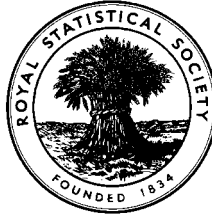# EXAMINATIONS OF THE ROYAL STATISTICAL SOCIETY

*(formerly the Examinations of the Institute of Statisticians)*



## ORDINARY CERTIFICATE IN STATISTICS, 1997

## Paper I

## Time Allowed: Three Hours

*There is no restriction on the number of questions that a candidate may attempt, nor on the order in which they are attempted. Candidates are not required to answer all the questions: they should answer as many as they can.*

*The number of marks allotted to each 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 silent, cordless, non-programmable electronic calculators.*

*Where a calculator is used the **method** of calculation should be stated in full.*

1. Government departments regularly perform major surveys to give detailed information about particular characteristics of their country's population. The General Household Survey and the National Travel Survey are examples of two such surveys performed in the United Kingdom. Similar surveys are performed in other countries.

   Describe six pieces of information about the *conduct* of the survey that you would expect to find within the *published report* of such a survey. [Note that you are *not* expected to quote examples of specific statistics themselves.]

   (12)

2. (a) State one advantage and one disadvantage of using *systematic sampling* rather than *simple random sampling*. (4)

   (b) A schoolteacher wishes to make a random selection of eight pupils to help her perform a science experiment. The names of all the pupils in her class are as follows:

   | | | | |
   |---|---|---|---|
   | Angela | Ben | Charlie | David |
   | Eric | Fiona | Graham | Hazel |
   | Ian | John | Karen | Linda |
   | Mike | Nancy | Olivia | Peter |
   | Quentin | Rachel | Sarah | Terry |
   | Una | Valerie | Wendy | Xavier |

   Use *systematic sampling* to obtain a sample of eight pupils for the teacher, explaining how you obtained your sample. (5)

3. *Cluster sampling* and *stratified random sampling* are two commonly used methods of sampling.

   (a) (i) Define the term *multi–stage sampling*.
       (ii) Give an example of a two–stage sample. (5)

   (b) With reference to the example you quoted in (a)(ii), define the terms
       (i) *stratified random sampling*;
       (ii) *cluster sampling*. (4)

   (c) State one advantage and one disadvantage of using *cluster sampling* rather than *stratified random sampling*. (4)

**Turn over**

4. An electricity generating company supplies electricity to 1,000,000 different properties in a city. Each of these properties is classified into one of three different types – "domestic", "commercial" or "institutional".

As part of an energy efficiency campaign, the company wishes to measure energy wastage (as a percentage of total consumption) in a random sample of 15,000 of these properties. To select which properties to visit, the company uses *stratified random sampling* where the sampling fractions in the strata are made proportional to the standard deviations of responses in that strata.

That is, $n_i$ *is made proportional to* $N_iS_i$,

where $n_i$ denotes sample size, $N_i$ denotes total number of properties and $S_i$ denotes standard deviation of responses in the $i$ th stratum.

In order to estimate the standard deviation of responses in each stratum the company conducted a pilot survey in which energy wastage was measured in random samples of 100 properties of each different type. The results of the pilot survey and the total number of properties of each type are given below.

| Type of Property | Number of Properties $N_i$ | Results from the pilot survey of 100 properties: | |
| --- | --- | --- | --- |
| | | $\sum x_{ij}$ | $\sum x_{ij}^2$ |
| Domestic | 800,000 | 1161.4 | 14107.25 |
| Commercial | 150,000 | 1229.7 | 16705.62 |
| Institutional | 50,000 | 1634.7 | 33058.44 |
| Total | 1,000,000 | 4025.8 | 63871.31 |

where $x_{ij}$ denotes energy wastage (as a percentage of total consumption) in the $j$ th property from the $i$ th stratum.

Using the information given, calculate the sample size required in each stratum to give the stratified random sample of 15,000 properties required. (13)

5. (a) What is the essential difference between *longitudinal* and *cross–sectional* surveys ? Give one example of each type of survey. (4)

(b) Discuss two advantages and two disadvantages of longitudinal surveys over cross–sectional surveys. (8)

2

6. A research group wishes to perform a large nationwide sample survey and has written to you to request statistical advice. You are told that all previous surveys performed by the research group have used only *personal (face–to–face) interviewing*. For this survey, however, the group wishes to consider the use of *telephone interviewing*.

   Write a letter to the research group to discuss the advantages and disadvantages of telephone interviewing, compared to personal interviewing.

   Divide your letter into three main sections:
   (i) describing differences in the method of sample selection between each method of interviewing;
   (ii) comparing the quality of information that would be obtained by each method of interviewing;
   (iii) discussing the potential biases in survey results involved with each method of interviewing. (18)

   Note: You may assume that the research group will understand any statistical terminology you use.

7. Bias in survey results can arise from *non–sampling errors*. Certain types of non–sampling error are known as *response errors*.

   (a) Define the term *non–sampling error*. (3)

   (b) Give three examples of possible sources of *response error* in a sample survey. For each example, state how this source of error may lead to bias, and how a researcher could eliminate or reduce this bias when planning his/her survey. (12)

8. Describe how *optical scanning devices* can be used in the conduct of a sample survey and discuss one advantage and one disadvantage of this type of equipment. (8)