Figure 3.2. SAMPLE SURVEY DESIGN/EXECUTION: Non-Sample Errors in Sample Surveys

The following article, reprinted (with slight adaptation) from a CSIRO Division of Mathematics and Statistics *Newsletter*, No. 32, June, 1977, pp. 1-3, provides an overview of the sources and nature of non-sample errors in sample surveys.

BY DENNIS TREWIN
Australian Bureau of Statistics
P.O. Box 10, Belconnen, A.C.T. 2614

The emphasis in training of sampling statisticians is on sampling techniques and methods of measuring sample error associated with these techniques. It is too easy to forget the other errors associated with statistical collections (usually referred to as non-sample errors) and concentrate on sample error. It is inefficient to have sample error reduced to 1% when errors from other sources (non-sample errors) may amount to 10%. This article gives a brief description of the source and nature of these errors. For interested readers, a good reference is Zarkovich.

Non-sample errors can affect both the accuracy (as measured by the bias) and the precision (as measured by the variance) of estimates. It is possible for the bias to be negligible, even though the errors in the individual observations are relatively large. The frequency distribution of these individual errors is important. If positive and negative errors are distributed at random about a zero mean, the estimates of totals and averages will be unbiased, although estimates of regression and correlation coefficients may be biased. But the variance of totals and averages will be increased by the addition of a "response variance" and, if there is correlation between the individual errors and the true value, by a term known as the "correlated response variance." A response variance can exist even though a response bias may not exist. However, in many cases, positive or negative errors predominate and there is a systematic bias. Zarkovich shows the above effects algebraically.

The following is an attempt at a conceptual framework for errors.

A. Errors resulting from inadequate preparations and methodology:
   - Biased procedures;
   - Biased tools.

B. Errors committed during data collection:
   - Missing data;
   - Response or observational errors;
   - Interviewer errors.

C. Processing errors:
   - Editing;
   - Coding;
   - Punching;
   - Tabulating.

This is not exhaustive or mutually exclusive but does give some framework for further discussion.

Biased Procedures

In all statistical surveys, some procedures are prescribed in order to achieve survey aims. A biased procedure is one whose repeated application results in biased data. Examples of biased procedures considered below are biased selecting, estimating, measuring and operational procedures.

The following is an example of a selecting procedure whose design is biased. The example refers to the use of sample surveys to estimate the yield of rice paddy in India by crop cutting from a sample of plots. The instructions issued for selecting fields were as follows:

"Against the name of each selected village are shown three random numbers smaller than the highest survey number in the village. Select the survey numbers corresponding to given random numbers for experiments. If the selected survey number does not grow paddy, select the next highest paddy growing survey number in its place." Fields not preceded by fields growing paddy have a higher chance of being selected, and as these are likely to be different in nature, a bias will result.

Selecting procedures may be well designed but still biased because the implementation is faulty. Zarkovich gives several examples.

Biased estimating procedures are of interest as long as they lead to more accurate estimates than the corresponding unbiased estimates. In other words, interest in biased estimates is based on grounds of efficiency. The concept of efficiency could be extended beyond accuracy. In some cases, the important factor is the convenience of the methods used, cost of computations, etc. For example, computations of variances may be replaced by some approximate method. The justification of these procedures is increased speed, more convenience, etc. Before a biased estimating procedure is used, the bias of the estimator should be assessed. Even for the commonly used ratio estimator, there is the possibility that the bias of the ratio estimator is of a magnitude that cannot be disregarded.

A measuring procedure which can lead to biased results is the use of aerial photographs to measure areas. This procedure is often used to measure areas under crops, forest, etc. A grid is placed over the country and a sample of grid points chosen. An aerial photograph is taken of selected areas and area under crop measured from the photograph. The procedure will be unbiased if the area photographed is on one level. On the other hand, it will be biased if topographic relief is not taken into account.

Operational procedures can also be biased. One example is the use of primary approach letters. Primary approach letters are generally regarded favourably by interviewers and can help to establish rapport between an interviewer and the respondent. As the following example shows, there are situations where a primary approach letter can have disastrous effects. A doctor from the Lidcombe Hospital wanted to survey the aged people in the district to find out their needs for geriatric services. We provided him with a sample of dwellings at which he was going to conduct a first-phase interview aimed primarily at discovering in which dwellings aged people lived. Before conducting this first-phase interview, he sent out a primary approach letter explaining that he was surveying aged people. In the first phase, he obtained less than half the expected number of aged people. He repeated the operation without the primary approach letter and his sample take was close to expectations.

