Sampling Bias

In: Encyclopedia of Survey Research Methods

By: Allan L. McCutcheon
Edited by: Paul J. Lavrakas
Book Title: Encyclopedia of Survey Research Methods
Chapter Title: "Sampling Bias"
Pub. Date: 2011
Access Date: November 30, 2020
Publishing Company: Sage Publications, Inc.
City: Thousand Oaks
Print ISBN: 9781412918084
Online ISBN: 9781412963947
DOI: https://dx.doi.org/10.4135/9781412963947
Print page: 785

Sampling bias occurs when a sample statistic does not accurately reflect the true value of the parameter in the target population, for example, when the average age for the sample observations does not accurately reflect the true average of the members of the target population. Typically, sampling bias focuses on one of two types of statistics: averages and ratios. The sources of sampling bias for these two types of statistics derive from different sources; consequently, these will be treated separately in this entry.

Sampling Bias For Averages

For survey researchers, sampling biases for averages derive from three sources: (1) imperfect sampling frames, (2) nonresponse bias, and (3) measurement error. Mathematical statisticians may also consider biases due to sources such as using the sample size (n) instead of n - 1, or using a sample statistic (e.g. s2) to estimate a population parameter (e.g. σ2), but these tend to be of academic interest and of less interest to practical research concerns; therefore, these will not be considered in this entry.

Imperfect sampling frames occur frequently in research and can be classified into four general categories: (1) frames where elements are missing, (2) frames where elements cluster, (3) frames that include foreign elements, and (4) frames with duplicate listings of elements. For example, household telephone surveys using random-digit dialing samples for landline telephones exclude households that are cell phone only (missing elements), include telephone numbers in some dwelling units that include more than a single household (element cluster), include some telephone numbers that are dedicated solely to fax machines (foreign elements), and include some households with more than a single landline telephone number (duplicate listings). All of these can cause sampling bias if they are not taken into account in the analysis stage.

Sampling bias due to nonresponse results from missing elements that should have been included in the sample but were not; these can be classified as noncontact (missing), unable to answer, or refusal. Noncontacts are those elements that are selected into the sample but cannot be located or for which no contact can be made; whereas the unable-to-answer either do not have the necessary information or the required health or skills to provide the answer, refusals are elements that, once located, decline to participate. Each will contribute to sampling bias if the sampled nonre-sponding elements differ from the sampled elements from which data are gathered. This sampling bias for averages can be characterized as follows:

$$\bar{Y}_{responders} - \bar{Y}_{population} = \frac{n_{nonresponders}}{n_{sample}} \left(\bar{Y}_{responders} - \bar{Y}_{nonresponders} \right).$$

As can be seen, when the number of nonresponse represents a small proportion of the total sample, or when there is a small difference between those who respond and those who do not, the resulting sampling bias will be small or modest. If it is possible to place bounds on the averages (such as with probabilities and proportions), researchers can quantify the possible range of sampling bias.

The third source of sampling bias for averages results from measurement error—that is, when what is measured among the sample elements differs from what researchers actually wished to measure for the target population. Measurement error, however, can be divided into random error and consistent bias; only consistent bias results in sampling bias, as random error will appear as sampling variance. These components of measurement error can be represented as

$$y_{ij} = \mu_i + \beta_i + \varepsilon_{ij},$$

where y ij represents the observed values of some variable y on j repeated observations of individual i, μirepresents the true value of what the researchers wish to measure for individual i, βi represents the consistent bias in individual i's response, and ϵij represents the random error associated with observation j for individual i. In sample surveys, consistent measurement bias can occur for a number of reasons, such as the questions focus on issues that are subject to a social desirability bias (e.g. illegal drug use).

Sampling Bias For Ratios

Unlike the estimation of sample averages, ratio estimators computed from samples are biased. In the 1950s, Herman Hartley and A. Ross showed that the absolute amount of bias in a ratio estimator is small relative to its standard deviation if the coefficient of variation (i.e. the ratio of the standard deviation to the average) of the ratio's denominator variable is small. In general terms, the amount of bias in ratio estimates will be small if the sample size is large, the sampling fraction is large, the mean of the denominator variable is large, the variance of the denominator variable is small, or if the correlation between numerator and denominator variables in the ratio is close to positive 1.0.

Allan L. McCutcheon

http://dx.doi.org/10.4135/9781412963947.n509
See also

- Measurement Error
- Nonresponse Bias
- Parameter
- Random Error
- Sampling Frame
- Social Desirability
- Statistic
- Target Population
- True Value

Further Readings

Cochran, W. G.(1977). Sampling techniques (3rd ed.). New York: Wiley.

HartleyH. O.RossA. Unbiased ratio estimators. *Nature*174(1954)270–271.

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