

16.4: Sampling Bias

Learning Objectives

- Recognize sampling bias
- Distinguish among self-selection bias, undercoverage bias, and survivorship bias

Descriptions of various types of sampling such as simple random sampling and stratified random sampling are covered in another section. This section discusses various types of sampling biases including self-selection bias and survivorship bias. Examples of other sampling biases that are not easily categorized will also be given.

It is important to keep in mind that sampling bias refers to the method of sampling, not the sample itself. There is no guarantee that random sampling will result in a sample representative of the population just as not every sample obtained using a biased sampling method will be greatly non-representative of the population.

Self-Selection Bias

Imagine that a university newspaper ran an ad asking for students to volunteer for a study in which intimate details of their sex lives would be discussed. Clearly the sample of students who would volunteer for such a study would not be representative of the students at the university. Similarly, an online survey about computer use is likely to attract people more interested in technology than is typical. In both of these examples, people who "self-select" themselves for the experiment are likely to differ in important ways from the population the experimenter wishes to draw conclusions about. Many of the admittedly "non-scientific" polls taken on television or web sites suffer greatly from self-selection bias.

A self-selection bias can result when the non-random component occurs after the potential subject has enlisted in the experiment. Considering again the hypothetical experiment in which subjects are to be asked intimate details of their sex lives, assume that the subjects did not know what the experiment was going to be about until they showed up. Many of the subjects would then likely leave the experiment resulting in a biased sample.

Undercoverage Bias

A common type of sampling bias is to sample too few observations from a segment of the population. A commonly-cited example of undercoverage is the poll taken by the Literary Digest in 1936 that indicated that Landon would win an election against Roosevelt by a large margin when, in fact, it was Roosevelt who won by a large margin. A common explanation is that poorer people were undercovered because they were less likely to have telephones and that this group was more likely to support Roosevelt.

A detailed analysis by Squire (1988) showed that it was not just an undercoverage bias that resulted in the faulty prediction of the election results. He concluded that, in addition to the undercoverage described above, there was a nonresponse bias (a form of self-selection bias) such that those favoring Landon were more likely to return their survey than were those favoring Roosevelt.

Survivorship Bias

Survivorship bias occurs when the observations recorded at the end of the investigation are a non-random set of those present at the beginning of the investigation. The gains in stock funds is an area in which survivorship bias often plays a role. The basic problem is that poorly-performing funds are often either eliminated or merged into other funds. Suppose one considers a sample of stock funds that exist in the present and then calculates the mean 10-year appreciation of those funds. Can these results be validly generalized to other stock funds of the same type? The problem is that the poorly-performing stock funds that are not still in existence (did not survive for 10 years) are not included and therefore there is a bias toward selecting better-performing funds. There is good evidence that this survivorship bias is substantial (Malkiel, 1995).

In World War II, the statistician Abraham Wald analyzed the distribution of hits from anti-aircraft fire on aircraft returning from missions. The idea was that this information would be useful for deciding where to place extra armor. A naive approach would be to put armor at locations that were frequently hit to reduce the damage there. However, this would ignore the survivorship bias occurring because only a subset of aircraft return. Wald's approach was the opposite: if there were few hits in a certain location on returning planes, then hits in that location were likely to bring a plane down. Therefore, he recommended that locations without hits on the returning planes should be given extra armor. A detailed and mathematical description of Wald's work can be found in Mangel and Samaniego (1984).

References

1. Malkiel, B. G. (1995) Returns from investing in equity mutual funds 1971 to 1991. *The Journal of Finance*, 50, 549-572.
2. Mangel, M. & Samaniego, F. J. (1984) Abraham Wald's work on aircraft survivability. *Journal of the American Statistical Association*, 79, 259-267.
3. Squire, P. (1988) Why the 1936 Literary Digest poll failed. *Public Opinion Quarterly*, 52, 125-133.

This page titled [16.4: Sampling Bias](#) is shared under a [Public Domain](#) license and was authored, remixed, and/or curated by [David Lane](#) via [source content](#) that was edited to the style and standards of the LibreTexts platform.

- **6.4: Sampling Bias** by [David Lane](#) is licensed [Public Domain](#). Original source: <https://onlinestatbook.com>.