Research Paper

Probability versus Non-Probability Sampling in Sample Surveys

Doherty, M. (1994) Probability versus Non-Probability Sampling in Sample Surveys, *The New Zealand Statistics Review* March 1994 issue, pp 21-28.

Probability versus Non-Probability Sampling in Sample Surveys

Introduction

We are sometimes asked "Why do official statistical agencies, such as Statistics New Zealand, insist on probability-based sampling methods for the surveys they run, while market research companies sometimes use quota sampling?" This article will attempt to answer that question.

Probability-based Versus Non-probability-based Selection

Researchers collect information by a wide variety of methods, ranging from the experimental designs used in the physical sciences through to the surveys more common in the social sciences. Many of these methods of gathering information involve a choice of experimental subject. For example, we may want to choose the patients to be examined in a medical study, or the respondents to be interviewed in a survey.

This choice can be made using probability-based methods, where the choice is by some "mechanical" procedure involving lists of random numbers, or the equivalent. Alternatively, the choice may be made by other methods, invoking some element of judgement. Methods involving judgement are sometimes referred to as purposive selection, judgement selection, or non-probability selection.

In this generality, it is difficult to say very much about the choice between probability and non-probability selection. I will narrow the focus to the two kinds of surveys mentioned in the introduction: probability-based sample surveys, and quota samples.

To set the scene, I first need to briefly describe what I mean by probability-based sampling, and what I mean by quota sampling.

Probability-based Sampling

In probability-based sampling, the first step is to decide on the population of interest, that is, the population we want the results about. This could be, for example, all persons aged 18 years or over who are resident in private households in New Zealand.

We then establish a frame - a listing, at least in principle - of all the units of that population. For our example of the persons in private households, we might use a geographic frame. Private dwellings would be listed according to the geographic area they are in, and people listed inside dwellings.

We select a sample from this frame using a probabilistic algorithm. It is important that every element of the frame has a known chance of being selected, and that we can calculate the probability of selecting the sample we end up with.

The sample might well be selected in several stages. In our example. geographical areas might first be selected, then dwellings inside these areas. Finally, people might be selected inside the dwellings.

In saying that we use a probabilistic algorithm to select the sample, one important feature is that interviewers will have no choice about who they are to interview. The algorithm specifies who is to be in the sample.

To produce our results, we combine the responses from the sample in a way which takes account of the selection probabilities. Our aim is that, if the sampling were to be repeated many times, the expected value of the results from the repeated samples would be the same as the result we would get if we surveyed the whole population.

Because we know the probability of getting each sample we select, we can also calculate a sampling error for the results. The sampling error tells us the amount of variation in the results due to the sampling alone. It gives a measure of the quality of the sample design, and of the survey results.

Quota Sampling

Just as there are many probability-based sample designs, quota sampling is not a single method. As Stephan and McCarthy say, "It is not sufficient to state that quota sampling was used in a survey, and expect anyone to have more than a very general idea how the sample was drawn."

A quota sample may be drawn in stages. The earlier stages may often select geographic areas, and might use probability-based methods, with only the last stage of sampling using quota methods.

The key idea in quota sampling is to produce a sample matching the target population on certain characteristics (eg age, sex) by filling quotas for each of these characteristics. The assumption is that if the sample matches the population on these characteristics, it *may* also match the target population on the quantities we are trying to measure.

Quota sampling resembles a method of probability-based sampling known as stratified sampling. It is different, because although interviewers are constrained by the quotas, they are still using some element of judgement in the choice of the sample. (The amount of freedom interviewers have varies from survey to survey.) Because there is this element of judgement, we have no way of knowing the probability of selecting a given unit of the population.

So, unlike probability-based surveys, there is no "randomisation distribution" to use in deriving results from the sample. Instead, we have to make use of some assumed probability model. This may be a model of the underlying population characteristics, or of the selection process. The validity of the model is an additional assumption, which must be made to derive results.

Note that the method requires good data on the whole population to be available to set quotas. For example, if we are setting age and sex quotas, we need to know the age and sex distribution of the population.

Historical Background

Probability-based sampling is a development of the last 60 to 70 years. Around the turn of the century, Kiar, in Norway, was an advocate for sampling. In the early work, purposive methods (ie non-probability sampling) predominated, but in 1934 Neyman published a paper which laid the basis of sampling theory, and explained the advantages of random sampling over purposive selection. (He used a number of examples in his paper, particularly an unsuccessful purposive sub-sample drawn from the 1921 Italian Census by the Italian census bureau.)

Over the next 20 or so years, the theory of probability-based sample design was further developed, and the major statistical offices were all won over to probability-based design. The first generation of sampling textbooks appeared around 1950.

Non-probability Samples in Official Statistical Agencies

In stating that major statistical agencies were won over to probability-based sample design, I do not mean that these agencies saw no place for non-probability samples. The standard example in a number of sampling textbooks is the case where we want to get information about the urban population, but can afford to sample in only one city. In that case, the textbooks are clear that it would be better not to use probability-based sampling to choose the city. They suggest using judgement to select a “representative" city.

In a number of European countries, major official sample surveys of businesses use purposive selection, because of severe problems in getting respondent cooperation.

In New Zealand, some parts of the Consumers Price Index use non-probability sampling. For example, we do not use probability methods in selecting the outlets where we price. (This is largely on cost grounds. This use of non-probability sampling has the approval of the Consumers Price Index Advisory Committee, a group of consumer and user representatives and experts who provide advice to the Government Statistician on the CPI.)

