EXAMINATIONS OF THE HONG KONG STATISTICAL SOCIETY

HIGHER CERTIFICATE IN STATISTICS, 2009

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. In medieval times, before the advent of printing, scribes copied books by hand, and errors naturally occurred in the process of copying. In studying a book and one particular copy of it, a random sample of 100 pages was examined and the numbers of errors per page were recorded. The data are summarised in the following table.

Number of errors, x, per page	*Frequency, f_x*
0	10
1	32
2	24
3	22
4	9
5	2
6	0
7	0
8	1

(i) Explain why it might be reasonable to assume that the number of errors per page would follow a Poisson distribution.

(2)

(ii) Calculate for these data the sample mean of the number of errors per page.

(2)

(iii) Carry out a χ^2 goodness-of-fit test to examine the null hypothesis that the number of errors per page has a Poisson distribution. Show all your working clearly and report your conclusion.

(16)

2. A random sample of twenty bottles of a medicine that comes in liquid form is taken and the contents of each bottle are accurately weighed. The sample mean weight (in grams) of the contents of a bottle is 99.43 and the sample variance is 0.4678. You may assume that the weights are Normally distributed.

(i) Calculate a 95% confidence interval for the underlying population mean weight of medicine in a bottle.

(5)

(ii) Calculate a 95% confidence interval for the underlying population variance of the weight of medicine in a bottle.

(5)

(iii) According to the manufacturer, the underlying mean weight of medicine in a bottle is exactly 100 grams. It is suspected that the underlying mean weight of medicine is in fact less than 100 grams. Perform an appropriate test of the manufacturer's claim at the 1% significance level, stating the null and alternative hypotheses. Report your conclusions.

(7)

(iv) State what is meant by the *p*-value of a test statistic. Using the appropriate statistical table, deduce the shortest range of values within which the *p*-value lies for the test statistic calculated in part (iii).

(3)

3. (i) A random sample of 50 households is taken in each of two districts, A and B, and for each household the presence or absence of digital television is recorded. The data are presented in the table below.

		District A	*District B*
Digital TV	*Yes*	38	27
	No	12	23

Investigate whether there is significant evidence of a difference between the two districts in the uptake of digital TV. Briefly state your conclusions.

(8)

(ii) Using the data of part (i), find an estimate and an approximate 95% confidence interval for the proportion of households in district B which have digital television.

(6)

(iii) In another study, 100 matched pairs of households have been selected such that the first household of each pair is from district C and the second is from district D. Matching has been done on the basis of certain socio-economic variables. For each household the presence or absence of broadband is recorded. The data are presented in the table below.

		Broadband in C	
		Yes	*No*
Broadband in D	*Yes*	57	9
	No	15	19

Investigate whether there is significant evidence of a difference between the two districts in the levels of uptake of broadband. Briefly state your conclusions.

(6)

4. Two examiners, A and B, independently mark student projects, in each case assigning a mark out of 100 for the project. The marks assigned by the examiners for each of 12 randomly selected student projects are recorded below.

Student	*Examiner A*	*Examiner B*
1	53	42
2	54	44
3	39	42
4	47	56
5	72	65
6	61	63
7	64	58
8	51	59
9	68	89
10	59	87
11	58	72
12	53	79

The issue has been raised as to whether one of the examiners tends to give higher marks than the other.

(i) Noting any assumptions made, carry out an appropriate parametric test and draw conclusions.

(9)

(ii) Specify an appropriate non-parametric test that may be used instead. Carry out this test and draw conclusions.

(8)

(iii) Comment briefly on the advantages and disadvantages of the above two tests. In the present case, does it make any difference to your conclusions which of these tests you use?

(3)