Biased Tools

The use of biased tools introduces various errors in the data collected. An important characteristic of errors resulting from the use of biased tools (random numbers, questionnaires, instructions, etc.) is that the errors tend to have a systematic character. If a particular question on a questionnaire is not properly worded and gives rise to faulty interpretation, the respondents will give the same type of inaccurate answer and the resulting bias may assume considerable magnitude. Examples of tools that could be biased are questionnaires, sample frames, random numbers and instructions and interviewer training.

The questionnaire can be considered as the most important tool in survey work. Any survey is based on some type of questionnaire. The success of a survey depends largely upon the quality of the questionnaire.

There are all sorts of ways in which questionnaires may be biased. Moser and Kalton have considerable discussion on questionnaire design.

Misinterpretation can exist because questions are not sufficiently specific. If a respondent is not able to understand some terms or questions, he may not be willing to ask for explanations and thereby show what he may consider to be his ignorance. This point is demonstrated by the following American study.

1. Do you recall that as a good citizen you voted last December in the special election for your State representative? (33)

(*continued overleaf*)

2. Have you ever heard of the Taft-Johnson-Pepper bill on veteran's housing? (53)
3. Have you ever heard of the Midwestern Life Magazine? (25)

None of these things mentioned above ever existed, but the percentage of "Yes" answers are shown in brackets after the question.

Incomplete alternatives may be given for questions requiring a dichotomous answer. A further common error is to ask a general question when an answer on a specific issue is wanted. Also, simple language should always be used. The aim of question wording is to communicate with respondents as nearly as possible in their own language. Technical terms and jargon should be avoided. Ambiguous questions should also be avoided as should loaded or leading questions. An example of a leading question is: "To what life insurance fund do you belong?" The question is leading because the respondent has the impression that he should be insured, and he may be reluctant to say "none". Leading sometimes takes very complicated forms and is difficult to detect. A study was conducted to determine respondents' opinions on whether the United States was spending too much money on foreign aid. Two different questionnaires were used – the first was made "internationalist" and the other "isolationist". In the internationalist context, 75% of interviewed persons answered that the United States was spending the right amount on foreign aid, while the corresponding percentage in the isolationist context was 20%. Questions and answers already given tended to influence answers to the remaining questions.

The number and order of questions can also have some effect. Experimental results indicate that respondents pay more attention to items listed at the beginning and at the end, than to those in the middle (Payne). Four ideas were presented to respondents in a different order and the respondents were asked to select one of them. Idea *A* was selected by 27% when at the top, 17% when in the middle, and 23% when at the bottom. It was a similar story with the other ideas.

Most factual questions to some extent involve the respondent in recalling information. His degree of success in doing this accurately is a basic determinant of the quality of his response. With certain questions such as sex and marital status, there is no such problem but, with a large range of survey questions, recalling information does provide a problem, the severity of which depends on what is to be recalled. Two factors of primary importance in memory are the length of time since the event took place and the event's importance to the respondent. There is a recent ABS experience of the problems of recall. A pilot study was conducted on a sample of accident victims. To test whether recall problems existed, a list of accident victims requiring hospitalization was obtained and included in the pilot study. Over 50% of respondents selected in this way failed to recall the accident with a 12-month recall period.

There are two types of memory errors. The first is the "recall loss" of which some examples are given above. The second occurs when the respondent reports an activity in the recall period when it actually took place outside that period. This is known as "telescoping" and is particularly prevalent in expenditure surveys, especially for large and prestige items.

The frame is the list of units used in sample surveys for the selection of samples. Since the composition of the sample is directly based on these lists, the frame is one of the most important tools in statistical work. An inaccurate frame will have some units listed more than once, some units omitted from the listings and will also contain some units which do not exist in the population. The different defects in frames are not equally important. As regards the omissions, it is clear that units omitted have no chance of being selected in the sample and could seriously affect estimated totals as the inflation of sample data is generally based on the data of only those units listed in the frame. Duplications mean that some units have "two" chances of selection. Their effect is the reverse of that of omissions, and there are means of estimating the extent of this over-estimation, although the precision of estimates will still be affected. Non-existent units do not cause much trouble if selected in the sample, although they will affect the precision of resulting estimates. They can be treated simply as a zero return.

Instructions and training are also important as far as the quality of data is concerned. An interesting study was made in England of the effect of training on interviewer performance. Two groups were compared. The first group received four hours of customary training only. The second group received a further four hours mainly devoted to practice interviews. To assess the relative effectiveness of the two systems of training, all completed questionnaires were rated in terms of errors. In terms of gross error rates, the second group did significantly better than the first group with an average error score of 21 compared to 36.

