

# **THE ROYAL STATISTICAL SOCIETY**

## **2001 EXAMINATIONS – SOLUTIONS**

### **ORDINARY CERTIFICATE**

#### **PAPER I**

The Society provides these solutions to assist candidates preparing for the examinations in future years and for the information of any other persons using the examinations.

The solutions should NOT be seen as "model answers". Rather, they have been written out in considerable detail and are intended as learning aids.

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, Paper I, 2001. Question 1

(i) A longitudinal study is one carried out over a period of time. A cohort is the sample of people chosen initially to take part in the study, who are then followed over the time period.

(ii) The sampling unit is an individual young person of school leaving age. Hence the sampling frame is all schools with pupils in the required year.

The sample design is to choose a random sample of pupils of the required age using the school lists of pupils.

[Note. The actual survey excludes special schools and those with less than 15 students.]

(iii)	Age	16:	response	77% of the original
		17:	$76 \times 77 =$	59% of the original
		18:	$76 \times 76 \times 77 =$	44% of the original
		23:	$62 \times 76 \times 76 \times 77 =$	28% of the original

The percentages decrease quite rapidly. Young people are likely to be quite mobile, so follow-up is likely to be difficult. The first response is only 77%, and this could be because a self-completion questionnaire was used; this would also be a factor in later years.

(iv) Government (the Department of Education actually sponsors the survey); educationalists (colleges, universities, training organisations); employers (supply of suitable skills and training).

Ordinary Certificate, Paper I, 2001. Question 2

ENTRY FORM FOR OUR PRIZE DRAW

Please complete the form in BLOCK CAPITALS or print your answers clearly

1. TITLE Mr / Mrs / Miss / Other \_\_\_\_\_ (please specify)  
Cross out those that do not apply
2. INITIALS \_\_\_\_\_ SURNAME \_\_\_\_\_
3. ADDRESS \_\_\_\_\_  
\_\_\_\_\_  
\_\_\_\_\_  
POSTCODE \_\_\_\_\_
4. TELEPHONE NUMBER DAYTIME \_\_\_\_\_  
EVENING \_\_\_\_\_
5. EMAIL ADDRESS (IF ANY) \_\_\_\_\_
6. PLEASE RING YOUR AGE GROUP  
18 – 25    26 – 35    36 – 45    46 – 55    56 – 65    OVER 65
7. DURING A TYPICAL WEEK, ON WHICH DAYS DO YOU BUY "THE NATIONAL DAILY"?  
  
MON    TUE    WED    THUR    FRI    SAT    (please ring)
8. DURING A TYPICAL WEEK, PLEASE SAY WHICH OTHER NATIONAL DAILY PAPERS YOU BUY (please write the names)  
  
Monday \_\_\_\_\_  
Tuesday \_\_\_\_\_  
Wednesday \_\_\_\_\_  
Thursday \_\_\_\_\_  
Friday \_\_\_\_\_  
Saturday \_\_\_\_\_
9. WHICH NATIONAL SUNDAY NEWSPAPERS ARE BOUGHT ONCE A MONTH OR MORE OFTEN ? Please list them (OR write NONE)  
  
\_\_\_\_\_  
\_\_\_\_\_

[Note. 8 and 9 could also be asked by giving a full list of all those available and asking for boxes to be ticked. This could take a lot more space]

Ordinary Certificate, Paper I, 2001. Question 3

(i) Simple random sampling, advantages:

- electoral roll can be used as a sample frame
- there is no personal bias in selecting units
- sampling variation can be estimated mathematically.

Simple random sampling, disadvantages:

- selection process tedious when the population is large
- some wards may be represented much better than others
- there may be non-response of selected units.

Stratification by ward, advantages:

- electoral roll can be used as sampling frame since it distinguishes between wards
- all wards represented, giving information about each ward as well as total population
- sampling variation for each ward can be estimated mathematically
- less tedious to select because each ward contains a manageable number of units.

Stratification by ward, disadvantages:

- non-response may occur
- preliminary calculation of sample sizes in each ward necessary, using guesses of relative variability
- calculation of overall sampling error less straightforward.

Quota sampling by ward, advantages:

- no difficulty over non-response
- only limited area to be covered, therefore quick
- different wards all represented satisfactorily.

Quota sampling by ward, disadvantages:

- no estimate of sampling variation can be made
- results may be biased through choice of individuals to be approached, and through willingness or not to reply
- appearance of interviewer may cause some people to respond, other not.

[Two comments required for each.]

**Continued on next page**

(ii) The roll will not be an up-to-date list of residents, due to deaths, removals into or away from area, including from one ward to another and from town to country or visa versa. Also any building, clearance or renovation schemes may have affected the structure of wards, of the numbers in them and the economic characteristics.

Ordinary Certificate, Paper I, 2001. Question 4

(i) The number of respondents, because the larger this is the smaller will be the standard error of the estimated percentage. This must be taken note of when assessing the meaning of the result.

