Data gathering 2

Preview

Introduction

This topic is a continuation of topic 6 and is concerned with the theory and methodology of sampling.

Sampling forms a foundation to methods of data gathering such as interviews and questionnaires. Here we will therefore introduce these key issues:

- the theory of sampling;
- determining the sample size;
- probability and non-probability sampling; and
- sampling error and systematic bias.

Theory of sampling

In daily life we commonly make judgments on the basis of an incomplete picture. For instance, we may buy a CD on the basis of only one song we hear on the radio. This is a sample of this singer's work, but how often have you found that it is not a representative sample, and that the rest of the music on the CD is in a completely different style or suggests a different mood altogether?

Serious research sampling is conducted differently, but is, nevertheless, based on the theory that it is possible to make judgments about the whole based on data gleaned from one part. In reality of course we do not need to be statisticians to know that there is no way that one sample unit could be perfectly representative of the entire population, or, as it is sometimes called, the 'sampling frame'. However, while there may be limitations, sampling does allow us to concentrate our attention upon a relatively small number of people and, hence, to devote more energy to ensuring that the information collected from them is accurate.

Formal research is therefore distinguished from our daily informal sampling by a serious concern to limit the possibilities of unrepresentative conclusions. Essentially this means being rigorous in our methodology, backed up by a sound understanding of how statistics work.

There are two important statistical laws upon which sampling is based. Namely:

- *the law of statistical regularity*, which states that 'any group of objects taken from a larger group will tend to possess the same characteristics as the larger group'; and

- *the law of inertia of large numbers*, which states that 'large groups are more stable than small ones'.

Applied to research, the first law means that a group of people selected at random from a defined population will tend to exhibit the same behaviour as the whole population. The second law means that the larger the samples, the smaller the differences will be between samples, and thus between a sample and the population.

The advantages of sampling are succinctly summed up by Kress (1988) in *Marketing Research*:

> *Reduced costs*. A much lower cost results from gathering information from only a portion of a population. It is less costly to survey 150 people than it is to survey 15 000 people.
>
> *Greater speed*. Data can be collected and summarized more quickly when a sample is used. This condition is especially important in projects where time is a key factor. Politicians frequently want immediate feedback on the public's reaction to major speeches or critical events. TV networks want 'next-day' feedback on audience response to new shows.
>
> *Greater accuracy*. This seems to be a contradiction. How can results taken from a sample be more accurate than those taken from the entire population? The increased accuracy occurs because a smaller, better trained, and better supervised workforce can be used to obtain information, if sampling is involved. (This assumes a rather large population is being studied.) If all adults of a large community had to be questioned about their television-viewing habits, a huge number of interviewers would be needed. This means people of marginal skills might have to be used as interviewers with a resulting decline in the quality of the data obtained.
>
> *Greater depth of information*. Since a smaller group of respondents is involved when a sample is used, the researcher may decide to get more information from each respondent. Consumer panels enable a limited number of housewives to be interviewed regularly about their attitudes toward certain products or stores. Because of the small number of the participants, these interviews can be quite lengthy, allowing in-depth information to be obtained.
>
> *Preservation of units*. There are some situations where the testing process involves the actual consumption or destruction of products. This applies to quality control studies involving such items as ammunition, matches, tires, and the like. Unless these types of tests were conducted on only a portion of the population, there would be no product remaining for eventual sale or distribution.
>
> (Kress 1988, pp. 156–7)

Probability and non-probability sampling

If you are using sampling for your own research project or commissioning work from others you need to know the difference between probability and non-probability sampling. This is important because the method you choose will determine whether or not you are able to generalise from the sample to the sample frame, based on the findings. For instance, the reason for choosing a probability design is to increase the probability that the sample statistics will be as close as possible to the population parameters.

Various techniques are used in sampling to increase the probability of greater accuracy with either method. These include random sampling, systematic sampling, stratified sampling, cluster sampling and so on. You should become familiar with the purpose of each of them.

Think carefully when designing your strategy, accounting for issues such as the extent of generalisability, the availability of time and resources, and the purpose for which the study is being done.

Sampling problems

Determining sample size

Determining sample size can be a problem in research you do yourself, especially if you must convince other people of the validity of the results. Generally, the larger the sample the better, but clearly the parameters of your project will determine what is reasonable. The basic principle is that a reliable and valid sample should enable the researcher to generalise the findings to the population being investigated.

Sampling error

Most readers of research results realise that it is impossible to totally eliminate error from sampling. They assume that the researcher has done their best to limit error rather than eradicate it. However, they do expect that, when the findings are presented, the researcher has clearly acknowledged the problems which should be considered when interpreting the results.

