Bias in Survey Sampling

In survey sampling, **bias** refers to the tendency of a sample [statistic](http://stattrek.com/Help/Glossary.aspx?Target=Statistic) to systematically over- or under-estimate a population [parameter](http://stattrek.com/Help/Glossary.aspx?Target=Parameter).

Bias Due to Unrepresentative Samples

A good [sample](http://stattrek.com/Help/Glossary.aspx?Target=Sample) is **representative**. This means that each sample point represents the attributes of a known number of [population](http://stattrek.com/Help/Glossary.aspx?Target=Population) elements.

Bias often occurs when the survey sample does not accurately represent the population. The bias that results from an unrepresentative sample is called **selection bias**. Some common examples of selection bias are described below.

* **Undercoverage**. Undercoverage occurs when some members of the population are inadequately represented in the sample. A classic example of undercoverage is the *Literary Digest* voter survey, which predicted that Alfred Landon would beat Franklin Roosevelt in the 1936 presidential election. The survey sample suffered from undercoverage of low-income voters, who tended to be Democrats.

How did this happen? The survey relied on a [convenience sample](http://stattrek.com/Help/Glossary.aspx?Target=Convenience%20sample), drawn from telephone directories and car registration lists. In 1936, people who owned cars and telephones tended to be more affluent. Undercoverage is often a problem with convenience samples.
* **Nonresponse bias**. Sometimes, individuals chosen for the sample are unwilling or unable to participate in the survey. Nonresponse bias is the bias that results when respondents differ in meaningful ways from nonrespondents. The *Literary Digest* survey illustrates this problem. Respondents tended to be Landon supporters; and nonrespondents, Roosevelt supporters. Since only 25% of the sampled voters actually completed the mail-in survey, survey results overestimated voter support for Alfred Landon.

The *Literary Digest* experience illustrates a common problem with mail surveys. Response rate is often low, making mail surveys vulnerable to nonresponse bias.
* **Voluntary response bias**. Voluntary response bias occurs when sample members are self-selected volunteers, as in [voluntary samples](http://stattrek.com/Help/Glossary.aspx?Target=Voluntary%20sample). An example would be call-in radio shows that solicit audience participation in surveys on controversial topics (abortion, affirmative action, gun control, etc.). The resulting sample tends to overrepresent individuals who have strong opinions.

**Random sampling** is a procedure for sampling from a population in which (a) the selection of a sample unit is based on chance and (b) every element of the population has a known, non-zero probability of being selected. Random sampling helps produce representative samples by eliminating voluntary response bias and guarding against undercoverage bias. All probability sampling methods rely on random sampling.

Bias Due to Measurement Error

A poor measurement process can also lead to bias. In survey research, the measurement process includes the environment in which the survey is conducted, the way that questions are asked, and the state of the survey respondent.

**Response bias** refers to the bias that results from problems in the measurement process. Some examples of response bias are given below.

* **Leading questions**. The wording of the question may be loaded in some way to unduly favor one response over another. For example, a satisfaction survey may ask the respondent to indicate where she is satisfied, dissatisfied, or very dissatified. By giving the respondent one response option to express satisfaction and two response options to express dissatisfaction, this survey question is biased toward getting a dissatisfied response.
* **Social desirability**. Most people like to present themselves in a favorable light, so they will be reluctant to admit to unsavory attitudes or illegal activities in a survey, particularly if survey results are not confidential. Instead, their responses may be biased toward what they believe is socially desirable.

Sampling Error and Survey Bias

A survey produces a sample statistic, which is used to estimate a population parameter. If you repeated a survey many times, using different samples each time, you might get a different sample statistic with each replication. And each of the different sample statistics would be an estimate for the*same* population parameter.

If the statistic is unbiased, the average of all the statistics from all possible samples will equal the true population parameter; even though any individual statistic may differ from the population parameter. The variability among statistics from different samples is called **sampling error**.

Increasing the sample size tends to reduce the sampling error; that is, it makes the sample statistic less variable. However, increasing sample size does not affect survey bias. A large sample size cannot correct for the methodological problems (undercoverage, nonresponse bias, etc.) that produce survey bias. The *Literary Digest* example discussed above illustrates this point. The sample size was very large - over 2 million surveys were completed; but the large sample size could not overcome problems with the sample - undercoverage and nonresponse bias.