

THE ROYAL STATISTICAL SOCIETY

2004 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, 2004. Question 1

NOTE. The question does not ask for definitions of the two sampling methods, so these can be assumed.

Advantages of quota sampling over random sampling include the following.

- (i) No sampling frame is needed, so the quota method can be used when a frame is not available.
- (ii) Very little preparatory work is required in the survey office.
- (iii) It is quick to carry out, since interviewers do not have to find specified persons for respondents [*this solution is written throughout in the context of a survey of people, but the ideas will apply in general for other contexts*], so call-back is not necessary.
- (iv) Because of speed, it can be used repeatedly for topical purposes such as election predictions.
- (v) The costs of planning and analysis are less than for random sampling with call-backs.
- (vi) The controls on quotas consist simply of finding men and women in fairly broad categories, such as age-group and type of occupation/work.
- (vii) A specified target sample size can be achieved, since it does not matter which individuals form the sample.

Disadvantages include the following.

- (i) There is no theoretical method of assessing sampling variability.
- (ii) Approximate methods have to be used for estimating variances.
- (iii) Interviewers can easily introduce bias when choosing people to form the sample.
- (iv) Refusals can be numerous if the same vicinity (e.g. a shopping area) is used for obtaining interviews for several surveys in succession.
- (v) Refusals may in any case be more common than with a random sample.
- (vi) It is difficult to control fieldwork, or to make checks while it is being done (and impossible after it finishes).
- (vii) Errors through bias etc form a "hidden" cost because of non-detection.

Ordinary Certificate, Paper I, 2004. Question 2

(i) Possible problems and suggestions for overcoming them are as follows. Credit was of course given in the examination for any other reasonable suggestions.

- (1) Some addresses will be non-residential; if this is recognised, they can be omitted from the sampling frame.
- (2) Extra "reserve" addresses need to be available to replace any that are empty or non-residential.
- (3) Some addresses may be listed more than once, by accident or error or in different categories, but no address should be used more than once. However, this will not get over the problem of repeated addresses having a higher probability of appearing in the sample.
- (4) Alternatively, repeated addresses may be deleted, but this could be a time-consuming task before a sample is selected.
- (5) If there is time, a list in some reasonable order will allow someone to travel round and check it for completeness before it is used – geographical order rather than, for example, alphabetical would be needed.
- (6) Some addresses will have more than one household living there (this should not apply to blocks of flats, where each household has its own number), and all households at such an address could be sampled to ensure proper representation of this type.

(ii) Each interviewer could have a quota of interviewees in age, sex and occupation groups, to be ascertained by the first few questions asked. Size of household(s) is a useful criterion also, and the whole of a target area must be covered by the team of interviewers between them. Time of interview should be varied (day, evening) to ensure all residents are available for interview.

Ordinary Certificate, Paper I, 2004. Question 3

A database might look like the following.

FIELD NAME	FIELD TYPE	WIDTH
Customer_Id	Auto-number	6
Title	Text	4
Surname	Text	24
Given_name	Text	24
Initials	Text	8
House_number	Numeric	4
Address_1	Text	36
Address_2	Text	36
Address_3	Text	36
Postcode	Text	8
Telephone	Numeric	16
Doctor	Text	24
AllergyA	Text	24
AllergyB	Text	24
MedicationA	Text	24
StartA	Date	6
FinishA	Date	6
MedicationB	Text	24
StartB	Date	6
FinishB	Date	6
MedicationC	Text	24
StartC	Date	6
FinishC	Date	6

Note. A separate table for medication would also be useful:

FIELD NAME	FIELD TYPE	WIDTH
Medication_Id	Auto-number	6
Customer_Id	Auto-number	6
StartDate	Date	6
FinishDate	Date	6

Ordinary Certificate, Paper I, 2004. Question 4

Please complete this questionnaire only if you are an employee working in the city of (X). If you do not work in (X), we apologise for bothering you.

Please give your Name _____
and Home Address _____
_____ Postcode _____

1. Where in (X) do you work? Please give the name of the road or building and its postcode. _____

2. How far is it from your home to work? Please tick ✓ the appropriate box.

Less than 2 miles	<input type="checkbox"/>
Between 2 and 5 miles	<input type="checkbox"/>
Between 5 and 10 miles	<input type="checkbox"/>
10 or more miles	<input type="checkbox"/>

3. On a normal day with no unusual delays, how long does it take you to travel from home to work? _____ hours _____ minutes

4. What method(s) of transport do you use? Please tick ✓ **all** the relevant boxes.

Foot	<input type="checkbox"/>
Bicycle	<input type="checkbox"/>
Car or Van	<input type="checkbox"/>
Bus, Tram or Coach	<input type="checkbox"/>
Train or Underground	<input type="checkbox"/>
Other (please specify)	<input type="checkbox"/> _____

5. For any of the methods of transport you have marked in question 4, please give the cost of a return trip.

Car or Van (cost of fuel only)	<input type="checkbox"/>
Bus, Tram or Coach	<input type="checkbox"/>
Train or Underground	<input type="checkbox"/>
Other	<input type="checkbox"/>

[**Note.** For many large cities, this question would need amending to allow for use of season or other multi-purchase tickets or "travelcards" that may cover more than one method of transport.]

6. Do you find any disadvantages in your present method(s) of travel?

Yes No

If you have answered Yes, please say what they are.

Thank you for completing this questionnaire.
Your answers will be kept confidential.

