OCR
RECOGNISING ACHIEVEMENT

Thursday 24 May 2012 – Morning

A2 GCE MATHEMATICS (MEI)

4769 Statistics 4

QUESTION PAPER

Candidates answer on the Printed Answer Book.

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

Other materials required:
- Scientific or graphical calculator

Duration: 1 hour 30 minutes

4715880612

INSTRUCTIONS TO CANDIDATES

These instructions are the same on the Printed Answer Book and the Question Paper.

- The Question Paper will be found in the centre of 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 destroyed.

Option 1: Estimation

1 In a certain country, any baby born is equally likely to be a boy or a girl, independently for all births. The birthweight of a baby boy is given by the continuous random variable X_B with probability density function (pdf) $f_B(x)$ and cumulative distribution function (cdf) $F_B(x)$. The birthweight of a baby girl is given by the continuous random variable X_G with pdf $f_G(x)$ and cdf $F_G(x)$.

The continuous random variable X denotes the birthweight of a baby selected at random.

(i) By considering

$$P(X \leqslant x) = P(X \leqslant x \,|\, \text{boy})\ P(\text{boy}) + P(X \leqslant x \,|\, \text{girl})\ P(\text{girl}),$$

find the cdf of X in terms of $F_B(x)$ and $F_G(x)$, and deduce that the pdf of X is

$$f(x) = \tfrac{1}{2}\{f_B(x) + f_G(x)\}. \quad \textbf{[3]}$$

(ii) The birthweights of baby boys and girls have means μ_B and μ_G respectively. Deduce that

$$E(X) = \tfrac{1}{2}(\mu_B + \mu_G). \quad \textbf{[1]}$$

(iii) The birthweights of baby boys and girls have common variance σ^2. Find an expression for $E(X^2)$ in terms of μ_B, μ_G and σ^2, and deduce that

$$\text{Var}(X) = \sigma^2 + \tfrac{1}{4}(\mu_B - \mu_G)^2. \quad \textbf{[7]}$$

(iv) A random sample of size $2n$ is taken from all the babies born in a certain period. The mean birthweight of the babies in this sample is $\overline{X}$. Write down an approximation to the sampling distribution of $\overline{X}$ if n is large. **[4]**

(v) Suppose instead that a stratified sample of size $2n$ is taken by selecting n baby boys at random and, independently, n baby girls at random. The mean birthweight of the $2n$ babies in this sample is $\overline{X}_{st}$. Write down the expected value of $\overline{X}_{st}$ and find the variance of $\overline{X}_{st}$. **[4]**

(vi) Deduce that both $\overline{X}$ and $\overline{X}_{st}$ are unbiased estimators of the population mean birthweight. Find which is the more efficient. **[5]**

Option 2: Generating Functions

2 The random variable X ($X = 1, 2, 3, 4, 5, 6$) denotes the score when a fair six-sided die is rolled.

(i) Write down the mean of X and show that $\text{Var}(X) = \frac{35}{12}$. **[3]**

(ii) Show that $\text{G}(t)$, the probability generating function (pgf) of X, is given by

$$\text{G}(t) = \frac{t(1 - t^6)}{6(1 - t)}.$$ **[2]**

The random variable N ($N = 0, 1, 2, \ldots$) denotes the number of heads obtained when an unbiased coin is tossed repeatedly until a tail is first obtained.

