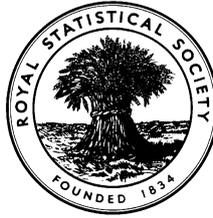


EXAMINATIONS OF THE ROYAL STATISTICAL SOCIETY
(formerly the Examinations of the Institute of Statisticians)



ORDINARY CERTIFICATE IN STATISTICS, 2001

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 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 silent, cordless, non-programmable electronic calculators.

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

1. In England and Wales, the government carries out a longitudinal study called the Youth Cohort Study. This work is designed to study young people from school leaving age (16) and tracks the routes they subsequently follow into further and higher education and the labour market, documenting the training and qualifications they receive.

For one cohort, self-completion questionnaires were issued at ages 16, 17, 18 and 23. The percentage responses were:

<i>Age</i>	<i>% response</i>
16	77
17	76
18	76
23	62

At the second and subsequent ages the figure was given as a percentage of those who responded in the previous survey.

- (i) Define the terms *longitudinal study* and *cohort*. (2)
- (ii) Explain how the initial choice of the cohort was, or could have been, made. You should suggest a suitable sampling unit, sampling frame and sampling design. (4)
- (iii) Calculate the percentage of the original sample responding at each age correct to the nearest whole number. Comment on, and give a possible reason for, the figures. (5)
- (iv) List three groups of people who might be interested in the results of such a survey. (3)

2. You are employed in the marketing department of a national newspaper "The National Daily" that publishes Monday to Saturday. Your paper is running a prize draw. In return for supplying data about themselves and their reading habits, readers aged 18 or over will be entered into a draw to win a valuable prize. You have been asked to design the form to be used in the newspaper as an entry to the draw.

The form has to elicit the following information:

Name, address including postcode (or zip-code), telephone number, email address, age group, on which days the reader usually buys "The National Daily", which other national daily newspapers are bought regularly during the week by the reader and which national Sunday newspapers are bought at least once a month.

Design a form that could be used for this purpose. Marks will be given for clarity of layout and ease of use both by the readers and by those processing the completed forms.

(12)

3. An opinion research company has been commissioned to conduct a survey of 2,000 voters in a town to estimate support for the candidates in a forthcoming parliamentary election. There are 80,000 persons listed on the electoral roll. They are divided geographically into six wards that have distinct characteristics in terms of economic well-being. The company is considering three ways of selecting the sample of potential voters:

- simple random sampling
- stratified random sampling (stratified by ward)
- quota sampling (quotas for each ward).

- (i) Give two advantages and two disadvantages for each of these three proposals.

(12)

- (ii) It is nearly a year since the electoral roll was compiled. Explain what consequences this may have for the survey.

(4)

4. A local radio station carries out regular polls of its listeners on items of current interest. In one such poll listeners were asked to telephone the station and just answer yes or no to the following question:

Do you think dogs should be allowed in public places only if on the lead?

The poll was carried out between 8 am and 9 am one morning. At 8.30 am the announcer said that the percentage yes vote was 63%. When the poll closed at 9 am he announced that the percentage yes vote was 52%.

- (i) What additional number should have been announced at the end of the poll in order to assess how accurate this percentage is? Explain briefly why this number is needed. (3)
- (ii) List three problems associated with this method of polling and suggest why each problem might cause misleading conclusions to be drawn. (6)
- (iii) If respondents could have been asked one question about themselves when they telephoned, suggest a suitable question which might be relevant to their response, and explain how conclusions from the survey could have been extended using the extra information. (3)

5. Explain what is meant by the term *sampling fraction*. (1)

A population has three strata A , B and C . The numbers in the three strata and the standard deviations of a variable of interest are given in the table.

	<i>Number in stratum</i>	<i>Standard deviation</i>
A	1000	10
B	4000	5
C	5000	2

You are required to draw a sample of size 400 from the population.

- (i) Calculate the required numbers in the sample from each of the strata using a uniform sampling fraction. (4)
- (ii) Calculate the required numbers in the sample from each of the strata using sampling fractions proportional to the standard deviation in each stratum. (6)
- (iii) Say, with reasons, which of the two sampling methods you would expect to lead to a better estimate of the mean of the variable. (2)

6. It has been suggested that shoppers are more likely to stay longer in a shop and spend more money if there is an agreeable aroma such as that of newly baked bread, freshly ground coffee or a floral perfume. A researcher is testing this hypothesis in a small gift shop. She has permission to observe customers in the shop on two days. On one day there will be no artificially induced scent but on the second day a floral perfume will pervade the store.

The researcher needs to know how long a customer spends in the store, how much inspection of goods takes place before a decision is made as to whether or not to purchase, and the value of any goods purchased. There are normally several customers in the shop at any one time but she cannot observe them all. Advise the researcher on how to carry out her observations. Your answer should include the selection of customers for observation and the recording of the behaviour on a suitably designed form, together with a mention of any practical difficulties you foresee in the data collection process.

(10)

7. (i) List three types of non-response which may arise in an interview survey to be conducted at the homes of pre-selected interviewees. (3)
- (ii) Explain why non-response may lead to bias in the survey results. (2)
- (iii) Outline why it may be necessary to use randomly chosen substitutes in the case of non-response and explain why using such substitutes does not necessarily overcome the problems of non-response bias. (4)
- (iv) Outline two other ways of trying to minimise non-response and one method of adjusting for non-response bias. (6)

8. A bank is redesigning its customer records system and intends to set up a new database to help it to identify customers who may be interested in new financial products. The database will contain data about the following customer attributes:

- Name and address including postcode
- Home telephone number
- Number and type of accounts (e.g. savings, current, loan) held with the bank.

Design a database for this purpose; your answer should list the fields, giving appropriate field names, together with field types and suitable widths.

(8)