EXAMINATIONS OF THE ROYAL STATISTICAL SOCIETY
(formerly the Examinations of the Institute of Statisticians)

HIGHER CERTIFICATE IN STATISTICS, 1997

CERTIFICATE IN OFFICIAL STATISTICS, 1997

Paper I : Statistical Theory

Time Allowed: Three Hours

Candidates should answer **FIVE** *questions.*

All questions carry equal marks.

Graph paper and Official tables are provided.

Candidates may use silent, cordless, non-programmable electronic calculators.

Where a calculator is used the **method** *of calculation should be stated in full.*

Note that $\binom{n}{r}$ *is the same as* ${}^nC_r$ *and that* $\ln$ *stands for* $\log_e$.

1. (a) (i) Four men wish to play doubles tennis. In how many different ways can they choose which two play against the other two?

If two men and two women wish to play mixed doubles (a pair consisting of one man and woman against the other man and woman), how many different mixed doubles matches can be formed from these four people?

(ii) How many different doubles matches can be arranged, choosing the participants from a pool of n players?

How many different mixed doubles matches can be arranged, choosing the participants from a pool of n_1 men and n_2 women?

(b) (i) In how many ways can 10 footballers be divided into 2 five-a-side teams?

(ii) In how many ways can the two teams be formed if there are only 2 goalkeepers among the 10 players and the goalkeepers have to play on opposing sides?

(iii) There are only 2 goalkeepers and 3 strikers (all different people) among the 10 players. In how many ways can the two teams be formed if each team must have one goalkeeper and at least one striker?

2. If a Ruritanian peasant farmer grows cereals, his profit X_1 in Ruritanian pounds (£*R*), is Normally distributed with mean 1750 and standard deviation 300. If he grows beans his profit X_2 is Normally distributed with mean 2000 and standard deviation 400. If he grows a proportion p of cereals and a proportion $1–p$ of beans, the profit, Y, is $pX_1 + (1-p)X_2$ where $0 \le p \le 1$.

(i) State the distribution of Y, assuming that X_1 and X_2 are independent.

(ii) State the value of p, p_1 say, which maximises the farmer's expected profit, $E(Y)$.

(iii) Find the value of p, p_2 say, which minimises the variance of the farmer's profit, that is, minimises $V(Y)$.

(iv) Calculate the expected profit when $p = p_1$ and when $p = p_2$.

(v) The farmer reckons that he will be ruined if his profit is less than £*R* 1480. Calculate the probability that the farmer will be ruined (a) if he adopts $p = p_1$, (b) if he adopts $p = p_2$. Which course of action do you think is better for the farmer, and why?

Turn over

3. The probability that any given child in a certain family has blue eyes is ¼, and this feature is independent of the eye colours of the other children in the family. There are 5 children in the family.

   (i) State the distribution of the number of children with blue eyes and find the probability that at least one child has blue eyes.

   (ii) Find the probability that at least three children have blue eyes, given that at least one child has blue eyes.

   (iii) Find the probability that at least three of the children have blue eyes, given that the youngest child has blue eyes.

   (iv) Calculate the expected number of children with blue eyes,

      (a) given that at least one child has blue eyes;

      (b) given that the youngest child has blue eyes.

   (v) Explain why the answers to (ii) and (iii) are not the same.

4. Matching nuts and bolts are manufactured in a production process. Bolt diameters X are $N(10.0\text{mm}, 0.0009\text{mm}^2)$ distributed and nut diameters Y are independently $N(10.1\text{mm}, 0.0016\text{mm}^2)$ distributed. In order to fit satisfactorily the nut must be at least 0.02 mm wider than the bolt (so that it is not too tight) and at most 0.2 mm wider than the bolt (so that it is not too loose).

   (i) Find the probability that a randomly chosen nut will fit a bolt of diameter 9.98 mm.

   (ii) Write down the distribution of the difference of the diameters $Y - X$ and hence find the probability that a nut and a bolt chosen at random from the production process fit satisfactorily.

   (iii) Holes of diameter Z_1 and Z_2 are drilled in each of two metal plates which are to be bolted together, the two values Z_1 and Z_2 being independently $N(10.3\text{mm}, 0.0144\text{mm}^2)$ distributed.

      (a) Find the probability that a bolt of diameter 10.06 mm will pass through the holes drilled in the two plates.

      (b) Find the probability that a bolt of diameter 10.06 mm will pass through the holes drilled in the two plates <u>and</u> be fitted satisfactorily by a randomly chosen nut.

(iv) As part of a quality control procedure, a random sample of 25 bolts is taken from the output, and the production process is stopped if the mean diameter of the bolts in the sample differs by more than 0.01 mm from 10 mm. Find the probability of this event.

5. It is required to carry out a blood test on a large number (N) of persons, to check for the presence or absence of a rare characteristic. It is assumed that the probability p of a positive result is the same for all persons, and is independent of the results of tests on other persons.

To reduce the work of testing, the blood samples of N persons are pooled into groups of size k, k being a factor of N, and tested together. If the group test is negative no further test is necessary for the k persons. If the group test is positive each person must be tested individually and so in all $(k + 1)$ tests are required for the group of k persons.

(i) Find the probability that the test for a pooled sample of k persons is positive.