(ii)

(1) The poll is only of listeners to the radio station; so it will not be representative of the whole area it serves. The audience may be biased to particular age-groups, economic characteristics, work and leisure habits, and to those who like the sort of entertainment the station gives.

(2) A telephone is required to answer the poll. There may be difficulties in getting through, or accessing a telephone.

(3) The timing will exclude several groups of people, perhaps even those out walking their dogs.

(4) The nature of the question may make dog owners more likely to respond.

(5) Giving a figure half-way through the hour will encourage more no-voters to respond as their view is in a minority (or vice versa if the announced percentage had been below 50).

(6) People can vote more than once if no identification is asked for and checked. This is a source of considerable bias.

[Three problems required.]

(iii) "Are you a dog-owner?", with the figures for Yes and No kept separate. This allows comparison of the views of the two groups. [It is possible that age or sex may be relevant also, but only one question is allowed.]

Ordinary Certificate, Paper I, 2001. Question 5

The sampling fraction is the proportion of the total population (or of a particular subgroup) that is used in the sample survey. If there are  $N$  in the group, of whom  $n$  are in the sample, the sampling fraction is  $n/N$ .

(i) A uniform sampling fraction requires  $400/10000=0.04$  to be selected from each stratum, i.e.

40 from  $A$ , 160 from  $B$ , 2000 from  $C$ .

[The population size is  $1000 + 4000 + 5000 = 10000$ .]

(ii) The sample fractions in  $A, B, C$  must be  $10k, 5k, 2k$  where  $k$  is a constant that will achieve the required total 400. The sample sizes then are  $1000 \times 10k, 4000 \times 5k$  and  $5000 \times 2k$  which add to  $40000k$ ; this has to be 400, so  $k = 0.01$ .

Hence 100 in  $A$ , 200 in  $B$ , 100 in  $C$ .

(iii) The second method should reduce the standard error of the estimate.

Ordinary Certificate, Paper I, 2001. Question 6

She should sit in a position where she can see the whole shop clearly, and can also see the cash desk/till to record the value of goods purchased. A pre-printed form should be used for each customer observed, recording sex, age (in the form of a very broad classification, young/middle-aged/old, since she cannot ask the customers), time of entry and exit. A floor-plan of the shop, printed on the form, would allow direction and pattern of movement around the shop to be recorded. The number of times an item is looked at, or picked up for examination, can be recorded on the plan.

Only one person can be observed at a time, so as soon as she is in her observation she should observe the first customer coming in, and when that customer has finally left she can take the next one to enter. This should ensure reliable records, and span the whole working day (with breaks taken at convenient times, e.g. for refreshment). Forms will be numbered and dated, and used in order, so comparisons between days, and of times in the same day, can be made.

Care must be taken not to be conspicuous, or to disturb the normal running of the shop; staff need to be fully aware of her task and to prevent customers asking her for assistance (so far as possible). Difficulty could arise if a shop cannot be seen fully and easily from one place – and probably should not be used for the study. Any groups of shoppers, perhaps looking for a single item, may be hard to record properly.

Ordinary Certificate, Paper I, 2001. Question 7

(i) People may simply refuse; may be away from home; may be out at the time of the call; may be unsuitable to be interviewed, for various reasons; or may be new occupants, not the persons on the available list. Pre-selected names should not normally be replaced by substitutes.

(ii) Non-respondents may well have different characteristics from those who do respond, and bias will depend on the extent of these differences and the amount of non-response.

(iii) It may be necessary to keep a sample up to the planned size, to provide sufficient data for analysis and for adequate estimation of sample variance. In stratified sampling, strata proportions need to be kept correct if response rates are likely to vary between strata. But since substitutions are necessarily responders, the difficulties in (ii) still remain.

(iv) Skilled professional interviewers can sometimes help with unsuitable interviewees and potential refusals. Brief, clear and well-designed questionnaires may overcome these difficulties also.

Repeat calling is used for those not at home, either choosing a different time of day or day of the week, in the light of any knowledge about age, sex, occupation etc of respondent, or if possible making a firm appointment.

Results could be weighted for characteristics such as age, sex, social class if it was thought, or there was information, that these differ between respondents and non-respondents.

Ordinary Certificate, Paper I, 2001. Question 8

A typical database might look like this:

FIELD NAME	FIELD TYPE	WIDTH
Title	Text	4
Surname	Text	24
Given_name	Text	24
Initials	Text	6
House_no	Numeric	4
Address_1	Text	36
Address_2	Text	36
Address_3	Text	36
Postcode	Text	8
Telephone	Numeric	16
Number	Numeric	2
Savings	Attribute	1
Current	Attribute	1
Loan	Attribute	1
Deposit	Attribute	1
Other	Attribute	1

[Attribute  $\equiv$  Yes/No]