

THE ROYAL STATISTICAL SOCIETY

2007 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, 2007. Question 1

Dear Customer,

We value the feedback from our customers on the holidays they take with us. So we hope you will spend a few moments answering the questions below. Please either tick the appropriate box when there is one or write your answer in the space provided. When you have completed the questionnaire, please hand it to your tour leader or, if you prefer, post it when you get home to the FREEPOST address given at the end.

1a. CODE NUMBER OF HOLIDAY _____

1b. DATES FROM _____ TO _____

2a. Are you Male Female

2b. Please indicate your age range 18-24 25-34 35-44
45-54 55-64 65-74 75+

3. How satisfied were you with:

	VERY SATISFIED	ONLY JUST SATISFIED	NOT AT ALL SATISFIED
(a) your accommodation?			
(b) the programme of excursions?			
(c) your tour leader?			

4. Would you recommend this holiday to other people?

YES NO

5. What did you like best about the holiday?

6. What was the worst thing about the holiday?

7. Please add here any other comments you would like to make

Solution continued on next page

Thank you for answering these questions. Please now hand this form to your tour leader, or remember to post it to the FREEPOST address, XXX, FREEPOST 6789, Townsname, YZ77 0XQ

[Note: most companies would include an undertaking not to pass on details to any other companies when people's names are on the form, but in this case there is no need to ask for a name.]

Ordinary Certificate, Paper I, 2007. Question 2

- (i) A large number of possible descriptions exist and, if they are to be classified, a considerable number of completed questionnaires will have to be gathered first so that the coders have a good idea of the sorts of responses that have been made.

Examples of responses include the following.

Beach activities. Swimming, surfing, etc. Sport – watching or playing. Golf, cricket, tennis, winter (eg skiing) etc. Evening activities. Excursions: coach or rail journeys, cruises (educational or other), city/region tours, sightseeing and/or educational. Walking/hiking/climbing. Safaris.

A coding system could either use groups of activities or individual activities having separate codes. If individual activities are each given a code, it might be useful to code them 01, 02, ... in order of popularity as shown by the questionnaires. A disadvantage of this is that similar activities would not have adjacent code numbers. Another disadvantage is that the popularity rankings might be found to change after the coding order has been laid down. Alternatively all activities of the same basic type could have a group of adjacent code numbers, e.g. sport – playing 11, watching 12, golf 13, cricket 14, tennis 15 and so on. However, with this system many responses would need several codes (e.g. a response of "playing golf and watching cricket" would need 11 13 12 14, with a database convention regarding the ordering); an alternative would be to have a code of say 11 for playing golf and a separate code of 12 for watching golf, similarly 13 and 14 for cricket, 15 and 16 for tennis, etc, but there would soon be very many codes. Either way, a response of simply (say) "golf" does not indicate whether it refers to playing or watching the sport. Any system of this kind would require a few basic types to be identified first, e.g. 0 for beach/water, 1 for sport, 2 for travel, and so on. Within the basic types those that turn out to be rare could be combined into the same code ("other") – but, once this has been done, it is very difficult to break the combined code down into its constituent parts again.

- (ii) In closed form, a list of the more popular activities could be given, each having a box to be ticked if the respondent liked doing that activity, while the less popular ones could be combined into related activities such as walking/hiking/climbing that would have just one box to tick.

Ordinary Certificate, Paper I, 2007. Question 3

- (i) A quota sample should contain a given number (quota) of people in each of a set of categories. These categories are based on sex, age-group and any other characteristics of interest, such as in this case whether people are in a group or are travelling independently. Assume he has been told what each category consists of (e.g. "males over 60 travelling independently"), and how many residents he must interview from each category.

He should walk round the hotel complex, visiting all its facilities at different times of the day, and also interview people in the restaurant and in their rooms if possible. There will be a few questions to ask first to identify which category a person is in, so that the manager knows whether he needs another member from that category or not. He should go on searching until all his quotas have been met – some people may not be easy to catch e.g. if they go on organised excursions most days.

