THE ROYAL STATISTICAL SOCIETY

2002 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, 2002. Question 1

A survey usually aims to estimate a mean or a proportion, e.g. the mean expenditure of a family per week or the proportion of the population holding a particular opinion. If a sample from the whole population is used for this purpose, it must "represent" the population so that the results from the sample can be applied to the wider population. Some methods of selecting a sample do not properly represent a population, e.g. if using a list of members that is not up to date. Some methods of obtaining information will cause non-response, the refusal of people to answer badly constructed questions or the failure to take part in enquiries when people have no interest in a topic. These, and other, errors in carrying out a sample survey rarely affect all sections or groups of a population in the same way or to the same extent, so the answers which are obtained are incomplete but the effect of this on estimated means or proportions cannot be measured. This leads to bias in the sense that the sample estimates, even from large samples, may differ systematically from the true (but unknown) value of the required mean or proportion in the whole population – due to, for example, a non-response group being systematically different from the rest. This bias cannot usually be corrected by statistical methods. It is a structural error in the sampling technique used.

In this survey, the response rate was so small that it gives negligible information on the whole population. It very likely represents the views of minorities or special interest groups who selected themselves by replying. No statistical selection took place, and no reminder to reply seems to have gone out. The 'Newsletter' is most unlikely to have been read in detail by more than a small proportion of the population, and even those who tried to telephone their responses may not always have got through. 52 out of 103456 residents, self-selected, cannot be taken as a good representation of residents in the borough.

Ordinary Certificate, Paper I, 2002. Question 2

It is quite likely that the same errors would occur if a repeat "survey" was done in the same way – responses would be very few and would represent only those with some special interest in the questions asked, very far from the population as a whole. Hence, in any repeat, the same sort of results might occur. Reliability in this sense is therefore a misuse of the word; it does not imply validity. If a sampling method is basically faulty, the same faults will influence any set of results.

Ordinary Certificate, Paper I, 2002. Question 3

These three methods all imply proper statistical selection, which is an improvement. Nevertheless we also need to select from all the population of residents, and none of the methods ensures this.

(i) A simple random sample, using the register of voters as its sampling frame, will exclude those residents newly moved in, and those who failed to return their registration forms for that year; and will include some who have moved out or died. In the UK, these lists become quite inaccurate by the end of a year. This is a source of bias, difficult to estimate. A postal survey, even with up to two reminders, rarely gains more than 50% response unless some form of visit can be organised for non-responders.

Advantages of simple random sampling, given an adequate frame, are that selection is made from the population with no personal bias by the survey organiser; that everyone on the list has the same probability of selection; that some valid statistical analysis of results can be done, and can be generalised to the whole population.

(ii) A stratified sample requires the population to be split into groups or "strata", which should ideally be homogenous within a group, the main differences occurring between groups. A simple random sample is taken within each of the strata. The advantage is that information is obtained for each stratum, and similarities or differences between them can be seen. A disadvantage is that the frame has to be split into strata, and the likely boundaries between strata are not always clear. Sometimes urban and rural may differ, sometimes type of housing is a useful division (and a borough council will have this information), especially if it also represents social class.

Personal interviewers will lead to a considerably better response rate, without the need for reminders, except that some revisiting is needed to those not at home on the first visit. It is more expensive, especially in less densely populated areas. Although it gives better quality data, a larger total sample may be needed if there are several strata, although in the present case that seems unlikely to be necessary.

(iii) The sampling frame is incomplete. Some people do not want their names and addresses published in a directory, some have mobile phones or unlisted cable services. This causes bias, since some groups of people are more likely than others to be in one of these categories. Also, people do not always respond to telephone interviews (unless some publicity beforehand has told the population that a telephone survey is to be carried out, and encouraged them to cooperate).

It is a quick method, even allowing for non-response, and enough reserves can be located to enable the desired size of sample to be achieved – whereas in other methods this depends on the response rate, which can be hard to guess in advance. However, the use of reserves instead of attempting to contact again those who are out at the first call can bias the response in favour of some parts of the population (older, different leisure interests, different hours of work).

This is a very cheap method and so is often used. Its response rate would be higher than for a postal survey.

Ordinary Certificate, Paper I, 2002. Question 4

(i) $N_1 + N_2 = 12000 + 80000 = 92000$ $n_1 + n_2 = 500$ $c_1 = £18$ $c_2 = £5$

$n_1 = \frac{12000}{92000} \times 500 = 65.217$; take $n_1 = 65$.

