

THE ROYAL STATISTICAL SOCIETY

2010 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, 2010. Question 1

- (i) Selection bias is bias due to the method of selecting the sample and arises when the members selected are in some way consistently atypical of the study population. It would result in estimates of population quantities that are systematically too low or too high.

Response rate is the proportion of those selected to take part in the survey who provide a reply. A low response rate could produce poor estimates of population quantities as those who respond might be atypical and not representative of the population, even if there was no selection bias when the sample was taken. Standard errors are also likely to be high, so precision of estimates will be low.

- (ii) In method A, there is no sampling scheme as such. Only those with a particular interest in the restaurant and with time to spare might reply, introducing both selection bias and a low response rate. These are both disadvantages. On the other hand, potentially anyone who approaches the pay desk could respond, which is an advantage, as is the publicity about the survey.

In method B, tables are selected rather than people, but there could well be more than one customer at a table. If the decision as to whom to interview at a table is left to the interviewer, there could be selection bias due to the interviewer. There could also be problems arising from the time of day (or day of the week), as the restaurant is likely to be much busier at some times than others. The response rate will depend on people's willingness to respond, and this might be low as they might not wish to be interviewed while, or just after, eating. The advantages of the method are the element of randomness involved and that a personal approach has the potential to increase the sample size compared with method A. The personal approach might also enable deeper questioning to be carried out ("probing").

Ordinary Certificate, Paper I, 2010. Question 2

- (i) The regions (which consist of small and large urban areas) are considered as clusters. The first stage is to take a simple random sample of clusters. This is cluster sampling. Having selected these clusters, the second stage should consist of stratified sampling, with each selected cluster stratified into small and large urban areas; it may be useful to further subdivide the large urban areas into those with (say) 2 or 3 outlets and those with 4 or 5 outlets. The stratified sampling would be conducted by simple random sampling within the strata, perhaps using proportional allocation. It is common practice that the selected sample of clusters contains only a small number of them (sometimes only one), and sometimes complete enumeration is then carried out within each selected cluster. This relies, of course, on each of the clusters being representative of the population as a whole.
- (ii) The tables need to be numbered. One method is to choose a simple random sample of tables and ask the interviewer to approach customers at these tables in turn in a specified order, returning to tables that were vacant at a later time. Another method is to take a systematic sample of tables and ask the interviewer to follow a similar procedure.

Bearing in mind that interviews will take time, say 10 minutes including time to approach customers and decide who to interview at a table, it might be reasonable to do six interviews in an hour. For restaurants with only 6 tables, all might be approached so that the sample size, in terms of tables, is 100% (complete enumeration); only a random order of tables to approach is needed. For restaurants with as many as 15 tables, a 50% sample of the tables might be appropriate (though perhaps somewhat ambitious in terms of the time taken). Restaurants with intermediate numbers of tables could reasonably have samples of between 50 and 100 per cent of the tables.

Ordinary Certificate, Paper I, 2010. Question 3

A suggested covering letter, to be on the headed paper of the organisation conducting the survey and signed by the chief researcher, is shown below. If the names of all the teachers are available, they could be inserted in the salutation. If the survey has been commissioned by a well-known organisation, the letter could start instead with "We have been commissioned by to undertake ...".

Dear teacher,

We are undertaking a survey to investigate whether teachers in colleges of students aged 16–19 feel stressed by their work and to investigate factors that might affect stress levels. You have been selected by a random process to take part in this survey and we hope that you will agree to do so. Your responses will be strictly confidential to our organisation.

The survey consists of a questionnaire which is enclosed with this letter. Please answer all the questions in the spaces provided. Many can be answered by ticking boxes. Please return the completed questionnaire to me at the address shown in the letter-heading. A reply-paid envelope is enclosed. Alternatively, if you prefer to answer the questionnaire electronically, please email me at [insert email address] and I will send it to you as a Word attachment which can be returned by email.

With our thanks in advance,

Yours sincerely,

[insert name]
Chief researcher

A suggested questionnaire is shown on the next page. Questions 1, 2, 5 and 10 are closed; questions 3 and 4 are open; questions 6, 7, 8 and 9 are rating scale.

Solution continued on next page

Q1. What is your sex?

Male	Female
------	--------

Q2. What is your age group?

Under 30	30 – 44	45 – 59	60 or over
----------	---------	---------	------------

Q3. What subject or subjects do you teach?

Q4. What position do you hold in your college?

