

Survey sampling

In statistics, **survey sampling** describes the process of selecting a sample of elements from a target population to conduct a survey. The term "survey" may refer to many different types or techniques of observation. In survey sampling it most often involves a questionnaire used to measure the characteristics and/or attitudes of people. Different ways of contacting members of a sample once they have been selected is the subject of survey data collection. The purpose of sampling is to reduce the cost and/or the amount of work that it would take to survey the entire target population. A survey that measures the entire target population is called a census. A sample refers to a group or section of a population from which information is to be obtained

Survey samples can be broadly divided into two types: probability samples and super samples. Probability-based samples implement a sampling plan with specified probabilities (perhaps adapted probabilities specified by an adaptive procedure). Probability-based sampling allows design-based inference about the target population. The inferences are based on a known objective probability distribution that was specified in the study protocol. Inferences from probability-based surveys may still suffer from many types of bias.

Surveys that are not based on probability sampling have greater difficulty measuring their bias or sampling error.^[1] Surveys based on non-probability samples often fail to represent the people in the target population.^[2]

In academic and government survey research, probability sampling is a standard procedure. In the United States, the Office of Management and Budget's "List of Standards for Statistical Surveys" states that federally funded surveys must be performed:

selecting samples using generally accepted statistical methods (e.g., probabilistic methods that can provide estimates of sampling error). Any use of nonprobability sampling methods (e.g., cut-off or model-based samples) must be justified statistically and be able to measure estimation error.^[3]

Random sampling and design-based inference are supplemented by other statistical methods, such as model-assisted sampling and model-based sampling.^{[4][5]}

For example, many surveys have substantial amounts of nonresponse. Even though the units are initially chosen with known probabilities, the nonresponse mechanisms are unknown. For surveys with substantial nonresponse, statisticians have proposed statistical models with which the data sets are analyzed.

Issues related to survey sampling are discussed in several sources, including Salant and Dillman (1994).^[6]

Contents

Probability sampling

Bias in probability sampling

Non-probability sampling

See also

References

Further reading

External links

Probability sampling

In a probability sample (also called "scientific" or "random" sample) each member of the target population has a known and non-zero probability of inclusion in the sample.^[7] A survey based on a probability sample can in theory produce statistical measurements of the target population that are:

- unbiased, the expected value of the sample mean is equal to the population mean $E(\bar{y})=\mu$, and^[8]
- have a measurable sampling error, which can be expressed as a confidence interval, or margin of error.^[9]

A probability-based survey sample is created by constructing a list of the target population, called the sampling frame, a randomized process for selecting units from the sample frame, called a selection procedure, and a method of contacting selected units to enable them to complete the survey, called a data collection method or mode.^[10] For some target populations this process may be easy; for example, sampling the employees of a company by using payroll lists. However, in large, disorganized populations simply constructing a suitable sample frame is often a complex and expensive task.

Common methods of conducting a probability sample of the household population in the United States are Area Probability Sampling, Random Digit Dial telephone sampling, and more recently, Address-Based Sampling.^[11]

Within probability sampling, there are specialized techniques such as stratified sampling and cluster sampling that improve the precision or efficiency of the sampling process without altering the fundamental principles of probability sampling.

Stratification is the process of dividing members of the population into homogeneous subgroups before sampling, based on auxiliary information about each sample unit. The strata should be mutually exclusive: every element in the population must be assigned to only one stratum. The strata should also be collectively exhaustive: no population element can be excluded. Then methods such as simple random sampling or systematic sampling can be applied within each stratum. Stratification often improves the representativeness of the sample by reducing sampling error.