(ii) Let m be the number of groups, so that $m=\frac{N}{k}$ and let S denote the total number of tests required for these people. Show that S can be written in terms of a Binomially distributed random variable X as

$$S = m+kX, \text{ where } X \sim B\left(m,\ 1-(1-p)^{k}\right).$$

(iii) Hence show that

$$E(S) = N\left\{\frac{1}{k}+1-(1-p)^{k}\right\};$$

$$V(S) = Nk(1-p)^{k}\left[1-(1-p)^{k}\right].$$

(iv) Treating k as if it were continuous, show by differentiation that the value of k which gives the minimum expected number of tests for the N persons satisfies the equation

$$1+k^{2}(1-p)^{k}\log_{e}(1-p) = 0. \qquad \text{(A)}$$

(You may assume that this equation has a unique solution and that it gives a minimum.)

(v) Given that $p = 0.01$ and an approximate solution of (A) is $k = 10.5$, find the value of k which minimises $E(S)$ when $N = 9900$.

Turn over

6. The lengths X of offcuts of timber in a carpenter's workshop follow the continuous uniform distribution with probability density function (pdf)

$$
\begin{aligned}
f(x) &= \frac{1}{\theta}, \quad 0 \le x \le \theta;\ \theta > 0; \\
&= 0 \quad \text{elsewhere.}
\end{aligned}
$$

(i) Obtain the cumulative distribution function (cdf), *F(x)* say, and sketch the graphs of *f(x)* and *F(x).* Also find *E(X)* and *V(X).*

(ii) The carpenter takes a random sample of offcuts with lengths $X_1, \ldots, X_n$. Explain why

$$P\text{ (length of longest offcut in sample} \le x) = \left(\frac{x}{\theta}\right)^n, \quad 0 \le x \le \theta,$$

and deduce the pdf of the sample maximum, $X_{(n)}$ say.

Show that

$$E(X_{(n)}) = \frac{n\theta}{n+1},$$

$$V(X_{(n)}) = \frac{n\theta^2}{(n+1)^2 \ (n+2)}.$$

Write down a multiple of $X_{(n)}$ which is an unbiased estimator of θ and obtain its variance.

Is $X_{(n)}$ the maximum likelihood estimate?

(iii) Show that $\frac{2}{n}\sum_{i=1}^{n} X_i$ is the method of moments estimator of θ and obtain its variance.

(iv) How would you advise the carpenter to estimate θ?

7. The data (x_i, y_i), where $x_i > 0$, $i = 1,...,n$, are thought to conform to a proportional regression model (linear regression through the origin)

$$Y_i = \beta x_i + e_i, \; i = 1,...,n,$$

where the e_i are independent Normally distributed error terms with mean zero.

(i) If the e_i have constant known variance σ^2, show that the maximum likelihood (ML) estimate of β minimises $\sum_{i=1}^{n} e_i^2$ and is given by

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n} x_i y_i}{\sum_{i=1}^{n} x_i^2}.$$

(ii) If instead the variance of e_i is given by $\sigma^2 x_i$, $i = 1, \ldots, n$, show that the ML estimate of β minimises $\sum_{i=1}^{n} \frac{e_i^2}{x_i}$ and is given by

$$\hat{\beta}_2 = \frac{\bar{y}}{\bar{x}},$$

where $\bar{x}$ and $\bar{y}$ are the sample mean values of $x_1,...,x_n$ and $y_1,...,y_n$ respectively.

(iii) Each time a motorcycle is filled with petrol, a record is kept of the amount of petrol in litres (x) used, and the distance travelled in miles (y) since the previous fill-up. Values of x and y recorded on the last 9 occasions were as follows:

x	4.3	4.9	5.7	6.5	7.2	8.3	8.4	9.6	10.1
y	123	156	183	183	204	234	270	273	324

Plot the data and calculate $\hat{\beta}_1$ and $\hat{\beta}_2$. Which of the models (i) or (ii) do you think better represents the data?

Turn over

8. (a) State the conditions under which

(i) the Poisson distribution may be used to approximate a Binomial distribution,

(ii) the Normal distribution may be used to approximate a Poisson distribution.

(b) The random variable X has probability mass function

$$f(x) = e^{-5}\frac{5^x}{x!}, \quad x = 0, 1, 2, \ldots,$$

$$= 0 \quad \text{otherwise.}$$

Find $P(4 \leq X \leq 6)$ (i) exactly, (ii) using a Normal approximation. Calculate the percentage error of this approximation, and comment briefly.

(c) Events happen in a Poisson process at rate λ per unit time so that the number of events in the interval $(0, t)$ is a Poisson random variable with mean λt. Write down the probability that there are no events in the interval $(0, t)$. By interpreting this probability as the chance that the time T to the next event in the process is greater than t, show that the probability density function (pdf) of T is given by

$$g(t) = \lambda e^{-\lambda t}, \; t \geq 0, \lambda > 0.$$

$$= 0 \quad \text{otherwise.}$$

Find the mean and variance of T.

(d) In the case $\lambda = 5$ a random sample of size 100 is taken from the pdf $g(t)$. Use the central limit theorem to approximate the probability that the sample mean is within 10% of the true mean of T.