Friday 24 June 2016 – Morning

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

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 The random variable X has a Cauchy distribution centred on m. Its probability density function (pdf) is $\mathrm{f}(x)$ where

$$\mathrm{f}(x) = \frac{1}{\pi}\frac{1}{1+(x-m)^2}, \qquad \text{for } -\infty < x < \infty\ .$$

(i) Sketch the pdf. Show that the mode and median are at $x = m$. **[6]**

(ii) A sample of size 1, consisting of the observation x_1, is taken from this distribution. Show that the maximum likelihood estimate (MLE) of m is x_1. **[3]**

(iii) Now suppose that a sample of size 2, consisting of observations x_1 and x_2, is taken from the distribution. By considering the logarithm of the likelihood function or otherwise, show that the MLE, $\hat{m}$, satisfies the cubic equation

$$\left(2\hat{m}-\left(x_1+x_2\right)\right)\left(\hat{m}^2-\left(x_1+x_2\right)\hat{m}+1+x_1x_2\right)=0.$$

[5]

(iv) Obtain expressions for the three roots of this equation. Show that if $|x_1 - x_2| < 2$ then only one root is real. How do you know, without doing further calculations, that in this case the real root will be the MLE of m? **[5]**

(v) Obtain the three possible values of $\hat{m}$ in the case $x_1 = -2$ and $x_2 = 2$. Evaluate the likelihood function for each value of $\hat{m}$ and comment on your answer. **[5]**

Option 2: Generating Functions

2 The random variable X has probability density function $\mathrm{f}(x)$ where

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

(i) Obtain the moment generating function (mgf) of X. **[3]**

(ii) Use the mgf to find $\mathrm{E}(X)$ and $\mathrm{Var}(X)$. **[6]**

The random variable Y is defined as follows:

$$Y = X_1 + \ldots + X_n,$$

where the X_i are independently and identically distributed as X.

(iii) Write down expressions for $\mathrm{E}(Y)$ and $\mathrm{Var}(Y)$.

Obtain the mgf of Y. **[4]**

(iv) Find the mgf of Z where $Z = \dfrac{Y - \dfrac{n}{\lambda}}{\dfrac{\sqrt{n}}{\lambda}}$. **[4]**

(v) By considering the logarithm of the mgf of Z, show that the distribution of Z tends to the standard Normal distribution as n tends to infinity. **[7]**

Option 3: Inference

3 A large department in a university wished to compare the standards of literacy and numeracy of its students. A random sample of 24 students was taken and sub-divided, randomly, into two groups of 12. The students in one group took a literacy assessment (scores denoted by x); the students in the other group took a numeracy assessment (scores denoted by y). The two assessments were designed to give the same distributions of scores when taken by random samples from the general population.

The scores obtained by the students on the two assessments are shown in the table.

x	23	42	43	46	48	48	50	54	58	59	62	65
y	44	36	63	55	53	58	63	80	61	57	83	54

$$\sum x = 598 \qquad \sum x^2 = 31196 \qquad \sum y = 707 \qquad \sum y^2 = 43\,543$$

(i) Carry out an appropriate t test, at the 5% level of significance, to compare the standards of literacy and numeracy. **[12]**

(ii) State the distributional assumptions required for the t test to be valid.

Name the test that you would use if the assumptions required for the t test are thought not to hold. State the hypotheses for this new test.

Explain, in general terms, which of the two tests is more powerful, and why. **[6]**

A statistician at the university looked at the data and commented that a paired sample design would have been better.

(iii) Explain how a paired sample design would be applied in this context, and how the data would be analysed. Explain also why it would be better than the design used. **[6]**

Option 4: Design and Analysis of Experiments

4 The cardiovascular unit of a hospital is studying the effect on patients' heart rates of three different light exercises, A, B and C. Patients are given an exercise to do and the increases in their pulse rates are measured after 5 minutes. There are 16 patients in the study: 5 are chosen randomly and allocated to exercise A, 6 to exercise B, and 5 to exercise C.

The data obtained are as follows.

A	B	C
63	69	56
41	72	44
42	52	65
51	64	48
47	54	53
	57	

	A	B	C
Sum of data	244	368	266
Sum of squares	12224	22910	14410

(i) State the usual one-way analysis of variance model.

Explain what the terms in the model mean in this context.

State the distributional assumptions required for the standard test.

Carry out the test at the 5% level of significance and report your conclusions. **[18]**

(ii) Someone unfamiliar with analysis of variance analysed these data. They used three t tests to compare A with B, B with C, and C with A. The test comparing A with B was significant at the 5% level; the other two tests were not significant at the 5% level.

Comment on this analysis, explaining whether it is better than, worse than or equivalent to the analysis carried out in part **(i)**. Your comments should include consideration of the independence of the t tests and the overall level of significance of the procedure. **[6]**

END OF QUESTION PAPER

BLANK PAGE

BLANK PAGE