Q5. Do you feel that the hours you work are excessively long?

Yes	No	Am not sure
-----	----	-------------

Please indicate the extent of your agreement with the statements in the next four questions.

Q6. My work makes a valuable contribution to society.

Strongly agree	Agree	Neither agree nor disagree	Disagree	Strongly disagree
----------------	-------	----------------------------	----------	-------------------

Q7. I feel valued at work.

Strongly agree	Agree	Neither agree nor disagree	Disagree	Strongly disagree
----------------	-------	----------------------------	----------	-------------------

Q 8. I feel stressed at work.

Strongly agree	Agree	Neither agree nor disagree	Disagree	Strongly disagree
----------------	-------	----------------------------	----------	-------------------

Q9. The pay is adequate.

Strongly agree	Agree	Neither agree nor disagree	Disagree	Strongly disagree
----------------	-------	----------------------------	----------	-------------------

Q 10. Are you likely to leave the sector during the next year?

Yes	No	Do not know
-----	----	-------------

Thank you for your time.

Ordinary Certificate, Paper I, 2010. Question 4

- (i) The total number of teachers is 610, so the required overall sampling fraction is about $150/610$. Working with a sampling fraction of exactly $150/610$ in each college gives

$$(150/610) \times 307 = 75.49 \text{ from A,}$$

$$(150/610) \times 200 = 49.18 \text{ from B,}$$

$$(150/610) \times 103 = 25.33 \text{ from C.}$$

Taking $75 + 49 + 25$ gives a total sample of 149 and a total cost (in £) of

$$(75 \times 5) + (49 \times 10) + (25 \times 7) = 1040.$$

Taking $76 + 49 + 25$ gives a total sample of 150 and a total cost (in £) of

$$(76 \times 5) + (49 \times 10) + (25 \times 7) = 1045.$$

[Note. In the examination, either answer was acceptable; candidates were not expected to give both.]

- (ii) Let n be the total sample size and n_1, n_2, n_3 the sample sizes for the three colleges. The required calculation is set out in the table below; the first four columns repeat the information given in the table in the question.

<i>College</i>	N_i	c_i	s_i	$N_i s_i / \sqrt{c_i}$	$\frac{N_i s_i / \sqrt{c_i}}{\sum N_i s_i / \sqrt{c_i}}$	<i>Cost</i>
A	307	5	7.5	1029.71	0.729	$0.729n \times 5 = 3.645n$
B	200	10	2.8	177.09	0.125	$0.125n \times 10 = 1.250n$
C	103	7	5.3	206.33	0.146	$0.146n \times 7 = 1.022n$
<i>Total</i>				1413.13		$5.917n$

The calculation shows that $n_1 = 0.729n$, $n_2 = 0.125n$, $n_3 = 0.146n$ and the total cost is $5.917n$.

So we require $5.917n \leq 1050$, which gives $n \leq 177.45$.

Using 177.45 as the value of n , we get $n_1 = 129.36$, $n_2 = 22.18$, $n_3 = 25.91$.

Taking 129, 22 and 26 respectively (with which $n = 177$) gives a total cost of 1047.

[Note. Slightly different decimal values might be found depending on rounding within the calculation, but these are unlikely to alter the integer values in the final answer.]

Solution continued on next page

- (iii) There are advantages and disadvantages of both methods. The optimum allocation method minimises the variance of the estimate of the mean number of years teachers have been at the colleges. However, we are told that the survey has several objectives, and there is no guarantee that this optimum allocation for the mean number of years will also be optimal in respect of any other objectives. Indeed, it almost certainly will not, depending on the standard deviations that would apply for other objectives. Using a uniform sampling fraction is safe in that it achieves representativeness across all variables that might need to be measured. Further, it does not rely on the standard deviations, which are only estimates. Compared with the optimum allocation, it uses a noticeably larger sample from college B and so will pick up more of the variation between teachers at B. Conversely, it has a smaller sample at A, but it remains quite a large sample, so A should be well covered. Its overall sample size is considerably smaller, which may have consequences for overall accuracy, but despite this it costs almost as much as the sample found by optimum allocation.

The overall decision is not clear-cut but, particularly as there are several objectives, perhaps on the whole the uniform sampling fraction method is to be preferred here.

Ordinary Certificate, Paper I, 2010. Question 5

(i) Advantages of this longitudinal study include the following.

Recent recruits are likely to be fairly interested in responding (initially at least).

