Wednesday 18 June 2014 – Afternoon

A2 GCE MATHEMATICS (MEI)

4769/01 Statistics 4

QUESTION PAPER

Candidates answer on the Printed Answer Book.

OCR supplied materials:
- Printed Answer Book 4769/01
- MEI Examination Formulae and Tables (MF2)

Other materials required:
- Scientific or graphical calculator

Duration: 1 hour 30 minutes

3249514045

INSTRUCTIONS TO CANDIDATES

These instructions are the same on the Printed Answer Book and the Question Paper.
- The Question Paper will be found inside the Printed Answer Book.
- Write your name, centre number and candidate number in the spaces provided on the Printed Answer Book. Please write clearly and in capital letters.
- **Write your answer to each question in the space provided in the Printed Answer Book**. Additional paper may be used if necessary but you must clearly show your candidate number, centre number and question number(s).
- Use black ink. HB pencil may be used for graphs and diagrams only.
- Read each question carefully. Make sure 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 scientific or graphical calculator in this paper.
- Final answers should be given to a degree of accuracy appropriate to the context.

INFORMATION FOR CANDIDATES

This information is the same on the Printed Answer Book and the Question Paper.
- The number of marks is given in brackets **[ ]** at the end of each question or part question on the Question Paper.
- 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**.
- The Printed Answer Book consists of **16** pages. The Question Paper consists of **8** pages. Any blank pages are indicated.

INSTRUCTION TO EXAMS OFFICER / INVIGILATOR

- Do not send this Question Paper for marking; it should be retained in the centre or recycled. Please contact OCR Copyright should you wish to re-use this document.

Option 1: Estimation

1 $X_1, X_2, \ldots, X_n$ represent n independent observations on the random variable X with probability density function

$$f(x) = \frac{\theta^3 x^2 e^{-\theta x}}{2}, \quad x > 0,$$

where θ is an unknown parameter $(\theta > 0)$. $\overline{X}$ denotes the sample mean of $X_1, X_2, \ldots, X_n$, ie $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$.

(i) Show that the maximum likelihood estimator of θ is $\hat{\theta} = \frac{3}{\overline{X}}$. **[9]**

(ii) Show that, in the case $n = 1$, $\hat{\theta}$ is a biased estimator of θ. **[8]**

(iii) For large n, the distribution of $\hat{\theta}$ is well approximated by $\mathrm{N}(\theta, \mathrm{H}(\theta))$ where

$$\mathrm{H}(\theta) = \frac{1}{\mathrm{E}\left(-\frac{\mathrm{d}^2 \ln \mathrm{L}}{\mathrm{d}\theta^2}\right)}$$

where L is the likelihood. Show that $\mathrm{H}(\theta) = \frac{\theta^2}{3n}$. For the case where $n = 100$ and the value of $\overline{X}$ is 5.0, evaluate $\hat{\theta}$ and $\mathrm{H}(\hat{\theta})$, and use these values to find an approximate 95% confidence interval for θ. **[7]**

Option 2: Generating Functions

2 **(i)** The probability density function of the random variable X is

$$f(x) = \frac{x^{k-1}e^{-x/\phi}}{\phi^k(k-1)!}, \quad x > 0,$$

where k is a known positive integer and ϕ is an unknown parameter ($\phi > 0$). Show that the moment generating function (mgf) of X is

$$M_X(\theta) = (1-\phi\theta)^{-k}$$

for $\theta < \frac{1}{\phi}$. **[12]**

(ii) Write down the mgf of the random variable $W = \sum_{i=1}^{n} X_i$ where $X_1, X_2, \ldots, X_n$ are independent random variables each with the same distribution as X. **[1]**

(iii) Write down the mgf of the random variable $Y = \frac{2W}{\phi}$. Given that the mgf of the random variable V having the χ_m^2 distribution is $M_V(\theta) = (1-2\theta)^{-m/2}$ (for $\theta < \frac{1}{2}$), deduce the distribution of Y. **[3]**

