

EXAMINATIONS OF THE HONG KONG STATISTICAL SOCIETY



HIGHER CERTIFICATE IN STATISTICS, 2008

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 7 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. Three fair six-sided dice are rolled independently.
- (i) Find the probability that
- (a) all three dice show the same number, (2)
- (b) all three dice show different numbers, (3)
- (c) exactly two of the numbers shown are the same. (4)

(ii) Let X be a random variable denoting the maximum number shown when the three dice are rolled.

(a) Explain clearly why

$$P(X \leq x) = \left(\frac{x}{6}\right)^3, \quad x = 1, 2, \dots, 6. \quad (4)$$

(b) Deduce the probability mass function of X , and hence find $E(X)$.

[Hint. You may note that $P(X \leq 0) = 0$, and use the fact that $P(X = x) = P(X \leq x) - P(X \leq x - 1)$ for $x = 1, 2, \dots, 6$.]

(7)

2. Tom, Dick and Harry are going for a hike on the moors, and each takes with him a mobile phone for communication in case of an emergency. However, the boys do not ensure that the batteries of their phones are fully charged before they set out. The respective probabilities that the batteries fail during the hike are 0.5, 0.6 and 0.7, and battery failures are assumed to be independent.

(i) Let X be the number of batteries that fail. Find the probability mass function of X .

(12)

(ii) Hence find $E(X)$ and $\text{Var}(X)$.

(8)

3. On the island of Newtopia, the height X in cm of a randomly chosen man has a Normal distribution with mean 178 and variance 100, that is $X \sim N(178, 100)$; and the height Y in cm of a randomly chosen woman is, in the same notation, distributed $N(172, 56.25)$.

(i) Write down the distribution of $X - Y$, and hence calculate $P(X < Y)$. (4)

(ii) The Newtopian Police Force has a minimum height requirement of 183 cm. For a group of 4 randomly chosen men on the island, find the probability that (a) all 4 men, (b) exactly 2 men, satisfy this requirement. Find also (c) the probability that the mean height of the group exceeds 183 cm. (7)

(iii) There are twice as many women as men on the island. Let Z denote the height in cm of a randomly chosen adult. Calculate $E(Z)$. By using the formula

$$E(Z^2) = (E(X^2) \times (1/3)) + (E(Y^2) \times (2/3)),$$

evaluate $E(Z^2)$ and hence deduce $\text{Var}(Z) = 78.8$. (5)

(iv) (a) State with a reason whether or not Z is Normally distributed. (2)

(b) Use the central limit theorem to obtain an approximate distribution for the mean height of a random sample of 100 of the adults. (2)

4. (i) Suppose that X and Y independently have Poisson distributions with means λ and μ respectively, and that $W = X + Y$. Use the relation

$$P(W = w) = \sum_{x=0}^w P(X = x)P(Y = w - x)$$

to show that W has a Poisson distribution, and write down its mean.

(4)

- (ii) An experienced secretary and a trainee work in an office. The number of typing errors per page of work made by the experienced secretary is assumed to be a Poisson random variable with mean 1. For the trainee the corresponding distribution is Poisson with mean 3. It may be assumed that errors made by the experienced secretary and by the trainee are independent.

- (a) Find the probability that a single page typed by the trainee contains exactly 2 errors.

(2)

- (b) Find the probability that a single page typed by the trainee contains more than 2 errors.

(2)

- (c) Using the result in part (i), write down the distribution of the number of errors in two pages typed by the experienced secretary, and calculate the probability that the two pages contain no errors.

(2)

- (iii) In 10 minutes, the experienced secretary types two pages and the trainee types one page.

- (a) Write down the distribution of the total number of errors in the three pages typed during this time, and calculate the probability that a total of 5 errors is made.

(5)

- (b) Given that the three pages contain exactly 5 errors, find the probability that all 5 errors are made by the trainee.

(5)

5. (i) The random variable X has the exponential distribution with probability density function (pdf)

$$f(x) = \lambda e^{-\lambda x}, \quad x > 0, \quad \lambda > 0.$$

Show that

(a) $\log[P(X > x)] = -\lambda x, \quad x > 0,$

(b) $E(X) = \frac{1}{\lambda},$

(c) $\text{Var}(X) = \frac{1}{\lambda^2}.$

Hence write down the coefficient of variation of X .

