OCR

RECOGNISING ACHIEVEMENT

OXFORD CAMBRIDGE AND RSA EXAMINATIONS

Advanced Subsidiary General Certificate of Education
Advanced General Certificate of Education

MEI STRUCTURED MATHEMATICS   **4769**

Statistics 4

Wednesday   **24 MAY 2006**   Afternoon   1 hour 30 minutes

Additional materials:
     8 page answer booklet
     Graph paper
     MEI Examination Formulae and Tables (MF2)

TIME   1 hour 30 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 72.

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

2

Option 1: Estimation

1 A parcel is weighed, independently, on two scales. The weights are given by the random variables W_1 and W_2 which have underlying Normal distributions as follows.

$$W_1 \sim N(\mu, \sigma_1^2), \qquad W_2 \sim N(\mu, \sigma_2^2),$$

where μ is an unknown parameter and σ_1^2 and σ_2^2 are taken as known.

  (i) Show that the maximum likelihood estimator of μ is

$$\hat{\mu} = \frac{\sigma_2^2}{\sigma_1^2 + \sigma_2^2}\, W_1 + \frac{\sigma_1^2}{\sigma_1^2 + \sigma_2^2}\, W_2. \qquad [11]$$

    [You may quote the probability density function of the general Normal distribution from page 9 in the MEI Examination Formulae and Tables Booklet (MF2).]

  (ii) Show that $\hat{\mu}$ is an unbiased estimator of μ. [2]

  (iii) Obtain the variance of $\hat{\mu}$. [2]

  (iv) A simpler estimator $T = \frac{1}{2}(W_1 + W_2)$ is proposed. Write down the variance of T and hence show that the relative efficiency of T with respect to $\hat{\mu}$ is

$$y = \left(\frac{2\sigma_1 \sigma_2}{\sigma_1^2 + \sigma_2^2} \right)^2. \qquad [5]$$

  (v) Show that $y \leqslant 1$ for all values of σ_1^2 and σ_2^2. Explain why this means that $\hat{\mu}$ is preferable to T as an estimator of μ. [4]

3

Option 2: Generating Functions

2   [In this question, you may use the result $\int_0^\infty u^m e^{-u} \, du = m!$ for any non-negative integer m.]

The random variable X has probability density function

$$f(x) = \begin{cases} \dfrac{\lambda^{k+1} x^k e^{-\lambda x}}{k!}, & x > 0, \\ 0, & \text{elsewhere,} \end{cases}$$

where $\lambda > 0$ and k is a non-negative integer.

   (i) Show that the moment generating function of X is $\left(\dfrac{\lambda}{\lambda - \theta} \right)^{k+1}$. [7]

   (ii) The random variable Y is the sum of n independent random variables each distributed as X. Find the moment generating function of Y and hence obtain the mean and variance of Y. [8]

   (iii) State the probability density function of Y. [3]

   (iv) For the case $\lambda = 1$, $k = 2$ and $n = 5$, it may be shown that the definite integral of the probability density function of Y between limits 10 and ∞ is 0.9165. Calculate the corresponding probability that would be given by a Normal approximation and comment briefly. [6]

[Turn over

4

Option 3: Inference

3 The human resources department of a large company is investigating two methods, A and B, for training employees to carry out a certain complicated and intricate task.

  (i) Two separate random samples of employees who have not previously performed the task are taken. The first sample is of size 10; each of the employees in it is trained by method A. The second sample is of size 12; each of the employees in it is trained by method B. After completing the training, the time for each employee to carry out the task is measured, in controlled conditions. The times are as follows, in minutes.

     Employees trained by method A: 35.2   47.8   25.8   38.0   53.6   31.0   33.9
                                                  35.4   21.6   42.5

     Employees trained by method B: 43.0   57.5   68.6   20.9   31.4   44.9   62.8
                                                27.6   41.8   46.1   39.8   61.6

     Stating appropriate assumptions concerning the underlying populations, use a *t* test at the 5% significance level to examine whether either training method is better in respect of leading, on the whole, to a lower time to carry out the task.   **[12]**

 (ii) A further trial of method B is carried out to see if the performance of experienced and skilled workers can be improved by re-training them. A random sample of 8 such workers is taken. The times in minutes, under controlled conditions, for each worker to carry out the task before and after re-training are as follows.

Worker	W_1	W_2	W_3	W_4	W_5	W_6	W_7	W_8
Time before	32.6	28.5	22.9	27.6	34.9	28.8	34.2	31.3
Time after	26.2	24.1	19.0	28.6	29.3	20.0	36.0	19.2

     Stating an appropriate assumption, use a *t* test at the 5% significance level to examine whether the re-training appears, on the whole, to lead to a lower time to carry out the task.   **[10]**

(iii) Explain how the test procedure in part **(ii)** is enhanced by designing it as a paired comparison.
                                                              [2]

Option 4: Design and Analysis of Experiments

4 An experiment is carried out to compare five industrial paints, A, B, C, D, E, that are intended to be used to protect exterior surfaces in polluted urban environments. Five different types of surface (I, II, III, IV, V) are to be used in the experiment, and five specimens of each type of surface are available. Five different external locations (1, 2, 3, 4, 5) are used in the experiment.

The paints are applied to the specimens of the surfaces which are then left in the locations for a period of six months. At the end of this period, a "score" is given to indicate how effective the paint has been in protecting the surface.

  (i) Name a suitable experimental design for this trial and give an example of an experimental layout. [3]

Initial analysis of the data indicates that any differences between the types of surface are negligible, as also are any differences between the locations. It is therefore decided to analyse the data by one-way analysis of variance.

 (ii) State the usual model, including the accompanying distributional assumptions, for the one-way analysis of variance. Interpret the terms in the model. [9]

(iii) The data for analysis are as follows. Higher scores indicate better performance.

Paint A	Paint B	Paint C	Paint D	Paint E
64	66	59	65	64
58	68	56	78	52
73	76	69	69	56
60	70	60	72	61
67	71	63	71	58

[The sum of these data items is 1626 and the sum of their squares is 106 838.]

Construct the usual one-way analysis of variance table. Carry out the appropriate test, using a 5% significance level. Report briefly on your conclusions. [12]