$n_2 = \frac{80000}{92000} \times 500 = 434.783$; take $n_2 = 435$.

Cost is $n_1c_1 + n_2c_2 = 1170 + 2175 = £3345$.

(ii) $n_i = k\frac{N_i s_i}{\sqrt{c_i}}$ where k is a proportionality constant to be found. Budget for sampling costs is £(3600 – 500) = £3100 $= n_1c_1 + n_2c_2$.

$\frac{n_1}{k} = \frac{12000 \times 11}{\sqrt{18}} = 31112.7$ $\frac{n_2}{k} = \frac{80000 \times 8}{\sqrt{5}} = 286216.7$

Hence $3100 = (31112.7k \times 18) + (286216.7k \times 5) = 1991112.1k$

Therefore $k = \frac{3100}{1991112.1}$ and so $n_1 = 31112.7k = 48.44$, say 48.

Then $n_2 = 286216.7k = 445.62$, say 446.

This gives as cost $(18 \times 48) + (5 \times 446) = 3094$.

(iii) The optimum method gives a total sample size of 494, allowing for costs.

Assuming that in proportional allocation £500 had also been required for fixed costs, the stratum sample sizes would have had to be reduced in the ratio $\frac{3100}{3345}$, giving n_1 = 60.44 and n_2 = 402.94 (take as 60 and 403); $n_1c_1 + n_2c_2$ is then £3095, so one more urban item could be take to give finally n_1 = 60, n_2 = 404. Total sample size would be 464.

The optimum method has given a sample 30 items larger, with 12 fewer in the smaller, more expensive stratum and 42 more in the larger, less expensive stratum.

The strata SDs from the pilot survey are not so very different, so the actual estimates in the full survey will not be much more variable on the simple method (i). Since we may be interested in stratum results as well as overall results, method (i) seems suitable in this case.

Ordinary Certificate, Paper I, 2002. Question 5

A pilot survey allows data collection equipment (such as questionnaires) and methods (such as use of interviewers) to be field-tested and improved, data processing to be checked and if necessary improved, explanation of purpose of survey for respondents to be refined. Sampling frames can be assessed for accuracy, completeness and suitability for use as a basis for whatever sampling method is planned (random, stratified, systematic, multi-stage). Training of enumerators (field workers) and supervisors is carried out in the pilot stage. Decisions on the sampling unit where necessary (e.g. individual or household) can be made, and whether to include or exclude any special types of unit that may exist. A pilot survey may indicate that other items of information need to be collected to make the final survey worthwhile, capable of answering important questions. It will give realistic ideas of the time needed for interviews and travel. Sources of variation, e.g. between towns/villages or parts of towns, can be determined, and/or previous knowledge checked for relevance to the present survey.

Ordinary Certificate, Paper I, 2002. Question 6

(i) (a) As many as possible of the questions should be capable of answers Yes/No, or a limited set of alternatives, with boxes to be ticked. All questions should be 'closed', not open-ended allowing imprecise verbal answers; those that ask for numerical information should have the units clearly stated and a format marked out for respondents to enter information in standard form, e.g. for a date:

D	D	M	M	Y	Y

or for a time:

Hours	Minutes

If decimal answers are possible, the position of the point should be marked clearly. Any questions that must be open-ended have to be copied into the database with answers abbreviated according to instructions given to the operator.

Solution to question 6 continued on next page

(b) A spreadsheet will use a row for each response, so these have to be numbered 1 to N, assuming N ($\leq$700) reply. A pattern could be constructed on the computer screen to make sure the data are transferred into the correct columns of the database and are entered in the right format; otherwise the spreadsheet must be completed carefully row-by-row.

(c) The data may be (partially) validated by providing upper and lower limits for some of the data, e.g. the column containing children's ages could be programmed to reject entries less than 5 years or more than, say, 19 years. Accuracy can be checked by a second person entering the data, and the program rejecting items which do not match. Each column requires the type of data to be specified, numerical or non-numerical.

(d) Usually a symbol such as * is entered where values are missing. When using the data, results for each column (i.e. each data item) may be used as they stand, or the rows with any item lost may be omitted completely. If two-way tables, using data from two columns, are to be produced, a row need only be omitted if one (or both) of these items is (are) lost, so minimising the number of missing values.

