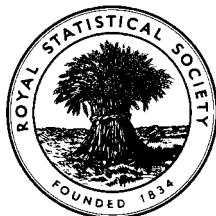# EXAMINATIONS OF THE ROYAL STATISTICAL SOCIETY

(formerly the Examinations of the Institute of Statisticians)

ORDINARY CERTIFICATE IN STATISTICS, 1998

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. The Family Expenditure Survey (FES) is a survey which has been conducted for about 40 years in the United Kingdom giving details of the expenditure of households in relation to their income and to regional factors. Participating households are visited, interviewed and asked to keep a diary of their expenditure over a short period of time.

   (i)   Suggest two uses for the statistics provided by the FES or another similar survey in a different country.

   (2)

   (ii) The sampling method used by the survey is a *stratified clustered random* sample drawn from a list of addresses by postal district.

         Give explanations of the terms *in italics* used in this quotation.

         What additional feature of the sampling method is needed to ensure that the survey reflects seasonal expenditure accurately?

   (5)

   (iii) The survey groups expenditure by a household into fourteen categories. Suggest what six of these categories are or could be.

   (3)

   (iv) The FES booklet lists the three types of error associated with the survey as sampling errors, non-response bias and incorrect reporting.

         How might each of these sources of error arise?

   (6)

Turn over

1

2. The winners of the twice-weekly lottery held in the United Kingdom are found by machines which choose six balls at random, without replacement, from a set of balls numbered 1 to 49. (The machines also choose a seventh bonus ball but this will not be considered here). The six numbers are then arranged in ascending numerical order.

    (i)   You wish to carry out a "pseudo-draw", or simulation, of the lottery using Table IV of your statistical tables.

          Suggest how you could use pairs of random digits to simulate the choosing of a ball. Your method should lead to discarding as few of the pairs of random digits as possible.

          How would your method be extended to simulate the drawing of 6 balls?

          (4)

    (ii) You are told to use digits drawn in sequence from the top row of Table IV ignoring all gaps. Your first three random numbers will thus be 12 00 58.

          Carry out three simulations of a draw using your method in (i). You should draw up a table showing the random pair of digits and the corresponding ball number, marking clearly any pairs which are to be discarded. Finally list the ball numbers for each draw in ascending numerical order.

          (6)

    (iii) It has been suggested in the media that the machines are producing too many consecutively numbered balls .

          How many of your draws give consecutively numbered balls?
          How could you extend your method to estimate the probability of consecutive numbers being drawn?

          (3)

2

3. As part of the annual audit process, a junior accountant has to select 25 company accounts out of a total of 4,000 accounts and to check them for errors. Details of the companies, such as number of employees and turnover, are kept on a database. The accountant is considering two sampling methods:

A. List the accounts by alphabetical order of company name. Select the first company in the list under each letter of the alphabet, ignoring any company names beginning with X.

B. Assign each company a four-digit number from 0000 to 3999. Use random number tables to choose 25 different four-digit numbers in the range 0000 to 3999 and select those companies matching these numbers.

(i) Which of these two methods would you recommend to the accountant? Give reasons for your answer.

(6)

(ii) Suggest another sampling method that could be used with advantage to select the accounts. Explain what benefits your method has over methods A and B.

(4)

4. A large company plans to conduct a survey amongst its employees to obtain opinions about relocation to another part of the country. It is important that views from all departments within the company are obtained.

(i) Describe how the survey could be conducted if
    (a) stratified random sampling
    (b) quota sampling
    were used.

(6)

(ii) There are six departments in the company:

	Number of employees
Administration	150
Finance and supply	250
Personnel	50
Manufacturing	1300
Sales and Marketing	50
Distribution	200

If stratified random sampling with sampling proportional to size is used to obtain a sample of total size 200, how many employees would be selected from each department?

(3)

Turn over

3

5. Explain the meaning of the *pilot survey* in relation to carrying out a sample survey using a questionnaire.

    (2)

    Describe six aspects of the implementation of a survey on which pilot surveys or pre-tests may provide guidance.

    (12)

6. A leading article in a recent edition of RSS News criticised two methods of carrying out surveys.

    The first survey was conducted as part of a television programme in which a studio audience and panel of experts discussed the future of the British Monarchy. At the end of the programme viewers were invited to phone in and vote for or against the continuation of the monarchy.

    (i)   What is the sampling frame from which this sample was drawn?

    (ii) What are the problems likely to be associated with this type of sample?

    The second type of survey concerns so-called "citizens' juries". In this a "jury" of 12-16 people, suitably chosen to be representative, discuss an issue for several days, before reaching a conclusion as to how to advise a decision-making body.

    (iii) What are the problems likely to be associated with this type of sample?

    (7)

7. A survey of a random sample of schoolchildren aged 11-16 is to be conducted over the whole country to determine the things on which children spend their money and how this is influenced by the area in which they live, the age of the child, the amount of spending money received by the child each week and whether this money is earned or given as pocket money. The children will be interviewed confidentially in school after parental consent has been obtained.

    (i)   Design a form for completion by the interviewer to collect the data sought.

    (10)

    (ii) The results of the questionnaire are to be stored in a database. Give the type of data and suggest a suitable field width for the answers to *six* of the questions in your questionnaire.

    (6)

8. Part of a hospital database, which was begun in 1990, consists of two dates, the date of a patient's admission to hospital and the date of his/her discharge. Each date is entered in the form dd/mm/yyyy where dd is a two digit number representing the day of the month (with leading zero if necessary), mm is a two digit number representing the month (again with leading zero if necessary), and, in anticipation of the millenium, yyyy is a four digit number representing the year.

Thus, a typical entry could be: 05/12/1995   18/01/1996

(a) Outline checks which should be made to verify that the data has been entered correctly. You should mention specifically checks on dd, mm, yyyy and the relationship between the two dates.

(8)

(b) The difference between the two dates gives the patient's length of stay in hospital.

Most lengths of stay are between 0 and 100 days.

How do you think "freak" values might arise in this variable?

What effect might such "freak" values have on measures of location and dispersion ?

Outline a method by which you could identify such "freak" values.

(7)