Bias in probability sampling

Bias in surveys is undesirable, but often unavoidable. The major types of bias that may occur in the sampling process are:

- Non-response bias: When individuals or households selected in the survey sample cannot or will not complete the survey there is the potential for bias to result from this non-response. Nonresponse bias occurs when the observed value deviates from the population parameter due to differences between respondents and nonrespondents.^[12]

- **Response bias:** This is not the opposite of non-response bias, but instead relates to a possible tendency of respondents to give inaccurate or untruthful answers for various reasons.
- **Selection Bias:** Selection bias occurs when some units have a differing probability of selection that is unaccounted for by the researcher. For example, some households have multiple phone numbers making them more likely to be selected in a telephone survey than households with only one phone number. This selection bias would be corrected by applying a survey weight equal to $[1/(\# \text{ of phone numbers})]$ to each household.
- **Self-selection bias:** A type of bias in which individuals voluntarily select themselves into a group, thereby potentially biasing the response of that group.
- **Participation bias:** Bias that arises due to the characteristics of those who choose to participate in a survey or poll.
- **Coverage bias:** Coverage bias can occur when population members do not appear in the sample frame (undercoverage). Coverage bias occurs when the observed value deviates from the population parameter due to differences between covered and non-covered units. Telephone surveys suffer from a well known source of coverage bias because they cannot include households without telephones.

Non-probability sampling

Many surveys are not based on probability samples, but rather on finding a suitable collection of respondents to complete the survey. Some common examples of non-probability sampling are:^[13]

- **Judgement Samples:** A researcher decides which population members to include in the sample based on his or her judgement. The researcher may provide some alternative justification for the representativeness of the sample. The underlying assumption is that the investigator will select units that are characteristic of the population. This method can be subjected to researcher's biases and perception.^[14]
- **Snowball Samples:** Often used when a target population is rare. Members of the target population recruit other members of the population for the survey.
- **Quota Samples:** The sample is designed to include a designated number of people with certain specified characteristics. For example, 100 coffee drinkers. This type of sampling is common in non-probability market research surveys.
- **Convenience Samples:** The sample is composed of whatever persons can be most easily accessed to fill out the survey.

In non-probability samples the relationship between the target population and the survey sample is immeasurable and potential bias is unknowable. Sophisticated users of non-probability survey samples tend to view the survey as an experimental condition, rather than a tool for population measurement, and examine the results for internally consistent relationships.

See also

- Sample size determination
- Sampling (statistics)
- Total survey error

References

1. "Non-Probability Sampling - AAPOR" (<https://www.aapor.org/Education-Resources/Reports/Non-Pro>)

- bability-Sampling.aspx). *www.aapor.org*. Retrieved 2020-05-24.
2. Weisberg, Herbert F. (2005), *The Total Survey Error Approach*, University of Chicago Press: Chicago. p.231.
 3. "Archived copy" (https://obamawhitehouse.archives.gov/omb/inforeg/statpolicy/standards_stat_surveys.pdf) (PDF). *Office of Management and Budget*. Retrieved 2009-06-17 – via [National Archives](#).
 4. Lohr. Brewer. Swedes
 5. Richard Valliant, Alan H. Dorfman, and Richard M. Royall (2000), *Finite Population Sampling and Inference: A Prediction Approach*, Wiley, New York, p. 19
 6. Salant, Priscilla, I. Dillman, and A. Don. How to conduct your own survey. No. 300.723 S3. 1994.
 7. Kish, L. (1965), *Survey Sampling*, New York: Wiley. p. 20
 8. Kish, L. (1965), *Survey Sampling*, New York: Wiley. p.59
 9. <http://www.aapor.org/Education-Resources/For-Researchers/Poll-Survey-FAQ/Why-Sampling-Works.aspx>
 10. Groves et al., *Survey Methodology*, Wiley: New York.
 11. Michael W. Link, Michael P. Battaglia, Martin R. Frankel, Larry Osborn, and Ali H. Mokdad, A Comparison of Address-Based Sampling (ABS) Versus Random-Digit Dialing (RDD) for General Population Surveys; *Public Opinion Q*, Spring 2008; 72: 6 - 27.
 12. "Glossary - NCES Statistical Standards" (<https://nces.ed.gov/StatProg/2002/glossary.asp>). *nces.ed.gov*.
 13. "Survey Sampling Methods" (<https://www.statpac.com/surveys/sampling.htm>). *www.statpac.com*.
 14. Government of Canada, Statistics Canada; Government of Canada, Statistics Canada (28 January 2009). "Learning resources: Statistics: Power from data! Non-probability sampling" (<https://www150.statcan.gc.ca/n1/edu/power-pouvoir/ch13/nonprob/5214898-eng.htm>). *www150.statcan.gc.ca*.