The procedures required will depend on which program is being used, and what computing equipment is available.

(ii) Given a good program, used properly, accuracy can be achieved quite quickly, since all the calculations after the checking stage can be done in the computer (column means etc, two-way tables), and if verbal answers can be categorised into a few classes these can also be summarised quickly. A disadvantage could be that the data are not scrutinised so closely as in an analysis by hand, so that some useful two-way tables are not calculated because a possible relation has not been spotted. Comments in words, from individuals or only a few people, may be lost in a purely mechanical analysis. On the other hand, standard statistical packages will note "odd" values (possible outliers) where these may be missed by hand. Computer analysis can carry out all the studies and comparisons that seem useful; this would often be impossible by hand.

Ordinary Certificate, Paper I, 2002. Question 7

(i) The UK Index groups items into broad categories, such as food, alcohol and tobacco, clothing and footwear, transport and vehicles, fuel and light, leisure activities. Household surveys provide estimates of the expenditure in each of these categories from a large number of participating households each year (or two years) and hence the weighting (i.e. the proportion of total expenditure) for each category is found. This weighting becomes out of date fairly quickly, as fashions change or some items suddenly increase in price so consumption is reduced. However, grouping into categories reduces the effect of these changes, as for example seasonal price rises in some food items will result in people buying less, and replacing them with alternatives. Only consumer items are used in the RPI (e.g. savings and investments are not).

(ii) Prices vary substantially in different parts of the country. Different groups of the population will buy somewhat different ranges of products and will react in different ways to prices changes. Some people will use more expensive small shops to avoid travelling; others will concentrate on supermarkets where prices are lower and less variable. Data need to be collected from the whole range of outlets, and a suitable form of "average" found from them.

(iii) Weights for the categories will, in practice, vary substantially from group to group in the population. Children's clothes need more frequent replacement, food consumption by the elderly differs in amount, and in types of food, from that by families with growing children; some types of occupation demand more energy-giving foods, as well as different types and strengths of clothing. Old people are more likely to need supportive footwear, for example, whereas teenage sportsmen and women want quite different special items. Only broad generalisations are possible between different types of household. However it is certainly possible to estimate the expenditure necessary for healthy living and eating as a basis for deciding what a "minimum income" should be. A single "cost of living" index is really a fallacy.

Ordinary Certificate, Paper I, 2002. Question 8

Quota sampling splits a population into a number of groups and samples a prescribed number (quota) of people in each group. Often these numbers will be in the same ratio as the totals in the population. Suppose that a college or university is split into male/female, home/overseas and three areas of subject study A, B, C. The numbers of each of the 12 groups in the population can be found from college records, and so a quota of each group (A/male/home), …, (C/female/overseas) can be specified: this may, for example, be 10% of each total. Interviewers now go in search of the appropriate numbers from each group, and ask the first suitable individuals they meet the questions in the survey they are conducting. As soon as the required number of (A/M/H) have been interviewed, no one else who is in that group will be asked the survey questions; and so on for all 12 groups. There is no question of randomness in the sampling, but if the survey is about opinions on some topic that is likely to affect people within the same group in much the same way, the answers can often be quite representative. However, there is no statistical theory that can be used to assess the results. Interviewers can be told not to concentrate on all the same type of unit, e.g. they can be warned not to go for the tallest males!

The population is the whole college; the frame is the college list; the method gives answers quickly, without the need to set up a randomised scheme, and unless the survey questions are very sensitive the answers are likely to be reasonable reliable.

Cluster sampling is useful where a number of similar large (primary) units, such as villages in an agricultural region, exist and a sample of individual farms or holdings (secondary units) is required. Time and cost can be saved by selecting some of the clusters, at random from all of the population of clusters – which are the villages. The remaining villages are not visited at all, and so the sampling frame only needs to exist for the chosen villages. Often it will have to be constructed as part of the survey, so considerable effort is saved in this way. From each of the chosen villages, a random sample is selected in the usual way, of farms or holdings to take part in the survey. (If there are not many farms in clusters, they can all be taken; but usually a sample of the same size would be taken for the survey.)

The sampling frame at the beginning needs to list all the villages. The population is all the holdings in all the villages. Assuming that differences between villages are not great, resources can be conserved by not having to visit all of them. This allows sampling within villages to be sufficient for a good estimate of variance to be found. A random method which required several villages to be visited for only a single unit, or very few units, to be studied would be inefficient by comparison.