(iv) Deduce that $P\left(l < \frac{2W}{\phi} < u\right) = 0.95$ where l and u are the lower and upper $2\frac{1}{2}\%$ points of the χ_{2nk}^2 distribution. Hence deduce that a 95% confidence interval for ϕ is given by $\left(\frac{2w}{u}, \frac{2w}{l}\right)$ where w is an observation on the random variable W. **[2]**

(v) For the case $k = 2$ and $n = 10$, use percentage points of the χ^2 distribution to write down, in terms of w, an expression for a 95% confidence interval for ϕ. By considering the mgf of W, find in terms of ϕ the expected length of this interval. **[6]**

Option 3: Inference

3 **(i)** Explain the meaning of the following terms in the context of hypothesis testing: Type I error, Type II error, operating characteristic, power. **[8]**

(ii) A chemical manufacturer is endeavouring to reduce the amount of a certain impurity in one of its bulk products by improving the production process. The amount of impurity is measured in a convenient unit of concentration, and this is modelled by the Normally distributed random variable X. In the old production process, the mean of X, denoted by μ, was 63 and the standard deviation of X was 3.7. Experimental batches of the product are to be made using the new process, and it is desired to examine the hypotheses $H_0: \mu = 63$ and $H_1: \mu < 63$ for the new process. Investigation of the variability in the new process has established that the standard deviation may be assumed unchanged.

The usual Normal test based on $\overline{X}$ is to be used, where $\overline{X}$ is the mean of X over n experimental batches (regarded as a random sample), with a critical value c such that H_0 is rejected if the value of $\overline{X}$ is less than c. The following criteria are set out.

- If in fact $\mu = 63$, the probability of concluding that $\mu < 63$ must be only 1%.
- If in fact $\mu = 60$, the probability of concluding that $\mu < 63$ must be 90%.

Find c and the smallest value of n that is required. With these values, what is the power of the test if in fact $\mu = 58.5$? **[16]**

Option 4: Design and Analysis of Experiments

4 A trial is being made of four experimental methods, A, B, C and D, for carrying out an industrial process. These are being compared with each other and with the standard method M. The trial is conducted according to a completely randomised design. The results, x, are as follows, in a suitable unit.

Method	Results x	Total	Mean
M	25.0 23.0 30.1 27.5 28.8 25.6 29.2 31.6	220.8	27.6
A	37.3 34.9 30.8 40.2	143.2	35.8
B	36.4 36.6 29.2 44.0 34.8	181.0	36.2
C	32.0 40.1 33.0 36.5	141.6	35.4
D	35.0 31.8 39.0 38.2	144.0	36.0
	Grand total	830.6	

You are also given that $\sum x^2 = 28\,260.18$.

(i) The usual statistical model underlying a one-way analysis of variance is given, in the usual notation, by

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

where x_{ij} denotes the jth observation on the ith treatment. State the properties that are assumed for the term that represents experimental error. **[3]**

(ii) Construct the usual analysis of variance table for these data. Stating your hypotheses carefully, test whether there is evidence of differences among the means for the five methods, using a 5% significance level. **[12]**

(iii) In each case using the residual mean square as the estimate of the variance of the experimental error, find a 95% confidence interval for the population mean for method M and a 95% confidence interval for the population mean for method A. What do these confidence intervals suggest about these population means? **[5]**

(iv) The residuals, calculated by subtracting the corresponding method mean from each observation, are given in the table below. For example the first residual for method M is 25.0 – 27.6 = –2.6. Each residual gives a measure of experimental error.

Method	Residuals
M	–2.6 –4.6 2.5 –0.1 1.2 –2.0 1.6 4.0
A	1.5 –0.9 –5.0 4.4
B	0.2 0.4 –7.0 7.8 –1.4
C	–3.4 4.7 –2.4 1.1
D	–1.0 –4.2 3.0 2.2

The diagram in the printed answer book shows a dotplot of the residuals for method M. Complete the diagram by adding the dotplots for the other methods.

Use these dotplots to comment briefly on the assumptions you have stated in part **(i)**. **[4]**

END OF QUESTION PAPER