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

HIGHER CERTIFICATE IN STATISTICS, 1999

CERTIFICATE IN OFFICIAL STATISTICS, 1999

Paper I : Statistical Theory

Time Allowed: Three Hours

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

All questions carry equal marks.
The number of marks allotted for each part-question is shown in brackets.

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* ${}^{n}C_{r}$ *and that* ln *stands for* $\log_e$.

This examination paper consists of 8 printed pages. This front cover is page 1. The reverse of the front cover, which is intentionally left blank, is page 2. Question 1 starts on page 3.

There are 8 questions altogether in the paper.

1. The game of Snip is played for marbles. In this game two players, *A* and *B* say, each have a coin which, unseen by the other player, they independently choose to display as Heads (*H*) or Tails (*T*). When *A* and *B* have made their choices, the coins are disclosed and the outcome of the game is decided in accordance with the table below.

		B displays	
		H	*T*
A displays	*H*	*A* wins 3 from *B*	*B* wins 2 from *A*
	T	*B* wins 2 from *A*	*A* wins 1 from *B*

(i) Suppose that *A* and *B* independently both randomly display *H* and *T* with probability $^1/_2$. Show that all four outcome combinations in the table above have the same probability, and deduce that the expected gain to *A* (or *B*) is zero.

(6)

(ii) Suppose now that *A* and *B* independently both randomly display *H* and *T*, but *A* displays *H* and *T* with respective probabilities $^2/_3$ and $^1/_3$ while *B* displays *H* and *T* with respective probabilities $^1/_3$ and $^2/_3$. Calculate the four outcome probabilities under these conditions and find the expected gain to *A*.

(6)

(iii) Finally, suppose that *A* and *B* independently and randomly display *H* and *T* with respective probabilities p_A and $1-p_A$; p_B and $1-p_B$. Calculate the four outcome probabilities and show that the expected gain to *A* may be written:

$$1-3p_A-3p_B+8p_A\,p_B\,.$$

Noting that this may be written as $1-p_A(3-8p_B)-3p_B$ or as $1-p_B(3-8p_A)-3p_A$, discuss whether you would prefer to play as *A* or *B* and how you would play.

(8)

2. (i) Suppose that a random variable X is the number of successes in n independent trials, each of which has a probability of success p. Derive the expected value of X, and write down the variance of X.

(5)

(ii) (a) On a journey to work, a cyclist has to pass through four sets of traffic lights. Assume that the lights operate independently and that at each light there is a probability of $^1/_2$ that the cyclist has to stop. Let Y be the number of lights at which the cyclist has to stop. Find $P(Y \geq 3)$ and $P(Y = 2)$.

(5)

(b) Use the results of part (i) above to obtain the mean and variance of Y.

(2)

(c) Assume now that the probabilities of stopping at the four lights are $^3/_4$, $^1/_3$, $^2/_3$, $^1/_4$, but that the lights still operate independently. Find the mean and variance of the number of lights at which the cyclist has to stop, and compare your results with those of part (ii)(b) above.

(8)

3. A year-group at the local secondary school consists of 100 boys and 81 girls. The heights of the boys may be assumed to be distributed Normally with mean 160 cm and variance 16 cm^2, and the heights of the girls to be distributed Normally with mean 150 cm and variance 9 cm^2.

(i) Find the probability

(a) that a randomly chosen boy is more than 156.0 cm tall, (2)
(b) that a randomly chosen girl is more than 156.0 cm tall, (2)
(c) that a randomly chosen student from the year-group is more than 156.0 cm tall. (2)

(ii) Four boys go to watch a football match. Making a suitable assumption (which should be stated), find the probability

(a) that all four boys are more than 156.0 cm tall, (2)
(b) that their mean height exceeds 156.0 cm. (2)

(iii) A boy and a girl go to a disco. Making a suitable assumption (which should be stated), find the probability that the boy is taller than the girl.

(4)

(iv) Find the probability that the mean height of the year-group is 156 cm to the nearest cm.

(6)

4. (i) The random variable X follows the Poisson distribution with mean $\mu > 0$, so that

$$P(X = x) = e^{-\mu} . \frac{\mu^x}{x!}, \qquad x = 0,1,2,...$$

Sketch the probability function of this distribution in the cases $\mu = 0.5$ and $\mu = 2$.

(7)

(ii) (a) The leaves on a plant are open to infection by a certain type of insect, and the number of insects per leaf has the Poisson distribution for some unknown value of μ. To estimate μ, a random sample of n leaves with insects on them is collected; let the random variable Y denote the number of insects on a randomly collected leaf. Noting that the sample contains no leaves with no insects on them, show that

$$P(Y = y) = \frac{e^{-\mu}}{1 - e^{-\mu}} . \frac{\mu^y}{y!}, \qquad y = 1,2,3...$$

and deduce that the mean number of insects on collected leaves is

$$E(Y) = \frac{\mu}{1 - e^{-\mu}} .$$

(7)

(b) Show further that

$$E(Y) - P(Y = 1) = \mu .$$

(3)

(c) Noting that the sample fraction of leaves with just one insect is an unbiased estimator of $P(Y = 1)$, suggest an unbiased estimator for μ given a random sample $Y_1,...,Y_n$.

(3)

Turn over

5. (i) The random variable X is exponentially distributed with probability density function

$$f(x) = \begin{cases} \lambda e^{-\lambda x}, & 0 \le x < \infty, \\ 0, & x < 0, \end{cases}$$

where $\lambda > 0$.

Obtain the distribution function of X, $F(x)$ say, and sketch the graphs of $f(x)$ and $F(x)$.

