5 ERROR, SAMPLE SIZE AND NON-RESPONSE

As has been mentioned in the discussion of Activity 3, there are two categories of error in survey research: sampling error and non-sampling error. Conceptually, the terms sampling error and non-sampling error refer to different entities, and it is theoretically important to consider them as such, but in practice, we can never have a true measure of sampling error, but only an estimate of it, and the influence of non-sampling error is hopelessly confounded within that estimate. Both researcher and research evaluator have to ensure that non-sampling error is avoided as far as possible, or is evenly balanced (non-systematic) and thus cancels in the calculation of population estimates, or else is brought under statistical control. As has been shown, the difference between sampling error and non-sampling error is that the extent of the former can be estimated from the sample variation, whereas the latter cannot. Further, we have seen that sampling error can only be reliably estimated if the selection of respondents has been random. At best, random sampling will allow unbiased estimates of sampling error; at worst, quota and opportunity sampling will provide little or none.

In practice, researchers often overlook or are unaware of these difficulties and quote standard errors, i.e., estimates of sampling error, even for samples where their use is not justified in theory. For example, a research project testing the effect of an innovation in method on the progress of children might use one school class for the innovation, and one other for control. Assume that these were the only classes available to the researcher, but that a coin had been tossed to decide which of the two classes received the innovation and which would be the control. Means could be calculated for the outcome measure and then standard errors could be calculated for these means. These standard errors could be used in a test to determine whether or not there was a significant difference between the means (as explained in Block 4), and the results could legitimately be reported. If you were evaluating this study you would note the positive finding, but also that the method used was 'purposive' or 'opportunity' sampling, and consequently that the standard errors would not be based on random components with a beneficial cancelling of bias. You would further note that for this reason the finding might not stand up to replication — that generalizability would be suspect. You would look for other information to take into account. You would be pleased if it were said, or shown from school progress records, that the two classes were comparable before the research began. You would be equally pleased to know that allocation to treatment and control groups had been at random.

**ACTIVITY II**

Write a brief note of possible sources of error in this study if the school head had insisted on which class should receive the innovation and which should be the control. Say if this would contribute to sampling error or non-sampling error, or both.

Assume, on the other hand, that only two pre-selected classes were available and that treatment and control children were randomly sampled from these. How far would this improve the possibility of obtaining good estimates of sampling error, if at all? What about non-sampling error? (Approx. 150 words)

In presenting standard errors in circumstances such as this, researchers are in effect saying:

OK, I know I haven’t got a random sample, and so can’t estimate sampling error. But this is the best I could do. It could be the case that it hasn’t mattered very much, and thus I have calculated the standard
errors and have used them in further tests. My finding has support from the literature, and looks useful. It's up to you dear reader, to decide how much reliance you will place on it. Perhaps you'll think that the result is important, and will be able to replicate it without the sampling difficulties which I have had, and have reported.

At least sampling error can be calculated, whether appropriately or not. The various sources of error grouped together as non-sampling errors are another matter—not because they will be necessarily greater in extent, although this could well be the case, but because they are difficult to control, or even detect. The great virtue of randomization is that it takes care of potential sources of bias both known and unknown. If it can be assumed that error, whatever its source, will be randomly spread across a sample, and will cancel when statistics are computed, then one does not even need to know what it is that is cancelled. The problem is systematic, non-random error, which will not cancel.

Non-sampling error is often overlooked when survey findings are evaluated, and if an estimate of sampling error is given, then it is often wrongly assumed that this shows the likelihood of total error. For example in the 1992 general election in the UK one survey predicted a Labour Party vote of 42 per cent ± 3 per cent. The figure of 3 per cent here will be derived from the estimated sampling error for the percentage of 42 per cent and the number of respondents sampled. Presumably it represents approximately twice the standard error, and thus the 95 per cent confidence range for this result would be from 39 per cent to 45 per cent. This says that if the same pollsters drew sample elements in exactly the same way, and questioned and recorded in exactly the same way, from the same population, a sample of the same size, then they could expect to obtain a value in that range 95 times for every 100 samples so drawn and tested.

