EXAMINATIONS OF THE ROYAL STATISTICAL SOCIETY

HIGHER CERTIFICATE IN STATISTICS, 2008

(Modular format)

MODULE 3 : Basic statistical methods

Time allowed: One and a half hours

Candidates should answer **THREE** *questions.*

Each question carries 20 marks.
The number of marks allotted for each part-question is shown in brackets.

Graph paper and Official tables are provided.

Candidates may use calculators in accordance with the regulations published in the Society's "Guide to Examinations" (document Ex1).

The notation log *denotes logarithm to base* ***e***.
Logarithms to any other base are explicitly identified, e.g. $\log_{10}$.

Note also that $\binom{n}{r}$ *is the same as* ${}^nC_r$.

This examination paper consists of 3 printed pages **each printed on one side only**.
This front cover is page 1.
Question 1 starts on page 2.

There are 4 questions altogether in the paper.

1. A population consists of the five values 1, 4, 9, 16 and 25.

(i) Calculate the population mean and variance. (3)

(ii) Write down all the samples of size two that may be drawn, with replacement, from this population, and calculate the sample mean of each. (6)

(iii) Let $\overline{X}$ denote the mean of a random sample of size two drawn, with replacement, from this population. Write down the expected value and variance of $\overline{X}$. (3)

(iv) For $\overline{X}$ as in part (iii), find $P(\overline{X} > 16.5)$. Find also an approximation to this quantity, using an appropriate Normal approximation [use of a continuity correction is not expected], and comment briefly on your results. (8)

2. (i) Two independent random samples are taken from separate populations. The first sample, of size n_1, is taken from a population in which the proportion of members with a certain attribute A is p_1. The second sample, of size n_2, is from a population containing proportion p_2 with attribute A. There are R_1 members of the first sample, and R_2 of the second, with attribute A.

Write down a Normal approximation to the distribution of the difference $D = \frac{R_1}{n_1} - \frac{R_2}{n_2}$, and deduce an approximate test of the null hypothesis that the proportion who possess attribute A is the same in each population, against a two-sided alternative hypothesis, when n_1 and n_2 are large.

(6)

(ii) Random samples of 1000 males and 1000 females were assessed for colour-blindness, and the results are shown in the following table.

	Males	*Females*
Colour-blind	80	6
Not colour-blind	920	994

Use the result of part (i) to test at the 0.1% level the hypothesis of equality of the proportions of colour-blind among all males and among all females, against the alternative that the proportions differ. State your conclusions clearly.

(7)

(iii) Use the data given in part (ii) to construct an approximate 95% confidence interval for the difference in the proportions of colour-blind amongst all the males and amongst all the females. Interpret this confidence interval.

(7)

Turn over

3. In a survey of borrowers of books from two public libraries, a random sample of borrowing transactions (all of independent individuals) is taken from each library and the borrowing times are noted. Data for each library are put in ascending order for convenience.

	Borrowing Times (days)	*Total*
Library 1	2, 3, 5, 10, 15, 21, 26, 38	120
Library 2	1, 4, 6, 7, 12, 18, 27, 36, 42	153

You are given that the respective sums of squares of the observations for libraries 1 and 2 are 2924 and 4359.

(i) A trainee librarian who once attended a statistics course suggests using a t test to examine the hypothesis that the mean borrowing times for the two libraries are equal, against the alternative that they differ. State the assumptions necessary for the validity of this test. Perform the test and report your conclusions.

(8)

(ii) Perform a suitable non-parametric test of the hypothesis that the median borrowing times for the two libraries are equal, against the alternative that they differ. Report your conclusions and state the assumptions necessary for the validity of this test.

(8)

(iii) Discuss critically which test you regard as more appropriate. Provide graphical and/or other evidence to support your argument.

(4)

4. The weights in grams of a random sample of 10 cherry tomatoes taken from the plants in a large greenhouse are 37, 39, 40, 33, 41, 39, 36, 46, 35 and 44.

(i) Calculate the mean and variance of the observed weights.

(4)

(ii) Assuming that the distribution underlying the data may be taken as approximately Normal, give 95% confidence intervals for (a) the mean and (b) the variance of the weights of cherry tomatoes in the greenhouse. State clearly any formulae you use for these calculations.

(9)

(iii) In order to be graded as acceptable for sale, the cherry tomatoes should not be too small or very variable in size. Test, in each case at the 5% level of significance, (a) the hypothesis that the mean weight is 42 g against the hypothesis that it is less than 42 g, (b) the hypothesis that the variance of weight is 12 g^2, against the hypothesis that it is greater than 12 g^2.

(7)