

# EXAMINATIONS OF THE HONG KONG STATISTICAL SOCIETY



## HIGHER CERTIFICATE IN STATISTICS, 2006

### Paper I : Statistical Theory

**Time Allowed: Three Hours**

*Candidates should answer **FIVE** questions.*

*All questions carry equal 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 6 printed pages **each printed on one side only**.

This front cover is page 1.

Question 1 starts on page 2.

There are 8 questions altogether in the paper.

1. Credit card holders with a certain bank are each assigned a 4-digit PIN (personal identity number). How many different possible PINs are there under each of the following conditions?
- (i) The first digit may not be zero. (3)
  - (ii) The first digit may not be zero, and the four digits may not all be the same. (3)
  - (iii) No zeros are allowed in any position and all four digits must be different. (4)
  - (iv) Zeros are allowed in all positions but sequences of four consecutive digits up (e.g. 2 3 4 5) or down (e.g. 3 2 1 0) are not allowed. (4)
  - (v) Zeros are allowed in all positions but no digit may occur more than twice. (6)
2. In a hi-tech company, the members of three research groups (A, B and C) are individually invited to enter a prize competition for the best solution to a technical problem. Group A has 2 staff, B has 3 and C has 5. It is assumed that all staff decide independently whether or not to enter. Members of groups A, B and C enter with respective probabilities  $1/2$ ,  $1/4$  and  $1/5$ .
- (i) For each group separately, find the probability of (a) no entries, (b) one entry. (8)
  - (ii) Given that there is just one entry in total, show that the probability that it comes from a member of group A is  $8/17$ . (6)
  - (iii) Explain (but without doing the calculations) the steps that are needed to calculate the probability that there are exactly two entries in total. (6)

3. The continuous random variable  $X$  has probability density function given by

$$f(x) = \begin{cases} kx^2(1-x)^2, & 0 \leq x \leq 1, \\ 0, & \text{elsewhere.} \end{cases}$$

(i) Find  $k$  and sketch the graph of  $f(x)$ . (7)

(ii) Find  $E(X)$  and  $\text{Var}(X)$ , and show that  $P\left(X \leq \frac{1}{3}\right) = \frac{17}{81}$ . (8)

(iii) A random sample of size 5 is taken from this distribution. Find, correct to 4 decimal places, the probability that all 5 observations exceed  $1/3$ . (3)

(iv) Find, correct to 4 decimal places, the variance of the mean of a random sample of size 5. (2)

4. My cycle journey to work is 3 km, and my cycling time (in minutes) if there are no delays is distributed  $N(15, 1)$ , i.e. Normally with mean 15 and variance 1.

(i) Find the probability that, if there are no delays, I get to work in at most 17 minutes. (2)

(ii) On my route there are three sets of traffic lights. Each time I meet a red traffic light, I am delayed by a random time that is distributed  $N(0.7, 0.09)$ . These lights operate independently. Find the probability of my getting to work in at most 17 minutes

(a) if just one light is set at red when I reach it,

(b) if just two lights are set at red when I reach them,

(c) if all three lights are set at red when I reach them. (9)

(iii) Suppose that, for each set of lights, the chance of delay is 0.5. Deduce that the mean value of  $T$ , my total journey time, is 16.05 minutes. (4)

(iv) Given that  $\text{Var}(T) = 1.5025$ , use a suitable approximation to calculate the probability that, over 10 journeys, my average journey time to work is at most 17 minutes. (5)

5. The random variable  $X$  follows a Poisson distribution with probability mass function

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

- (i) Show that  $E(X) = \text{Var}(X) = \lambda$ . (5)

- (ii) Given a random sample of values  $x_1, x_2, \dots, x_n$  from this distribution, obtain the maximum likelihood estimator (MLE) of  $\lambda$ ,  $\hat{\lambda}$  say. (4)

