THE ROYAL STATISTICAL SOCIETY

2016 EXAMINATIONS − SOLUTIONS

ORDINARY CERTIFICATE – MODULE 1

The Society has published 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.

Ordinary Certificate Module 1 – 2016 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) Simple random sampling is a scheme in which each sample (of the same size) has the same probability of selection.

   [2 marks, allow 1 mark for "each item has the same probability of occurring"].

   In order to obtain a simple random sample, a list of all sampling units within the target population is required and the required sample size is obtained by random selection without replacement.

   [3 marks - a list is required [1], select items at random [1], without replacement [1]].

   In order to obtain a systematic sample, the first sample member is selected at random, then every nth sample unit is selected from the list or arrangement of units to obtain the required sample size.

   [2 marks - starting point at random [1], then every nth item [1]].

   Systematic sample will behave as though simple random in the absence of periodicities [2 marks].

   (9)

(ii) A sample of 100 customers is required and we know the estimated number of calls is 400 per day. If the call-centre is open Mondays to Saturdays only, then we may need to recruit 16.67 (100/6) callers each day (assuming the number of calls received each day are similar).

To estimate *n* where *n* represents every *nth* caller to be selected:

$$\frac{total\ no.callers\ per\ day \times no.days}{sample\ size\ required} = \frac{400 \times 6}{100} = 24$$

To achieve a sample of approximately 16.67 callers each day, every 24th caller would systematically be included into the sample. In practice, a random number between 1 and 17 would be selected as the first to be surveyed, and then every 24th caller after that would be included in the sample e.g. if the number 8 was randomly selected, caller numbers 8, 32, 56 etc. would be selected for inclusion in the sample.

By recruiting our sample over a weekly period i.e. Monday to Saturday, we would hope to reduce sampling bias due to different callers calling on different days of the week etc.

The systematic sampling method should include callers throughout the whole day, reducing bias due to some callers being more likely to call at a particular time of the day e.g. in the evenings.

[5 marks for explaining how to calculate 'n' where 'n' represents the nth caller; 4 marks for explaining how to reduce bias].

(9)

2. (i) Two advantages and two disadvantages of each method may include any of the following (alternative solutions may be given) :

(a) Advantages of using interviews as a data collection method are (1) interviewers can repeat or explain the meaning of questions to respondents face-to-face if necessary, minimising errors in the data collection or incomplete responses (2) respondents are more likely to participate if invited personally to take part in a survey (3) the interviewer can be reassured that the correct or appropriate person invited to take part is actually the person answering the questions in the survey, increasing the reliability of the study findings.

Disadvantages of face-to-face interviews are (1) they can be costly to implement due to the interviewer's time and any travel costs etc. incurred (2) it can take a long time to gather the data by face-to-face interviews and therefore it will take a long time before the data can be analysed and findings presented (3) face-to-face interviews are prone to interviewer bias to some degree.

(b) Advantages of postal questionnaires are (1) a wide geographical area can be easily reached (2) they are reasonably cost effective (compared to interviews) (3) respondents are more likely to answer sensitive questions when completing the questions in privacy i.e. when not face-to-face.

Disadvantages of postal questionnaires are (1) it can take a long time to collect the data e.g. if reminders need to be sent out (2) response rates tend to be low leading to a reduced sample size and nonresponse bias (3) we cannot be sure if the target respondent is filling out the questionnaire e.g. it could be another member of the household thereby compromising the reliability of the findings.

(c) Advantages of an internet survey are (1) it is cost effective (2) a large number of respondents can be reached quickly (3) data can be stored electronically which minimises errors associated with entering data.

Disadvantages are (1) not everyone in the target population may have access to the internet leading to sampling bias (2) it is not always possible to implement random sampling as a list of email addresses for the target population would first be required (3) Response rates tend to be low leading to response bias.

[1 mark for each two advantages and each two disadvantages for each of the three data collection methods in parts a to c].

(12)

(3) Advantages of the use of secondary data include (1) data are often readily available e.g. U.K. census data, so the data can be obtained quickly, (2) if good quality data are available for use, savings related to the costs of collecting, collating, imputing, formatting and cleaning the data are made, (3) good quality secondary data sets e.g. from large government surveys, often have a technical manual which explains the sampling and data collection methods etc., so researchers can assess the reliability and quality of the secondary data set before using it, (4) it helps to make primary data collection more specific, as the researcher can determine the gaps and deficiencies in the secondary data and then decide what additional information needs to be collected.

