Sampling Error

Taking **probability samples** has become common practice for market researchers and business professionals alike. While a random sample selection process is generally the best way to create a representative sample of a population, it does not guarantee a perfect sample. Even the best random samples will always be a little different from the true population of interest.

This difference is referred to as sampling error.

When Does Sampling Error Occur?

Sampling error occurs when researchers take a random sample instead of observing every individual subject that comprises a population.

While dealing with large populations this process becomes the only option, and so sampling error is extremely difficult to avoid. In some sense, sampling error can be considered an occupational hazard.

An Example of Sampling Error

Let's pretend that we are a group of researchers administering a survey with the goal of learning how much money a specific group of people spends while purchasing a vehicle.

To kickstart the study, we distribute our survey to 1,000 randomly selected United States residents.

By dumb luck, respondent #347 happens to be Mark Cuban — billionaire businessman and investor. While it's unlikely that someone with the status of Mark Cuban would complete our survey, it's still possible.

Let's also say that, again by chance, respondent #789 is Elon Musk — another billionaire. He also decides to fill out our survey.

Yes, it's unlikely that billionaires with jam-packed schedules would choose to fill out our survey, but that doesn't mean that it's out of the realm of possibility.

This scenario represents a more common occurrence that researchers encounter while distributing surveys.

While interested in something directly related to a person's income, such as how much individuals spend while purchasing a vehicle, by chance we put ourselves at risk of collecting data from significant outliers of the population.

In this case, billionaire businessmen Mark Cuban and Elon Musk do not accurately represent average members of the target population we are interested in, and therefore the accuracy of our results would be negatively affected.

The same goes for if we were to collect a significant amount of data from individuals that fall below the poverty line.

If too many of our respondents are either too wealthy or struggling financially, our sample will look different than the true nature of the real-world population.

This difference is the sampling error.

Sampling Error is Unavoidable

Although sampling error is unavoidable when collecting a random sample, we can take measures to estimate and reduce sampling error.

The margin of error that you commonly see with survey results is in fact an estimate of sampling error. Because it is just an estimate, there is a small chance (typically five percent or less) that the margin of error is actually larger than stated in a report.

Reducing Sampling Error by Increasing Sample Size

One way to reduce sampling error is to increase the size of your sample by selecting more subjects to observe.

Sampling error and sample size have an inversely correlated relationship, meaning that as sample size grows, sampling error decreases.

However, it's important to note that increasing sample size usually results in an increase in cost.

The more people that you want to survey in your study, the more expensive your study will be, as there are costs associated with identifying respondents or participants. There will also be an increased cost from a time usage perspective.

We've found that after bringing sample size to 1,000 participants, researchers generally start to get less bang for their buck. This is due to the relationship between sample size and margin of error. Once you have a sample size of 1,000, even if you more than double your sample size to 2,500 you are only decreasing your margin of error by one percent.

Reducing Sampling Error with Solid Sample Design

Sampling error can also be reduced by ensuring that you have a solid sample design.

For example, if your target population is made up of defined subpopulations, then you could reduce margin of error by sampling each subpopulation independently.

The tactics above only reduce sampling error — they do not eliminate it.

The only true way to eliminate sampling error entirely is to examine each and every individual member of your target population. This is often impractical, and in many cases impossible.

But that's not to say that generating random samples is an ineffective means of investigating a population. It's still a convenient and effective way to examine a large-scale, complex population.