EXAMINATIONS OF THE HONG KONG STATISTICAL SOCIETY

ORDINARY CERTIFICATE IN STATISTICS, 2014

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. A commuter group took a random sample of 312 train journeys on a particular route, and recorded for each journey whether it departed on time or late, and whether it arrived on time or late. Some of their figures are shown in the table.

		Arrived		
		Late	On time	Total
Departed	Late	33		43
	On time			
	Total	52		312

(i) By completing the table find, correct to 2 decimal places, the proportions of all the journeys that

(a) departed late,

(b) arrived on time,

(c) departed late and arrived on time.

Find also the proportion of journeys that arrived on time given that they departed late.

(5)

(ii) What can you deduce about the independence or dependence of the events "Departed late" and "Arrived on time"?

(2)

2. A census in Manhattan found that about 50% of households consisted of just one person. A newspaper reported this as "Half the residents of Manhattan live alone".

Explain clearly the misunderstanding shown in the newspaper report.

(2)

Show that the correct figure for the proportion of residents of Manhattan who live alone will be less than $\frac{1}{3}$.

(3)

3. A survey of 5194 households in the UK collected data on net household wealth, defined as the value of savings and investments minus non-mortgage debts. (The net value of a house and any mortgage debt was dealt with separately in the survey.) Percentiles and net household wealth in £ were as follows.

Percentile	10	25	50	75	90
Net household wealth (£)	–9164	–394	1309	19 660	70 620

(i) Explain the meaning of the figures in the shaded part of the table.

(2)

(ii) Sketch a graph of the likely shape of the distribution of net household wealth. State, with a reason, whether you would expect the mean of this distribution to be less than, approximately equal to, or greater than the median.

(5)

(iii) Calculate the inter-quartile range (IQR) for these data.

(1)

(iv) One definition of an outlier is any item of data less than L or more than U, where L is 1.5 × IQR below the lower quartile, and U is 1.5 × IQR above the upper quartile. What can you say about whether or not there are outliers in this data set?

(4)

4. The table shows the years in which the human population reached 1, 2, …, 7 billion.

Population in billions	1	2	3	4	5	6	7
Year	1804	1927	1960	1974	1987	1999	2012

(i) Calculate the rate of growth of the human population between 1927 and 1960

(a) as the average increase per year,

(b) as the average percentage increase per year.

(5)

(ii) Calculate the corresponding figures for the period from 1999 to 2012.

(3)

(iii) Comment briefly on the figures you have calculated.

(2)

5. The following table comes from a survey of people's weight in Mexico. It shows the number of men and women in the survey, with the percentages who are normal, overweight and obese. (People who are underweight are classified as normal in this survey.)

	Number	*Normal*	*Overweight*	*Obese*
Men	19 798	27.3%	35.8%	36.9%
Women	13 225	31.8%	44.7%	23.5%
Total	33 023			

(i) Calculate the numbers of men and women in the survey who are in each of the categories normal, overweight and obese. Hence obtain the total percentages of people in the survey who are normal, overweight and obese. (6)

Further information from the same survey is given in the table below.

		Number	*Normal*	*Overweight*	*Obese*
Age	20–29	*x*	44.2%	34.1%	21.7%
	30–39	9 052	26.4%	42.1%	31.5%
	40–49	6 788	20.5%	41.8%	37.7%
	50–59	4 187	19.8%	41.0%	39.2%
	60+	5 917	29.9%	39.8%	30.3%
Location	rural	9 530	34.1%	*y*	26.6%
	urban	23 493	27.7%	39.6%	32.7%
Economic Status	low	12 840	34.2%	39.3%	26.5%
	medium	11 451	26.1%	39.9%	34.0%
	high	8 732	27.4%	39.4%	33.2%
Level of Education	low	3 791	32.5%	38.8%	28.7%
	medium	16 176	25.7%	39.7%	34.6%
	high	13 056	32.2%	39.6%	28.2%

(ii) Calculate the missing values, *x* and *y*. (2)

(iii) Describe briefly how patterns of being obese vary with the classification factors

(a) age,

(b) location,

(c) economic status,

(d) level of education. (4)

(iv) Describe briefly how being overweight (but not obese) varies with the different classification factors. (2)

6. The figures below represent the numbers of 'hits' that a blogger gets on his new website in successive time intervals.

126	356	408	404	420	425	176	167	398	433	446	419	431	189
120	431	390	433	495	454	215	117	451	426	466	477	457	210

(i) Explain why it seems likely that these are daily figures (rather than, say, hourly or weekly figures).

(2)

On the assumption that these are daily figures, a series of 7-point simple moving averages is calculated, each being correct to 3 significant figures. This series begins as shown.

331	337	343	346	352	352	353	355	348	353	347	345	356	359
363	362	365	370	375									

(ii) Show how the first moving average value is calculated. Calculate the three remaining values, after the value 375, in the moving average series.

(4)

(iii) Draw a graph showing the original series and the moving averages.

