

**EXAMINATIONS OF THE HONG KONG STATISTICAL SOCIETY**



**ORDINARY CERTIFICATE IN STATISTICS, 2009**

**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 8 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. In each of the following cases write down the lowest possible value of the italicised statistic and explain for what values of the data this lowest value will be achieved.
- (i) The *mean* of a sample of non-negative data values. (2)
  - (ii) The *variance* of a sample of data values. (2)
  - (iii) *Spearman's rank correlation coefficient* for two sets of rankings of the same objects. (2)
  - (iv) The *product-moment correlation coefficient* between two variables measured on the same objects. (3)

2. A recent telephone bill included the following summary of the number and the cost of calls made.

<i>Type of call</i>	<i>Cost as % of total cost</i>	<i>Number as % of total number of calls</i>
Daytime	61	50
Evening and weekend	13	33
Mobile	17	7
0845 numbers	6	8
All others	2	1
Total	99	99

Give a possible reason why the total of each column is not 100%. (1)

To represent these data, either bar charts or pie charts could be used. Give two advantages of using a bar chart. Give two advantages of using a pie chart. (4)

Draw pie charts to represent the data. (8)

Comment on two of the main features shown by the pie charts. (2)

3. In a comparison of the workloads of two call centres, the numbers of calls received per day were recorded over a 15 day period and the results are shown in the table.

**Daily calls**

<i>Day</i>	<i>Centre A</i>	<i>Centre B</i>
1	583	615
2	518	571
3	574	623
4	554	601
5	564	518
6	592	455
7	508	527
8	521	519
9	546	572
10	529	546
11	509	558
12	591	667
13	546	436
14	503	418
15	582	523

Calculate the range, the median, the lower quartile and the upper quartile of the numbers of calls for each centre.

(6)

Write a concise paragraph comparing the workloads, incorporating the statistics you have calculated.

(4)

4. Each member of a sample of 130 university students was asked to record how long had been spent on academic work in the previous week. The responses are summarised in the table below.

<i>Working time (hours)</i>	<i>Number of students</i>
Less than 5.5	0
5.5 but less than 10.5	6
10.5 but less than 20.5	12
20.5 but less than 30.5	22
30.5 but less than 35.5	22
35.5 but less than 40.5	20
40.5 but less than 50.5	24
50.5 but less than 70.5	16
70.5 but less than 90.5	8
90.5 or more	0

You wish to draw a histogram, in which the class "30.5 but less than 35.5" will have height 22 centimetres. What would be the corresponding heights for the classes "20.5 but less than 30.5" and "70.5 but less than 90.5"?

(5)

Construct a table that shows the values in hours and the frequencies that you would use to estimate the mean of the data. Calculate this mean.

(4)

Construct a table that shows the values in hours and the cumulative frequencies that you would use to plot the points for a cumulative frequency graph.

(2)

5. At the Men's European Cup Athletics Competition held in Munich on 23 June 2007 the results in the 100m, Hammer and High Jump events are given below.

<i>Finishing Position</i>	<i>100m</i>	<i>Hammer</i>	<i>High Jump</i>
1	GBR	POL	GER
2	FRA	GER	RUS
3	GER	GRE	POL
4	UKR	RUS	FRA
5	POL	FRA	GRE
6	GRE	UKR	UKR
7	RUS	GBR	BEL
8	BEL	BEL	GBR

BEL Belgium FRA France GBR Great Britain GER Germany GRE Greece  
POL Poland RUS Russia UKR Ukraine

Copy and complete the following table giving the rankings of the countries in each event.

<i>Ranking</i>	<i>100m</i>	<i>Hammer</i>	<i>High Jump</i>
GBR	1		
FRA	2		
GER	3		
UKR	4		
POL	5		
GRE	6		
RUS	7		
BEL	8		

(3)

Calculate the three Spearman rank correlation coefficients between pairs of the events.  
(7)

Comment on each coefficient.

(3)

6. In a television game show, there are 22 boxes each containing a different sum of money, the smallest being 1p and the largest £250,000. There are 11 "blue" sums of money (less than £1,000) and 11 "red" sums of money (of £1,000 or more). The contestant is assigned one box chosen at random, and takes it to a central table away from the other 21 boxes before playing the game. This is known as the central box.
- (i) Suppose that in 99 consecutive shows no contestant had £250,000 in the central box. Write down an expression for the probability of this event, and show that it is slightly less than 0.01. (3)
- (ii) (a) In the first round of the game, the contestant opens five boxes, chosen at random from the other 21 boxes. Find the probability that the contestant opens boxes all containing "blue" sums if he or she has a "blue" sum in the central box. Find the probability of the same event if he or she has a "red" sum in the central box. Hence show that the probability that the contestant opens five boxes all containing "blue" sums when the "colour" of the sum in the central box is unknown is 0.0175 correct to 4 decimal places. (9)
- (b) You are given that the probability that 4 "blue" sums and 1 "red" sum are found in the first round is 0.1379 when the "colour" of the sum in the central box is unknown. Find the probability that 3 "blue" sums and 2 "red" sums are found. (3)

7. The table shows the numbers of properties purchased by Britons in Spain over a 10-year period. Draw a time chart of the data. (4)

**Number of properties bought by Britons in Spain**

<i>Year</i>	<i>Number</i>
1998	2634
1999	2385
2000	2920
2001	4148
2002	3683
2003	3947
2004	5114
2005	5336
2006	5438
2007	5412

Estimate the trend of the data by using a 3-year moving average and plot the trend on your chart. (4)

The equation of the least squares trend line for the data is

$$y = 2421 + 373x$$

where  $y$  is the number of properties and  $x = (\text{Year} - 1998)$ . Plot this line on the chart. (4)

Making reference to your chart, give one advantage and one disadvantage of using each of the two methods for estimating trend. (4)

State with a reason which method you believe gives the better estimate of the trend for these data. (1)

8. A private investor has held shares in five companies over a period of years. The shareholding and market price per share on two dates are shown in the table.

**Investor's shareholding**

<i>Company</i>	<i>Number of shares held</i>	<i>Price in pence on 1 Dec 1998</i>	<i>Price in pence on 1 Dec 2008</i>
A	200	765	1566
B	400	636	1050
C	350	813	1231
D	500	62	283
E	250	153	767

Which company's share price has increased by the greatest amount of money? State the increase per share for this company.

(1)

Calculate price relatives, rounded to the nearest whole number, showing the price on 1 Dec 2008 with 1 Dec 1998 as base. Which company's share price has increased by the greatest percentage? State the percentage increase.

(2)

The investor wants to use an index number to work out the overall change in price of his shares over the period. Give a reason for using a weighted index number rather than an unweighted one.

(1)

Calculate a current weighted share price relative index for 1 Dec 2008 with 1 Dec 1998 as base.

(5)

Comment on the value of the index.

(1)