



## Don't Let Your Research Investment Be Torpedoed By These Common Sampling Problems

**The value of survey research findings depends heavily on the quality of the research sample. Avoid the serious pitfalls of non-response and self-selection that can easily destroy the value of your market research investments.**

---

When you wish to measure attitudes from a large population, sampling is called for because it is too expensive to collect input from ALL members of the population. But the sample must accurately represent the population, or else the research findings are of questionable value. This paper explains two common sources of sampling problems, "non-response error" and "self-selection," and tells how to avoid them.

### Example

Let's say e-commerce company XYZ places a survey on its website asking visitors to take a brief satisfaction survey. After a week a total of 2,000 visitors have taken the survey. XYZ's objective in running this study is to learn how its typical customers rate their online shopping experience. It wants to know what percentage of customers are very satisfied, what percentage are neutral, and what percentage are dissatisfied. The company asks its research advisor, "How accurately will this sample of 2,000 provide this information?"

### The Error of Non-Response

The diligent researcher responds, "How many visitors were exposed to the invitation and declined to take the survey?" Let's say 200,000 visitors saw the invite and of these 2,000 or 1% chose to take the survey. That low percentage is rather common in web surveys as described.

You don't want to base your decisions on this type of sample. Why not? Because it is very likely that the 1% who chose to respond have different satisfaction than the 99% who chose NOT to respond. In general people who are UNHAPPY with their experience and people who are VERY HAPPY with their experience are more likely to respond. The e-commerce company's sample will OVER-represent the unhappy and the very happy visitors and UNDER-represent the typical visitors. The sample will not be representative of the population.

## The Error of Self-Selected Sample

But there's more. It is not possible to measure this sample's quality, which is another serious problem. That's because the sample was "self-selected." Participants decided whether they wanted to take the survey and only 1% chose to participate. If 60% or 70% or 80% had participated, we could be quite confident the sample would represent the population. But the e-commerce company would probably not want to bother up to 160,000 of its website visitors with taking the survey.

## The Right Way to Select the Sample

What is the right way to draw the sample? XYZ needs to create a RANDOM sample with a high degree of participation. For example, let's say XYZ draws a sample of 400 by offering the survey to every 100th customer and providing them with a coupon good for an amount off their next purchase, upon completion of the questionnaire. Now if 600 were offered the survey and 400 actually completed it, we have a simple random sample with a 33% (200/600) non-response rate. That's a pretty good sample.

The sample is "good" in part because its quality can be measured. Because it was randomly selected (there are various ways to select samples randomly - selecting each "nth" item is usually a reliable method), it is by definition representative of the overall population of 200,000. Because it is a random sample we can now apply the laws of statistics. A random sample of 400 from a population of 200,000 produces a maximum margin of sampling error of +/- 5% for variables expressed as percentages. This means that if 50% of respondents were "very satisfied" we can say with 95% confidence that between 45% and 55% of the overall population of 200,000 being analyzed would also have selected "very satisfied." The +/- 5% is a measure of the QUALITY of this sample.

So the properly drawn sample of 400 provides more useful information than the self-selected sample of 2,000. In fact, the self-selected sample doesn't provide much at all that is useful. Even if 10,000 had responded it wouldn't be very useful. Even worse, because the sample of 2,000 involves quite a few customers it may masquerade as a high quality sample and cause XYZ's management to take unwarranted actions.

## Tips for Proper Sample Selection

- (1) To minimize the error of non-response, be sure a high percentage of survey invitees actually take the survey. Look for the RESPONSE RATE in the survey report.
- (2) Use RANDOM SAMPLING in inviting participants, so you can be sure your sample represents the population.
- (3) Don't be swayed by the number of respondents. Remember, in the example 2,000 respondents produced worthless data while 400 produced high quality data.
- (4) Look for the MARGIN OF SAMPLING ERROR, which is the measure of sample quality. The quality of a random sample can be measured, and the quality of a self-selected or other non-random sample cannot.