EXAMINATIONS OF THE HONG KONG STATISTICAL SOCIETY

HIGHER CERTIFICATE IN STATISTICS, 2011

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. Ten plants of a certain type, grown under standard conditions and treated with a new brand of fertiliser, attain the following heights (in cm).

    25   28   24   23   27   30   24   21   28   30

    The sample mean is 26.0 and the sample variance is 9.33. From extensive previous experiments, it is known that plants of the same type, grown under similar conditions but treated with a standard fertiliser, attain a mean height of 25.0 cm.

    (i)   Write down a statistical model for the distribution of plant heights in the population from which the sample is assumed to be drawn, including the definition of any unknown parameters.

    (3)

    (ii) Obtain a 99% confidence interval for the population mean height of plants treated with the new fertiliser.

    (4)

    (iii) Carefully specifying your hypotheses, conduct a test at the 10% significance level to decide whether there is evidence to suggest that the new fertiliser produces plants of greater mean height than the standard fertiliser. State your conclusions.

    (10)

    (iv) Provide a 95% confidence interval for the population variance of the heights of plants treated with the new brand of fertiliser.

    (3)

2

Turn over

2. Twenty cows were used in an experiment to compare two types of feed, Feed A and Feed B. Half of the cows, chosen at random, were fed with Feed A over a certain period of time and the other half with Feed B. In fact two of the cows on Feed B were wrongly fed for part of the period and they had to be removed from the experiment. The gains in weight, in pounds, of the remaining cows over the period of the experiment are listed below. The corresponding sample means and variances are also given.

Feed A	30	26	30	19	25	37	27	38	26	31
Feed B	40	34	28	29	26	36	28	37		

	sample mean	sample variance
Feed A	28.9	32.1
Feed B	32.3	26.5

(i)   Defining any unknown parameters, write down a statistical model for the distributions of weight gains in the two populations of cows given the two feeds, assuming that the variances of these two distributions are equal.

(3)

(ii) Carefully specifying your hypotheses, test at the 5% significance level whether there is a difference between the mean weight increases for the two feeds.

(10)

(iii) It is now required to test at the 5% significance level the assumption that the variances of the two weight gain distributions are equal. State any adjustments that are needed to the model in part (i), carefully specify the hypotheses to be tested, carry out the test and give your conclusions.

(7)

3. The data below give the daily numbers of homicides in London for the 1095 days from 1 April 2004 to 31 March 2007.

Daily number of homicides	Frequency
0	713
1	299
2	66
3	16
4	1
≥ 5	0

(i) Calculate, to 4 decimal places, the mean number of homicides per day.

(2)

(ii) If homicides are assumed to happen as random events, then the daily numbers of homicides should follow a Poisson distribution. Under the hypothesis that the data are a random sample from a Poisson distribution, calculate the expected frequencies that correspond to the observed frequencies in the table above.

(6)

(iii) Carry out a formal test at the 5% significance level to investigate whether or not homicides occurred as random events.

(8)

(iv) For a random sample of 139 homicides where the method of killing was known, 29 of the homicides were by shooting. Find a 95% confidence interval for the underlying population proportion of homicides that are by shooting.

(4)

4

4. It is commonly assumed in biblical scholarship that the authors of the gospels of Matthew and Luke independently of each other made use of the earlier gospel of Mark, and it has been suggested that, if this was so, then Matthew tended to follow the text of Mark more closely than did Luke. To test whether this suggestion is supported by statistical evidence, a random sample of 15 sections of text were taken from Mark and compared with each of the parallel sections in Matthew and Luke. A numerical measure of similarity, which takes values between 0 and 1, was used to construct the two variables, "MtSim", which measures the similarity between Matthew and Mark, and "LkSim", which measures the similarity between Luke and Mark. The values of these two variables are given below for each of the selected sections of text, larger values indicating greater similarity.

MtSim	LkSim
0.106	0.199
0.499	0.604
0.465	0.382
0.299	0.301
0.472	0.100
0.642	0.529
0.590	0.386
0.233	0.322
0.308	0.094
0.319	0.128
0.402	0.459
0.672	0.252
0.772	0.483
0.408	0.314
0.337	0.293

(i) Explain why it may be deemed appropriate to use the Wilcoxon signed-rank test for these data and state what null and alternative hypotheses are being tested.

(4)

(ii) Carry out the Wilcoxon signed-rank test at the 5% significance level and state your conclusions.

(8)

(iii) As an alternative to the signed-rank test, carry out the sign test at the 5% significance level and state your conclusions.

(5)

(iv) Comment briefly on why the conclusions differ when the sign test is used instead of the signed-rank test.

(3)