HONG KONG STATISTICAL SOCIETY

2015 EXAMINATIONS − SOLUTIONS

ORDINARY CERTIFICATE – MODULE 1

The Society is providing these solutions to assist candidates preparing for the examinations in 2017.

The solutions are intended as learning aids and should not be seen as "model answers".

Users of the solutions should always be aware that in many cases there are valid alternative methods. Also, in the many cases where discussion is called for, there may be other valid points that could be made.

While every care has been taken with the preparation of these solutions, the Society will not be responsible for any errors or omissions.

The Society will not enter into any correspondence in respect of these solutions.

The OC1 paper is, by its nature, open-ended: the questions expect candidates to determine their own approach to analysis and interpretation. For this reason it is not always possible to give definitive solutions. The mark scheme below therefore gives, for each question, one possible approach and the corresponding mark allocation. Other approaches will require the marker to exercise judgement.

1.

i) In a census everyone in the population is intended to be included [1]; in a sample survey only a subset of the population is included. [1]

ii) Missing members of the population [1]
   Duplicate members of the population [1]
   Inaccurate information about members of the population [1]
   All can lead to bias i.e. results systematically different from what they should be [1]

iii) Identification of country and purpose of census/survey [1]
   Description of population [1]
   Identification of sampling frame [1]
   Description of two checks etc carried out [2] e.g cross-checking with other lists;
Piloting to estimate accuracy; procedures to keep list up-to-date.
(11 marks)

2.

i) If the houses are evenly spaced (at 10 degrees intervals) around a circle centred on where the bottle lies [1]

ii) there are groups of houses e.g. one behind another in some directions [1]

iii) A table of two-digit numbers where at any point in the table any number in the range 00 to 99 is equally likely. [1] Independence is also necessary for pairs of digits to be truly random. [1]

iv) Identify each of the houses with a number 01 to 36. The allocation order does not matter [1] (or other valid range of numbers)
Select a starting point in a table of two-digit random numbers [1]
Work systematically through the table selecting numbers in the range 01 to 36,and discarding all other numbers (00 and 37-99), providing the number has not already been chosen [1], until you have six [1].
As the numbers are random [1] and every house has been allocated an identifying number that is in the table [1], each house has an equal chance of being chosen irrespective of where it is situated. [1]
(11 marks)

3.

i) The three areas are non-overlapping subsets of the overall population which cover the population, so these are strata [1] Nursing homes are sampled within each area rather than from the overall population, so this is stratified sampling [1]

ii) Nursing homes are chosen non-randomly because by the deliberate choice of the researchers (purposely or by convenience, although candidates may not know these terms) [1]

i)   Nurses are busy people and will not have much time to give to research [1]
If they already know the researchers then it is true that they will be more likely to participate [1] So the claim is valid [1]
ii)   The total sample cannot be random [1] because of nurses' views are likely to vary according to the home in which they work [1] homes are not even chosen randomly [1] so some geographical regions cannot be represented [1] and also there are disproportionate numbers in the 3 areas [1]
iii)   Given that the aim is to compare the three areas [1] it is sensible to have equal numbers in each area [1] unless it is known that the variability of answers is likely to be very different between the 3 different areas [1]

(14 marks)

4.

i)

This is an open question as there are no pre-specified answer options.   [1]

Closed questions are much easier and quicker to code and to process [1]

However, if the options do not include the answer that the participant would want to choose, then the results are unreliable [1] or the respondent may fail to complete the question [1]. Moreover, it is difficult to obtain explanations in closed questions. [1]

ii)
Sunny Restaurant

Thank you for booking a meal here. We are always interested in customer feedback and ask you to spend a few minutes completing this questionnaire. We are very happy to offer you a free coffee while you do this.

John Smith, Manager

1. How satisfied were you with the range of choice on the menu? (please tick one box)

Very dissatisfied	Dissatisfied	Satisfied	Very satisfied

Please explain your answer below.

