EXAMINATIONS OF THE ROYAL STATISTICAL SOCIETY

HIGHER CERTIFICATE IN STATISTICS, 2010

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* ${}^{n}C_{r}$.

This examination paper consists of 5 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. At the beginning of an academic year, students studying mathematics in their final year at a certain school were split into two tutorial groups, Group A (7 students) and Group B (11 students), with different teachers for the two groups. They all took the same mathematics examination at the end of the year and achieved the percentage marks tabulated below, where the marks for each group have been sorted into increasing order.

The question to be investigated is whether there is significant evidence of a difference in the overall levels of performance between the two groups.

Examination marks

	Group A	*Group B*
	12	36
	31	39
	34	46
	35	48
	55	58
	57	66
	61	69
		70
		77
		84
		85
Sample mean	40.71	61.64
Sample variance	312.90	303.85

(i) Stating carefully your null and alternative hypotheses, carry out an appropriate *t* test and draw conclusions.

(8)

(ii) State the assumptions made in carrying out the *t* test and comment briefly on whether they appear to be satisfied. If instead the Wilcoxon rank sum test is to be used, specify what assumptions have to be made for this test to be valid.

(5)

(iii) Re-analyse the data using the Wilcoxon rank sum test. Again state carefully your null and alternative hypotheses and draw conclusions.

(7)

2. Manuscripts of a certain biblical text were produced in two forms, Type A and Type B. In archaeological excavations over a number of Egyptian sites, 10 copies of this text have been found, 2 of which are of Type A and 8 of Type B. Assuming that these 10 copies may be regarded as a random sample from a large number of copies of the text that were originally produced, it is to be decided whether the observed numbers provide significant evidence that the proportion of Type A manuscripts originally produced differed from the proportion of Type B manuscripts produced.

(i) Under the hypothesis that Type A manuscripts constituted a proportion θ of the manuscripts produced, what is the distribution of the number of Type A manuscripts found in a random sample of size 10?

(2)

(ii) Specify the null and alternative hypotheses to be tested in terms of the parameters of the distribution specified in part (i).

(2)

(iii) Using the appropriate statistical table, find the p-value that corresponds to the observed data and draw conclusions.

(6)

After further excavations, out of a total of 100 copies of the text that have been discovered, it is found that 39 are of Type A and 61 are of Type B. It is to be considered again whether there is significant evidence that the proportion of Type A manuscripts originally produced differed from the proportion of Type B manuscripts produced.

(iv) In the larger sample of size 100, again assuming that Type A manuscripts constitute a proportion θ of the manuscripts produced, what is the approximate distribution of the number of Type A manuscripts found in the sample?

(3)

(v) Using the same null and alternative hypotheses as specified in part (ii) and the appropriate statistical table, find the p-value that corresponds to the observed data and draw conclusions.

(7)

3. The pseudorandom number generator in a spreadsheet package is used to produce a sequence of random digits. If the pseudorandom number generator is behaving as it should, each member of the sequence is equally likely to take any of the ten values 0, 1, …, 9. A sequence of 100 such random digits is produced and the observed frequency distribution of the ten digits is given in the table below.

digit	0	1	2	3	4	5	6	7	8	9
frequency	10	7	8	12	10	13	7	4	15	14

(i) Calculate the expected frequency for each digit if the pseudorandom number generator is behaving as it should, explaining your reasoning.

(3)

(ii) Calculate an appropriate test statistic to decide whether the observed frequencies are consistent with the assumed properties of the pseudorandom number generator, and draw conclusions.

(8)

The pseudorandom number generator can also be used to generate standard Normal deviates. To test how well it simulates the tails of the Normal distribution, a random sample of 1000 standard Normal values was simulated. It was found that 54 of these values exceeded 2 in absolute value.

(iii) Use an appropriate statistical table to evaluate the probability that a standard Normal variate exceeds 2 in absolute value.

(2)

(iv) Carry out a statistical test to decide whether the observed number of 54 simulated values that exceeded 2 in absolute value provides significant evidence that the pseudorandom number generator is not performing as it should.

(7)

4. (a) Explain what is meant by the *sampling distribution* of a sample statistic. State the sampling distribution of the sample mean $\bar{X}$ and sample variance S^2 in a random sample of size n from a Normal distribution with mean μ and variance σ^2.

(6)

(b) A random sample of 100 bags of flour of Brand A is weighed. It is found that the sample mean is 453.08 gm and the sample standard deviation is 5.42 gm. It is assumed that the weights of the bags of flour are Normally distributed with unknown mean μ and unknown variance σ^2.

(i) Find a 95% confidence interval for μ.

(4)

(ii) Find a 95% confidence interval for σ^2.

(5)

(iii) It is suspected that the weight of flour in bags of Brand B is more variable from bag to bag than for Brand A. A random sample of 20 bags of Brand B is weighed, and it is found that the sample standard deviation is 11.15 gm.

Describing briefly the underlying distribution theory, carry out a test to determine whether there is significant evidence that the variance of all weights of bags is greater for Brand B than for Brand A.

(5)