RECOGNISING ACHIEVEMENT

OXFORD CAMBRIDGE AND RSA EXAMINATIONS

Advanced Subsidiary General Certificate of Education
Advanced General Certificate of Education

MEI STRUCTURED MATHEMATICS 2616

Statistics 4

Tuesday **10 JANUARY 2006** Afternoon 1 hour 20 minutes

Additional materials:
8 page 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.
- 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 60.

This question paper consists of 5 printed pages and 3 blank pages.

1 The lengths of offcuts of timber at a builders' merchant are modelled by the random variable X with the rectangular (continuous uniform) distribution having probability density function

$$f(x) = \frac{1}{\theta}, \qquad 0 \leqslant x \leqslant \theta,$$

where $\theta\ (\theta > 0)$ is an unknown parameter. You may use the results

$$E(X) = \frac{\theta}{2},$$
$$\mathrm{Var}(X) = \frac{\theta^2}{12}.$$

(i) The lengths of a random sample of n offcuts are given by the independent random variables $X_1, X_2, \ldots, X_n$ where each X_i is distributed as X. Explain why

$$\mathrm{P}(\text{length of longest offcut} \leqslant y) = \left(\frac{y}{\theta}\right)^n,$$

and deduce that the probability density function of the sample maximum, denoted here by $Y\ (Y = \max(X_1, X_2, \ldots, X_n))$, is

$$g(y) = \frac{ny^{n-1}}{\theta^n}$$

for $0 \leqslant y \leqslant \theta$. [5]

(ii) Show that

$$E(Y) = \frac{n\theta}{n+1}$$

and

$$\mathrm{Var}(Y) = \frac{n\theta^2}{(n+1)^2(n+2)}.$$ [6]

(iii) Deduce the multiple of Y that is an unbiased estimator of θ, and give its variance. [3]

(iv) Show that $2\overline{X}$ is also an unbiased estimator of θ, and write down its variance. [3]

(v) Explain why the estimator in part **(iii)** should be preferred to $2\overline{X}$ for each fixed $n > 1$. [3]

2 **(i)** Engineers are testing the capacities of two types of industrial pump. This capacity is measured for a random sample of 50 pumps of type P and it is found that the sample mean, measured in a conventional unit, is 26.2. The capacity is also measured for a random sample of 40 pumps of type Q and it is found that the sample mean, measured in the same unit, is 28.1. The standard deviations of the underlying populations of capacities are known from long experience to be 4.6 for pumps of type P and 5.4 for pumps of type Q.

Test at the 5% level of significance whether there is evidence that the true mean capacities for the two types of pump differ, stating carefully the null and alternative hypotheses you are testing. Provide a two-sided 90% confidence interval for the true mean difference. [12]

(ii) Explain why no distributional assumption for the underlying populations of capacities is needed for your analysis in part **(i)**. [2]

(iii) Suppose that the information given in part **(i)** is changed so that the quoted standard deviations are the sample values (using divisor $n - 1$) for the two samples, but with no other changes. Explain why the analysis in part **(i)** is still valid. [2]

(iv) Suppose that all the **original** information in part **(i)** is still available **except that** the samples are much smaller (say of size about 10). Further, suppose that it is known that the underlying populations of capacities are Normally distributed. Explain why an analysis of the same kind as in part **(i)** can still be applied. [2]

(v) Suppose that the situation is as in part **(iv)** **except that** the underlying population standard deviations are not known but are assumed to be equal. What test can be applied if the sample standard deviations are known? [1]

(vi) Suppose that the situation is as in part **(v)** **except that** it is known that the underlying populations of capacities are not Normally distributed. What test can be applied if the capacity measurements for all the pumps in the samples are known? [1]

3 Two chefs are taking part in a competition. A set of ten menus, $A, B, \ldots, J$, is given to each chef, and each is asked to prepare meals according to each menu. The meals are rated by a tasting panel, an overall quality score being given to each. The scores are as follows.

Menu	A	B	C	D	E	F	G	H	I	J
Chef 1	63	68	72	77	79	62	63	60	52	70
Chef 2	61	74	76	72	70	73	80	76	65	88

These scores may be taken as random samples from appropriate underlying populations.

(i) It is decided to undertake a paired analysis using the customary Wilcoxon procedure at the 10% level of significance. Carry out this test, stating the null and alternative hypotheses and your conclusion. [10]

(ii) Find the actual level of significance of the data as given by the Normal approximation to the distribution of the Wilcoxon statistic under the null hypothesis. Comment on your answer in relation to the result of the test in part **(i)**. [10]

[You are reminded that the parameters of the Normal approximation are

$$\mu = \frac{n(n+1)}{4} \text{ and } \sigma^2 = \frac{n(n+1)(2n+1)}{24}.]$$

4 A statistician is investigating whether a reasonable model for a particular population is the random variable X having the exponential distribution with parameter λ ($\lambda > 0$), i.e. having the probability density function

$$f(x) = \lambda e^{-\lambda x}, \qquad x \geqslant 0.$$

(i) Show that $P(a \leqslant X < b) = e^{-\lambda a} - e^{-\lambda b}$ where $b > a \geqslant 0$. [2]

(ii) Use the fact that $E(X) = \dfrac{1}{\lambda}$ to explain why a reasonable estimate of λ is $\hat{\lambda} = \dfrac{1}{\bar{x}}$ where $\bar{x}$ is the mean of a random sample of observations from X. [2]

(iii) A random sample of 60 observations is available, and is recorded as the following frequency distribution.

Range	$0 \leqslant x < 5$	$5 \leqslant x < 10$	$10 \leqslant x < 15$	$15 \leqslant x < 20$
Frequency	14	16	16	14

Use the value of $\bar{x}$ given by this frequency distribution to obtain the value $\frac{1}{10}$ for $\hat{\lambda}$. [1]

(iv) Using this value of $\hat{\lambda}$, set up the appropriate table of estimated expected frequencies. [6]

(v) Use the customary χ^2 test at the 5% level of significance to show that the model X appears not to fit the data well. Discuss briefly the principal differences between the model and the data. [9]

BLANK PAGE

BLANK PAGE

BLANK PAGE