

EXAMINATIONS OF THE HONG KONG STATISTICAL SOCIETY



ORDINARY CERTIFICATE IN STATISTICS, 2004

Paper II

Time Allowed: Three Hours

*Candidates may attempt **all** the questions.*

*The number of marks allotted to each question or part-question is shown in brackets.*

*The total for the whole paper is 100.*

*A pass may be obtained by scoring at least 50 marks.*

*Graph paper and Official tables are provided.*

*Candidates may use silent, cordless, non-programmable electronic calculators.*

*Where a calculator is used the **method** of calculation should be stated in full.*

This examination paper consists of 7 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. Explain what the terms *quantitative* and *qualitative* mean when referring to variables in a set of data. (2)

Suggest two categories into which qualitative variables can be sub-divided and give an example of a variable of each category. (4)

Quantitative variables may be divided into the categories *discrete* and *continuous*. Give one example of each of these categories. (2)

For each of the four examples you have given, state which of a bar chart or a histogram you would use to graph its frequency distribution. In each case, indicate how you would order the categories across a page. (4)

2. A consumer magazine has noted the prices of a random sample of 50 pairs of ladies' shoes as advertised in catalogues from two mail order companies. The prices in £, arranged in ascending order, are shown.

Seromba catalogue

10	16	16	18	22	23	24	24	25	25
25	26	26	30	30	30	30	30	30	35
35	35	37	38	38	40	40	40	40	40
40	40	40	45	45	45	45	45	45	47
50	50	50	50	52	55	60	60	60	60

Etouderal catalogue

15	19	23	25	25	26	26	29	29	29
29	35	35	35	35	35	35	35	35	35
35	35	39	39	39	39	39	39	39	39
39	45	45	45	45	45	45	49	49	49
55	55	55	55	59	59	59	65	65	65

- (i) Draw two stem and leaf diagrams, one for each company. (6)
- (ii) Find the median and quartile prices in each sample and use your results to draw two boxplots, one for each company. (10)
- (iii) Use your results in (i) and (ii) to make a comparison of prices and pricing policies for ladies' shoes in the two catalogues. (4)

3. A sample of data is given as 2.7, 3.8, 3.0, 4.4, each value being correct to 1 decimal place.
- (i) Draw up a table showing the minimum and maximum possible values of each of the data points. (4)
- (ii) Using your table, calculate the minimum and maximum possible values for the mean of the four data points. (2)
- (iii) You are told that the minimum and maximum possible values of the standard deviation of the four points are 0.7182 and 0.8261, respectively. Show that the coefficient of variation must lie between 20.4% and 24.1%. (2)
4. Manpool City Council are responsible for a number of parks. These parks may contain some or all of the following amenities: swings, boating lake and café. 84% contain one or more of these amenities, 36% contain two or more and 16% contain all three.
- (i) Draw up a table showing the proportions of parks containing 0, 1, 2 and 3 of the amenities. Calculate the mean number of amenities. (3)
- (ii) You are now told that the proportion of the parks having just a café is equal to the proportion having just a boating lake and is 0.06 less than the proportion having just swings. The proportion of the parks having just a café and a boating lake is equal to the proportion having just a café and swings and is 0.02 less than the proportion having just swings and a boating lake. Draw a suitable Venn diagram and mark the appropriate proportion in each of the regions. (6)

5. Daily takings (£000) of a supermarket over three successive weeks are shown in the table.

**Supermarket takings (£000)**

<i>Day</i>	<i>Week 1</i>	<i>Week 2</i>	<i>Week 3</i>
Monday	92	89	87
Tuesday	98	93	90
Wednesday	104	100	97
Thursday	106	103	99
Friday	120	115	114
Saturday	132	125	121
Sunday	101	99	97

- (i) Draw a time chart of the data. (4)
- (ii) Explain the term *seasonal variation* as used in time series analysis. Comment on any seasonal variation here. (2)
- (iii) Use an appropriate moving average to find the trend of the series and plot it on your time chart. (6)
- (iv) Comment on the trend. (2)

6. A UK holiday company has a brochure listing the prices of holidays to certain destinations in continental Europe. For each of eight destinations (coded as A to H), one holiday is chosen at random from the list, and the length and price are shown in the table. All holidays include accommodation and air travel, and are in hotels of comparable standard.

<i>Destination</i>	<i>x = length of holiday (days)</i>	<i>y = price (£) per person</i>
A	8	895
B	9	1495
C	14	895
D	8	1095
E	9	945
F	15	1575
G	13	1495
H	8	1095

You are given that  $\Sigma x = 84$ ,  $\Sigma y = 9490$ ,  $\Sigma x^2 = 944$ ,  $\Sigma xy = 102230$ .

- (i) Calculate the sample mean and standard deviation of the holiday lengths. (3)
- (ii) Draw a scatter diagram showing the relationship between price ( $y$ ) and length of holiday ( $x$ ) and comment on what it shows. (5)
- (iii) Calculate the equation of the regression line of  $y$  on  $x$  and plot the line on your scatter diagram. (7)
- (iv) Use your equation to estimate the price of a similar holiday lasting 21 days. Say, with reasons, whether this is likely to be a good prediction. (3)

7. Every Sunday I check my car and this check may be thorough or cursory. The kind of check I do one Sunday depends only on the kind I did the previous Sunday. If I do a thorough check one week, the probability of a thorough check the following week is 0.2. If I do a cursory check one week, the probability of a cursory check the next week is 0.3. I do a cursory check in week 1.
- (i) Draw a tree diagram illustrating the situation for weeks 1, 2, 3 and 4. Mark the appropriate probabilities on the branches. (4)
- (ii) Calculate the following probabilities:
- (a) that I check my car thoroughly in week 2;
  - (b) that I check my car thoroughly in week 3;
  - (c) that I check my car thoroughly in week 4. (6)
- (iii) Calculate the probability that I check my car thoroughly in week 2, given that I check it thoroughly in week 4. (3)

8. Standardised mortality ratios (SMRs) are used to compare death rates in areas whose populations may have different structures of age and sex. An SMR is calculated as the ratio of the actual number of deaths in an area to the expected number if the national age and sex specific mortality rates were to apply. This ratio is then multiplied by 100 to form an index number. Thus, an SMR of 100 suggests that local mortality rates are the same as national mortality rates when age and sex differences in the two populations are taken into account.

The SMRs in the table are for each electoral ward in the area of South Trafford in Greater Manchester, England.

<i>Electoral ward</i>	<i>SMR</i>
Altrincham	85
Bowdon	73
Broadheath	97
Brooklands	82
Hale	70
Mersey St Mary's	72
Priory	99
Sale Moor	122
St Martin's	121
Timperley	70
Village	73

The SMR for South Trafford as a whole is 87.

Write a report interpreting this table for the local mayor who wants to know how mortality in South Trafford varies across the area and how it compares to the national SMR.

(6)