Solution continued on the next page

Ordinary Certificate, Paper I, 2004. Question 4 continued

Additional questions would be on relevant important topics, such as delays on public transport, overcrowding, reliability of services, cost, congestion on roads, problems caused by flexible working times.

Ordinary Certificate, Paper I, 2004. Question 5

The advantages of a "diary" include the following.

- (i) There is a much more accurate record of what was eaten and when.
- (ii) Answers do not depend on long- or medium-term memory.
- (iii) A diary form could be designed, with suitable headings and definitions, to make accurate and correct recording easier.
- (iv) A further improvement may be to record quantities in some convenient way.
- (v) Regularity of diet can be included by having carefully specified "time" boxes.

Disadvantages include the following.

- (i) It takes time, and may become tedious, for a diary to be fully completed over a reasonable period.
- (ii) There is no guarantee that it is completed fully accurately, at the time food or drinks are consumed or very soon afterwards.
- (iii) Diets will vary somewhat according to seasonal availability of some items, requiring repetition of the exercise a few times during a year.
- (iv) People may actually change their regular habits during the time they are keeping a diary.
- (v) "Snacks" between main meals may not be recorded unless clear instructions are given – and not always then.

Ordinary Certificate, Paper I, 2004. Question 6

What is to be estimated – in particular, will the interest lie in mean values of measurements or in proportions? This determines which variance formula is used. Proportions give much less information per item and so need much larger samples.

Are estimates required for subgroups? If so, stratification is required, and each subgroup has to be sampled adequately.

How accurate are estimates required to be (how close to the "true" population value), and are the available resources (time, money, staff) sufficient to collect and process sufficient data to achieve this accuracy?

Does the person planning the survey have any information on the variability of measurements to be taken or the size of the proportion to be estimated? If not, few of these questions can be answered satisfactorily and some preliminary work or a pilot survey will be needed.

Is the sample multi-purpose, i.e. required to estimate several things, either several measurements or proportions or a mixture of the two? If so, the sample size must be large enough to meet the requirements of precision for all of them. Assess this in the light of available resources.

What level of non-response may be expected? Allowance for this will be needed in deciding sample size.

How quickly are results needed, and is this realistic with the available resources?

Ordinary Certificate, Paper I, 2004. Question 7

(i) The total number of farms is 400, so we can take a 10% sample (40 farms) and therefore 10% in each size group. Rounded to the nearest whole number, this gives 20, 12, 8.

(ii)

Size group	Number \times SD	Sample size (see calculation below table)
Small	$203 \times 6.4 = 1299.2$	10.67
Medium	$115 \times 11.6 = 1334.0$	10.95
Large	$82 \times 27.3 = 2238.6$	18.38
	4871.8	

The sample sizes are $\frac{1299.2}{4871.8} \times 40$, $\frac{1334.0}{4871.8} \times 40$, $\frac{2238.6}{4871.8} \times 40$.

Rounding, these will be taken as 11, 11, 18.

(iii) Method (i) is proportional allocation, which is easy to plan and does not need estimates of standard deviations in groups. The groups (or strata) are represented in the same proportions as in the population, so the method gives reasonable estimates valid for the whole population without further adjustment.

Method (ii) is optimal allocation, sampling more intensively in the more variable parts of the population and in the larger strata. Its estimates have minimum variance for fixed total sample size (provided the available information on SDs is good). The recorded data have to be kept in the correct strata during the estimation calculations.

Ordinary Certificate, Paper I, 2004. Question 8

- (i) (a) For systematic sampling, number the books in order of positions on shelves, beginning with (say) the top shelf and then move to the second, then to the third and so on until all N books are identifiable. (Since a library will have many sets of shelves, the shelves in one set will be completed first, then move to the next set; this will usually be easier than completing all the top shelves first.)

Calculate $k = N/n$ and round it to the nearest whole number. Choose at random a number between 1 and k (inclusive), say j . Locate the j th book along from the starting point; this is the first member of the sample. Then take every k th book after that.

- (b) For cluster sampling, the clusters could be taken as the sets of shelves or, alternatively, individual shelves could be used. Number these clusters 1 to L . To obtain an approximate value of the number of books in each cluster, use $M = N/L$. Then choose n/M clusters to form the sample, and take every book in each cluster chosen.

Advantages of systematic sampling are that it is easy to carry out and would be very much quicker than a random sampling scheme. Since N is known, no complete count is necessary at the beginning. Most likely there would be no periodic variation in the ages of books, so age can be assumed to be a random variable when based on systematic sampling. All the stock of books would be covered, so long as no shelves were missed in the initial count. There is no theoretical basis on which to study systematic sampling, but either a form of cluster sample analysis can be used or one based on assuming simple random sampling.

Cluster sampling requires the shelves, or sets of shelves, to be numbered first. Care needs to be taken when sampling each cluster to look at every book in it, once and once only. It is a good method if the distribution of ages of books is similar in each cluster and the clusters are similar to the overall distribution in the library. This may not happen if books are arranged by subjects, some of which will have more recent books and others older books.

- (ii) Suitable strata would be very hard to define – books of similar age would take a long time to locate (unless there is a computer listing of stock – in which case average age could probably be calculated directly for the population without sampling). Even if some other strata (not age) were used, the process would not be easy; for example, if shelves (or sets of shelves) were strata, every book would need numbering so that random samples could be taken. This would be very time-consuming.

Stratified sampling does not seem a good idea for this purpose.