Missing Data

In all surveys there is likely to be some missing data. If data from part of the sample is missing, the problem arises as to what conclusions can be made. There is plenty of evidence to show that, in many surveys, the non-responding units are different from the responding units, but the bias can still be small if the non-response rate is low. Missing data may exist because of:

1. Non-contacts;
2. Refusals;
3. Other reasons such as language difficulties, a recent and serious illness in the family, etc.

In ABS Surveys, 1 is the major problem, particularly for surveys of the population. These problems can be minimized if wide publicity is given to the survey, if there is a legal obligation to cooperate, and if there is adequate control on enumeration with suitable follow-up procedures. With a well-designed survey, it is usually possible to keep non-response down to a reasonable level and to estimate roughly what biasing effects it might have on the results. Techniques exist which range from adjustments at the data collection stage to post hoc adjustments to estimates because of biases in missing data. Hawkins gives a good discussion of techniques for dealing with missing data.

Response Errors

The respondent is also an important source of errors in data. Many have been discussed above with questionnaires as it is often the interaction of the respondent and the questionnaire that provides the error. When the respondent starts to answer questions included in the survey program, an extremely complicated machinery of psychological processes often starts working. Thus, even if they know accurate replies to the questions, they may deviate from the truth for various reasons and present an inaccurate response. Respondent errors may be of several types. Memory errors and telescoping are discussed above. Other types of errors are rounding off, prestige errors, and conditioning.

A study conducted in the USA is a good example of rounding errors. Farmers were asked to report for area harvested as well as bushells harvested. The figures on bushells harvested were divided by areas and the following results obtained.

"In 55% of cases for oats and 49% of cases for soy beans, the quotient was exactly an integer. 29% of oats reports gave yields of 40, 50 or 60 bushells/acre. 28% of soy bean reports showed yields of 20, 25 or 30 bushells/-acre. It is a reasonable inference that, in these cases, production was obtained by multiplying acreage by estimated yield per acre."

Inaccurate data are often the result of the respondent's desire to impress (prestige errors). Many examples are known. Women do not like to disclose their age and often declare themselves younger. Some people raise their grade of occupation – medical assistants become medical practitioners, etc. Expenditure on alcohol and tobacco is understated.

Conditioning can be a problem in longitu-

(*continued*)

Figure 3.2. SAMPLE SURVEY DESIGN/EXECUTION: Non-Sample Errors ..... (continued 1)

dinal and panel surveys. Conditioning refers to those changes in the respondent's reaction to surveys in general and to various questions in the survey program in particular that result from previous exposure to requests for the same or similar material. Conditioning may be the explanation for a phenomenon found in labour force surveys in the United States, Canada and, to a lesser extent, in Australia. In all countries, the same respondents are kept in the sample for a number of surveys (eight in the United States and Australia), and the unemployment rate is significantly higher in the first survey than in the following surveys. Conditioning is a likely explanation as, after the first survey, respondents can "prepare" the answers they will give to interviewers.

Interviewer Errors

The most usual method of collecting data in population surveys is by personal interviewing. The use of interviewers usually increases the response rate over what would be possible with mail questionnaires, but they still may be a possible source of error. This can happen in several ways.

1. A number of studies have shown that the personal characteristics of interviewers can influence the answers obtained, either because of the impression she made on respondents or the way that she had asked questions.
2. Several studies have found that interviewers' opinions are related to the responses they have obtained. A sample of Christian respondents were selected with the aim of obtaining their opinion on the influence of Jews in the business world. In the part of the sample where Christian enumerators were used, 56% of respondents declared that Jews had too much influence in the business world. In the part of the sample interviewed by Jewish enumerators, the same answer was obtained from only 22% of respondents.
3. It has been found that interviewer effects arise not so much from interviewers' characteristics and ideology as from their expectations of the respondent's views and behaviour. This may happen in several ways. Questions early in the interview give the interviewer an impression of the respondent. She may interpret answers later in the interview in the light of these expectations.
4. Carelessness in recording is another potential source of error.
5. Cheating does not seem to be a major problem. Complete falsification of entire questionnaires seems to be rare. This problem is reduced if a regular reinterview program is employed.

Processing Errors

I will not say much about processing errors. It is easy to see how errors can occur in coding, punching and tabulating. As these errors will tend to be systematic, it is important to have some form of quality control (*e.g.*, acceptance sampling) to ensure that such errors are under control.