Disadvantages are that (1) secondary data may not provide complete coverage e.g. geographically for the purpose in hand, (2) the data may have been collected several years ago and therefore the findings from a statistical analysis of the data may also be out of date and not relevant, (3) the researcher may have difficulty checking or determining exactly which definitions e.g. of geographical areas or industrial classifications have been applied during data collection, (4) the researcher must assess whether the secondary data is reliable and does not contain too many errors or missing values, in order to be confident that the statistical findings will be valid and reliable.

[1 mark for each of four advantages and four disadvantages discussed].

(8)

(4) (i) A researcher who wishes to collect information on the amount of time that red deer spend grazing, sleeping and resting may use observation to collect this information as the deer are unable to answer questions for themselves and this is a feasible approach to record the information.

(2)

(ii) The main problems with observing the behaviours of the wild red deer and the time they spend on each behaviour are (1) it is difficult to keep track of individual deer within the herd as they are moving around, (2) researcher bias may be a problem as the researcher must interpret the deer behaviours and record the data accurately, (3) the presence of the observer may influence the deer behaviour (even though the researcher is well away, deer have a very good sense of smell).

[2 marks each for explaining 3 problems].

(6)

4

5.   (i)

---

1. Please select your age group:

19 years or less   ☐

20-24 years   ☐

25-29 years   ☐

30 years or over   ☐

2. Please select your sex:

Male   ☐

Female   ☐

3. Do you do any paid work during term time whilst studying at University?

Yes   ☐

No   ☐

Sometimes   ☐

4. On average, how many hours of paid work do you do per week?

None   ☐

1-10 hours   ☐

11-20 hours   ☐

More than 20 hours   ☐

5. On average, how many hours do you spend per week on your university studies, outside of the timetabled classes?

None   ☐

1-10 hours   ☐

11-20 hours   ☐

More than 20 hours   ☐

[2 marks for each of the 5 questions included - of which 1 mark to be awarded for the clarity of the question and 1 mark awarded for appropriate category options;

2 marks for the overall layout and consistency of instructions etc.].

(12)

5.   (ii)   A short study information sheet:

<div style="border:1px solid black; padding:10px;">

Research Study Information Sheet

A research study exploring the length of time students spend on their university studies and how much paid work they do is being undertaken. You have been selected to take part in this research study and are invited to complete the attached survey questionnaire. Your participation is voluntary and you are under no obligation to take part. However, your participation would be much appreciated and we can assure you that your responses will be kept confidential and secure and will not be used for any other purpose beyond this research study.

</div>

[1 mark for brief sentence explain the research study;
1 mark for explaining that participation is voluntary;
1 mark for assuring confidentiality;
1 mark for assuring data security].

(4)

(iii)   There will be many different courses and student cohorts, lecturers and tutors etc. across the university, and if the questionnaire is to be administered during class time it would be desirable to cause the least disruption possible, but at the same time obtaining a representative sample of students across the university. Cluster sampling may involve selecting a random sample of student courses or cohorts (as the clusters), then students can be selected from within these clusters as required. Cluster sampling would therefore cause less disruption as only the classes selected as clusters would be involved in the data collection process. Cluster sampling would also enable the data collection to be less time consuming and easier to administer as not every class and tutor/ lecturer would be involved.

[2 marks maximum for each advantage discussed (maximum 4 marks)]

(4)

(iv)   It would be unethical if staff were able to observe the student response on the completed questionnaires, as this could potentially bias their perceptions of students' abilities or potential achievements. Also, students may not be willing to complete the questions honestly if they feel that their tutors have access to their responses.

This may be overcome by having an independent person come to the classroom to administer the questionnaire whilst the tutor leaves the room

during the data collection. Also, students should be given assurance that their data will be held securely and only used for the purpose of the research study.

[2 marks for discussing ethical issue(s);
2 marks for explaining how the issue(s) may be overcome].