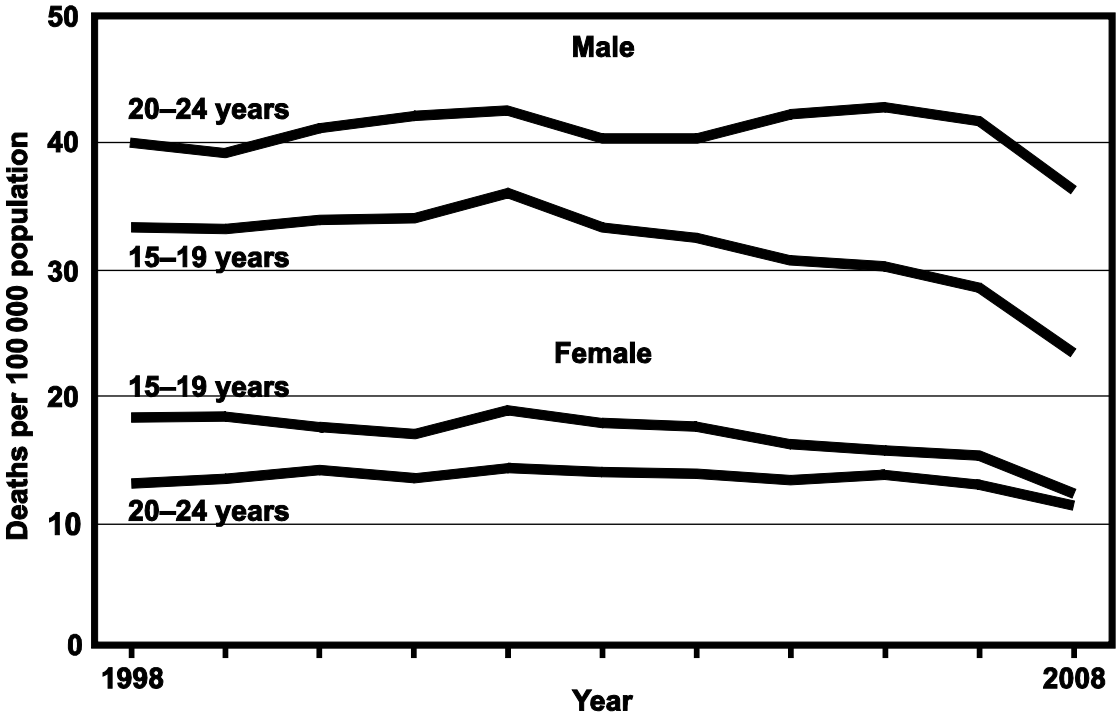
(5)

(iv) Describe the trend and variation in the data, and say what they indicate about the pattern of 'hits' on the blogger's website.

(4)

7. The following diagram is taken from a statistical report on health, injury and death in the USA. It shows the motor vehicle-related death rates for four populations: males aged 15–19, males aged 20–24, females aged 15–19 and females aged 20–24. In 2003 there were approximately 10 million people in each of these four populations.

Motor vehicle-related deaths among persons 15–24 years of age, by sex and age: United States, 1998–2008



(i) Calculate an estimate of the total number of motor vehicle-related deaths among persons 15–24 years of age in 2003.

(5)

(ii) State, clearly and concisely, what the graphs indicate about the following.

(a) Comparative death rates for males and females.

(b) Comparative death rates in the two different age groups

- for males,
- for females,
- for males and females combined.

(c) Death rates over time in the four groups.

(7)

8. I eat an apple every day, and I keep careful records of the types of apple I eat and what they cost. Figures for the years 2001 and 2011 are in the table.

Type of apple	*Price in pence per apple in 2001*	*Quantity of apples in 2001*	*Price in pence per apple in 2011*	*Quantity of apples in 2011*
Granny Smith	17	137	29	107
Braeburn	20	77	30	120
Gala	15	75	26	43
Golden Delicious	19	43	27	20
Cox	21	33	29	75

(i) Taking 2001 as the base year, calculate the simple price relatives for 2011 of the five types of apple. Give also the arithmetic mean of these price relatives.

(4)

(ii) Taking 2001 as the base year, calculate an index for my total expenditure on apples in 2011.

(3)

(iii) Calculate the Laspeyres price index for the data.

(2)

(iv) Give a possible reason for the fact that, in 2001, I ate more Granny Smiths than any other type of apple, but in 2011 I ate more Braeburns than any other type of apple.

(2)

9. The heights, x inches, and weights, y pounds, were measured for a random sample of 50 professional football players. Summary statistics are as follows.

$n = 50 \quad \Sigma x = 3681.2 \quad \Sigma x^2 = 271\,137 \quad \Sigma y = 9973.3 \quad \Sigma y^2 = 2\,011\,873 \quad \Sigma xy = 734\,756$

(i) Calculate the means and standard deviations for these data.

(6)

(ii) Calculate the product moment correlation coefficient for these data.

(3)

(iii) Suppose the data had been converted to metric units, centimetres and kilograms, before analysis. There are 2.54 cm per inch and 2.20 pounds per kg. Find the values in metric units of the figures calculated in parts (i) and (ii).

(5)

BLANK PAGE