

EXAMINATIONS OF THE HONG KONG STATISTICAL SOCIETY



ORDINARY CERTIFICATE IN STATISTICS, 2013

MODULE 2 : Analysis and presentation of data

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 printed pages.

This front cover is page 1.

Question 1 starts on page 2.

There are 9 questions altogether in the paper.

1. Data are handled on computer networks in packets. The size of a packet is measured in bytes. The number of bytes in a packet is a power of 2.

In one 24-hour period, a particular computer network handled packets of data as follows.

<i>Size of packet in bytes</i>	64	128	256	512	1024	2048	4096
<i>Number of packets</i>	12 700	11 920	1840	1030	370	320	3560

- (i) Write down the modal packet size. (1)
- (ii) Find the median packet size. (1)
- (iii) Calculate the mean packet size. (2)
- (iv) Describe in words how the packet sizes vary. Discuss briefly whether the mode, median and mean are useful measures in this case. (5)
2. A health website contains the following information.

<p>Our bodies need their daily calorie intake to maintain body weight. Foods that are high in calories provide us with energy. Your daily calorie intake requirements change depending on age.</p> <p>Recommended daily calories allowance Men: 2550 Women: 1940</p>

Explain why the given information is limited in its usefulness. (3)

Suggest *three* additional items of information a statistician might wish to see as part of this kind of advice. (3)

3. In the United Kingdom, the percentages of people with the four different blood types, O, A, B, AB, are as follows.

<i>Blood type</i>	O	A	B	AB
<i>Percentage</i>	44%	42%	10%	4%

- (i) Two people are chosen at random. Find the probability they have the same blood type.

Find the probability that they have different blood types.

(4)

- (ii) Four people are chosen at random. Find the probability that they all have different blood types.

(3)

- (iii) A random sample of n people is now taken. Write down an expression for the probability that none of them is of blood type AB.

Find how large n must be to give more than a 25% chance of the sample including someone of blood type AB.

(4)

4. The table shows the average price, in pence, of a litre of petrol in England on 31 December each year from 2003 to 2011.

<i>Year</i>	2003	2004	2005	2006	2007	2008	2009	2010	2011
<i>Price (p)</i>	76.02	84.17	87.35	85.47	102.13	92.79	108.73	119.21	133.38

- (i) Draw a suitable graph of the data and describe briefly what the graph shows.

(6)

- (ii) Calculate the percentage change in the price of petrol from the previous year, for the years to December 2005 and December 2008.

(2)

- (iii) Taking 2003 as base year, express the price of petrol in 2004 and 2011 as index numbers.

(2)

- (iv) Use linear interpolation to estimate the price of a litre of petrol at the end of March, June and September in 2008. Comment briefly on the accuracy of these estimates.

(4)

5. Each of the 10 workers, P, Q, ..., Y, was required to complete two tasks, A and B. The supervisor observed the workers and gave them a score for the quality of their work on each task, where higher scores indicate better quality work. These scores are shown in the following table.

<i>Worker</i>	P	Q	R	S	T	U	V	W	X	Y
<i>Task A</i>	43	86	87	92	102	103	108	109	111	113
<i>Task B</i>	41	93	108	120	88	105	110	94	122	107

- (i) Calculate the mean and standard deviation of the scores on task A. (4)
- (ii) Comment on worker P's performance on the two tasks. (1)
- (iii) Calculate Spearman's rank correlation coefficient for the data. (3)
- (iv) Without doing any further calculations, but taking into account your comment in part (ii), explain why Spearman's coefficient may be preferable to the product-moment correlation coefficient in this case. (3)

6. A householder estimated that the electricity he used was consumed by different household appliances and devices as follows.

<i>Appliance or device</i>	<i>Percentage of total</i>
Water heater	32%
Refrigerator	18%
Cooker	21%
Washing machine	6%
Lighting	8%
Audio-visual equipment	7%
Computer	2%
Air conditioner	6%

- (i) Represent this information on a suitable graph. Explain why the type of graph you choose is appropriate.

(4)

In order to save money, the householder cuts the amount he spends on heating water by a quarter, and he decides to stop using the air conditioner. However, he doubles his computer usage.

- (ii) Calculate the overall percentage reduction in his electricity usage.

(1)

- (iii) Calculate the new percentages of total electricity usage for each appliance or device.

(3)

7. The table below shows the fuel consumption in litres per kilometre travelled for 9 types of aircraft labelled A–I. It also shows the maximum number of passengers that the type of aircraft carries.

<i>Aircraft type</i>	A	B	C	D	E	F	G	H	I
<i>Fuel consumption (l/km)</i>	0.59	3.02	3.05	3.12	2.41	10.2	3.40	3.15	3.81
<i>Number of passengers</i>	6	132	105	118	102	286	69	59	116
<i>Fuel economy (l/km/passenger)</i>	0.0983			0.0264	0.0236		0.0493	0.0534	0.0328

The fuel economy for an aircraft takes into account the number of passengers it carries. It is defined as the fuel consumption in litres per kilometre per passenger.