2. How satisfied were you with the quality of the service (please tick one box)

Very dissatisfied	Dissatisfied	Satisfied	Very satisfied

Please explain your answer below.

3. What did you think about the environment in which you had your meal? (Please tick one box)

Very unacceptable	Unacceptable	Acceptable	Very acceptable

Please explain your answer below.

4. What is your opinion about the value for money of your meal? (please tick one box)

Very poor value for money	Poor value for money	Good value for money	Very good value for money

Please explain your answer below.

5. Please add any other comments you would like us to consider.

Thank you very much for taking the time to complete this questionnaire. Please place it in the box at reception. We hope to see you again soon.

Information at the start [1]

Thank you [1]

Instructions about where to put the questionnaire [1]

Phrasing of questions [1]

Layout [1]

Use of bothclosed and open questions [1]

Justification for closed questions – used for ratings to allow easy summarisation of results [1]

Justification for open questions – used to obtain reasons and opinions which cannot be pre-specified [1]

(13 marks)

5.

i) Missing values occur when a participant fails to give an answer [1]

Errors occur when the participant gives a wrong answer [1] (which may not always be detectable) [1]

Errors usually end up as missing values [1] and if these and the other missing values are not 'at random' [1] then there is likely to be bias in the results [1] Candidates may mention transcription errors here, and would be credited accordingly.

ii) There is 1 missing value [1] – the other one is 'not applicable' [1]

There is an incorrect code of G in Sex [1]

In the last line there are no children, but an age given for the eldest [1]. It is not clear which value is wrong [1]

The age 10 looks wrong if there are 2 children the eldest of which is 6 [1]

The zero for age of child for the 14-year old is probably 'not applicable' [1]

NB The age zero for a child is not necessarily an error as the child could be a young baby

iii) Some errors occur in transcription so it would be worthwhile checking against the original source of the data [1]

(iv) Age, number of children and age of eldest child: coded as number with special codes, which cannot be a valid data value (e.g. a negative value), for 'missing' and 'error'.[1] The value zero can be retained for babies under 1 year. [1]

There also needs to be a code for 'not applicable' for the Age of oldest child [1]

Sex: coded as 0 and 1 or as M and F with special codes for missing and error. [1]

Marital status coded as 0,1,2 or M S D with special codes for missing and error [1]

Not applicable, error and missing are all different and need to be distinguished for the analysis [1]

(20 marks)

6.

i) In an experiment factors of interest are controlled by the experimenter [1]

In an observational study the researcher observes items and events without changing any of them [1]

ii) It is possible to randomise participants to the two different treatments in an experiment [1] thereby making the two groups as similar as possible apart from the factor of interest i.e. the treatments [1]

In an observational study there may be many factors which differ between the groups [1] some of which are not even known [1] and so it is difficult to attribute cause and effect [1]

iii)
a) As routine practice is not changed this is an observational study [1]

 b) Other characteristics of the mothers (e.g. age) [1]

and of the labour (e.g. duration) [1]
and maybe of the seniority/experience of the clinician in charge [1]
These variables could all affect outcome irrespective of the method used [1]

c) If the data are not routinely collected then given the pressures in the situation people may forget to record them [1]

 People may complete the data retrospectively and inaccurately [1]

Some mothers might get omitted from the dataset [1]

Some mothers may not want their data to be used in the study [1]

d) data need to be

checked for errors [1]
cleaned [1]
coded [1]
entered into computer [1]

(20 marks)

7.

i) Questions are sensitive [1] and face-to-face interviews difficult to arrange and expensive [1]

Students are likely to have mobile phones but not easy access to landlines and so it would be impractical to get a reliable list of phone numbers. [1]

So they are sensible decisions [1]

ii) Postal questionnaires are feasible as addresses are known [1], but they are expensive [1] and have a poor response rate [1]

All students use the computers [1] and internet surveys are easy to set up [1] and it is easy to check that all questions have been answered [1].

Internet survey is preferred [1]

(11 marks)