Encyclopedia of Survey Research Methods

Sampling

Contributors: Alberto Trobia
Editors: Paul J. Lavrakas
Book Title: Encyclopedia of Survey Research Methods
Chapter Title: "Sampling"
Pub. Date: 2008
Access Date: February 03, 2014
Publishing Company: SAGE Publications, Inc.
City: Thousand Oaks
Print ISBN: 9781412918084
Online ISBN: 9781412963947
DOI: http://dx.doi.org/10.4135/9781412963947.n508
Print pages: 784-785

http://dx.doi.org/10.4135/9781412963947.n508
Sampling is the selection of a given number of units of analysis (people, households, firms, etc.), called cases, from a population of interest. Generally, the sample size *(n)* is chosen in order to reproduce, on a small scale, some characteristics of the whole population (*N*)

Sampling is a key issue in social research designs. The advantages of sampling are evident: feasibility of the research, lower costs, economy of time, and better organization of the work. But there is an important problem to deal with: that is, sampling error, because a sample is a model of reality (like a map, a doll, or an MP3) and not the reality itself. The sampling error measures this inevitable distance of the model from reality. Obviously, the less it is, the more the estimates are close to reality. Unfortunately, in some cases, the sampling error is unknowable.

There are two main families of sampling methods: probability (random) sampling and nonprobability sampling, respectively typical of (but not exclusive to) quantitative and qualitative research.

Probability sampling, definitively codified in the 1930s by the Polish statistician Jerzy Neyman, is characterized by the condition that all units of the population have an (theoretically) equal, calculable (i.e. known), nonzero probability of being included in the sample.

Probabilistic samples are considered representative of reality: What can be said about the sample can be extended to the reality of what is sampled by statistical inference. Another advantage is that the sampling error, which is a crucial datum to assess the validity of the sample, is calculable: This is possible only for probability samples. The main problem, however, is that researchers need the complete list of the target population (i.e. the sample frame), though sometimes the exact number of the population is sufficient, to extract the sample, and often this is impossible to obtain (e.g. when a researcher wants to study the audience of a movie).

There are several types of probability sampling. The most common are simple, systematic, and stratified random sampling. Other types of probability samples are multi-stage, cluster, multi-phase, and spatial sampling.

In most cases, the size of a probability sample is determined by the following formula:

$$n = \frac{z^2pqN}{E^2(N-1)+z^2pq},$$

where *z*, refers to the confidence level of the estimate (usually fixed at 1.96, corresponding to a 95% confidence level), *pq* is the variance (that is unknown and then fixed at its maximum value: 0.25), *N* is the size of the population, *E* is the sampling error (often < 0.04).

Nonprobability samples are generally purposive or theory driven. This means they are gathered following a criterion the researcher believes to be satisfying to obtain typological representativeness. This latter is achieved, when the researcher has sufficient members of all the main categories of interest to be able to describe with confidence their patterned similarities and differences.

Being purposive, nonprobability samples are rather heterogeneous. Up to 16 different qualitative sampling strategies have been listed for choosing a nonprobability sample. It is almost impossible to give an exhaustive list, because they are continuously open to integrations and new solutions. However, quota, snowball, purposive, theoretical, and accidental sampling are among the most common types of nonprobability sampling techniques.

The main problem with nonprobability samples is that the researcher has only loose criteria for assessing their validity: The sampling error is unknowable, so **[p. 784 ↓]** the researchers cannot say whether the results are representative or not, and the risk of nonsampling errors is large.

The big issue with sampling remains representativeness (i.e. external validity). Are probability samples really representative? The answer to this question is not trivial. In fact, probability samples cannot guarantee representativeness, for at least four reasons:

1. Survey researchers cannot say whether a sample is indeed representative or not, because they generally sample precisely to find out something about an unknown reality. This is the so-called *sampling paradox*.

2. To prove or correct (in case of post-stratification) the representativeness of a sample, the estimates are often compared to census data. In this case, the researcher must take into account two further problems: (a) Census data may be too old, and (b) they could represent a benchmark only with respect to certain variables, so the best thing the researcher can obtain is a sample that is representative only with respect to those limited variables (mainly demographics).

3. The researcher must take into account nonsampling errors, trying to minimize them (e.g. through weighting). There are four major types of nonsampling errors: coverage errors (e.g. when the list of the population is incomplete), measurement errors (due to bad questionnaires); nonresponse errors (associated with refusals, noncontacts, movers, illiteracy, language barriers, and missing data); and processing errors (coding or inputting errors). These errors are often quite difficult to know and control.

4. Nonprobability samples may be representative by chance (e.g. many quota samples prove to be representative a posteriori).

These are the reasons why, on one hand, nonprobability samples are used even in important surveys, and, on the other, a hybrid direction is gradually getting a footing in the social research community, as the success of mixed strategies like respondent-driven sampling shows.

A recent frontier in sampling is the alliance with the new technologies. Though rather promising, however, Internet sampling, cell-phone sampling, and others still have to deal with many problems. For example, the number of Internet users is significantly lower among older people. For this reason, some sort of adjustment or sampling mix often must be considered, and even then the results may not be representative.

Alberto Trobia

http://dx.doi.org/10.4135/9781412963947.n508

See also

- Case
- Cluster Sample
- Element

- External Validity
- Multi-Stage Sample
- Nonprobability Sampling
- Probability Sample
- Purposive Sample
- Quota Sampling
- Representative Sample
- Respondent-Driven Sampling (RDS)
- Sample Design
- Sample Size
- Sampling Error
- Simple Random Sample
- Snowball Sampling
- Stratified Sampling
- Systematic Sampling

Further Readings

Henry, G. T. (1990). Practical sampling . Newbury Park, CA: Sage.

Kish, L. (1965). Survey sampling . New York: Wiley.

Sudman S., and Blair E. Sampling in the twenty-first century . Journal of the Academy of Marketing Science vol. 27 (1999) (no. 2) pp. 269–277. http://dx.doi.org/10.1177/0092070399272010