(9)

- (ii) A random sample of size n is drawn from a population with the above pdf. Obtain the maximum likelihood estimator of λ .

(7)

- (iii) A random sample of components was taken from a large batch of a certain type of electronic component and the lifetimes in hours of the selected components were measured, yielding a sample mean and sample standard deviation of 1050 and 150 hours respectively. Before collecting the data an engineer suggested that the lifetimes of the components might be exponentially distributed. Calculate the sample coefficient of variation, and comment briefly on the suitability of the exponential model for these data.

(4)

6. A geometric random variable X has probability mass function given by

$$P(X = x) = q^{x-1} p, \quad x = 1, 2, 3, \dots, \quad \text{where } 0 < p < 1 \text{ and } q = 1 - p.$$

(i) Show that $P(X > x) = q^x$. (4)

(ii) A random sample of size 10 is taken from this distribution, and the results are summarised in the following table.

x	1	2	3	4	>4
<i>Frequency</i>	4	1	2	1	2

Explain clearly why the likelihood of the data can be written as $L(p) = p^8 q^{16}$, and hence obtain the maximum likelihood estimate \hat{p}_{ML} of p . (9)

(iii) You are given that \hat{p}_{ML} is approximately Normally distributed with mean p and variance $\frac{p^2 q}{10(1 - q^4)}$. By using \hat{p}_{ML} in estimating the variance of this distribution, obtain an approximate 90% confidence interval for p . (7)

7. The discrete random variables X and Y have the joint probability mass function (pmf) specified in terms of an unknown parameter θ in the following table.

		<i>Values of Y</i>		
		-1	0	1
<i>Values of X</i>	-1	$3k + \theta$	$4k - 2\theta$	$5k + \theta$
	0	$6k - 2\theta$	$8k + 4\theta$	$10k - 2\theta$
	1	$9k + \theta$	$12k - 2\theta$	$15k + \theta$

(i) Show that $k = 1/72$. Find also the range of possible values of θ for which the pmf is valid. (4)

(ii) Find the marginal distributions of X and Y and $\text{Cov}(X, Y)$. (7)

(iii) Is there a value of θ such that X and Y are independent? If there is, write down this value. Explain your reasoning clearly. (3)

(iv) In the case when $\theta = k$, find $P(X = 0 | Y = 0)$. (6)

8. Write down the model for the simple linear regression of a random variable Y on an explanatory variable x , and state the usual distributional assumptions for this model. (2)

A statistician obtains the following data for the weight of her baby son.

Age (weeks), t	1	2	3	4	5	6	7	8	9	10
Weight (grams), w	3250	3350	3550	3800	4000	4300	4400	4550	4800	5000

You are given that $\sum t = 55$, $\sum t^2 = 385$. However, you may find it easier to work in terms of the coded variable $v = w/50$. In terms of v , $\sum v = 820$, $\sum v^2 = 68568$, $\sum tv = 4840$.

- (i) Plot these data on a graph and comment on the suitability of a linear regression model. Assuming that such a model is appropriate, fit a linear regression to the data by the method of least squares. State but do not prove any formulae that you use. Also calculate the residual mean square, s^2 , and the coefficient of determination (R^2). (9)
- (ii) Use the fitted model to obtain point estimates of the baby's weight at (a) 11 weeks, (b) 20 weeks, noting any reservations you may have about either or both of these estimates. (4)
- (iii) Unfortunately there is no record of the baby's weight at birth. Write down a point estimate of this quantity based on your fitted model. Given that

$$s \sqrt{1 + \left\{ \frac{\left(\sum_{i=1}^{10} t^2 \right)}{\left(10 \sum_{i=1}^{10} (t - \bar{t})^2 \right)} \right\}}$$

is a suitable estimate of the standard error of the boy's weight at birth, where \bar{t} is the sample mean of the t values, provide a 95% prediction interval for the true birthweight. (5)