THE ROYAL STATISTICAL SOCIETY

2005 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, 2005. Question 1

(i) Using stratified sampling followed by simple random sampling, we could proceed as follows.

Stage 1: stratify schools into size groups and select samples from the strata.

Stage 2: using the lists of students, select simple random samples from the chosen schools.

Using cluster sampling followed by systematic sampling:

Stage 1: take the schools in different geographical regions as clusters and select some clusters at random from these;

Stage 2: use the student lists as the basis for a systematic sample of individuals.

(ii) For the first scheme above, we are bound to obtain schools of all size groups; but we have to construct the strata first. Then it is easy to estimate means, proportions etc from a simple random sample, but the method is rather tedious to plan and carry out. Also occasionally a very "untypical" sample can result. For the second scheme, clusters are administratively easier to handle, for example by using local offices as bases to cut down travel, but the chosen clusters may not be typical of the whole country and the schools within a cluster may be quite similar to one another. Then systematic samples from lists are easy to carry out, and can ensure a good balance of leaving dates, but there is no theoretical basis for estimating variability.

NOTE that other combinations are possible, and will show similar advantages and disadvantages.

Ordinary Certificate, Paper I, 2005. Question 2

A possible questionnaire might be as follows.

Please give the date you left school: ☐☐ ☐☐☐☐
Month Year

Your age then: ☐☐ ☐☐
Years Months

Your date of birth: ☐☐/☐☐/☐☐
D D M M Y Y

We are interested to know what your occupation has been in each year since you left school. This could be study, or work (paid or unpaid), or something quite different such as travel. For this survey the main occupation is defined as one on which you spent 4 or more hours per week and which lasted at least one month. If you changed your occupation within a year, please give the details of the months spent in each. (If you have any queries about how to complete this question, please contact the survey team.)

Calendar Year 2004 [Repeat all of this for 2005]

Study? Yes ☐ No ☐ (please tick ✓ relevant box)

If YES please give the following information:

Time for which it lasted (months) _____

Name and location of institution _____

[NOTE – it would be wise to give two spaces for answering these questions in case there has been a change]

Work? Yes ☐ No ☐ (please tick ✓ relevant box)

If YES please give the following information:

Paid ☐ Unpaid ☐

Time for which it lasted (months) _____

Name and location of employer _____

Type of work done _____

[Give a second space for answers here also]

Solution continued on next page

If you had a main occupation other than study or work, please give details here.

_____

Please tell us what you hope to be doing in five years time from now:

_____

_____

And in ten years:

_____

_____

(These could be the same as now or something quite different).

Thank you for responding to this survey.

Ordinary Certificate, Paper I, 2005. Question 3

A database might look like the following, continuing with as many occupations as necessary.

The details of the occupations would be coded to indicate names (if appropriate) and locations. Career plans would also be coded.

FIELD NAME	FIELD TYPE	WIDTH
Respondent_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
Date_left	Date	6
Age	Numeric	2
Year_1_occ1	Numeric	2
Date_occ11	Date	6
Details_occ11	Numeric	3
Year_1_occ2	Numeric	2
Date_occ12	Date	6
Details_occ12	Numeric	3
Year_1_occ3	Numeric	2
Date_occ13	Date	6
Details_occ13	Numeric	3
Year_2_occ1	Numeric	2
:		
Career_plans	Numeric	3

It would also be useful to have separate tables for occupations:

FIELD NAME	FIELD TYPE	WIDTH
Occupation_ID	Auto-number	6
Respondent_ID	Auto-number	6
StartDate	Date	6
FinishDate	Date	6
Details	Numeric	3

A spreadsheet would have variable names similar to the field names, and cell widths the same as the field widths. Coding could be the same as the above. Either extra columns can be used for occupations or separate sheets for each. Types are number, date or text. Auto-number ≡ number.

Ordinary Certificate, Paper I, 2005. Question 4

(i) A non-random method could:

(*A*) ensure that regions of both types are sampled, without the need to stratify;
(*B*) use regions that are easy to get to as often as necessary;
(*C*) choose regions that are likely to have a variety of types of market;
(*D*) give what may be thought a "representative" sample.

Disadvantages include:

(*A*) possibility of bias in choice;
(*B*) no valid estimate of sampling variation can be found;
(*C*) areas thought "not typical" may not be included;
(*D*) those least easy to reach may not be used.

(ii) (a) Merchants change from day to day; the same merchant may bring a different set of produce at each visit; some merchants may have access to very little land and so visit rarely; a "sampling frame" of merchants is not likely to exist.

Sampling of stalls, at random or perhaps as a systematic sample, may be satisfactory. Local representatives may be able to say who is a regular stall-holder, if required. Those sampled could be asked how often they come to the market, and some balance between those who are very often there and those rarely there could perhaps be achieved. Supplementary questions about their background and sources of produce could be asked.

