MAKING SENSE OF DATA

Essentials series

SURVEYS AND SAMPLING

Copyright © 2012 by City of Bradford MDC

Survey research involves utilising interviews or questionnaires to obtain quantitative information in fields such as marketing, politics, and social science. Utilizing surveys is an efficient, though sometimes costly way of collecting data from a large number of respondents, accurately representing a whole population. Surveys also have the benefit of providing data that is relatively free from errors.

In this topic key issues are considered, used to either evaluate or plan a survey.

The target population

The *target population* is the entire group of individuals that you are interested in studying. For example, suppose you want to know what the young people in England think of reality TV. The target population here is all "young people" resident in England. How should we define "young people"? Should we include those aged 16, 17 or only those aged 18 to 25?

If the target population isn't well defined, the survey results are likely to be biased. The sample that's actually studied may contain people outside the intended population, or the survey may exclude people who should have been included.

Sample matches target population

When conducting a survey, you typically can't ask every single member of the target population to provide the information you're looking for. The best you can do is to select a good *sample* (a subset of individuals from the population) and get the information from them. A good sample *represents* the target population. The sample doesn't systematically favour certain groups within the target population, and it doesn't systematically exclude certain people, either.

The best scenario for selecting a representative sample is to obtain a *sampling frame* — a list of all the members of the target population — and draw randomly from that. If a list isn't possible, you need some mechanism that gives everyone in the population an equal opportunity to be chosen to participate in the survey. For example, if a house-to-house survey of a city is needed, an up-to-date map with all households could act as the sampling frame.

Randomly selecting the sample

An important feature of a good survey is that the sample is randomly selected from the target population. “Randomly” means every member of the target population has an equal chance of being included in the sample. In other words, the process used for selecting a sample can’t be biased.

Simple random

In a simple random sample of a given size, all subsets of the frame are given an equal probability. Each element of the frame has an equal chance of selection: the frame is not subdivided or partitioned. This minimises bias and simplifies analysis of results. In particular, the variance between individual results within the sample is a good indicator of variance in the overall population. However, this method can be vulnerable to sampling error because the randomness of the selection may result in a sample that doesn't reflect the makeup of the population.

Systematic

Systematic sampling relies on arranging the target population according to some ordering scheme and then selecting elements at regular intervals through that ordered list. Systematic sampling involves a random start and then proceeds with the selection of every kth element from then onwards. In this case, k=population size/sample size. It is important that the starting point is not simply the first in the list, but is instead randomly chosen from the first to the kth element in the list. For example, starting at the 3rd entry then selecting every 10th name from a telephone directory.

Stratified

Where the population embraces a number of distinct categories, the frame can be organized by these categories into separate "strata." Each *stratum* is then sampled as an independent sub-population, out of which individual elements can be randomly selected. There are several potential benefits to stratified sampling:

- dividing the population into distinct, independent strata can enable researchers to draw inferences about specific subgroups;
- utilizing a stratified sampling method can lead to more efficient statistical estimates;
- sometimes data are more readily available for individual, pre-existing strata within a population;
- as each stratum is treated as an independent population, different sampling approaches can be applied to different strata.

Cluster

Sometimes it is more cost-effective to select respondents in groups ('clusters'). Sampling is often clustered by geography, for instance, if surveying households within a city, we might choose to select particular neighbourhoods and then interview every household. Clustering can reduce travel and administrative costs. This method generally increases the variability of sample estimates above that of simple random sampling, depending on how the clusters differ between themselves, as compared with the within-cluster variation. For this reason, cluster sampling requires a larger sample size.

Quota

In quota sampling, the population is first segmented into mutually exclusive sub-groups, as in stratified sampling. Then judgement is used to select the subjects or units from each segment based on a specified proportion. For example, an interviewer may be told to sample 200 females and 300 males between the age of 45 and 60. In quota sampling the selection of the sample is non-random, for example interviewers might be tempted to interview those who look most helpful.

Sample size

If you have a large sample size, and the sample is representative of the target population (meaning randomly selected), you can count on that information to be accurate. Exactly how accurate depends on the sample size, but in general a bigger sample leads to more accurate information.

As a rough estimate, we can calculate how accurate a survey is by taking the reciprocal of the square root of the sample size. For example, a survey of 1,000 (randomly selected) people is accurate to within $\pm 1/\sqrt{1000}$, which is 0.032 or 3.2%. This percentage is called the *margin of error*.

Beware of surveys that have a large sample size but are not randomly selected, for example internet surveys. A company might say that 50,000 people logged on to its Web site to answer a survey, but that information is biased, because it represents opinions of those who had access to the Internet, went to the Web site, and chose to complete the survey.

Non-response issues

After the sample size has been chosen and the sample of individuals has been randomly selected from the target population, you have to get the information you need from the people in the sample. If you've ever thrown away a survey or refused to answer a few questions over the phone, you know that getting people to participate in a survey isn't easy.

If a researcher wants to minimize bias, the best way to handle non-response is to follow up one, two, or even three times, offering incentives for participating such as stamped return envelopes, chances to win prizes, and so on. The response of a survey is found by taking the number of respondents divided by the total sample size. The ideal response rate is anything over 70% however, most fall well short of this.

Types of survey

Surveys come in many types: mail surveys, telephone surveys, Internet surveys, house-to-house interviews, and man-on-the-street surveys (in which someone asks, "Do you have a few minutes to participate in a survey?"). One very important yet sometimes overlooked criterion of a good survey is whether the type of survey being used is appropriate for the situation. For example, if the target population is the population of people who are visually impaired, sending them a survey in the mail that has a tiny font isn't a good idea!

When looking at the results of a survey, be sure to find out what type of survey was used and reflect on whether this was appropriate.

Wording questions

The way in which a question is worded in a survey can affect the results. One big problem is the use of misleading questions (in other words, questions that are worded in such a way that you know how the researcher wants you to answer).

An example of a misleading question is, "Do you agree that the Council should have more local powers to eliminate street waste?" This question could be worded in an indirect way, such as "Does the Council do enough about street waste?" Then give a scale from 1 to 5 where 1 = strongly disagree and 5 = strongly agree.

Appropriate timing

The timing of a survey is everything. Current events shape people's opinions, and while some pollsters try to determine how people really feel, others take advantage of these situations. For example, polls regarding the death penalty often come out right after a series of murders are reported by the national media. Timing of any survey, regardless of the subject matter, can still cause bias. Check the date when a survey was conducted and see whether you can determine any relevant events that may have temporarily influenced the results.

Administration and personnel

The people who actually carry out surveys have tough jobs. They have to deal with hang ups, take-us-off-your-list responses, and answering machines. After they do get a live respondent at the other end of the phone or face to face, the job becomes even harder. For example, if the respondent doesn't understand the question and needs more information, how much can you say, while still remaining neutral?

For a survey to be successful, the survey personnel must be trained to collect data in an accurate and unbiased way. The key is to be clear and consistent about every possible scenario that may come up, discuss how they should be handled, and have this discussion well before participants are ever contacted.

You can also avoid problems by running a pilot study (a practice run with only a few respondents) to make sure the survey is clear and consistent and that the personnel are handling responses appropriately. Any problems identified can be fixed before the real survey starts.

Conclusions made

Even if a survey is done correctly, researchers can misinterpret or over-interpret results so that they say more than they really should. Here are some of the most common errors made in drawing conclusions from surveys:

- ✓ Making projections to a larger population than the study actually represents
- ✓ Claiming a difference exists between two groups when a difference isn't really there
- ✓ Saying that "these results aren't scientific, but . . ." and then presenting the results as if they are scientific