RECOGNISING ACHIEVEMENT

ADVANCED GCE
MATHEMATICS (MEI) 4769
Statistics 4

Candidates answer on the Answer Booklet

OCR Supplied Materials:
- 8 page Answer Booklet
- Graph paper
- MEI Examination Formulae and Tables (MF2)

Other Materials Required:
- Scientific or graphical calculator

Friday 18 June 2010
Afternoon

Duration: 1 hour 30 minutes

INSTRUCTIONS TO CANDIDATES

- Write your name clearly in capital letters, your Centre Number and Candidate Number in the spaces provided on the Answer Booklet.
- Use black ink. Pencil may be used for graphs and diagrams only.
- Read each question carefully and make sure that you know what you have to do before starting your answer.
- Answer any **three** questions.
- Do **not** write in the bar codes.
- 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 document consists of **4** pages. Any blank pages are indicated.

Option 1: Estimation

1 The random variable X has probability density function

$$f(x) = \frac{xe^{-x/\lambda}}{\lambda^2} \qquad (x > 0),$$

where λ is a parameter ($\lambda > 0$). $X_1, X_2, \ldots, X_n$ are n independent observations on X, and $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ is their mean.

(i) Obtain $E(X)$ and deduce that $\hat{\lambda} = \frac{1}{2}\bar{X}$ is an unbiased estimator of λ. **[7]**

(ii) Obtain $\text{Var}(\hat{\lambda})$. **[7]**

(iii) Explain why the results in parts **(i)** and **(ii)** indicate that $\hat{\lambda}$ is a good estimator of λ in large samples. **[2]**

(iv) Suppose that $n = 3$ and consider the alternative estimator

$$\tilde{\lambda} = \tfrac{1}{8}X_1 + \tfrac{1}{4}X_2 + \tfrac{1}{8}X_3.$$

Show that $\tilde{\lambda}$ is an unbiased estimator of λ. Find the relative efficiency of $\tilde{\lambda}$ compared with $\hat{\lambda}$. Which estimator do you prefer in this case? **[8]**

Option 2: Generating Functions

2 The random variable X has the Poisson distribution with parameter λ.

(i) Show that the probability generating function of X is $G(t) = e^{\lambda(t-1)}$. **[3]**

(ii) Hence obtain the mean μ and variance σ^2 of X. **[5]**

(iii) Write down the mean and variance of the random variable $Z = \frac{X - \mu}{\sigma}$. **[2]**

(iv) Write down the moment generating function of X. State the linear transformation result for moment generating functions and use it to show that the moment generating function of Z is

$$M_Z(\theta) = e^{f(\theta)} \quad \text{where } f(\theta) = \lambda\left(e^{\theta/\sqrt{\lambda}} - \frac{\theta}{\sqrt{\lambda}} - 1\right).$$ **[7]**

(v) Show that the limit of $M_Z(\theta)$ as $\lambda \to \infty$ is $e^{\theta^2/2}$. **[4]**

(vi) Explain briefly why this implies that the distribution of Z tends to $N(0, 1)$ as $\lambda \to \infty$. What does this imply about the distribution of X as $\lambda \to \infty$? **[3]**

Option 3: Inference

3 At a factory, two production lines are in use for making steel rods. A critical dimension is the diameter of a rod. For the first production line, it is assumed from experience that the diameters are Normally distributed with standard deviation 1.2 mm. For the second production line, it is assumed from experience that the diameters are Normally distributed with standard deviation 1.4 mm. It is desired to test whether the mean diameters for the two production lines, μ_1 and μ_2, are equal. A random sample of 8 rods is taken from the first production line and, independently, a random sample of 10 rods is taken from the second production line.

(i) Find the acceptance region for the customary test based on the Normal distribution for the null hypothesis $\mu_1 = \mu_2$, against the alternative hypothesis $\mu_1 \neq \mu_2$, at the 5% level of significance. **[6]**

(ii) The sample means are found to be 25.8 mm and 24.4 mm respectively. What is the result of the test? Provide a two-sided 99% confidence interval for $\mu_1 - \mu_2$. **[7]**

The production lines are modified so that the diameters may be assumed to be of equal (but unknown) variance. However, they may no longer be Normally distributed. A two-sided test of the equality of the population medians is required, at the 5% significance level.

(iii) The diameters in independent random samples of sizes 6 and 8 are as follows, in mm.

First production line	25.9	25.8	25.3	24.7	24.4	25.4		
Second production line	23.8	25.6	24.0	23.5	24.1	24.5	24.3	25.1

Use an appropriate procedure to carry out the test. **[11]**

[Question 4 is printed overleaf.]

Option 4: Design and Analysis of Experiments

4 At an agricultural research station, a trial is made of four varieties (A, B, C, D) of a certain crop in an experimental field. The varieties are grown on plots in the field and their yields are measured in a standard unit.

(i) It is at first thought that there may be a consistent trend in the natural fertility of the soil in the field from the west side to the east, though no other trends are known. Name an experimental design that should be used in these circumstances and give an example of an experimental layout. **[5]**

Initial analysis suggests that any natural fertility trend may in fact be ignored, so the data from the trial are analysed by one-way analysis of variance.

(ii) The usual model for one-way analysis of variance of the yields y_{ij} may be written as

$$y_{ij} = \mu + \alpha_i + e_{ij}$$

where the e_{ij} represent the experimental errors. Interpret the other terms in the model. State the usual distributional assumptions for the e_{ij}. **[7]**

(iii) The data for the yields are as follows, each variety having been used on 5 plots.

Variety			
A	B	C	D
12.3	14.2	14.1	13.6
11.9	13.1	13.2	12.8
12.8	13.1	14.6	13.3
12.2	12.5	13.7	14.3
13.5	12.7	13.4	13.8

$[\Sigma\Sigma y_{ij} = 265.1, \quad \Sigma\Sigma y_{ij}^2 = 3524.31.]$

Construct the usual one-way analysis of variance table and carry out the usual test, at the 5% significance level. Report briefly on your conclusions. **[12]**