(7)

(ii) Given a random sample $x_1, ..., x_n$ from this distribution, write down the likelihood function and show that the maximum likelihood estimate of λ, $\hat{\lambda}$ say, is the reciprocal of the sample mean.

(6)

(iii) Show that the second derivative of the log-likelihood with respect to λ is given by

$$\frac{d^2 \ln L}{d\lambda^2} = -\frac{n}{\lambda^2} \; .$$

You are given that, if n is large,

$$\hat{\lambda} \sim N(\lambda, \; -\left[\frac{d^2 \ln L}{d\lambda^2}\right]^{-1}) \text{ approximately.}$$

By replacing λ by $\hat{\lambda}$ in the variance of this distribution, obtain an approximate 95% confidence interval for λ.

(7)

6. An electronic device consists of two identical components which each have, independently, probability p of being faulty. The device works unless both components are faulty. A random sample of n devices is checked.

(i) Write down the probability distribution of R, the number of devices found to work.

(4)

(ii) There are two separate production lines. Line A uses components from factory A for which $p = ½$ and line B uses components from factory B for which $p = {}^1/_3$. Calculate the probability that a randomly chosen device from the output of line A is found to be faulty, and make the corresponding calculation for line B.

(4)

(Question continued on next page)

(iii) A device is chosen at random from the combined output of both production lines and is found to be faulty. Find the probability that the faulty device came from line *A*,

(a) supposing that lines *A* and *B* produce equal numbers of devices, (4)

(b) supposing that line *B* produces three times as many as line *A*. (4)

Comment on the difference between your answers to (a) and (b). (4)

7. The continuous random variables *X* and *Y* are distributed with a joint probability density function (pdf)

$$f(x,y) = \begin{cases} \dfrac{2}{a^2}, & x \ge 0,\ \ y \ge 0,\ \ x + y \le a \text{ for some positive } a, \\ 0, & \text{otherwise,} \end{cases}$$

that is to say, (X,Y) follows a continuous uniform distribution over the triangle bounded by the lines $x = 0$, $y = 0$, $x+y = a$.

(i) Sketch a graph to show the region in which $f(x,y) > 0$. (3)

(ii) Show that the marginal pdf of *X* is given by

$$f_X(x) = \begin{cases} \dfrac{2(a-x)}{a^2}, & 0 \le x \le a, \\ 0, & \text{otherwise}, \end{cases}$$

and hence obtain the marginal cumulative distribution function of *X*, $F_X(x) = P(X \le x)$. (5)

(iii) Show that $E(X) = \dfrac{a}{3}$ and write down $E(Y)$. (3)

(iv) Show that $E(X^2) = \dfrac{a^2}{6}$ and hence obtain $V(X)$ and $V(Y)$. (4)

(v) Show that $E(XY) = \dfrac{a^2}{12}$ and hence obtain the correlation coefficient between *X* and *Y*. (5)

8. (i) In the multiple linear regression model

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + e_i, \qquad i = 1,\dots,n,$$

state the standard assumptions which are made about the error terms $e_1,\dots,e_n$.

Least squares estimates of β_0, β_1 and β_2, denoted respectively by $\hat{\beta}_0$, $\hat{\beta}_1$ and $\hat{\beta}_2$, are obtained by minimising the quantity

$$S = \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_{1i} - \beta_2 x_{2i})^2.$$

Obtain the equation $\dfrac{\partial S}{\partial \beta_0} = 0$ and deduce that

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}_1 - \hat{\beta}_2 \bar{x}_2,$$

where $\bar{y}$, $\bar{x}_1$ and $\bar{x}_2$ denote sample mean values.

By substituting for $\hat{\beta}_0$ in the equations $\dfrac{\partial S}{\partial \beta_1} = 0$, $\dfrac{\partial S}{\partial \beta_2} = 0$, show that $\hat{\beta}_1$ and $\hat{\beta}_2$ satisfy the equations

$$\sum_{i=1}^{n} (y_i - \bar{y})(x_{1i} - \bar{x}_1) = \hat{\beta}_1 \sum_{i=1}^{n} (x_{1i} - \bar{x}_1)^2 + \hat{\beta}_2 \sum_{i=1}^{n} (x_{1i} - \bar{x}_1)(x_{2i} - \bar{x}_2)$$

$$\sum_{i=1}^{n} (y_i - \bar{y})(x_{2i} - \bar{x}_2) = \hat{\beta}_1 \sum_{i=1}^{n} (x_{1i} - \bar{x}_1)(x_{2i} - \bar{x}_2) + \hat{\beta}_2 \sum_{i=1}^{n} (x_{2i} - \bar{x}_2)^2.$$

(8)

(ii) Carry out a multiple linear regression of Computer Aptitude Score on Verbal Ability and Arithmetic Ability using the data below.

Student:	1	2	3	4	5	6	7	8	Mean
Computer aptitude score (y):	22	24	34	48	58	76	82	96	55
Verbal ability (x_1):	53	52	58	60	57	61	66	65	59
Arithmetic ability (x_2):	50	45	51	55	62	70	73	74	60

You may assume that

$\sum y_i^2 = 29600$, $\sum x_{1i}^2 = 28028$, $\sum x_{2i}^2 = 29700$,

$\sum x_{1i} x_{2i} = 28680$, $\sum x_{1i} y_i = 26860$, $\sum x_{2i} y_i = 28560$.

Given that the conventional unbiased estimate of var(e_i) is 36 and the estimated sampling variances of the two regression coefficients are 1 and 0.2 respectively, test the whole regression for significance and make partial t-tests of the two coefficients. Comment briefly on your results.

(12)