Encyclopedia of Survey Research Methods

Sampling Error

Contributors: Karol Krotki
Editors: Paul J. Lavrakas
Book Title: Encyclopedia of Survey Research Methods
Chapter Title: "Sampling Error"
Pub. Date: 2008
Access Date: October 15, 2013
Publishing Company: Sage Publications, Inc.
City: Thousand Oaks
Print ISBN: 9781412918084
Online ISBN: 9781412963947
DOI: http://dx.doi.org/10.4135/9781412963947.n510
Print pages: 786-790

http://dx.doi.org/10.4135/9781412963947.n510
Sampling error consists of two components: *sampling variance* and *sampling bias*. Sometimes overall sampling error is referred to as *sampling mean squared* **[p. 786 ↓]** *error* (MSE), which can be decomposed as in the following formula:

$$\begin{aligned} MSE(p) &= E(p-P) \\ &= E[(p-p')+(p'-P)] \qquad (1) \\ &= \mathrm{Var}(y)+\mathrm{Bias}^2, \end{aligned}$$

where *P* is the true population value, *p* is the measured sample estimate, and *p′* is the hypothetical mean value of realizations of *p* averaged across all possible replications of the sampling process producing *p*.

Sampling variance is the part that can be controlled by sample design factors such as sample size, clustering strategies, stratification, and estimation procedures. It is the error that reflects the extent to which repeated replications of the sampling process result in different estimates. Sampling variance is the random component of sampling error since it results from "luck of the draw" and the specifie population elements that are included in each sample. The presence of sampling bias, on the other hand, indicates that there is a systematic error that is present no matter how many times the sample is drawn.

Using an analogy with archery, when all the arrows are clustered tightly around the bull's-eye we say we have low variance and low bias. At the other extreme, if the arrows are widely scattered over the target and the midpoint of the arrows is off-center, we say we have high variance and high bias. In-between situations occur when the arrows are tightly clustered but far off-target, which is a situation of low variance and high bias. Finally, if the arrows are on-target but widely scattered, we have high variance coupled with low bias.

Efficient samples that result in estimates that are close to each other and to the corresponding population value are said to have low sampling variance, low sampling bias, and low overall sampling error. At the other extreme, samples that yield estimates that fluctuate widely and vary significantly from the corresponding population values are said to have high sampling variance, high sampling bias, and high overall sampling error. By the same token, samples can have average level sampling error by achieving high levels of sampling variance combined with low levels of sampling bias, or vice versa. (In this discussion it is assumed, for the sake of explanation, that the samples are drawn repeatedly and measurements are made for each drawn sample. In practice, of course, this is not feasible, but the repeated measurement scenario serves as a heuristic tool to help explain the concept of sampling variance.)

Sampling Variance

Sampling variance can be measured, and there exist extensive theory and software that allow for its calculation. All random samples are subject to sampling variance that is due to the fact that not all elements in the population are included in the sample and each random sample will consist of a different combination of population elements and thus will produce different estimates. The extent to which these estimates differ across all possible estimates is known as sampling variance. Inefficient designs that employ no or weak stratification will result in samples and estimates that fluctuate widely. On the other hand, if the design incorporates effective stratification strategies and minimal clustering, it is possible to have samples whose estimates are very similar, thereby generating low variance between estimates, thus achieving high levels of sampling precision.

The main design feature that influences sampling variance is sample size. This can be seen readily from the following formula for the sampling variance to estimate a proportion based on a simple random sample design:

$$var(p) = pq/n, \qquad (2)$$

where ρ is the sample estimate of the population proportion, *q - l—p*, and *n* is the sample size. (Formula 2 and subsequent formulae are relevant for proportions. Similar formulae are available for other statistics such as means, but they are more complicated.)

It can be easily seen from this formula that as *n* increases, the variance decreases in direct and inverse proportion. Because sampling variance is usually measured in terms of the confidence interval and standard error (which is the square root of the sampling variance), we usually refer to the impact of an increase in sample size in terms of the square root of that increase. Thus, to double the sampling precision, that is, reduce the sampling variance by 50%, we would have to increase the sample size by a factor of 4.

Sampling variance, or its inverse, sampling precision, is usually reported in terms of the standard error, confidence interval, or more popularly, the margin of error. Under a simple random sample design, the **[p. 787 ↓]** mathematical formula for the standard error (3) and the 95% confidence interval for a proportion *p* (4) are

$$\mathrm{se}(p) = [\mathrm{var}(p)]^2 \qquad (3)$$

$$ci(p) = [p - 1.96\mathrm{se}(p), p + 1.96\mathrm{se}(p)]. \qquad (4)$$

The margin of sampling error is equal to half the length of the confidence interval as defined in Formula 4. For example, a proportion of 50% from a sample size of 1,000 would have a margin of error of "plus or minus 3%," meaning that if we were to draw 100 simple random samples of approximate size of 1,000, for about 95 of the samples, the sample value would differ by no more than 3 percentage points in either direction from the true population value.

