OCR
Oxford Cambridge and RSA

Tuesday 16 June 2015 – 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

* 4 4 2 7 7 2 3 6 0 4 *

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.

Turn over

Option 1: Estimation

1 The random variable X has the following probability density function, in which a is a (positive) parameter.

$$f(x) = \frac{2}{a} x e^{-x^2/a}, \quad x \geqslant 0.$$

(i) Verify that $\int_0^\infty f(x)\,dx = 1$. **[1]**

(ii) Show that $E(X^2) = a$ and $E(X^4) = 2a^2$. **[7]**

The parameter a is to be estimated by maximum likelihood based on an independent random sample from the distribution, $X_1, X_2, \ldots, X_n$.

(iii) Show that the logarithm of the likelihood function is

$$n \ln 2 - n \ln a + \sum_{i=1}^{n} \ln X_i - \frac{1}{a} \sum_{i=1}^{n} X_i^2.$$

Hence obtain the maximum likelihood estimator, $\hat{a}$, for a.

[You are not required to verify that any turning point you find is a maximum.] **[7]**

(iv) Using the results from part **(ii)**, show that $\hat{a}$ is unbiased for a and find the variance of $\hat{a}$. **[5]**

(v) In a particular random sample from this distribution, $n = 100$ and $\sum x_i^2 = 147.1$. Obtain an approximate 95% confidence interval for a. (You may assume that the Central Limit Theorem holds in this case.) **[4]**

Option 2: Generating Functions

2 The random variable Z has the standard Normal distribution. The random variable Y is defined by $Y = Z^2$.

You are given that Y has the following probability density function.

$$f(y) = \frac{1}{\sqrt{2\pi y}} e^{-\frac{1}{2}y}, \quad y > 0.$$

(i) Show that the moment generating function (mgf) of Y is given by

$$M_Y(\theta) = (1 - 2\theta)^{-\frac{1}{2}}.$$ **[6]**

(ii) Use the mgf to obtain $E(Y)$ and $\text{Var}(Y)$. **[5]**

The random variable U is defined by

$$U = Z_1^2 + Z_2^2 + \ldots + Z_n^2,$$

where $Z_1, Z_2, \ldots, Z_n$ are independent standard Normal random variables.

(iii) State an appropriate general theorem for mgfs and hence write down the mgf of U. State the values of $E(U)$ and $\text{Var}(U)$. **[4]**

The random variable W is defined by

$$W = \frac{U - n}{\sqrt{2n}}.$$

(iv) Show that the logarithm of the mgf of W is

$$-\sqrt{\frac{n}{2}}\theta - \frac{n}{2}\ln\left(1 - \sqrt{\frac{2}{n}}\theta\right).$$

Use the series expansion of $\ln(1 - t)$ to show that, as $n \to \infty$, this expression tends to $\frac{1}{2}\theta^2$.

State what this implies about the distribution of W for large n. **[9]**

Option 3: Inference

3 At an agricultural research station, trials are being carried out to compare a standard variety of tomato with one that has been genetically modified (GM). The trials are concerned with the mean weight of the tomatoes and also with the aesthetic appearance of the tomatoes.

(a) **(i)** Tomatoes of the standard and GM varieties are grown under similar conditions. The tomatoes are weighed and the data are summarised as follows.

Variety	Sample size	Sum of weights (g)	Sum of squares of weights (g^2)
Standard	30	3218.3	349 257
GM	26	2954.1	338 691

Carry out a test, using the Normal distribution, to investigate whether there is evidence, at the 5% level of significance, that the two varieties of tomato differ in mean weight.

State one assumption required for this test to be valid. **[10]**

(ii) The data in part **(i)** could have been used to carry out a test for the equality of means based on the *t* distribution. State **two** additional assumptions required for this test to be valid.

Discuss briefly which test would be preferable in this case. **[4]**

(b) In order to judge whether, on the whole, GM tomatoes have a better aesthetic appearance than standard tomatoes, a trial is carried out as follows. 10 of each variety are chosen and a consumer panel is asked to arrange the 20 tomatoes in order according to their appearance.

(i) State **two** important features of the way in which this trial should be designed.

Comment briefly on how reliable the evidence from the trial is likely to be. **[3]**

(ii) The order in which the consumer panel arranges the tomatoes is as follows. The tomato with best appearance is listed first. *G* and *S* denote GM and standard tomatoes respectively.

$$G \quad G \quad G \quad S \quad G \quad G \quad G \quad S \quad G \quad S \quad S \quad S \quad G \quad G \quad S \quad G \quad S \quad S \quad S \quad S$$

Carry out an appropriate test at the 1% level of significance. **[7]**

Option 4: Design and Analysis of Experiments

4 **(a)** The standard one-way Analysis of Variance (ANOVA) model is expressed in the usual notation as follows.

$$Y_{ij} = \mu + \alpha_i + \varepsilon_{ij}$$

(i) Explain what the terms Y_{ij}, μ, α_i and ε_{ij} represent. **[4]**

(ii) State a distributional assumption about ε_{ij} and explain briefly why this assumption is required. **[4]**

(iii) State the null and alternative hypotheses for the usual one-way ANOVA test. Explain clearly how to interpret the two possible outcomes of the ANOVA test. **[4]**

(b) I drive frequently between two cities, A and B. There are k different routes that I can take. On each of these routes the journey time varies according to time of day, traffic conditions and so on.

In order to test whether or not there are any differences between the mean journey times on the k routes, I chose a route at random for each of N journeys. I recorded the time for each journey, entered the data into a spreadsheet, and carried out an ANOVA analysis. Part of the output was as follows.

Source of variation	Sum of squares	Degrees of freedom
Between groups	333.77	
Within groups		15
Total	752.96	18

(i) State the values of k and N. **[2]**

(ii) Complete the analysis using a 5% significance level. **[10]**

END OF QUESTION PAPER

BLANK PAGE

BLANK PAGE