A systematic sample can be taken once it is known how many people are needed and how many registrations there are in that particular week. If n interviews are needed and N people have registered during the week, he can take every (N/n) th from the registration list, the starting point being chosen at random between 1 and N/n . [If most arrivals are in groups, the sampling fraction for independent travellers might need to be larger than that for groups to obtain adequate precision.]

- (ii) The systematic method is easy to carry out, and should be effectively random (unless the "sampling interval" N/n unfortunately coincides with any cyclical pattern among the registrations). Everyone who has registered has the same chance of being selected (but this may need adjusting as mentioned above). Some may of course be difficult to locate, or may refuse, as in any survey. A quota sample ought to be representative of the population of residents but the interviewer may be selective in which people are approached – those who look less likely to refuse – and there is a danger of missing altogether some types of residents, e.g. those who breakfast early or dine late because they spend a lot of time out of the hotel. There is no theoretical support for quota sampling because there is no element of random selection in it.

Ordinary Certificate, Paper I, 2007. Question 4

- (i) (A) is two-stage cluster sampling; (B) is stratified random sampling.
- (ii) For cluster sampling (method A) to work well, the views of the patients in the sampled wards have to be representative of those in the hospital on the whole, i.e. in each chosen ward the variability in patients' views must reflect variability in the whole hospital.

For stratified sampling (method B) to work well, the views of patients in sampled wards of a given type (O, SC or IC) should not vary much between wards of that type, but the types may differ noticeably – in which case results for each particular type are more useful than a single "overall" result.

- (iii) (A) Advantages include: sample relatively quick to choose; require details of patients only for selected wards; less effort to visit only a few wards than the whole hospital.

Disadvantages include: may not have all three types (O, SC, IC) represented adequately in the sample; variances of estimators tend to be high.

- (B) Advantages include: easy to get results for each type of ward; possibility of including different questions relevant to each type of ward; estimates are often more precise.

Disadvantages include: takes time to select samples; takes time to visit several wards of each type; details are needed to trace individual patients.

Ordinary Certificate, Paper I, 2007. Question 5

- (i) $N = \sum_i N_i = 85$. So the sample sizes n_i have to be in the ratio $\frac{50}{85} : \frac{25}{85} : \frac{10}{85}$, multiplied by 36 to find the actual numbers. Thus we get 21.2 in ward A, 10.6 in B and 4.2 in C, so we take 21, 11 and 4 respectively.

The cost of this is $(5 \times 21) + (5 \times 11) + (10 \times 4) = 200$ dollars.

- (ii) For A, $\frac{N_i s_i}{\sqrt{c_i}} = \frac{50 \times 1.81}{\sqrt{5}} = 40.47$; for B, $\frac{25 \times 3.23}{\sqrt{5}} = 36.11$; for C, $\frac{10 \times 2.18}{\sqrt{10}} = 6.89$.

$\sum \frac{N_i s_i}{\sqrt{c_i}} = 83.47$, so the ratio $n_A : n_B : n_C$ is to be $\frac{40.47}{83.47}n : \frac{36.11}{83.47}n : \frac{6.89}{83.47}n$, where n is the total sample size.

This gives $0.485n : 0.433n : 0.083n$ and the total cost is then

$$(0.485 \times 5n) + (0.433 \times 5n) + (0.083 \times 10n) = 5.42n.$$

Hence we require $5.42n \leq 200$, or $n \leq 36.9$.

Taking n as 36.9 gives

$n_A = 0.485 \times 36.9 = 17.90$;	take $n_A = 18$
$n_B = 0.433 \times 36.9 = 15.98$;	take $n_B = 16$
$n_C = 0.083 \times 36.9 = 3.06$;	take $n_C = 3$.

The total cost will then be $(5 \times 18) + (5 \times 16) + (10 \times 3) = 200$ dollars, so this is satisfactory.

[Note: because the sample size for the more expensive ward (C) is rounded down to 3, we save just enough to take n as 37. Otherwise we would need to consider $n = 36$ and find suitable n_A, n_B, n_C .]

