EXAMINATIONS OF THE HONG KONG STATISTICAL SOCIETY

ORDINARY CERTIFICATE IN STATISTICS, 2011

Module 1 : Collection and compilation 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).

1. Describe a price index in a country of your choice, for instance the Retail Prices Index in Great Britain.

   Your answer should include comments on the following.

   - How often it is published
   - How the data are obtained
   - How the index is constructed
   - How and why the constitution of the index is changed over time

   (9)

   Give three ways in which the index is used by different groups for different purposes.

   (6)

2. You are interested in the mean value of a measure in a given population. The population is large and you decide to estimate this mean using a simple random sample from the population.

   (i)   What would you compute to estimate the population mean?

   (1)

   (ii) Explain in detail what is meant by *bias* in sampling.

   (2)

   (iii) Why are the sample members chosen randomly?

   (1)

   (iv) Explain what is meant by sampling *accuracy* and how it is measured.

   (3)

   (v)   How can sampling accuracy be improved?

   (1)

   (vi) Describe the circumstances in which the following sampling methods might be used, and say what advantage they each have over simple random sampling.

       (a)   Stratified random sampling.

       (b)   Cluster sampling.

       (7)

Turn over

3. An estimate is required of the mean value of a variable defined on a large population.

   The population contains two strata in proportions $p : (1 - p)$, where $0 < p < 1$.

   (i)   Explain what is meant by *strata*.

   (3)

   (ii)   Give an example of such a population.

   (2)

   A random sample of size n is to be drawn, comprising n_1 from stratum 1 and n_2 from stratum 2, where $n_1 + n_2 = n$. For each stratum, the sample means, $\bar{x}_1$ and $\bar{x}_2$, and the sample standard deviations, s_1 and s_2, will be computed.

   The formula for the estimate of the population mean is $\bar{x} = p\bar{x}_1 + (1-p)\bar{x}_2$.

   An estimate for the standard deviation of $\bar{x}$ (the standard error of the mean) is

   $$se = \sqrt{\frac{p^2 s_1^2}{n_1} + \frac{(1-p)^2 s_2^2}{n_2}}.$$

   (iii)   Supposing that $n_1 = 20$, $n_2 = 10$, $p = \frac{2}{3}$, $\bar{x}_1 = 5$, $\bar{x}_2 = 8$, $s_1 = 1$ and $s_2 = 3$, compute $\bar{x}$ and se.

   (4)

   (iv)   Supposing that $n_1 = 12$, $n_2 = 18$, $p = \frac{2}{3}$, $\bar{x}_1 = 5$, $\bar{x}_2 = 8$, $s_1 = 1$ and $s_2 = 3$, compute $\bar{x}$ and se.

   (4)

   (v)   Given the evidence of your computations, should strata sample sizes be chosen in proportion to strata sizes in this example? Justify your answer.

   (2)

Turn over

4.   The management of a package tour company wishes to gather information from customers who have completed their holidays. You are to advise on how this may best be achieved.

You should specify the questions to be asked and the forms of the answers that are required, but you need not produce a design for how the questionnaire might appear.

Your answer should include the following.

- Your recommendation regarding the mechanism by which the data should be collected

(3)

- No more than 5 questions to determine appropriate personal details

(3)

- No more than 5 questions to determine appropriate details about the holiday

(3)

- No more than 10 questions to determine customer experiences and level of satisfaction with appropriate aspects of the holiday

(4)

- How you might persuade customers to complete questionnaires, and an explanation of why this is important

(2)

Why might it not be necessary to collect personal and holiday details? What would be the advantage of not collecting these details?

(2)

5.   A restaurant in an Alpine ski resort has an outdoor seating area on the slopes, separated from the restaurant by a service road. To improve efficiency, the waiters and waitresses are equipped with hand-held electronic devices with wireless capability. Staff enter customer orders on their handsets, and the orders are instantly transmitted to the kitchen.

Apart from shortening the time between a customer placing an order and receiving his or her meal, the handsets provide electronic sales data that could help the business of the restaurant in various ways. Give four areas of the business in which such data might be useful, and describe briefly how the data could be used in each area.

(8)

4

6. A researcher is interested in studying the side effects of a particular brand of pain killer. She can question individuals who use this brand about their experiences, or she can set up a clinical trial to measure any observable side effects.

Describe what is involved in following each of these approaches. Compare and contrast them.

(15)

7. Identify a potential source of bias error in each of the following proposed sampling methodologies. You should also specify the type of sampling that is involved in each case.

(i) To investigate energy consumption, a researcher sends out interviewers to 25% of the houses in a neighbourhood. He selects the sample by choosing every fourth house listed on the electoral register.

(ii) A market research company employs interviewers to conduct telephone surveys. Each interviewer selects residential numbers at random to ring during working hours until that interviewer's quota of respondents have agreed to be interviewed.

(iii) Employees of a large corporation are e-mailed invitations to take part in job satisfaction surveys.

(iv) Machines in a factory are chosen at random on two separate days and the amount of wear in a specific component is measured. Components are replaced when wear reaches a certain level but some components reach that level more quickly than others. The intention is to estimate the average rate of wear.

(15)