- (iii) Write down  $\text{Var}(\hat{\lambda})$  as a function of  $\lambda$ , and hence find the MLE of  $\text{Var}(\hat{\lambda})$ . Show that a large-sample approximate 95% confidence interval (CI) for  $\lambda$  is given by

$$\hat{\lambda} - 1.96\sqrt{\frac{\hat{\lambda}}{n}} \leq \lambda \leq \hat{\lambda} + 1.96\sqrt{\frac{\hat{\lambda}}{n}}. \quad (4)$$

- (iv) Assume that the numbers of books,  $x_1, x_2, x_3, \dots$ , that go missing each month from the local library follow a Poisson distribution with unknown mean  $\lambda$ . The monthly numbers of missing books in 2005 were

3   7   2   5   8   2   4   5   4   4   1   3.

Use these data to calculate  $\hat{\lambda}$  and an approximate 95% CI for  $\lambda$ . Also compute the sample variance of the data; discuss briefly whether this computation supports, or throws doubt on, the Poisson model suggested (no formal test is required).

(7)

6. The random variable  $X$  has the distribution with probability density function

$$f(x) = \frac{\lambda}{2} e^{-\lambda|x|}, \quad -\infty < x < \infty.$$

Sketch a graph of this density function.

(4)

Write down  $E(X)$  and show that  $\text{Var}(X) = \frac{2}{\lambda^2}$ . Find also the semi-interquartile range of  $X$ .

(9)

A random sample  $x_1, x_2, \dots, x_n$  is taken from this distribution. Show that the maximum likelihood estimate of  $\lambda$  is given by

$$\hat{\lambda} = \frac{n}{\sum_{i=1}^n |x_i|}.$$

(7)

7. The table below shows the joint distribution of two random variables,  $X$  and  $Y$ .

		<i>Values of Y</i>			
		1	2	3	4
<i>Values of X</i>	1	$6c$	$3c$	$2c$	$4c$
	2	$4c$	$2c$	$4c$	0
	3	$2c$	$c$	0	$2c$

- (i) Find  $c$ .

(2)

- (ii) Calculate the marginal distributions of  $X$  and  $Y$ .

(5)

- (iii) Calculate  $E(X)$  and  $\text{Var}(X)$ , and show that the covariance  $\text{Cov}(X, Y) = 0$ .

(6)

- (iv) State, with a reason, whether or not  $X$  and  $Y$  are independent.

(2)

- (v) The random variables  $U$  and  $V$  are defined by

$$U = 1 \text{ if } X = 1 \text{ or } 3, \quad U = 0 \text{ if } X = 2,$$

$$V = 1 \text{ if } Y = 1 \text{ or } 3, \quad V = 0 \text{ if } Y = 2 \text{ or } 4.$$

Tabulate the joint distribution of  $U$  and  $V$  and state with a reason whether or not  $U$  and  $V$  are independent.

(5)

8. (i) Write down the model for, and standard assumptions of, simple linear regression analysis. State a condition under which the method of least squares is equivalent to the method of maximum likelihood for estimating the regression coefficients.

(4)

- (ii) (a) Suppose now that the intercept parameter in the regression model is known to be zero, so that the model becomes

$$y_i = \beta x_i + e_i,$$

where the usual assumptions apply to  $e_i$ . Show that the least squares estimator of  $\beta$  is  $\frac{\sum x_i y_i}{\sum x_i^2}$ .

(6)

- (b) Over a period of one month, a survey was made on each of ten main roads in a large city. Each road was observed for a one-hour period randomly chosen during the working day. For each road, the mean traffic flow,  $x_i$  (in vehicles per minute), and the number of speed limit violations,  $y_i$ ,  $i = 1, 2, \dots, 10$ , were recorded. Plot the data shown below on a suitable graph and comment on the suitability of the above model. Fit the model to the data and hence estimate the expected number of violations on a road with an average traffic flow of 20 vehicles per minute. *Without* any further calculation, comment on the suggestion that an intercept should be included in the model.

Flow, $x$	5	5	5	10	10	15	25	25	30	50
Violations, $y$	2	1	1	4	2	5	8	2	5	10

(10)