This statement about error tells us nothing whatsoever about whether the sampling frame truly represented the voters of the UK overall, let alone the more restricted set of those who actually did vote. It tells us nothing about interviewer bias, or procedural reactivity, or untruthfulness on the part of respondents. If one took a guess and allowed another 3 per cent for all of these, then the predicted range would increase to 36–48 per cent, which would greatly decrease the usefulness of the survey finding, since in a moderately close election it would be very unlikely to predict successfully which way the outcome would go, because the estimates for the two major parties would always overlap. An advantage the pollsters do have, however, is replication. Many polls are taken, and by different organizations. Taking all into account might give some possibility of balancing some sources of non-sampling error—but not all. It could, of course, be the case that all the polls suffered from similar defects in which case pooling would not cancel the bias and prediction would be highly unreliable.

Major sources of non-sampling error related to the sampling process itself include: sampling-frame defects, non-response, inaccurate or incomplete response, defective measuring instruments (e.g. questionnaires or interview schedules), and defective data collection or management. Some of these are the subject of other units in this course, but their relevance here should also be kept in mind. Many of these effects are, or can be, controlled by proper randomization in sampling. For example, in a large survey the error related to small differences in technique on the part of interviewers (perhaps consequent on personality differences) will be randomly spread across the data, and will cancel out. Any residual effect should be small and would be lost in the estimates of standard errors, possibly here balancing with other small residual effects.

5.1 SAMPLE SIZE

Often selecting an appropriate sample size has been a hit and miss business of choosing a size which can be managed within the resources available, or a size similar to that used in earlier published work. There is a misconception that sample size should be related to the size of the population under study. As has been
shown above, the precision of sample estimates depends very much on sample size (the sample SD is divided by the square root of the sample \( n \)) and no reference is made to the size of the population sampled.

Assuming that for a sample survey the 95 per cent level of confidence is required \((p < 0.05)\), and the maximum error is set to 5 units on the scale of measurement, then the following formula will provide the estimated sample size.

\[
\text{sample size} = 2 \times \frac{1.96(\text{SD})^2}{5^2}
\]

If the estimated SD is 10 then the required sample size would be approximately 16. If the limit for the difference of interest was reduced from 5 to 2 points, then estimated sample size would increase to close to 100 assuming that the SD remains unchanged. Note that the researcher had to provide an estimate of the SD, although the actual value will not be known until the research is concluded.

This is a very simple account of what might appear to be a simple subject, but which in fact is a complex one. Just how big a sample should be is a matter of balancing cost against the level of precision required. True, as sample size increases the size of the standard error of any estimate of a statistic does decrease. But this needs to be qualified by knowledge that large samples may introduce more non-sampling error (as mentioned in the answer to Activity 3) than smaller ones, where measurements and management problems may be smaller. Also the power of the statistical test to be used must be taken into account and tables have been published for doing this (Cohen, 1969; Lipsey, 1990) But, in theory at least, the precision with which a sample statistic, such as the mean, has been estimated will depend on anything which makes the standard error smaller.

5.2 NON-RESPONSE

Estimating the required sample size needed for a stated level of precision has been discussed. There is, however, little point in reporting that sample size was formally determined to achieve maximum precision, if a sizable proportion of the sample were subsequently lost through non-response, or because items of data were missing. This is a major source of error in many surveys.

Procedures for dealing with non-response and missing data have to be established when the research is being planned, and not left to desperate post hoc remedy. In establishing such procedures total non-response should be distinguished from failure to respond to individual items in a questionnaire, and both should be distinguished from data which are simply missing, i.e., lost or inadequately recorded. Preliminary data analysis will also lead to further data loss, usually due to the dropping of elements (individuals or cases) found to have been included in the sampling frame by mistake, but which do not belong to the population studied, or because inconsistent or highly improbable values have been found on crucial variables for the elements dropped.

Final reports should contain information on the extent of sample loss and missing data, which amounts at least in part, and sometimes completely, to the same thing. Non-response rates as high as 50 per cent or more have frequently been reported. Some elements of the sample simply will not be found, others will refuse to participate either completely or in part. In addition data, and sometimes whole subjects, will be lost due to clerical inaccuracy. The extent of data lost for this reason alone is seldom reported, but is usually surprisingly high, e.g., as much as 8 per cent (Schofield et al., 1992) Response rate is influenced by such design matters as the appearance of a questionnaire, its layout, length and readability. The details of these topics are dealt with elsewhere in this course, but they are introduced briefly here because of the important influence on error in sample survey research. Information on such matters will be sought in pilot studies, in which different versions of a survey instrument can be tested. Sample loss for these reasons is likely to introduce bias, because it might increase the proportion of more persistent or better educated respondents.
If the survey involves home interviews, non-response might be related to time of day at which the interview was sought. From Table 3 it can be seen that a higher proportion of persons over 14 years of age are at home in the early hours of the evening on Sunday, Monday and Tuesday than at any other time. This, however, is also evening meal time, and a busy time for families with young children. Again sample loss could be systematic, and it could introduce bias.

