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

HIGHER CERTIFICATE IN STATISTICS, 1998

CERTIFICATE IN OFFICIAL STATISTICS, 1998

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

1. A company runs a competition advertised in three national daily newspapers A, B and C, which have large readerships in the proportions 2:3:1 respectively. Each reader reads one newspaper only. The proportions of the readerships of A, B and C who enter the competition are 0.02, 0.01 and 0.05 respectively. You may assume that all entries to the competition are correct, and if drawn would receive a prize. All entries are equally likely to be drawn.

   (i) What are the probabilities that the top prize is won by a reader of paper A, B or C respectively?

   (ii) What is the probability that the top prize and the second prize are both won by readers of paper A?

   (iii) What is the probability that the top prize and the second prize are won by readers of two different papers?

   (iv) What is the probability that the top three prizes are won by readers of all three different papers?

2. Paget's disease is estimated to affect 10% of persons over 65 years old. It is known that, among persons over 65 without the disease, a blood measure X in suitable units is distributed N(7,9) (i.e. Normally, with mean 7 and variance 9) but for persons over 65 suffering from the disease, X is distributed N(19,36). A value of $X \geq 10$ is taken as a cause for further investigation.

   (i) What proportion of non-sufferers over 65 will be investigated because of their value of X?

   (ii) What proportion of sufferers over 65 will go undiagnosed?

   (iii) At what critical level of X (instead of 10) would the false positive rate be 5% (i.e. 5% of non-sufferers would have X above this level)?

   (iv) If all people over 65 are screened using $X \geq 10$ as the criterion for further investigation, what proportion will be investigated?

   (v) Of those investigated further, what is the proportion of individuals that will actually have the disease?

Turn over

3. Manufactured items are submitted for inspection in large batches. Any item can have a major defect or a minor defect or both; these defects occur independently with probabilities p_1 and p_2 respectively. Occurrences of defects in different items are also independent. As part of the inspection process, a random sample of 10 items is selected from each batch. If a major defect is found the batch is rejected and subjected to full inspection. If the sample contains only one defect, a minor one, a second sample of 10 items is taken and if this contains any defect the batch is again rejected. Derive an expression for the total probability of rejection under this scheme, and evaluate this probability for the four cases (p_1, p_2) where p_1 = 1% and 2% and p_2=2% and 5%.

   Compare the above with a scheme where a single sample of 20 items is taken and the batch is rejected if a major defect or more than one minor defect is found.

4. (a) A manufacturer of colour television sets offers a one year warranty which covers free replacement if the picture tube fails and requires replacement. He estimates the time in years until the tube fails, T, to be a random variable with probability density function

$$\begin{aligned} f(t) &= \frac{1}{5}.\exp(-t/5),\ t \geq 0 \\ &= 0 \quad \text{otherwise.} \end{aligned}$$

   (i) Find the probability that a randomly chosen set will require at least one replacement tube under the warranty.

   (ii) If the gross profit per sale is £100 and the replacement of a tube costs £70, find the expected net profit per set sold, assuming that no set requires more than one replacement under the warranty.

   (b) The number of days required to complete a project is denoted by X, where X is assumed to have the distribution

x	10	11	12	13	14	otherwise
p(X= x)	0.2	0.4	0.2	0.1	0.1	0

   The return from the project is Y = £2000 (12 − X).

   (i) Find the probability distribution of Y and obtain the probability of making a loss.

   (ii) Find E(X), V(X), E(Y) and V(Y).

5. The random variable X has a Poisson distribution with parameter λ, so that

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

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

(i) Find $E(X)$.

(ii) If $p(X = k) = p(X = k + 1)$, where k is some integer, show that λ must be the integer $k + 1$.

(iii) If λ is not an integer, show that the mode, m, of the distribution satisfies the inequality

$$\lambda - 1 < m < \lambda .$$

.

(iv) In the manufacture of curtain material, small flaws occur at random in the material at a mean rate of 1 per 5m^2. Find the probability that in a randomly selected 5m^2 area of this material

(a) there are no faults,

(b) there are at most 2 faults.

(v) A new hotel requires curtains for 50 rooms. All the curtains will be made from the material described above, and each room's curtains will use 20m^2 of material.

(a) Find the probability that the curtains in any given room will contain 3 or more faults.

(b) Use a suitable approximation to find the probability that more than 40 of the rooms will have curtains containing 3 or more faults.

Turn over

6. (a) The simple linear regression model

$$y = a + b\,x + \text{error}$$

is often fitted by least squares. Under what circumstances, which are often assumed, will doing so have conventional optimal properties?

In a simple linear regression analysis the data (x_i, y_i), $i = 1, \ldots, n$ are transformed to coded variables

$$Y = y - y_o,$$
$$X = k\,(x - x_o),$$

where x_o, y_o and k are known constants chosen for convenience of analysis. Given that the fitted regression in coded form is

$$\hat{Y} = \hat{A} + \hat{B}\,X$$

with estimated mean squared error S^2, obtain the estimates $\hat{a}, \hat{b}$ and mean square error s^2 relating to the original data in terms of $\hat{A}, \hat{B}$, S^2, x_o, y_o and k.

(b) The weight of a rabbit is measured at different ages, the results being as shown.

Age (days)	84	91	98	105	112	119
Weight (grams)	527	555	585	615	640	666

Fit a simple linear regression to these data and estimate the residual mean squared error. Also compute and plot the estimated residuals, and comment on the fit.

7. The joint distribution of the random variables X and Y is specified in the table below, the values tabulated being the probabilities p(X=x, Y=y),

x y:	1	2	3	4
1	$3k$	$10k$	$21k$	$36k$
2	$10k$	$36k$	$78k$	$136k$
3	$21k$	$78k$	$171k$	$300k$

where k is a constant.

(a) Find k.

(b) Obtain the marginal distributions of X and Y.

(c) Find E(X), V(X), E(Y), V(Y) and the correlation between X and Y.

(d) Find the conditional distribution of X given $Y = 1$.

(e) Find $E(X \mid Y = 1)$.

8. The random variable X follows the Geometric distribution such that

$$p(X = x) = q^{x-1}p, \quad x = 1,2,3, \ldots ,$$

$$= 0 \text{ otherwise,}$$

where $0 < p < 1$ and $q = 1 - p$. Sketch the probability function of this distribution and, by using the result

$$1+2q + 3q^2 + 4q^3 + \ldots = \frac{1}{(1-q)^2},$$

or otherwise, find E(X).

Obtain the cumulative distribution function of X and deduce that the median of X is the smallest integer not less than $-\dfrac{\ln 2}{\ln q}$.

The random variable Y has the same distribution as X and is independent of X. Show that

$$P(Y = X) = \frac{p}{1+q} .$$