- (ii) The main difference is in B, where the second method gives a larger sample size. Since B is the most variable ward, this should lead to more precise overall results. A sample as large as 21 in A, as in the first method, is rather wasteful, although C does have 4 on that scheme but only 3 on the second scheme. On balance, the second scheme is likely to be preferred.

Ordinary Certificate, Paper I, 2007. Question 6

- (i) Telephone interviews are relatively cheap – no travel costs, and hardly any wasted time through finding selected sample members are not at home or if the respondent says "call back later". Interviewers' performance is easily monitored as conversations will usually be recorded (it is important to make the respondents aware of this). Sometimes seeing an interviewer can put potential respondents off, whereas hearing may not do so. Answers to sensitive questions may be better in telephone interviews than face to face.

- (ii) However, refusals may be more likely, especially if people have been interrupted at a busy or inconvenient time. Conversations must be kept short. Background noise and the possibility of being overheard are possible. And not everyone is accessible by telephone.

Ordinary Certificate, Paper I, 2007. Question 7

- (i) A longitudinal study follows the same group of people through the whole period of the study. One advantage is that a sample will be relatively easy to choose from the list of last year's graduates, which is likely to be complete and very nearly fully up-to-date in terms of addresses etc, and thus provides a good sampling frame. Recent graduates are likely to be interested in responding, so there should be a good response rate at least initially – though this may fall off over time. It is useful to be able to follow a group through an extended period of time, noting changes in occupations and reasons for them.

A disadvantage is that occupations which are recruiting in one year may not be recruiting every year, so the pattern of jobs that are obtained immediately after graduation may change year by year. Another disadvantage is that contact may be lost as people change jobs or move to other addresses. People may also lose interest in the survey, leading to problems of reduced sample size and possible non-response bias. On the other hand, it is even possible that participation in the survey, with the feeling of being "watched", might make some respondents change some aspects of their occupations in the hope of creating a good impression. A further disadvantage is that the study must obviously take ten years (plus time for analysis) before full results are available.

- (ii) A major advantage of using samples from graduates of five and ten years ago as well as current new graduates is timeliness. Results will immediately be available for people in varying stages of their careers and with different experiences of the initial jobs market.

However, unless the university has kept a good database of its graduates, with sufficient contact to keep addresses etc up to date, the sample frames available at five and ten years will not be as good as the recent one. If contact has not been reasonably regular, even those who do receive an inquiry may not be very interested in responding. These disadvantages would make it wise to select a larger sample to allow for non-response, but this would increase the cost of the survey. Another disadvantage is that it would not be possible to find out how the careers of individual graduates had changed over time (unless the respondents were also asked to record their progress throughout five or ten years – this could be done if it was thought appropriate).

Ordinary Certificate, Paper I, 2007. Question 8

- (i)
 - (1) People who respond might be systematically different from those who do not. For example, they might be very interested in the topic of the survey while most other people are not, or they might hold strong views which in no way represent those of the whole population. This can bias results very seriously.
 - (2) When some of the selected people fail to respond, the achieved sample size becomes smaller than was planned and so the results have lower precision than was aimed for. This can be very serious if there are many non-respondents.
 - (3) Following up non-response (as opposed to outright refusal) can be done by telephone or by having an interviewer visit people, but this costs resources (time and money) and may not be possible in a short time-scale when results are needed quickly.

- (ii) Reasons include the following.
 - (1) The available address is no longer the correct one, as the person has moved.
 - (2) The questionnaire may not get delivered.
 - (3) The questionnaire may be regarded as junk mail and destroyed.
 - (4) The intended respondent may have died.
 - (5) A questionnaire may not reach the survey organiser even if it is returned.
 - (6) People may be too busy to reply, or not interested, or simply set the survey aside until they are less busy – when it is too late.

It may be possible to improve the look of a questionnaire to make it seem more worthwhile answering; to remind people (also by post) at intervals; to send a pre-paid reply envelope with the questionnaire; to offer gifts or inducements, such as taking part in a draw, as an incentive to reply; to take care that the introductory explanation to the questionnaire makes it seem interesting and relevant, and is expressed in simple form rather than unduly "official".

Also selected non-respondents could be visited or telephoned.