Editing of survey schedules is intended to detect and, as far as possible, eliminate errors in completed questionnaires. Even though editing is designed to improve the quality of data, it can provide a source of error in its own right. One of the greatest problems a sampling statistician has to face in sample estimating is that of an outlying observation, particularly when the observation contributes a significant amount to the sample estimate or the sample error. The usual technique is to decide whether or not such a sample observation is really representative of the parent population. The usual procedure is that if it is decided that the observation is representative, no adjustment is made, but otherwise it is deleted from the sample. Such a procedure is biased, and can only be justified on the grounds that the mean square error is reduced. But wholesale undertaking of this type of deletion can lead to serious biases and this has been shown to be the case in several surveys run by the ABS.

References

Hawkins, Darnell F. (1975) "Estimation of Nonresponse Bias". *Sociological Methods and Research* **3**, 4 (May), 461-488.

Moser, C.A. and Kalton, E. (1971) Survey Methods in Social Investigation, *Heinemann Educational Books*, London.

Payne, S.L.B. (1951) The Art of Asking Questions. *Princeton University Press*, Princeton.

Zarkovich, S.S. (1966) Quality of Statistical Data, FAO, Rome.

1 The word *error* or *errors* is used at 45 places in the text of the article reprinted in this Figure 3.2, in the following contexts:

- ..... sampling techniques and methods of measuring sample *error* ..... [Column 1]
- It is too easy to forget the other *errors* ..... [Column 1]
- ..... (usually referred to as non-sampling *errors*) and concentrate on sample *error*. [Column 1]
- It is inefficient to have sample *error* reduced to 1% when *errors* from other sources (non-sampling *errors* ..... [Column 1]
- ..... the source and nature of these *errors*. [Column 1]
- Non-sampling *errors* can affect both the accuracy ....6 [Column 1]
- ..... even though the *errors* in the individual observations are relatively large. [Column 1]
- The frequency distribution of these individual *errors* is important. [Column 1]
- If positive and negative *errors* are distributed at random about a zero mean, ... [Column 1]
- ..... if there is correlation between the individual *errors* and the true value, ... [Column 1]
- ..... in many cases, positive or negative *errors* predominate ..... [Column 1]
- The following is an attempt at a conceptual framework for *errors*. [Column 1]
- *Errors* resulting from inadequate preparations and methodology: [Column 1]
- *Errors* committed during data collection: [Column 1]
- Processing *errors*: [Column 1]
- The use of biased tools introduces various *errors* in the data collected. [Column 3]
- An important characteristic of *errors* resulting from the use of biased tools ..... [Column 3]
- ..... the *errors* tend to have a systematic character. [Column 3]
- A further common *error* is to ask a general question when an answer on a specific issue is wanted. [Column 4]
- There are two types of memory *errors*. [Column 5]

(*continued overleaf*)

1 ● ..... all completed questionnaires were rated in terms of *errors*. [Column 5]

- In terms of gross *error* rates, the second group did significantly better than the first group with an average *error* score of 21 compared to 36. [Column 5]
- **Response *Errors*** [Column 6]
- The respondent is also an important source of *errors* in data. [Column 6]
- Respondent *errors* may be of several types. Memory *errors* and telescoping are discussed above. Other types of *errors* are rounding off, prestige *errors*, and conditioning. [Column 6]
- A study conducted in the USA is a good example of rounding *errors*. [Column 6]
- Inaccurate data are often the result of the respondent's desire to impress (prestige *errors*). [Column 6]
- The use of interviewers ..... still may be a possible source of *error*. [Column 7]
- Carelessness in recording is another potential source of *error*. [Column 8]
- **Processing *Errors*** [Column 8]
- I will not say much about processing *errors*. [Column 8]
- It is easy to see how *errors* can occur in coding, punching and tabulating. [Column 8]
- As these *errors* will tend to be systematic, it is important to have some form of quality control (*e.g.*, acceptance sampling) to ensure that such *errors* are under control. [Column 8]
- Editing of survey schedules ..... eliminate *errors* in completed questionnaires. [Columns 8 and 9]
- ..... it can provide a source of *error* in its own right. [Column 9]
- ..... the observation contributes a significant amount to the sample estimate or the sample *error*. [Column 9]
- Such a procedure is biased, and can only be justified on the grounds that the mean square *error* is reduced. [Column 9]

Our meaning of *error* is *the* difference *between what is stated* [*e.g.*, *in an Answer*] *or assumed* [*e.g.*, *in a response model*] *and the* actual *state of affairs*.

– For each occurrence of the word *error* (as catalogued above) in the article, decide from the context if its meaning is the *same* as ours; if it is *not*, indicate briefly its meaning.

