OCR
Oxford Cambridge and RSA

# Friday 22 June 2018 – 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

*7315555603*

## 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 barcodes.
- 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 **12** 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** A biased coin has probability $p$ that it gives a tail when it is tossed. It is known that $0 < p < 1$. The random variable $T$ is the number of tosses up to and including the second tail.

**(i)** Show that, for $t \geqslant 2$, $\mathrm{P}(T = t) = (t-1)(1-p)^{t-2}p^2$. **[3]**

**(ii)** A single observation of $T$ takes the value $t$. Find the maximum likelihood estimate for $p$ based on this observation. You may assume that any stationary point that you locate is a maximum. **[4]**

**(iii)** Show that $Z = \dfrac{1}{T-1}$ is an unbiased estimator of $p$. **[5]**

An alternative method for finding an unbiased estimator of $p$ is to consider the random variable $F$, the number of tosses up to and including the first tail.

**(iv)** Show that $\dfrac{1}{F}$ is not an unbiased estimator of $p$. You may use the fact that

$$\sum_{n=1}^{\infty} \frac{x^n}{n} = -\ln(1-x) \text{ for } |x| < 1.$$ **[4]**

**(v)** The random variable $Y$ is defined by

$$Y = \begin{cases} 1 & \text{if } F = 1, \\ 0 & \text{otherwise}. \end{cases}$$

Show that $Y$ is an unbiased estimator of $p$. **[3]**

You are given that, for the estimator $Z$ defined in part **(iii)**, $\mathrm{Var}(Z) = p^2\left(\dfrac{1}{1-p}\ln\left(\dfrac{1}{p}\right) - 1\right)$.

**(vi)** Show that if $p = 0.5$ then $\mathrm{Var}(Y) > \mathrm{Var}(Z)$. State, with a reason, which of $Y$ or $Z$ is the more efficient estimator of $p$ in this case. **[5]**

*Option 2: Generating Functions*

**2** The random variable $X$ has probability density function

$$f(x) = \begin{cases} \frac{1}{2b}e^{\frac{x}{b}} & \text{for } x \leqslant 0, \\ \frac{1}{2b}e^{-\frac{x}{b}} & \text{for } x > 0, \end{cases}$$

where $b$ is a positive parameter.

**(i)** Show that, for $-\frac{1}{b} < t < \frac{1}{b}$, the moment generating function $M_X(t)$ of $X$ is $\frac{1}{1-b^2t^2}$. **[9]**

**(ii)** Hence find the expectation and variance of $X$ in terms of $b$. **[6]**

A random variable $Y$ with moment generating function $M_Y(t) = \frac{1}{1-bt}$ is said to have an exponential distribution with parameter $b$, denoted by Exp($b$).

**(iii)** Two independent observations of $Y$ are denoted by $Y_1$ and $Y_2$. Show that $Y_1 - Y_2$ has the same distribution as $X$. **[5]**

**(iv)** A radioactive substance emits beta particles. The time, in seconds, between successive emissions is called the 'waiting time' and has an Exp(4) distribution. Find the probability that the absolute value of the difference between two independent waiting times is more than 0.1 seconds. **[4]**

*Option 3: Inference*

**3** A manufacturer produces large numbers of a particular type of power supply which is specified as delivering its output at 15 volts.

**(a)** It is important that the actual voltage should not be too high, as that could damage the equipment to which the power supply is attached. So the manufacturer's quality control department routinely tests each batch of power supplies. The null and alternative hypotheses under test are

$$H_0: \mu = 15.0, \qquad H_1: \mu > 15.0,$$

where $\mu$ volts is the true mean voltage delivered by the power supplies in the batch.

The test consists in drawing a random sample of size $n$ from the batch, and comparing the mean voltage, $\bar{x}$, with a critical value $c$. The test is required to have a Type I error of size no greater than 0.05. The size of the Type II error, when $\mu = 15.1$, is to be no greater than 0.01.

Previous data indicate that the voltages are Normally distributed with a standard deviation of 0.2.

**(i)** Determine suitable values for $c$ and $n$. **[9]**

**(ii)** Without doing any further calculations, sketch the power function for this test. **[3]**

**(b)** One part of the manufacturer's quality control procedure involves studying how the power supplies perform after a lengthy period of usage. A random sample of 10 power supplies are used for a year and then returned to the manufacturer for testing. A control set of 12 power supplies is chosen randomly and kept unused for a year. Each of these power supplies then has its output voltage measured. The data, ordered by voltage, are as follows.

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Used power supplies | 14.50 | 14.67 | 14.71 | 14.93 | 14.99 | 15.00 | 15.02 | 15.05 | 15.16 | 15.18 | | |
| Unused power supplies | 14.77 | 14.81 | 14.92 | 14.94 | 15.04 | 15.07 | 15.10 | 15.11 | 15.14 | 15.26 | 15.28 | 15.39 |

Because of the differing conditions under which some of the power supplies have been used, it is not thought safe to assume underlying Normality.

**(i)** Carry out a suitable test, at the 5% level of significance, to investigate whether or not usage affects the median output voltage of the power supplies. **[9]**

**(ii)** What assumption about the variability of voltages is made in the test in part **(i)**? Discuss briefly (but without carrying out any calculations) whether or not this assumption is reasonable. **[3]**

*Option 4: Design and Analysis of Experiments*

**4** **(a)** What is randomised block design and what is its purpose? Describe an experimental situation where randomised block design is appropriate. You should identify the main effect and the blocking factor in your description. **[5]**

**(b)** What are the distributional assumptions required to conduct an ANOVA test? **[5]**

**(c)** A sample of 15 people take part in a medical trial to compare 3 different drugs for controlling blood pressure. The sample is split into 3 groups, with each group taking a different drug. Each group contains 5 people. The decrease in the blood pressure of each person is measured (in arbitrary units).

A partially completed ANOVA table is shown below.

| **Source of variation** | **Sum of squares** | **Degrees of freedom** |
|---|---|---|
| Between treatments | $x$ | $a$ |
| Residual | | $b$ |
| Total | 1000 | |

**(i)** State the standard null and alternative hypotheses for ANOVA, stating the meaning of any symbols used. **[2]**

**(ii)** Write down the values of $a$ and $b$. **[2]**

**(iii)** Find the minimum value of $x$ which results in rejection of the null hypothesis using a 5% significance level. **[10]**

**END OF QUESTION PAPER**

**BLANK PAGE**

**BLANK PAGE**