

EXAMINATIONS OF THE HONG KONG STATISTICAL SOCIETY



ORDINARY CERTIFICATE IN STATISTICS, 2006

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 calculators in accordance with the regulations published in the Society's "Guide to Examinations" (document Ex1).

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. (a) In a large sample, 39.0% of the items are less than 14.5 and 64.3% of them are less than 29.5. Estimate the median of the sample by linear interpolation. (4)
- (b) A table gives the critical values of a particular statistic as 43.77 for n equal to 30, and 55.76 for n equal to 40.
- (i) Using linear interpolation, estimate the corresponding value for n equal to 35. (1)
- (ii) Draw up in a table the values of $1/n$ for $n = 30, 35$ and 40 correct to 5 decimal places and on another line write in the given critical values of the statistic for $n = 30$ and $n = 40$. Use these last two values to estimate the critical value for $n = 35$, assuming that the linear interpolation is to be done using the values of $1/n$. (4)

2. (i) "Driver sleepiness is thought to account for one in ten of all UK road accidents. On motorways and dual carriageways that figure drops to one in five." Rewrite this quotation, expressing both rates as percentages, and correcting the error. (2)
- (ii) "People often have problems with percentages; so say '1 in 10,000' rather than '0.0001 per cent'." What should be the correct percentage in this quotation? (2)
- (iii) "Last year a total of 75,440 taxpayers filed on-line. This year 280,000 have done so – an increase of 400 per cent on the previous year." Rewrite this quotation putting in the correct percentage. (2)
- (iv) The table is said to show the changes over a three-year period in the proportions of different types of stopper used in the Australian bottled wine market.

Screw cap	3 per cent	increased
Synthetic cork	25 per cent	remained the same
Composite cork	25 per cent	increased
Natural cork	40 per cent	remained the same
Sparkling wine cork	7 per cent	remained the same

Explain what is wrong with this table.

(2)

3. As part of its quality control procedures, a company that manufactures light bulbs tested a sample of 100 bulbs to destruction. The lifetimes of the bulbs are recorded in the table.

Lifetimes, in completed hours, of 100 light bulbs

199	270	366	423	513	554	624	675	752	757
790	792	843	849	856	869	872	874	886	948
976	997	1001	1003	1004	1005	1008	1025	1025	1037
1052	1055	1080	1087	1088	1091	1116	1126	1135	1135
1170	1176	1180	1184	1225	1233	1234	1241	1248	1249
1251	1253	1261	1267	1278	1302	1313	1321	1344	1347
1349	1355	1361	1362	1363	1366	1368	1370	1371	1402
1405	1406	1406	1408	1412	1415	1422	1424	1430	1438
1443	1455	1465	1466	1467	1479	1485	1486	1490	1494
1517	1536	1620	1653	1671	1683	1692	1707	1724	1891

- (i) Draw a stem and leaf diagram with stem unit 100 hours and leaf unit 10 hours. (6)

- (ii) Form a grouped frequency distribution with classes (in hours) as follows.

0 but under 500
 500 but under 1000
 1000 but under 1100
 1100 but under 1200
 1200 but under 1300
 1300 but under 1400
 1400 but under 1500
 1500 but under 2000

(3)

- (iii) Draw a histogram of the grouped frequency distribution.

(6)

- (iv) Which of the two diagrams is more useful for assessing the quality of production? Give an explanation for your answer.

(3)

4. (i) Write down a formula for $P(A|B)$, the conditional probability of an event A given an event B . (You may assume that the probability of B is non-zero.) (2)
- (ii) State what is meant by saying that two events A and B are independent
- (a) in terms of $P(A \text{ and } B)$;
- (b) in terms of $P(A|B)$. (2)
- (iii) Two components X and Y have respective probabilities $3/4$ and $7/8$ of functioning correctly. They function independently of one another. Two devices are constructed using such components. Device 1 works only if both X and Y function correctly, Device 2 works whenever at least one of X and Y functions correctly.
- (a) Find the probability that Device 1 works.
- (b) Find the probability that Device 2 works.
- [There is no need to convert fractions to decimals.] (4)
- (iv) Suppose Device 1 works. Find the three probabilities
- (a) that X is functioning correctly,
- (b) that only X is functioning correctly,
- (c) that both X and Y are functioning correctly.
- Find the same probabilities, supposing instead that Device 2 works. (7)

5. The pre-tax profits, in thousands of pounds, for ten football clubs in the season 2002/03 are shown in the table. (A minus sign indicates that the club made a loss.) The clubs are arranged in decreasing order of turnover.

<i>Club</i>	<i>Pre-tax profit</i>	<i>League position</i>
Manchester United	39345	1
Liverpool	3641	5
Arsenal	4529	2
Newcastle United	4369	3
Chelsea	-26290	4
Tottenham Hotspur	-7118	10
Manchester City	-15425	9
Southampton	-4841	8
Everton	-12980	7
Blackburn Rovers	-12005	6

- (i) Calculate Spearman's rank correlation coefficient between pre-tax profit and position in the league. Comment on your result. (8)
- (ii) Give one reason why it would be inappropriate to calculate Pearson's product moment correlation coefficient for these data. (1)
6. The numbers of heart valve and heart bypass operations performed in a random sample of eight UK hospitals in the years 2001–2003 are shown in the table.

<i>Hospital</i>	<i>Number of heart valve operations</i>	<i>Number of heart bypass operations</i>
A	191	1420
B	331	3197
C	231	1721
D	199	1615
E	117	1005
F	247	1993
G	314	1391
H	160	1944

- (i) Calculate the mean number of heart valve operations performed and the standard deviation. Hence find the coefficient of variation. (4)
- (ii) Calculate the mean number of heart bypass operations performed and the standard deviation. Hence find the coefficient of variation. (4)
- (iii) Compare the results found in (i) and (ii). (3)

7. A comparison of house price changes in the North and the South of England is shown in the graph, taken from a newspaper. The lines show the percentage changes in price over the previous quarter from the second quarter of 2003 to the fourth quarter of 2004.

House price changes 2003 - 4



- (i) Using the graph, draw up a table showing chain base house price index numbers for the North and the South over this period. (4)
- (ii) Convert the chain base index numbers to fixed index numbers based on the first quarter of 2003. (7)
- (iii) Describe and compare the house price changes in the North and the South over the period 2003–4. (2)

8. The numbers of new members joining a family history society each quarter are shown in the table.

<i>Period</i>	<i>New members</i>	<i>Period</i>	<i>New members</i>	<i>Period</i>	<i>New members</i>	<i>Period</i>	<i>New members</i>	<i>Period</i>	<i>New Members</i>
2000 Q1	242	2001 Q1	278	2002 Q1	227	2003 Q1	173	2004 Q1	177
Q2	357	Q2	214	Q2	186	Q2	175	Q2	166
Q3	168	Q3	210	Q3	172	Q3	186	Q3	239
Q4	186	Q4	205	Q4	168	Q4	188	Q4	132

- (i) Draw a time chart of the data and comment on what it shows. (6)
- (ii) Calculate the centred four-quarterly moving average trend of the data and plot it on the chart. (4)
- (iii) Calculate the seven-quarterly moving average trend of the data and plot it on the chart. (4)
- (iv) State, with reasons, whether either of these trend plots would be helpful to the society in predicting future membership trends. (3)