– Identify, as a separate category, any instances where *error* is used as a synonym for *mistake*.

2 The word *bias* or *biased* is used at 34 places in the text of the article reprinted in this Figure 3.2, in the following contexts:

- Non-sample errors can affect both the accuracy (as measured by the *bias*) ..... [Column 1]
- It is possible for the *bias* to be negligible, even though the errors in the individual observations are relatively large. [Column 1]
- ..... the estimates of totals and averages will be *unbiased*, although estimates of regression and correlation coefficients may be *biased*. [Column 1]
- A response variance can exist even though a response *bias* may not exist. [Column 1]
- However, in many cases, positive or negative errors predominate and there is a systematic *bias*. [Column 1]
- *Biased* procedures; [Column 1]
- *Biased* tools; [Column 1]
- ***Biased* Procedures** [Column 1]
- A *biased* procedure is one whose repeated application results in *biased* data. [Column 2]
- Examples of *biased* procedures considered below are *biased* selecting, estimating, measuring and operational procedures. [Column 2]
- The following is an example of a selecting procedure whose design is *biased*. [Column 2]
- ..... and as these are likely to be different in nature, a *bias* will result. [Column 2]
- *Biased* estimating procedures are of interest as long as they lead to more accurate estimates than the corresponding *unbiased* estimates. [Column 2]
- In other words, interest in *biased* estimates is based on grounds of efficiency. [Column 2]
- Before a *biased* estimating procedure is used, the *bias* of the estimator should be assessed. [Column 2]
- Even for the commonly used ratio estimator, there is the possibility that the *bias* ..... is of a magnitude that cannot be disregarded. [Column 2]
- A measuring procedure which can lead to *biased* results is the use of aerial photographs to measure areas. [Column 2]
- The procedure will be *unbiased* if the area photographed is on one level. [Column 2]
- On the other hand, it will be *biased* if topographic relief is not taken into account. [Column 2]
- Operational procedures can also be *biased*. [Column 3]
- ***Biased* Tools** [Column 3]
- The use of *biased* tools introduces various errors in the data collected. [Column 3]
- An important characteristic of errors resulting from the use of *biased* tools (random numbers, questionnaires, instructions, etc.) ..... [Column 3]

(*continued*)

Figure 3.2. SAMPLE SURVEY DESIGN/EXECUTION: Non-Sample Errors ..... (continued 2)

2
- .... and the resulting *bias* may assume considerable magnitude. [Column 3]
- Examples of tools that could be *biased* are questionnaires, sample frames, random numbers and instructions and interviewer training. [Column 3]
- There are all sorts of ways in which questionnaires may be *biased*. [Column 3]
- Techniques exist which range from adjustments at the data collection stage to post hoc adjustments to estimates because of *biases* in missing data. [Column 6]
- Such a procedure is *biased*, and can only be justified on the grounds that the mean square error is reduced. [Column 9]
- But wholesale undertaking of this type of deletion can lead to serious *biases* and this has been shown to be the case in several surveys run by the ABS. [Column 9]

Our meaning of *bias* is *the model quantity representing inaccuracy*, which is *average error – its* systematic *component – under repetition.*

– For each occurrence of the word *bias* (as catalogued above) in the article, decide from the context if its meaning is the *same* as our usual one; if it is *not*, indicate briefly its meaning.

3 The word *variance* or *variances* is used at 6 places in the text of the article reprinted in this Figure 3.2, in the following contexts:
- .... and the precision (as measured by the *variance*) of estimates. [Column 1]
- But the *variance* of totals and averages will be increased by the addition of a "response *variance*" and, if there is correlation between the individual errors and the true value, by a term known as the "correlated response *variance*." [Column 1]
- A response *variance* can exist even though a response bias may not exist. [Column 1]
- For example, computations of *variances* may be replaced by some approximate method. [Column 2]

The term *variance* usually occurs in probability theory and there refers to a measure of the *variation* of a random variable – its mean square minus its squared mean.

– [*We* reserve the related term *variability* for *the model quantity representing imprecision*, which is *the standard deviation of error – its* haphazard *component exhibited as variation – under repetition.*]

– For each occurrence of *variance* (as catalogued above) in the article, decide from the context if its meaning is this one from probability theory; if it is *not*, indicate briefly its meaning.

– In which, if any, of its 6 occurrences could *variance* be replaced by *standard deviation*? In those cases where it could be replaced, would this usage refer to a *probabilistic* or a *data* standard deviation? Explain briefly in each case.

The 1977 article reprinted in this Figure 3.2 is also used in Figure 8.5 in the STAT 220 Course Materials.

Blank page