In general, the main drivers of sampling variance are stratification and clustering. Stratification usually results in a lower sampling variance because the number of possible samples is reduced in comparison with an unrestricted simple random sample. Not only is the number of possible samples reduced, but potential outliers are eliminated. For example, suppose we wanted to sample households in the United States. An unrestricted random sample might contain households that are all located in the Northeast—the probability is not high, but it is not zero. However, if we stratify by region, then we reduce the probability of such a skewed sample to zero. To the extent that the variable of interest is related to our stratification variables, in this case geography, stratification will reduce the overall sampling variance. In setting up the design, therefore, it is important to strive to define strata that are relatively homogeneous with respect to the variables of interest.

Clustering, on the other hand, works in a very different way. Clustering plays a role in sample designs that are used for surveys in which the data are collected in person, for example, via household visits. Clustering is used to control field costs, especially those related to travel, which often represent a significant portion of the overall survey budget. However, this results in fewer degrees of freedom in the sense that the sample now focuses on a smaller number of sampling units, that is, the first-stage clusters, often referred to as primary sampling units. For example, selecting an unclustered sample of 1,000 households throughout the United States would mean that the households could be located anywhere in the country and, of course, this would result in large travel costs. Restricting the sample first to 100 clusters (e.g. counties), and then taking 10 households within each cluster, reduces the travel costs but reduces our ability to spread the sample effectively over the entire country. This reduction in efficiency is further exacerbated by the fact that within each cluster, usually a geographically contiguous area, households tend to be more alike than households across these units. This phenomenon, called *intraclass homogeneity*, tends to drive up the sampling variance because efficiency is lost and the original sample of 1,000 might, in effect, have only the impact of 100 if, in the extreme case, the clusters are perfectly homogeneous.

Thus, in summary, with respect to sample design optimization, stratification is beneficial in that it reduces sampling variance, whereas clustering is to be avoided when possible or at least minimized as its effect is to increase sampling variance. Usually the effect of clustering is more marked than that of stratification. In many situations, though, clustering is necessary for cost reasons; thus, the best clustered design strategy involves finding a compromise between the cost savings and the penalty to be paid in terms of lower precision.

Another important factor that influences sampling variance is weighting. Weighting refers to adjustment factors that account for design deviations such as unequal probabilities of selection, variable nonre-sponse rates, and the unavoidable introduction of bias at various steps in the survey process that are corrected for through a process called *post-stratification*. The effect of weighting is to increase the sampling variance, and the extent of this increase is proportional to the variance among the weights.

A useful concept that quantifies and summarizes the impact of stratification, clustering, and weighting on sampling variance is the *design effect*, usually abbreviated as *deff*. It is the ratio of the true sampling variance taking into account all the complexities of the design to the variance that would have been achieved if the sample had been drawn using a simple random sample, incorporating no stratification, clustering, or weighting. A value of 1.00 indicates that the complexity of the design had no measurable impact on the sampling variance. Values less than 1.00 are rare; values larger than 5.00 are generally considered to be high.

The design effect is closely related to *rho (#)*, the intraclass correlation, mentioned previously. The following formula shows the relationship between the two: **[p. 788 ↓]**

$$deff = 1 + (b - 1)\rho, \qquad (5)$$

where *deff* is the design effect, ρ is the intraclass correlation, and *b* is the average cluster size.

The correct calculation of sampling variance, incorporating all the complexities of the design, is not straightforward. However, there is extensive software currently available that uses either the empirical bootstrap replication approach or the more theoretically based Taylor Series expansion. These systems typically allow for many types of stratification, clustering, and weighting although the onus is always on the user or the data producer to ensure that relevant information, such as the stratum identifier, cluster identifier, and weight, are present in the data set.

Sampling Bias

This component of sampling error results from a systematic source that causes the sampling estimates, averaged over all realizations of the sample, to differ consistently from their true target population values. Whereas sampling variance can be controlled through design features such as sample size, stratification, and clustering, we need to turn to other methods to control and reduce bias as much as possible.

Sampling bias can only be measured if we have access to corresponding population values. Of course, the skeptic will point out that if such information were available, there would be little point in drawing a sample and implementing a survey. However, there are situations in which we can approximate sampling bias by comparing underlying information such as basic demographics for the sample with corresponding data from another, more reliable, source (e.g. census or large national survey) to identify areas in the data space for which the sample might be underrepresented or overrepresented.

