D - \bar{d}$ scores M1A0, but next 4 marks still available.	
	Refer to t_9.	M1	No ft from here if wrong.	
	Double tail 5% point is 2·262.	A1	No ft from here if wrong.	
	Not significant.	E1	ft only c's test statistic.	
	Seems mean yield using experimental fertilizer is same as for standard.	E1	ft only c's test statistic. Expect reference to mean(s) and context.	11
(b)	Now need Normality for yields using experimental fertilizer.	B1		
	For these yields, $\bar{x} = 20 \cdot 43,\ s_{n-1} = 4 \cdot 0803,\ s_{n-1}{}^2 = 16 \cdot 649$	B1	Accept $s_n = 3 \cdot 8709,\ s_n{}^2 = 14.9841$ ONLY if correctly used in sequel.	
	One-sided CI (lower confidence bound) is given by 20·43	M1	Mean. Allow c's $\bar{x}$.	
	–	M1	Minus.	
	1·833	B1	From t_9.	
	$\times \dfrac{4 \cdot 0803}{\sqrt{10}}$	M1	Allow c's s_{n-1}, or $s_n / \sqrt{9}$ (see above).	
	= 20·43 – 2·36(51) = 18·06(49)	A1	Depends on all 4 preceding marks.	
	In repeated sampling, lower confidence bounds obtained in this way would fall below the true mean on 95% of occasions.	E2	(E0, E1, E2). Comment should refer to lower bound rather than just the confidence interval.	9
				20

Q4				
(a)	Data 29 32 34 38 40 46 51 52 59 63 71 95 Median 60 Difference –31 –28 –26 –22 –20 –14 –9 –8 –1 3 11 35 Rank of |diff| 11 10 9 8 7 6 4 3 1 2 5 12			
		M1	For differences. ZERO in this section if differences not used.	
		M1	For ranks of |difference|.	
		A1	All correct.	
	$T = 2 + 5 + 12 = \underline{19}$	B1	ft from here if ranks wrong. Or 1 + 3 + 4 + 6 + 7 + 8 + 9 + 10 + 11 = 59	
	Refer to tables of Wilcoxon single sample (/paired) statistic.	M1	No ft from here if wrong.	
	Lower (or upper if 59 used) 2½% tail is needed.	M1	No ft from here if wrong.	
	Value for n = 12 is <u>13</u> (or 65 if 59 used).	A1	No ft from here if wrong.	
	Result is not significant.	E1	ft only c's test statistic.	
	No real evidence that median is not 60.	E1	ft only c's test statistic.	9
(b) (i)	$P(80 < N(62, \sigma = 27\cdot3) \le 100)$ $= P(0\cdot6593(4) < N(0, 1) \le 1\cdot3919(4))$ $= 0\cdot9180 - 0\cdot7452 = 0\cdot1728$	B1		
	$\therefore$ expected frequency = 200 × 0·1728 = 34·6	B1	BEWARE printed answer.	2
(ii)	Grouping the last two cells, X^2 = 5·6903 + 0·1946 + 18·3265 + 5·2024 + 8 9526 + 5·6195	M1	Allow without grouping. This becomes … + 0·0769 + 21·7529.	
	= 43.98(59)	A1	X^2 becomes 60·19(62). Then <u>must</u> have χ^2_4 below.	
	Refer to χ^2_3.	M1	<u>NEXT</u> mark not available if not χ^2_3.	
	Extremely highly significant – overwhelming evidence that Normal model does not fit data.	A1	ft only c's test statistic.	
	The fit is not particularly good in most of the intervals, but the main points are that the modal class is perhaps "half an interval lower" than expected, that there are many fewer low values than expected, and that there a lot of upper outliers.	E2	(E0, E1, E2)	6
(iii)	Part (a) has a small sample and it appears that the underlying distribution is not Normal – could be dangerous to use a t test.	E2	(E0, E1, E2)	
	There is also the point that, in the absence of Normality (or at least of symmetry), we could not use the t test for the mean as a proxy test for the median.	E1		3
				20

2616 - Statistics 4

General Comments

There were 93 candidates from 20 centres (June 2004: 82 from 20). The overall standard of the scripts seen was pleasing: many candidates were clearly well prepared for this paper. Routine calculations were carried out well but the candidates' ability to comment and interpret were a little disappointing at this level.

Question 1 was by far the least popular question with only about 15 candidates attempting it. Every candidate attempted Question 2; Questions 3 and 4 were equally popular.

Comments on Individual Questions

1) **Estimation theory**

Although this was the least popular question it seemed to have the highest mean mark, with most of those attempting it scoring full or almost full marks. Those who were prepared to try it were likely to be successful as long as their algebra was up to the task. Sometimes the algebra arrived at the correct destination by brute force rather than elegance.
There were just two places where marks seemed likely to be lost: part (iv) where some neglected to verify that the required value of k did indeed give a minimum and part (vi) where there was a temptation for some to use the converse argument.

2) **Two sample *t* test and confidence interval; the strengths of steel rods**

This was the most popular question being attempted by all candidates. It was also a very high scoring question: about half of the entry scored full or almost full marks.

(i) The hypotheses were usually stated correctly but there was rather less care in providing verbal definitions of the population means. Similarly, the required assumptions were sometimes less than ideal.

(ii) Most candidates carried out the test competently. There was rarely any problem over finding and using the pooled variance. The critical value was almost always correct but on a number of occasions the conclusion was badly expressed.

(iii) As in part (ii) most candidates had little difficulty here. Just occasionally the standard error (which had been correctly constructed in part (ii)) became "pooled $s \times \frac{1}{\sqrt{17}}$".

(iv) This part was almost always correct.

3) **Paired sample *t* test and one-sided confidence interval; comparing fertilizers**

(a) The hypotheses were usually stated correctly but candidates were not as careful about defining the symbol μ. Nor were they sufficiently careful when it came to the distributional assumption.
However there were only a very few candidates who did not realise that they should carry out a paired test. The vast majority made good progress with the test itself, and only the final conclusion left room for improvement.

(b) As above, most realised what to do here and the correct value for the lower bound was usually found. A small minority tried to construct the confidence interval using the information from the paired test. There was some uncertainty again with the distributional assumption.
The main area of difficulty was with the interpretation of the interval. Very many comments revealed a flawed understanding of a confidence interval to quite a worrying extent.

4) **Wilcoxon rank sum test for the median; Chi-squared test for goodness of fit; waiting times in an airport**

(a) This part of the question was almost always answered well. Many fully correct solutions were seen.

(b) (i) This part was frequently done correctly.

(ii) Most candidates calculated a correct value of X^2 (with or without grouping) but relatively few were able to identify the correct Chi-squared distribution to look up. Most of those who got this second aspect wrong made no allowance for estimated parameters while a few thought that there were 200 degrees of freedom. Hardly any commented on the fact that the test statistic was significant at **any** level available to them in the tables.
Disappointingly few candidates took the trouble to comment at all on the reasons for the poor quality of fit.

(iii) In this part of the question very few candidates realised that they could refer back to the previous part for evidence that the assumption of background Normality was not viable. They knew that Normality was required, but often chose to look at the sample data in part (a), sometimes with the aid of a dot plot. Hardly any candidates included in their discussion the small sample size which might prompt the use of a *t* test.
No more than a handful of candidates picked up on the fact that a *t* test examines the population mean whereas the Wilcoxon test in part (a) examined the median.