Table 3  Proportion of households with at least one person over 14 years of age at home

(Source: Weeks et al., 1980)

If, when a survey is being planned, it seems likely that response rate will be low due to the nature of the information sought, or the accuracy of the sampling frame, or the method used to contact respondents, then sample size could be increased. This might seem to be an obvious and easy solution, but will be costly in terms of management and material, and in any case, would be unlikely to solve the problem.

ACTIVITY 12

In the planning of a sample survey by questionnaire sent by post it has been calculated that a sample of $n = 200$ will give the required level of precision for estimating population means at the 95 per cent confidence level or better for most items of interest. But only about 40 per cent to 50 per cent of those sent questionnaires are expected to return them. The researchers propose simple random sampling with sample size increased to $n = 400$.

Comment on this proposal. If you were an adviser, what advice would you give? (Approx. 75 words)

Increasing sample size to cover expected non-response would, in fact, be more likely to increase than to decrease bias. More money would be spent, to no avail. Studies have shown that non-responders tend to be the elderly; those who are withdrawn; urban rather than suburban, or rural, dwellers; individuals who fear they will not give the information adequately in comparison to others, or who fear...
that they might expose themselves, and be judged in some way by the responses they make. To lose such individuals selectively would very likely reduce the representativeness of a survey sample. To increase sample size whilst continuing to lose such individuals would in no way help, and could lead to comparatively stronger influence from, perhaps, initially small biasing groups.

Whether information is collected by questionnaire or by interview, positive effort should be made to follow up non-responders. Even when second copies of a questionnaire are sent out, or repeat interviews arranged, response rates above about 80 per cent are seldom achieved. The task for the researcher, who wants sample results which truly represent the population studied, plus information which will help evaluate how far this objective has been achieved, is to get as much information as possible on those individuals who are still missing when all possible action has been taken to maximize response rate.

For this reason the records of individuals who have made only partial, or even nil, response should never be dropped from a data set. Usually information will be available on some variables, for example geographical region, home address, perhaps age or sex. Analyses can be made to see if the missing individuals are at least randomly distributed throughout the sample in terms of these measures, or grouped in some way which might help identify the possible direction and extent of bias on other measures for which there is no data.

Even better would be a small follow-up survey of a random sample of non-responders possibly involving home visits and/or the offer of incentives, so that reliable predictions can be made about the likely characteristics of all non-responders. In some circumstances this could be counter-productive in that interviewer/respondent reactivity might be increased. One way or another, however, the problem of non-response has to be tackled. Vagueness or, worse, total lack of information on this topic, is no longer permissible.

6 CONCLUDING REMARKS

This unit has dealt with methods and problems of designing sample surveys, and has related these to the wider research context where ultimately the validity of findings will rest on how well the sample represents the population being researched. We have seen that the quality of the inferences being made from a sample will be related to both sample size and sampling method. We have seen that provided a probabilistic method has been used then a reliable estimate can be made of the extent to which the sample results will differ from the true population values, and that error of this type is known as sampling error. The methods discussed included both simple and stratified random sampling, systematic sampling, and cluster sampling, and also non-probabilistic methods such as quota sampling. Selecting the best method for any particular research will usually involve compromise, and will be a matter of balancing the level of precision required, in terms of the width of the error estimates, against feasibility and cost.

We have also seen that error from many other sources — non-sampling error — will have to be taken into account when planning survey research and when evaluating results. Major sources of non-sampling error which have been discussed in this unit include faulty selection of the sampling frame, as in the 1936 Presidential election survey, and non-response. There are many others, such as the instruments used for collecting the information, e.g. schedules, questionnaires and observation techniques. The problem for researchers is that however well they plan the technical side of sampling and calculate estimates of sampling error of known precision, non-sampling error will always be present, inflating overall error, i.e. reducing representativeness. Estimating the extent of this is a matter not of mathematical calculation, although statistical procedures can help, but of scientific judgement, based on an awareness of what problems are likely, as well as common sense.
ANSWERS TO ACTIVITIES

ACTIVITIES 1 AND 2

In Sections 1 and 2 of this unit restricted meanings have been given to the following words:

Population: for Activity 1 you probably noted that a population was a collection of elements and gave such examples as those given in the section on sampling, e.g. all the children under 5 years old in the UK. For Activity 2 you possibly amended this to say that a population was a collection of elements about which inferences were to be made, but of course the examples would be unchanged.