(iii) Show that $\text{P}(N = r) = \left(\frac{1}{2}\right)^{r+1}$ for $r = 0, 1, 2, \ldots$. **[1]**

(iv) Hence show that $\text{H}(t)$, the pgf of N, is given by $\text{H}(t) = (2 - t)^{-1}$. **[2]**

(v) Use $\text{H}(t)$ to find the mean and variance of N. **[4]**

A game consists of tossing an unbiased coin repeatedly until a tail is first obtained and, each time a head is obtained in this sequence of tosses, rolling a fair six-sided die. The die is not rolled on the first occasion that a tail is obtained and the game ends at that point. The random variable Q ($Q = 0, 1, 2, \ldots$) denotes the total score on all the rolls of the die. Thus, in the notation above, $Q = X_1 + X_2 + \ldots + X_N$ where the X_i are independent random variables each distributed as X, with $Q = 0$ if $N = 0$. The pgf of Q is denoted by $\text{K}(t)$. The familiar result that the pgf of a sum of independent random variables is the product of their pgfs does **not** apply to $\text{K}(t)$ because N is a random variable and not a fixed number; you should instead **use without proof** the result that $\text{K}(t) = \text{H}(\text{G}(t))$.

(vi) Show that $\text{K}(t) = 6(12 - t - t^2 - \ldots - t^6)^{-1}$. **[4]**

[Hint. $(1 - t^6) = (1 - t)(1 + t + t^2 + \ldots + t^5)$.]

(vii) Use $\text{K}(t)$ to find the mean and variance of Q. **[6]**

(viii) Using your results from parts **(i)**, **(v)** and **(vii)**, verify the result that (in the usual notation for means and variances)

$$\sigma_Q^2 = \sigma_N^2 \mu_X^2 + \mu_N \sigma_X^2.$$ **[2]**

Option 3: Inference

3 At an agricultural research station, trials are being made of two fertilisers, A and B, to see whether they differ in their effects on the yield of a crop. Preliminary investigations have established that the underlying variances of the distributions of yields using the two fertilisers may be assumed equal. Scientific analysis of the fertilisers has suggested that fertiliser A may be inferior in that it leads, on the whole, to lower yield. A statistical analysis is being carried out to investigate this.

The crop is grown in carefully controlled conditions in 14 experimental plots, 6 with fertiliser A and 8 with fertiliser B. The yields, in kg per plot, are as follows, arranged in ascending order for each fertiliser.

Fertiliser A	9.8	10.2	10.9	11.5	12.7	13.3		
Fertiliser B	10.8	11.9	12.0	12.2	12.9	13.5	13.6	13.7

(i) Carry out a Wilcoxon rank sum test at the 5% significance level to examine appropriate hypotheses. **[9]**

(ii) Carry out a *t* test at the 5% significance level to examine appropriate hypotheses. **[11]**

(iii) Goodness of fit tests based on more extensive data sets from other trials with these fertilisers have failed to reject hypotheses of underlying Normal distributions. Discuss the relative merits of the analyses in parts **(i)** and **(ii)**. **[4]**

Option 4: Design and Analysis of Experiments

4 **(i)** In an engineering research laboratory, a study is being made of the strength of steel girders supplied by four different manufacturers. Four techniques for casting the girders are to be used, as are four slightly different chemical compositions of the steel. Sixteen girders are to be supplied for testing purposes, four by each manufacturer.

Name an experimental design that should be used for allocating the work to the manufacturers in such a way that any differences in strength of girders between the different manufacturers can be studied, whether or not there are consistent differences resulting from the casting techniques or from the chemical compositions. Give an example of a suitable layout of the experiment. **[5]**

(ii) After initial investigation, it is decided that differences in strength resulting from the casting techniques or the chemical compositions can be ignored. A one-way analysis of variance is therefore carried out on the results, which are as follows, measured in a convenient unit.

Strength of girder

Manufacturer			
A	B	C	D
109.4	114.4	114.8	115.1
110.0	113.1	113.7	114.0
110.9	113.5	115.4	114.7
110.3	112.5	114.3	115.6

[The sum of these data items is 1811.7 and the sum of their squares is 205 202.57.]

Construct the usual one-way analysis of variance table. Carry out the appropriate test and report your conclusion. **[12]**

(iii) Using the customary notation, write down the usual model underlying the one-way analysis of variance. Carefully interpret the terms in this model. State the assumptions that are usually made for the error term in the model. **[7]**

BLANK PAGE

BLANK PAGE