



EXAMINATIONS OF THE HONG KONG STATISTICAL SOCIETY

HIGHER CERTIFICATE IN STATISTICS, 2014

MODULE 1 : Data collection and interpretation

Time allowed: One and a half hours

*Candidates should answer **THREE** questions.*

Each question carries 20 marks.

The number of marks allotted for each part-question is shown in brackets.

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).

The notation \log denotes logarithm to base e .

Logarithms to any other base are explicitly identified, e.g. \log_{10} .

Note also that $\binom{n}{r}$ is the same as ${}^n C_r$.

This examination paper consists of 8 printed pages.

This front cover is page 1.

Question 1 starts on page 2.

There are 4 questions altogether in the paper.

1. A highly selective university has an intake of several thousand students per year. In order to monitor the overall ability of the intake, each year a group of new students is given a general test of academic ability lasting 1 hour. The test is taken by the first 200 students to volunteer. The students take the test under supervision, and they are given a small sum of money for their time. The test is designed so that, if it were taken by the whole adult population of the country, the scores would be distributed Normally with mean 100 and standard deviation 15.

Table 1 gives summary data on the scores from the last 5 years.

<i>Year</i>	<i>2009</i>	<i>2010</i>	<i>2011</i>	<i>2012</i>	<i>2013</i>
Mean	132.8	132.6	130.1	127.9	125.7
Standard deviation	8.6	8.1	8.8	9.4	9.5
Minimum	108.9	114.2	105.5	94.5	97.6
5th percentile	118.9	119.1	115.9	113.0	109.6
25th percentile	126.6	126.6	124.7	121.4	120.1
50th percentile	132.6	132.7	130.4	128.8	126.0
75th percentile	139.8	138.6	135.9	134.9	131.6
95th percentile	146.9	145.6	144.4	141.8	141.4
Maximum	153.6	151.7	150.0	150.0	148.0

Table 1

Table 2 gives the cumulative percentages in the population as a whole for various scores on the test.

Score	100	105	110	115	120	125	130	135	140	145	150
Population cumulative percentage	50.0	63.1	74.8	84.1	90.9	95.2	97.7	99.0	99.6	99.9	100.0

Table 2

- (i) Write a report summarising what the data indicate about the ability of the new intakes of students over the last 5 years, comparing them with each other and, as appropriate, with the population as a whole. Your report should include summary data and statistical diagrams where appropriate, but should not assume any specialist statistical knowledge on the part of the reader. (10)
- (ii) Justify briefly how you have chosen to present the data in your report. If there are particular data items that you have chosen not to use, explain your reasoning. (4)
- (iii) Comment on the university's procedure in gathering this information, recommending changes if you think they would be appropriate. (6)

2. The ship *Titanic* sank in 1912 with the loss of many lives among all three passenger classes and the crew. Since that time, it has often been said that first class passengers had greater access to lifeboats than third class passengers. There have also been numerous stories of male passengers pushing their way onto lifeboats ahead of women and children.

The table shows the numbers who survived and died among the various categories of passengers and crew.

<i>Class</i>	<i>Age</i>	<i>Sex</i>	<i>Died</i>	<i>Survived</i>	<i>Total</i>
1st	adult	female	4	140	144
		male	118	57	175
	child	female		1	1
		male		5	5
2nd	adult	female	13	80	93
		male	154	14	168
	child	female		13	13
		male		11	11
3rd	adult	female	89	76	165
		male	387	75	462
	child	female	17	14	31
		male	35	13	48
crew	adult	female	3	20	23
		male	670	192	862
Total			1490	711	2201

Analyse these data in order to answer the following questions. In each case, if appropriate, you should illustrate your findings by means of suitable statistical diagrams.

- (i) How did survival rate vary between classes? (Note: here and in (ii), the crew counts as a class.) (6)
- (ii) Identify any consistent patterns and any anomalies within classes. (9)
- (iii) What evidence is there for or against the existence of a 'women and children first' policy in getting people into the lifeboats? (5)

- 3 A sample survey investigated the weights of men aged 65 or more in the Valladolid region of Spain. There were about 42 000 men in the region within this age range, and a sample of 508 men was taken. They were categorised by age and by type of residence (own home, publicly funded care home, privately funded care home).

Table 1 shows the numbers in each category.

<i>Age (years)</i>	<i>Own home</i>	<i>Public care home</i>	<i>Private care home</i>	<i>Total</i>
65–69	41	16	18	75
70–74	41	21	21	83
75–79	36	28	27	91
80–84	34	20	26	80
85–89	37	17	23	77
90–94	25	14	16	55
95+	23	10	14	47
Total	237	126	145	508

Table 1

Table 2 shows the mean weights, in kg, of the men in each category.

<i>Age (years)</i>	<i>Own home</i>	<i>Public care home</i>	<i>Private care home</i>	<i>Overall</i>
65–69	66.1	68.0	67.4	66.8
70–74	65.0	67.5	68.1	66.4
75–79	63.8	66.2	66.5	65.3
80–84	61.7	65.3	64.0	63.3
85–89	60.7	63.4	62.8	<i>a</i>
90–94	58.0	62.0	60.1	59.6
95+	58.8	61.4	58.9	59.4
Overall	62.5	65.3	<i>b</i>	63.7

Table 2

- (i) Find the numbers shown as *a* and *b* in Table 2. (2)
- (ii) The survey report states that the standard deviation for the weights of all 508 men surveyed was 9.7 kg. Estimate the standard error of the mean for all 508 men.

Assuming the same standard deviation of 9.7 kg within each of the categories, estimate the standard error of the mean for

- (a) those aged 65–69 living in their own homes,
 (b) those living in publicly funded care homes.

Explain how standard errors should be used in interpreting the data.

(6)

Question continued on the next page

- (iii) Explain how stratified sampling could have been used to select the 508 men. State one advantage of stratified sampling.

Explain what cluster sampling is and state one advantage of this type of sampling. Why would cluster sampling be difficult to implement in this case?

(8)

- (iv) Describe the main patterns evident in the data, drawing diagrams where appropriate to illustrate comparisons and contrasts between subsets of the data.

(4)

- 4 A local volunteer group is hoping to launch a community radio station. The station would be a 'not for profit' enterprise, and it has very limited funding, but it does have a website. The station would be broadcast locally and available on the internet.

The volunteer group is carrying out market research in its target area in order to identify the features that potential listeners would value most. It also hopes that conducting market research will raise awareness of the radio station, make people more likely to listen when it is launched, and encourage others to volunteer to help.

- (i) Design a questionnaire suitable for use by a volunteer interviewing passers-by in the town centre. The questionnaire should gather information, in a suitable form for summarising and processing, on the following topics.

- Age and sex of respondent
- Current radio listening – e.g. times of day, location, stations currently listened to
- Preferred content – e.g. news, sport, weather, traffic (local and national in each case); talk shows, phone-ins; types of music liked and disliked
- General reaction to the idea of a local community station
- Attitude to religious programming (the station has a possible source of funding from a local religious organisation if it carries religious programming)

(12)

- (ii) A version of the questionnaire is to be made available on the station's website for people to complete online. Discuss briefly any additional design issues that need to be considered for the online version.

(4)

- (iii) Explain briefly the possible sources of bias and error in gathering the required information by means of interviews and online.

(4)

BLANK PAGE

BLANK PAGE

BLANK PAGE