Element: an element is the object, quality, process or individual on which a measurement has been taken, e.g. individual children in a population sampled.

Sample: a sample is a set of elements selected in some way from a population, and which is to be used to make inferences about that population. More precisely, a sample is a collection of sampling units (each containing one or more elements) drawn from a sampling frame.

Population parameter. population parameters are statistics, such as the mean, or a proportion, and their standard deviations, calculated using all of the elements of the population, and not just a sample of these. Such parameters are usually unknown.

Sample statistic: a sample statistic is a statistic calculated from the data in a sample, e.g. a mean, or a proportion, or standard deviation, and which is used to estimate the value for the same statistic in the population from which the sample was drawn.

Sampling error: developing the concept of sampling error more fully is a major task in this unit, but at this stage it can be thought of as the error introduced by estimating, for example, a mean from the data in a sample, rather than from measurement of every element in a population.

Sampling units: sampling units are collections of elements which do not overlap, and which exhaust the entire population. For example in a national survey sampling units could be geographical regions.

Sampling frame: a sampling frame is a list of sampling units. It provides access to the individual elements of the population under study, either via sampling units, or directly when the population elements and the sampling units are identical.

Bias: bias in this unit is defined as any effect on data which moves the calculated value of a statistic (such as the mean) further from the population value than would have been the case if that effect were not present such that repeated samples influenced by the same bias would not be centred around the population value.

It is important to learn and understand the technical definition of these words, and many others which follow later in the unit, so that sampling methods and problems in sampling can be established and discussed without ambiguity.

ACTIVITY 3

The main advantages of a sample survey over a full census is that it will be easier and cheaper to set up, manage and analyse than a full census. Although the results based on a sample will, in theory, be less accurate than if the whole population had been included (assuming this to be possible), this might not be the case in practice. Many sources of bias — for example management problems, faulty measurement, lost or corrupted data — will potentially be present whichever method is used, and will be easier to control in a tightly constructed and managed survey than in a full census.
ACTIVITY 4

The table, with the missing entries added, is shown below. The first of these, in the fourth column, is the sample size \( (n) \) for proportionate sampling. This was found by calculating 45.5 per cent of the total sample of 400. This gave a sample proportion of \( n = 182 \) for representation of 16-year-old school leavers. Similar calculations were made to find the other sample proportions. For the disproportionate method the total sample was divided into three equal groups, one for each school leaving age, without taking into account the differing incidence in the population of each of these groups.

Table 1 (completed)

<table>
<thead>
<tr>
<th>School leaving age</th>
<th>Population size</th>
<th>Percentage total in each stratum</th>
<th>Proportionate sample size</th>
<th>Proportionate sampling fraction</th>
<th>Disproportionate sample size</th>
<th>Disproportionate sampling fraction</th>
</tr>
</thead>
<tbody>
<tr>
<td>16</td>
<td>2,730</td>
<td>45.5</td>
<td>182</td>
<td>1/15</td>
<td>134</td>
<td>1/20</td>
</tr>
<tr>
<td>17</td>
<td>1,950</td>
<td>32.5</td>
<td>130</td>
<td>1/15</td>
<td>134</td>
<td>1/15</td>
</tr>
<tr>
<td>18 and over</td>
<td>1,320</td>
<td>22.0</td>
<td>88</td>
<td>1/15</td>
<td>134</td>
<td>1/10</td>
</tr>
<tr>
<td>Total</td>
<td>6,000</td>
<td>100.0</td>
<td>400</td>
<td>1/15</td>
<td>402</td>
<td>1/15</td>
</tr>
</tbody>
</table>

*a The denominators for the disproportionate sampling fractions have been rounded to give whole numbers.

Compare your note on the sampling method you would choose with the explanation given in the three paragraphs following the activity in the text, where several non-statistical reasons are given for balancing the various alternatives. If you have not included the points discussed in your note, it would be a good idea to add a summary of them now to aid your end of course revision.