- (i) Calculate the fuel economy for the aircraft of types B, C and F.

Draw a scatter diagram showing fuel economy on one axis and the number of passengers carried on the other.

Identify two outlying observations on your scatter diagram and state which aircraft type they represent. Briefly discuss the fuel consumptions of these two types of aircraft.

(11)

- (ii) You are given that the product-moment correlation coefficient for fuel economy and number of passengers is -0.56 . Comment on this figure in the light of your scatter diagram.

Calculate the product-moment correlation coefficient after the two outlying observations are removed.

Briefly discuss the relationship between fuel economy and number of passengers carried.

(7)

8. The first table shows the monthly numbers of rooms booked in a small hotel for the years 2010–2012.

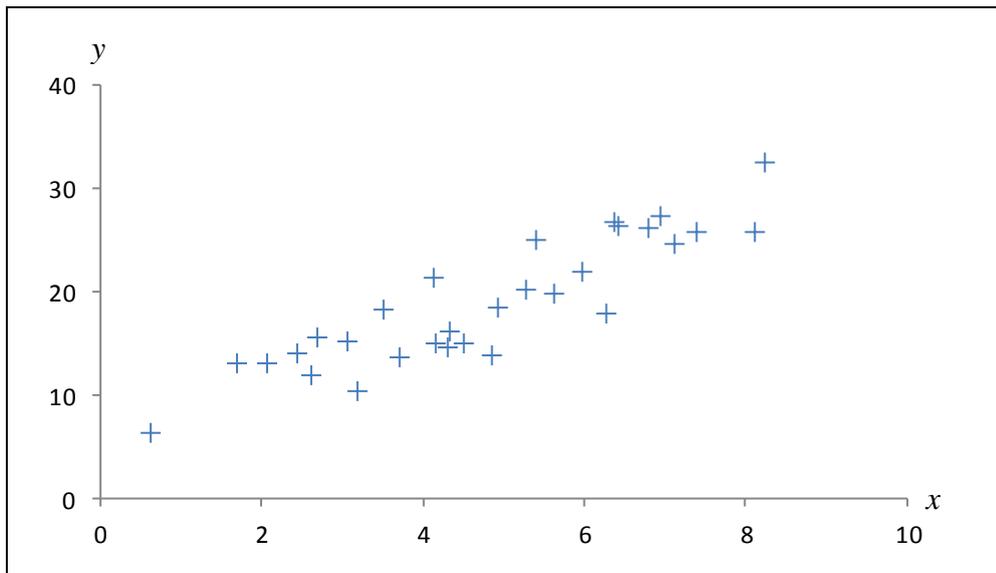
	<i>Jan</i>	<i>Feb</i>	<i>Mar</i>	<i>Apr</i>	<i>May</i>	<i>Jun</i>	<i>Jul</i>	<i>Aug</i>	<i>Sep</i>	<i>Oct</i>	<i>Nov</i>	<i>Dec</i>
2010	501	488	504	578	545	632	728	725	585	542	480	530
2011	518	489	528	599	572	659	739	758	602	587	497	558
2012	555	523	532	623	598	683	774	780	609	604	531	592

The second table shows, correct to 1 decimal place, the 12-monthly centred moving average.

	<i>Jan</i>	<i>Feb</i>	<i>Mar</i>	<i>Apr</i>	<i>May</i>	<i>Jun</i>	<i>Jul</i>	<i>Aug</i>	<i>Sep</i>	<i>Oct</i>	<i>Nov</i>	<i>Dec</i>
2010							570.5	571.3	572.3	574.2	576.2	578.5
2011	580.0	581.9	584.0	586.5	589.1	x	593.7	596.7	598.3	599.4	601.5	603.6
2012	606.0	608.4	609.6	610.6	612.8	y						

- (i) Calculate the missing values, shown as x and y , in the second table. (4)
- (ii) Calculate the average difference from the trend as estimated by the 12-monthly centred moving average for
- (a) January,
- (b) July. (4)
- (iii) Describe briefly the patterns of trend and seasonal variation in the data. (4)

9. An experiment gave a set of data points (x, y) as shown in the scatter diagram.



The data are summarised in the following table.

n	Σx	Σy	Σx^2	Σy^2	Σxy
30	142.678	567.024	791.188	11 838.229	3014.726

- (i) Find the least squares regression line for y on x , expressing your answer in the form $y = a + bx$. You should show your working clearly. (6)
- (ii) Estimate the mean value of y when
- (a) $x = 3$,
- (b) $x = 10$.

Comment briefly on the reliability of these two estimates.

(5)