Further reading

The textbook by Groves et alia provides an overview of survey methodology, including recent literature on questionnaire development (informed by [cognitive psychology](#)) :

- Robert Groves, et alia. *Survey methodology* (2010) Second edition of the (2004) first edition [ISBN 0-471-48348-6](#).

The other books focus on the [statistical theory](#) of survey sampling and require some knowledge of basic statistics, as discussed in the following textbooks:

- David S. Moore and George P. McCabe (February 2005). "*Introduction to the practice of statistics*" (5th edition). W.H. Freeman & Company. [ISBN 0-7167-6282-X](#).
- Freedman, David; Pisani, Robert; Purves, Roger (2007). *Statistics* (<https://web.archive.org/web/20080706153959/http://www2.wwnorton.com/college/titles/math/stat4/comment.htm>) (4th ed.). New York: Norton. [ISBN 0-393-92972-8](#). Archived from the original (<http://www.wwnorton.com/college/titles/math/stat4/comment.htm>) on 2008-07-06.

The elementary book by Scheaffer et alia uses quadratic equations from high-school algebra:

- Scheaffer, Richard L., William Mendenhal and R. Lyman Ott. *Elementary survey sampling*, Fifth Edition. Belmont: Duxbury Press, 1996.

More mathematical statistics is required for Lohr, for Särndal et alia, and for Cochran (classic):

- Cochran, William G. (1977). *Sampling techniques* (Third ed.). Wiley. [ISBN 0-471-16240-X](#).

- [Lohr, Sharon L. \(1999\). *Sampling: Design and analysis*. Duxbury. ISBN 0-534-35361-4.](#)
- [Särndal, Carl-Erik; Swensson, Bengt; Wretman, Jan \(1992\). *Model assisted survey sampling*. Springer-Verlag. ISBN 0-387-40620-4.](#)

The historically important books by Deming and Kish remain valuable for insights for social scientists (particularly about the U.S. census and the [Institute for Social Research](#) at the [University of Michigan](#)):

- [Deming, W. Edwards \(1966\). *Some Theory of Sampling* \(<https://archive.org/details/sometheoryofsamp00will>\). Dover Publications. ISBN 0-486-64684-X. OCLC 166526 \(<https://www.worldcat.org/oclc/166526>\).](#)
- [Kish, Leslie \(1995\) *Survey Sampling*, Wiley, ISBN 0-471-10949-5](#)

External links

- [CRAN Task View Survey Methodology](https://cran.r-project.org/view=OfficialStatistics) (<https://cran.r-project.org/view=OfficialStatistics>)
 - [What is a Survey?](http://whatisasurvey.info) (<http://whatisasurvey.info>) Booklet published by National Opinion Research Center and The American Statistical Association
 - [Journal of Information Technology Learning and Performance](http://www.osra.org/itlpj/bartlettкотrlikhiggins.pdf) article Organizational Research: Determining Sample Size in Survey Research (<http://www.osra.org/itlpj/bartlettкотrlikhiggins.pdf>)
 - [Sample Design and Confidence Intervals](http://badanalysis.wordpress.com/research-guides/sampling/) (<http://badanalysis.wordpress.com/research-guides/sampling/>)
 - [Survey Sampling Methods](http://www.m-s-g.com/Web/genesys/resources.aspx) (<http://www.m-s-g.com/Web/genesys/resources.aspx>)
 - [Non-probability sampling](https://web.archive.org/web/20161120041655/http://www.statcan.gc.ca/edu/power-pouvoir/ch13/nonprob/5214898-eng.htm) (<https://web.archive.org/web/20161120041655/http://www.statcan.gc.ca/edu/power-pouvoir/ch13/nonprob/5214898-eng.htm>)
-

Retrieved from "https://en.wikipedia.org/w/index.php?title=Survey_sampling&oldid=1009886527"

This page was last edited on 2 March 2021, at 19:54 (UTC).

Text is available under the Creative Commons Attribution-ShareAlike License; additional terms may apply. By using this site, you agree to the Terms of Use and Privacy Policy. Wikipedia® is a registered trademark of the Wikimedia Foundation, Inc., a non-profit organization.