Quota Sampling may be appropriate when there is no suitable list of the population we are surveying.

For example, a survey of expenditure by international visitors may depend on quota samples (based on country, length of trip, etc) of visitors departing from international airports.

Comparing Probability-based and Quota Sampling

The main differences between probability-based and quota sampling are the following:

- If probability-based sampling is properly carried out, there will be none of the bias which can arise from subjective judgements in sample selection. There is the possibility of such bias, however, in quota samples. For example, interviewers may consciously or unconsciously choose non-threatening or easy-to-approach respondents, or those who are easy to contact.
- As Deville says "... the quota method demands the formulation of a hypothetical model to fit the data. On the other hand, a probabilistic survey does not, in principle, depend upon any model."
- The validity of the model underlying quota sampling may be open to question, and difficult to verify.
- With probability sampling, we use the randomisation distribution to draw conclusions from the sample, and to obtain sampling errors. In a quota sample, we cannot get comparable estimates of precision.
- In general, non-response in a quota sample is handled by selection of another respondent fitting the quota. We can be more flexible in the way we handle non-response in a probability-based sample (although this means the use of some form of modelling).
- In general, the costs of a quota sample will be lower than a probability-based sample of the same size. In this regard, it might be worth repeating the comment of Deming who wrote sternly, "There is no way to compare the cost of a probability sample with the cost of a judgement sample, because the two types of sample are used for different purposes. Cost has no meaning without a measure of quality, and there is no way to appraise objectively the quality of a judgement sample as there is with a probability sample."

Recent Approaches

I have discussed the earlier history of random sampling, and the use of models in understanding quota sampling. Samplers are not as unanimous as they once were on the foundations of their craft, and the topics we have been discussing bear on their differences. Over the last 20 or so years, sampling theorists and practitioners have debated the role of models, and the place randomisation should play in sample design. Some of this discussion of the foundations has been heated.

The dust seems to have settled a little now. For example, last year Statistics New Zealand played host to C. Sarndal and B. Swensson, two of the three authors of the best recent text on survey sampling. During their visit, they gave a seminar which addressed, among other things, the differing views on the foundations of statistics.

To summarise (from memory) my impressions of what they said, they pointed out that no existing survey exactly matches the ideal picture of probability-based sampling painted in the classical textbooks. In real-world surveys there are problems, like non-response, which can only be dealt with by assuming some kind of model. For example, we typically handle non-response by identifying "similar" respondents whose data we can borrow. Implicit in that is a model of response behaviour. Like it or not, there is an element of modelling in all real surveys. None the less, the practitioners of classical sampling theory are not misguided in their endeavours to establish good frames, monitor and control non-response, etc. A compromise position is possible.

Sarndal and Swensson embody this compromise in the title of their text, "Model Assisted Survey Sampling", that is, use models to assist us in developing probability-based designs, and the methods we use to produce estimates.

In this more relaxed climate, two authors sympathetic to model-based ideas have recently published important papers on quota sampling (Smith 1983 and Deville 1991). Smith's paper uses a model-based approach. He gives technical conditions for drawing conclusions from a sample selected by non-random methods. He argues that the model-based approach is appropriate is situations like

the following:

- there is a single client;
- the results are for the use only of that client, and will not be published more widely; and
- the statistician and client agree that the required assumptions hold, at least approximately.

His summary is "... in the public sector ... there is no simple well-defined user and it is reasonable to ask that the sampling method used should have wide acceptability. Random sampling methods provide that wide acceptability."

Deville, more recently, takes a fresh and detailed look at the forms of modelling which underlie quota sampling. In terms of his theory, he lays some emphasis on the size of the sample in determining the choice between a probability-based and a quota sample. He considers that for a small sample, "the bias of a quota sample will be more tolerable than the lack of precision of a probabilistic survey."

However, he reaches the same conclusions as previous authorities about the place of quota samples in official surveys. His final conclusions are: "In a survey, the use of any speculative model represents methodological risk-taking. This may be perfectly reasonable if the users are aware of it, and if they have ratified the speculations leading to the specification of the model ... Official statisticians, on the other hand, are responsible for generating data that can be used by the entire society; **...** Official statistics should not tolerate any uncontrollable bias in its products. It should carry out sample surveys using probabilistic methods."

Summary

This article has briefly explored why official statistical agencies prefer probability-based surveys. The main reason is that , as both theory and previous experience shows, using other forms of sampling carries risks that the underlying assumptions are not appropriate. These are not risks official statisticians generally feel comfortable taking.

Bibliography

Deming WE (1960), "Sample Design in Business Research", John Wiley and Sons, New York.

Deville Jean-Claude (1991), "A Theory of Quota Surveys", *Survey Methodology,* Vol. 17, pp 163-181.

Neyman J, "On the two different aspects of the representative method: the method of stratified sampling and the method of purposive selection", *Journal of the Royal Statistical Society,* Vol. 97, pp 558-606.

Sarndal CE, Swensson B, and Wretman J (1992), "Model Assisted Survey Sampling", Springer-Verlag, New York.

Smith TFM (1983), "On the validity of Inferences from Non-random Samples", *Journal of the Royal Statistical Society,* Vol. 146, pp 394-403.

Stephan FP and McCarthy PJ (1958), "Sampling Opinions", John Wiley and Sons, New York.