(b) It should be possible to construct a comprehensive list of the fruits and vegetables likely to be available in the region at the times of visits. Staple varieties of these should generally be on sale in season, but a few merchants may also have less common specialities. Prices of the staple foodstuffs should be collected if possible (it is not always easy to do so if an actual purchase is not made) and the proportion of merchants lacking any of these should be noted. Interviewers need skill and tact to achieve reliable results. Local representatives can probably advise about shortages, or poor quality of produce, due to adverse weather.

Ordinary Certificate, Paper I, 2005. Question 5

(i) Some of the local representatives may be chosen at random from those with responsibilities in, and/or knowledge of, the community served by the market. Some may have an official position which makes them well known, and able to obtain the required information. Merchants may have to pay a hire charge, and if so the numbers doing so on any particular day could be obtained from the authority receiving the charge. Times of start and end of activity each day could be found by regular visits, and the chosen representatives must be prepared to do this. Numbers could also be estimated in this way. Merchants could be asked exactly where they come from, and their range of produce recorded. The amount they bring will determine the length of time they stay. Representatives have to be taught how to obtain accurate, reliable information.

(ii) Few people are likely to remember last year's prices at any particular season, even if they have a rough idea that things are cheaper/more expensive this year. Merchants could perhaps be asked whether supplies are more or less plentiful than last year, which may be related to price, but any attempt at numerical estimates is probably not worthwhile. Central figures on price and quantity of staple foods may be available but regions will vary.

Ordinary Certificate, Paper I, 2005. Question 6

Interviewers can encourage respondents to give particular answers by asking questions in a particular way, loaded to a particular answer. Also the question may not be fully understood by the interviewer, so it is not answered as it was meant to be asked. Respondents may not be shown the questionnaire upon which answers are recorded, so they do not know all the possible answers expected; or the interviewer may take advantage of illiteracy to enter inaccurate answers. A trained interviewer should ask all questions in a neutral way, not depart from the wording of them, and be careful to record the answer as closely as possible to what the respondent says. A friendly attitude, helping but not forcing people towards answers, is necessary.

When collecting prices from markets, it may be wise to record prices quoted to potential buyers instead of asking directly. A direct answer may be one thought likely to please the interviewer, or biased in either direction according to why the merchant thinks the question is being asked – such as if he fears higher stall charges. Interviewers should explain the purpose of the survey, to avoid suspicions over the reason for it.

Ordinary Certificate, Paper I, 2005. Question 7

(i) There are 80 markets. Hence a 10% sample is required. This requires 2.7 and 5.3 markets, or 3 large and 5 small as the nearest whole numbers.

This will cost $(3 \times 15) + (5 \times 12) = 105$ currency units.

(ii) Using n_1 large and n_2 small markets,

$$n_1 \propto \frac{27 \times 0.05}{\sqrt{15}} = 0.3486 \quad \text{and} \quad n_2 \propto \frac{53 \times 0.07}{\sqrt{12}} = 1.0710$$

so that $n = n_1 + n_2 \propto 1.4196$.

Hence $n_1 = \dfrac{0.3486}{1.4196} = 0.2456n$ and $n_2 = 0.7544n$.

Total cost $= c_1 n_1 + c_2 n_2 = \left(0.2456n \times 15\right) + \left(0.7544n \times 12\right) = 12.7368n$.

This should not exceed 105, so $n \leq \dfrac{105}{12.7368} = 8.24$.

So $n_1 = 0.2456n = 2.02$ and $n_2 = 0.7544n = 6.22$.

Therefore take 2 large and 6 small markets, cost 102.

(iii) If the data refer to a very important vegetable, the optimal allocation may be preferred as it gives a minimum-variance estimate. However, there is no guarantee that it will do the same for any or all of the other vegetables and fruit on sale. Using a uniform sampling fraction should achieve representative results for items of produce as a whole, and it does not need estimated variances to make any calculations. The cost is marginally higher but this is not very serious. The slightly larger number of large markets in uniform sampling could also help in assessing the variation among the large markets.

Ordinary Certificate, Paper I, 2005. Question 8

Index numbers for prices require a set of fruits and vegetables to be specified to go into the index. This is the first important decision to make. Prices (p_i) have to be found for each chosen item, and also the quantities (q_i) of these that are consumed in the population. Decisions on how to collect (p_i) and (q_i) have to be made. The index number is generally calculated as a weighted average based on $\sum p_i q_i$.

Assuming that no index at present exists, the prices from as many markets (or other outlets) as possible should be obtained. These may vary with season, but a decision must be made on a "typical" p_i for the index. Some combination, possibly weighted, of local prices is likely to be best. In the same way, q_i for each item has to be constructed. This is often done by a consumer survey, separately from the price survey, though care should be taken to see whether p_i and q_i for particular parts of the country are related (scarcity in a region may lead to high prices there). Town and rural consumption patterns may differ. The first year's data would be used as "baseline" information, and subsequent years' data compared with that baseline – usually by changing only the prices. Quantities are updated less often. Whatever method has been used to construct the (p_i) should go on being used for subsequent years to provide valid comparisons. (If any serious error is found in the method after later use, a new base may need to be set up using a modified version of the collection method.) Although indices are typically quoted as annual figures, food prices are almost bound to vary by season and the data collection has to allow for this.