One major source of sampling bias is *frame coverage*; that is, the frame from which the sample is drawn is defective in that it fails to include all elements in the population. This is a serious error because it cannot be detected, and in some cases its impact cannot even be measured. This issue is referred to as *under-coverage* because the frame is missing elements that it should contain. The opposite phenomenon, *overcoverage*, is less serious. Overcoverage occurs when the frame includes foreign elements, that is, elements that do not belong to the target population. However, these elements, if sampled, can be identified during the field operation and excluded from further processing. A third potential source of frame bias is *duplication*. If certain elements appear several times on the frame, their probabilities of selection are higher and thus they might be overrepresented in the sample. Furthermore, it is not always known how many times the elements occur on the frame, in which case it is impossible to ascertain the extent of the problem and thus the size of the bias.

Sampling bias can also occur as a result of flaws in the sample selection process, errors in the sample implementation, and programming missteps during the sample processing stage. An example of bias occurring during sample selection would be a systematic sample of every fifth unit when, in fact, there is a repeating pattern in the list and every fifth unit belongs to a special group. An example of how sampling bias can occur during the sample implementation process is the method interviewers use to visit households in the field. Field instructions might indicate "every 10th household," and the interviewer might instead elect to visit households that appear more likely to generate an interview. This could, and often does, lead to sampling bias. Finally, sampling or estimation bias can occur during the sample processing stage, for example, by incorrect calculation of the weighting adjustment factors, giving excessive importance to certain subpopulations.

One severe challenge faced by all survey practitioners is how to measure bias. Whereas the estimation of sampling variance emanates from statistical theory (see Formula 2, presented earlier), the only way to measure sampling bias is to compare the resulting empirical value with the true target population value. Of course, this is problematic because we seldom possess the population value and thus must use indirect methods to estimate bias. One approach uses data from other sources, such as the census or large national samples, as surrogates for the population being sampled. The problem with this strategy is that even the census is subject to error, in terms of both variance and bias.

It was pointed out previously that weighting tends to increase sampling variance and reduce precision. The reason weighting is implemented in survey research, in spite of its negative effect on variance, is that in many cases it can be used to reduce bias by bringing the sampling distributions more in line with known population distributions. For example, it is often possible to weight to basic census distributions **[p. 789 ↓]** by gender and age, even for minor geographical subdivisions such as tracts. To take a hypothetical example, suppose the sample distribution by gender turns out to be 40% male, 60% female, a not uncommon result in a typical random-digit dialing telephone survey. Furthermore, assume that the corresponding census numbers are close to 50:50. Weighting would assign relative adjustment factors of 50/40 to males and 50/60 to females, thus removing the possible bias due to an overrepresentation of females in the sample.

Challenges

Overall sampling error needs to be viewed in terms of a combination of sampling variance and sample bias. The ultimate goal is to minimize the mean squared error. Survey researchers know how to measure sampling variance, and they have a good handle on how it can be reduced. Sampling bias represents more of a challenge as it is often difficult to measure and even if it is measurable, bias reduction is often expensive and problematic to achieve.

It is illustrative to discuss surveys that are based on nonprobability judgment, quota, or convenience samples—that is, samples that are not based on probability-based design. One currently prominent example is the Internet-based panel, which consists of members who choose (self-select) to belong to these panels. That is, the panel members are not selected randomly and then invited to join the panel, but rather, the members themselves decide to join the panels, hence the term *opt-in* populations. This means that the underlying frame suffers from undercoverage and many potential types of bias, only some of which are known. These samples might be appropriate for certain studies (e.g. focus groups), in which generalizing with confidence to the population is not an absolute prerequisite. But, in general, these surveys fall short of required methodological rigor on two counts. In the first place, the probabilities of selection are usually unknown and often unknowable, thus precluding any chance of calculating sampling variance. Second, these surveys suffer from coverage and selection bias issues that, in many cases, are not even measurable.

With the advent of relevant software, surveys now regularly produce large-scale sampling variance results showing not only standard errors and confidence intervals but also design effects and measures of intraclass correlation. The results typically are presented for the entire sample and also for important subpopulations that are relevant for data users. These are useful not only to shed light on the quality of the data but also to inform future sample designs. The choice of estimates and subpopulations for which to publish sampling errors is not simple, and some researchers have developed "generalized variance functions" that allow users to estimate their own sampling errors based on the type of variables in question, the sample size, and level of clustering. However, these results are usually limited to sampling variance, and much less is calculated, produced, and disseminated with respect to sampling bias. This is due largely to the difficulty of calculating these measures and to the challenge of separating sampling bias from other sources of bias, such as nonresponse bias and response bias.

Karol Krotki

http://dx.doi.org/10.4135/9781412963947.n510
See also

Further Readings

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

Lohr, S. L. (1999). Sampling: Design and analysis . Pacific Grove, CA: Duxbury Press.