ACTIVITY 5

Your advice in answer to the first question would probably involve pointing out the difficulty of conducting random sampling across two school years in all the infant and junior schools in a Local Education Authority. You would mention the time and financial cost, and the difficulty of following up those absent from school when the survey was conducted if these were spread across the whole school area. You would probably then suggest cluster sampling, and ask for more information on what the objectives of the survey were to be.

To the researcher proposing a multivitamin intervention you would agree that random sampling within the cluster would be the best procedure from a theoretical point of view, assuming an appropriate research design. But you would advise the researcher intending to introduce a computer into a classroom for some children to use (those randomly sampled), and others not to use, that this would be impracticable, and certain to bias population estimates and to inflate sampling error. You would suggest that cluster sampling should be used throughout, i.e. that the intervention should be given to intact classes, rather than individual children, and that the classes to receive the intervention should be chosen at random.

ACTIVITY 6

Clearly a probabilistic method would be preferable since this would permit a valid estimate of the extent of sampling error. As the population of interest is all the adults in the UK, a simple random sample would be costly and difficult. Precision relative to sample size could be increased by appropriate stratification, and thus you would recommend a stratified random sample.

105
ACTIVITY 8

The mean of the first set of figures is 3, and the mean of the second is also 3. The two standard deviations are, respectively, 1.414 and 4.517. In other words, the two sets have the same mean but very different standard deviations because they differ greatly in the way the individual values are distributed about the mean. The average of this dispersion (the standard deviation, SD) is much greater for the second set than for the first.

ACTIVITY 10

The standard errors given in Table 4 have been calculated by dividing the standard deviations in Table 2 by the square root of the sample size.

The lower extreme of the 95 per cent confidence interval was found by subtracting 1.96 times the standard error from the sample mean. The upper limit was found by adding the same value to the sample mean. Strictly the multiplier of 1.96 is only accurate for larger samples than those in the table, but small sampling theory is not dealt with in this unit. If to simplify the calculations you rounded 1.96 up to 2.0 then your interval may be slightly wider than those given in Table 4.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Standard error, ( S )</th>
<th>95% confidence interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>0.10 years</td>
<td>0.97 to 1.37 years</td>
</tr>
<tr>
<td>Birth weight</td>
<td>0.09 kg</td>
<td>2.47 to 2.83 kg</td>
</tr>
<tr>
<td>Weight</td>
<td>0.26 kg</td>
<td>6.02 to 7.04 kg</td>
</tr>
<tr>
<td>Mother's age</td>
<td>1.13 years</td>
<td>21.63 to 26.06 years</td>
</tr>
</tbody>
</table>

ACTIVITY 11

As randomization would be totally missing no valid estimate could be made of sampling error. In addition there would be a high risk of non-sampling error from some special feature of the classes, not known to the researcher, but presumably motivating the head teacher's insistence. In these circumstances the study would be better not undertaken at all.

Random sampling from two pre-selected classes would certainly improve the situation in so far as sampling error is concerned, but the sample so obtained would only be representative of those two pre-selected classes. Generalization of findings to a wider context could only be on the basis of judgement outside the research itself. The influence of non-sampling error (which would in turn inflate the estimate of sampling error) would depend, amongst other things, on the nature of the intervention, and would be great if there were reactivity between sampled or non-sampled children or from classroom staff in relation to this.

ACTIVITY 12

You will probably want to accept the decision to use random sampling, provided that an appropriate sample frame is available, and also that there is sufficient finance to cover the cost of obtaining a sample of the size needed for the required precision. You will then point out that with an expected response rate of 40-50 per cent the sample is not likely to be representative of the population of interest as the non-responders are likely to differ in important ways from those who do respond. Merely increasing sample size will be costly, and will not help. You would suggest that the additional money should be spent instead on making a second, and even a third, approach to non-responders; doing analyses on whatever limited data is available for the non-responder to see how they differ from
those who do respond, or setting up a small random study of the characteristics of non-responders, perhaps by visiting their homes, or offering an incentive for participation.

FURTHER READING


Although somewhat dated, this remains a standard text for material covered in this unit.


A further elementary text, recently revised, which includes some of the mathematical derivations of sampling methods, but with many practical examples of surveys and methods. Useful later if you have the task of designing a survey.


A short and fairly readable text on statistical power in social science research. It includes charts for determining sample size, and is mentioned here so that you have a reference to a source of information on this.

REFERENCES


ACKNOWLEDGEMENT

Grateful acknowledgement is made to the following source for permission to reproduce material in this unit:

TABLE