'Sampling errors' are the errors that are unavoidable. They relate primarily to that inherent problem of sampling; the fact that it is highly unlikely for any sample to be a perfect miniature of the whole population. As Kotler and Kotler have written:

> No sample is likely to produce results that are precisely those of the entire population. It is always possible to pick, strictly by chance, a group whose members happen to be different in some important attribute from the population as a whole. When differences between the sample data and the population data result purely by chance, they are known as sampling error. Statistical findings from smaller samples are more likely than findings from larger samples to differ from population data, so smaller samples have more sampling error. The larger the sampling error, the lower the reliability; so the smaller the sample, the lower the reliability of the data.
>
> (Kotler & Kotler 1998, pp. 165–6)

Systematic bias

Examples of 'systematic bias' are of much greater interest to us because we can do our best to try to avoid them. There is no way of avoiding them all; most professional market researchers consider some of them to be almost as 'inherent' in the sample as are 'sampling errors'. That might sound cynical, but the reason it is mentioned here is to remind you not to say anything in your report like this: 'In order to eliminate systematic bias, the researcher did the following … '. Say instead: 'In order to keep systematic bias to a minimum, the researcher did the following … '. It is more realistic, and will thus give you more credibility.

Here is a good list of examples of 'systematic bias' from Kotler and Kotler (1998), each with an appropriate heading and a short explanation:

1. *Frame bias.* A probability sample is drawn from a wrongly chosen population. For example, estimating the proportion of people who like contemporary art by sampling solely spectators at a football game is likely to systematically underestimate the true number, because many people who like contemporary art may not be avid sports enthusiasts.

2 *Selection bias*. The procedure for drawing actual sample members always excludes or underrepresents certain types of population members. This occurs, for example, when telephone interviewers only telephone during the day, thereby underrepresenting the people who work. It also occurs in field surveys where the interviewers pick whom to interview. If all the field workers tended to interview only people who looked friendly, the findings would be biased.

3 *Non-response*. The contacted people decline to participate out of a lack of interest, antagonism, or busyness.

4 *Interviewer bias*. An untrained interviewer deliberately or inadvertently leads the respondent to a particular response. This can occur, for example, when interviewers read or 'clarify' a question in a way that suggests a particular answer is preferred.

5 *Questionnaire bias*. Poor or confusing wording, leading questions, identification of the research sponsor when this information is to be withheld, and omission of important possible responses result in bias.

6 *Respondent bias*. The respondent lies or unwittingly distorts an answer. Often this source of bias manifests itself when respondents exaggerate their income, educational level, or frequency of attendance.

7 *Processing bias*. The interviewer or respondent writes down the wrong answers, the office enters the answers into the computer incorrectly, or the computer analysis is programmed incorrectly.

(Kotler & Kotler 1998, p. 166)

Kenneth Hudson's credentials, as listed on the first page of the article, help considerably to allow him to do as he wishes. But you should certainly take heart from his commonsense approach. Do not let research 'rules' get in your way if you feel that a certain aspect of your business or organisation ought to be examined in a certain way, and that way does not easily lend itself to statistical proofs or formal survey techniques.

In the same way, there is room for 'gut feeling' in decision-making. In fact, there *must* be room for it. Studies like Hudson's (1985) should be carried out. When presenting the results of such work, however, be aware that your own conviction and persuasive powers will have to take the place of research 'proof'. Your Board, or the people to whom you are presenting your research, may be cautious about trusting their own 'gut feeling' about it. (If you can refer to research principles or prior research it can make things a bit easier for everyone to justify their 'gut feelings'.)

Perhaps you might even use an article like Hudson's in such a situation. You could say something like: 'Hudson's research in the British Isles was, as he himself admitted, not properly scientific, but what he discovered was certainly of use when … '. Be creative. If you feel you are right, go for it.

(Obviously, in university studies such as your research study, you are expected to use valid research methodology to underpin all your conclusions. In your research study your understanding of the basic principles of research methodology must be demonstrated. There is no place for gut feelings.)

Review

Summary

In this topic we have looked at the theory and methodology of sampling. Clearly, it is crucial to understand how methodological issues such as the selection of a sample can affect the data-gathering process and the findings of your research. A good understanding of sampling theory will also help in the assessment of reports and surveys relevant to your profession.

References

Hudson, K. 1985, 'Museums and their customers', *Museums are for People*, Scottish Museums Council, Edinburgh.

Kotler, N. & Kotler, P. 1998, *Museum Strategy and Marketing: Designing Missions, Building Audiences, Generating Revenue and Resources*, Jossey-Bass, San Francisco.

Kress, G. 1988, *Marketing Research*, 3rd edn, Prentice Hall, Englewood Cliffs, NJ.