As the same group is followed, any changes can be related directly to the teachers.

As the sampling frame is recent, it is likely to be fairly accurate.

It is only necessary to look at one year's list of teachers to select the sample.

Disadvantages include the following.

The results relate to one particular group only (things might be different for those joining in other years).

Sample members might get conditioned to responding and change some of their views because of an impression they want to create.

It is necessary to wait five or more years to get results for those who have been at the college for five years.

Members of the sample might leave the college or, even if they stay, might get tired of responding and drop out of the study (leading to reduced sample size and/or likely bias due to non-response).

(ii) Advantages of this method of sample surveys include the following.

Information is obtained from teachers with different experiences time-wise.

Results relating to one, two and five years are obtained in one survey and immediately.

It is not too much of a burden on respondents.

Disadvantages include the following.

It could be difficult to select the samples required as records have to be searched for appropriate details or a preliminary "census" done to find out when teachers joined.

Teachers might have forgotten how they felt about stress in the past.

It is not easy to find how views of individual teachers have changed over time.

Ordinary Certificate, Paper I, 2010. Question 6

A pilot survey is a small scale initial survey done with similar procedures to a proposed survey.

It is done to test various aspects of the proposed survey and to help in the design of the survey and to train personnel. In particular:

- Variability and costs can be estimated to help in the determination of sample size
- Decisions on the sampling units can be made
- Sample frames can be drawn up and/or tested for accuracy, completeness etc
- Questionnaires can be tested and improved, including the introduction made to potential respondents
- Interviewers, if used, can be field-tested and given further training if necessary
- Office procedures can be developed and staff can be trained
- Coding and analysis procedures can be pre-tested
- Background information useful to the full-scale survey can be obtained
- The times needed for the different stages can be assessed.

Ordinary Certificate, Paper I, 2010. Question 7

- (i) Several potential problems are listed here with suggestions as to how they might be overcome. [In the examination, candidates were only asked to discuss three problems. Other reasonable suggestions were of course accepted.]

Some addresses might be missing from the list. Could travel round the area and add these to the list before taking a sample, or could take an additional sample from the extra addresses found while doing the survey.

Some addresses might no longer exist. Could take a further sample to compensate for this loss of sample members.

Some addresses might be listed more than once. Delete duplicates from the list before taking the sample if they are spotted. Do not include the same address more than once in the sample.

Some addresses will be non-residential. Might overcome by dropping these from the sample, but would need to choose more addresses than the required number of households to avoid too big a reduction in planned sample size.

Some addresses with more than one household living at them could be under-represented (for example if the address does not identify individual households living at it). Might always include all households living at any selected address.

Some households might have more than one address. Do not include them more than once in the sample (but it might be difficult to identify duplicates of this type).

Some households might not have an address. Perhaps supplement the sample by using other lists.

- (ii) The interviewers should be given quotas to tell them the numbers of households of each size and type that should be interviewed, perhaps also including quotas for the ages and sex of people interviewed. An alternative might be to instruct the interviewers how to take a systematic sample of residential dwellings.

Geographical coverage should be ensured, either by making this part of each quota or by telling interviewers in which area of the community they should interview.

Interviewers should be instructed to interview in evenings as well as during the day, and on all days of the week.

Interviewers might perhaps be advised to consider other ways of finding members of households: for example, as well as knocking on doors, they could stand in shopping centres.

Ordinary Certificate, Paper I, 2010. Question 8

The researcher should observe customer characteristics such as sex, broad age group, ethnicity and whether the customer is with others such as children or other adults. The researcher should observe what goods the customers look at and the approximate time for which they do so, and what they put in their basket or trolley.

A variety of shoppers should be selected for observation, in all the weeks of the survey and at different times of day. This might be done continuously, selecting a further shopper as soon as observation of one has been completed.

The information would be best recorded on a check form so that as far as possible the researcher just has to tick boxes.

One difficulty is that the researcher needs to be unobtrusive, and must avoid being mistaken for a member of the supermarket staff. It is difficult to hide a clip-board. It might be possible to stand at the end of an aisle.

Another difficulty is that, if there are a lot of people around the shelves, it will be hard to tell who is doing what.

Selecting a variety of shoppers might also be difficult as time goes on. To begin with most shoppers will be suitable, but later it might be difficult to fill some quotas, especially in the rarer groups.

It will be difficult to time accurately how long people look at goods; indeed, it may be difficult to tell whether they have looked at them at all.