(4)

6. In an experimental study, the investigators apply a treatment or an intervention to the experimental units e.g. people, animals, plants etc., then any effects following the treatment or the intervention are measured. For example, one group of children may be exposed to a new teaching approach whilst another group of children continue with the existing approach. Researchers may then measure or assess whether the new teaching approach is more successful than the existing approach.

In an observational study, the investigators do not apply any intervention or treatment, but observe subjects and measure variables or outcomes. In an observational study, the treatment or intervention that each subject receives is beyond the control of the investigator. For example, the researcher may be interested in whether smoking adversely affects fitness levels. The researcher will observe and record levels of smoking and fitness levels in individuals and may assess whether there is a relationship between smoking and fitness.

[1 mark for defining experimental study; 1 mark for giving an example of an experimental study; 1 mark for defining observational study; 1 mark for giving an example of an observational study; 2 marks for explaining the main difference between the two studies].

(6)

7. (i) In order to calculate the response rate in a survey, you need to know how many people were invited to particulate in the survey and how many people responded.

The response rate $= \dfrac{number\ of\ responses}{total\ number\ invited\ to\ participate}$

(2)

(ii) Non-response bias occurs when those in the sample who do not respond differ systematically from those who do respond. For example it may be that those who choose to respond are more interested in the survey topic, have more time available, or any other reason which may make them more likely to respond than those who do not. Non-response bias may cause a problem because the survey findings may not be representative of the whole study population if there are (often unknown or unmeasurable) systematic differences between respondents and non-respondents.

[2 marks for explaining non-response bias;

2 marks for explaining why it is problematic].

(4)

(iii)   Carrying out a pilot questionnaire survey allows the researcher to try the survey procedure in practice. Based on the feedback and experience of pre-testing the survey, this gives the opportunity for the survey data collection to be improved to ensure that the survey instructions, questions and response options are clear and appropriate before administering the final survey. This will help to minimise errors in the data as respondents will understand the questions and instructions and will be more likely to complete the questionnaire. The pilot survey may also help to test the process of inviting respondents and for the completed questionnaires to be returned. Data from large pilot surveys may also be coded ready for analysis and this will act as a check that the survey questions are designed in such a way that they can be coded appropriately in preparation for analysis.

[2 marks for showing an understanding of what a pilot survey is;

4 marks for explaining the benefits of carrying out a pilot survey (with a maximum of 2 marks for each benefit discussed)].

(6)

8.(i) The aim of the Scottish Household Survey (SHS) is to inform Government policies in areas including social issues and households. The SHS provides information on the characteristics and attitudes of individuals in Scotland, the composition of private households and the condition of Scotland's homes.

The target population is the general population in Scotland and approximately 11,000 household interviews are completed annually.

A multi-stage stratified sampling design to select households and individuals is used. For the first stage, a sample of households is selected. The second stage involves sampling to select addresses within each local authority. Once the overall sample is selected, a subsample is selected for a physical house condition survey.

In a selected household, the highest income householder or their partner/ spouse is interviewed (face-to-face) to gather data on the demographic breakdown of the household members and their economic situation details relating to the dwelling. This also includes questions on childcare and questions about a randomly selected schoolchild (if there are any schoolchildren in the household). The data for the physical house condition survey is collected by a trained surveyor who visits the property and records his/ her observations.

Weighting procedures are applied to the results to correct for the unequal probabilities of selection of households and individuals, as well as variations in response rates. Results are mainly presented as counts and percentages in tables, and confidence intervals are also provided.

[2 marks for describing the survey aim or purpose; 1 mark for defining the target population; 2 marks for explaining the sampling approach; 2 marks for explaining the data collection method(s); 2 marks for describing the methods of analysis and presentation of findings].

(9)

(ii) Potential sources of error may be due to non-response bias as some people refuse to take part in the survey and these non-respondents may systematically differ from those who do respond.

The sampling frame only covers private households and does not cover the target population because accommodation in hospitals, prisons, military bases, large student halls of residence etc. are not included. Therefore the results of the survey may not be representative of the complete population as these non-private accommodation types are not included.

There will be natural sampling variability as data are collected on a sample and not the complete population. However, confidence intervals have been presented to try to estimate the sampling error.

[1 mark for explaining each of three sources of error].   (3)