

---

This tutorial covers selection bias. Selection bias goes by several other names. It's also called selection effect, or undercoverage bias. With selection bias, it occurs when some subjects are systematically included from the possibility of being part of the sample. The key to this definition is this "systematically."

Now, the systematic exclusion is because there's some particular thing that's causing people to not be included. Now, obviously you're taking a sample, so not every member of the population is going to end up being included. So people will be randomly selected, and random chance should be what determines whether or not people are in the sample or not.

If there's something else that's involved, there's some sort of systematic difference between the people who are being included and not, then you have an issue with bias-- and in particular, an issue with selection bias, also known as undercoverage bias. This type of bias could result in an unrepresentative sample. Here's an example.

If you're looking at telephone numbers and you choose to use a phone book, you'd end up with an unrepresentative sample. The numbers in a phone book right now tend to be older people. Most people have cell phones that are younger, and don't have a landline, and are not listed in any phone book. And of those older people, people who are women living alone tend not to list their phone numbers in a phone book.

So because there's systematic exclusion of young and of people who are afraid to list their phone numbers, then we're not going to have a representative sample, because there are people who are systematically excluded. Now, the way to fix this is through something called random digit dialing. With random digit dialing, the call list includes cell phones, landlines, and unlisted numbers.

The way this happens is a computer randomly selects an area code-- so for example, Minneapolis, 612. And then the computer randomly generates the following seven digits in order to come up with a telephone number. Now, because the digits are randomized and not reliant on a particular list, then it's possible to reach any phone number that exists, so the cell phones and those unlisted numbers are also going to be included. With random digit dialing, we no longer have the issue of systematically excluding the younger or people who are choosing not to list their numbers, and we have a more representative sample, and we don't have an issue of bias.

One organization that uses this technique-- the random digit dialing-- nowadays is Gallup. Gallup

conducts public opinion polls. Gallup produces public opinion polls, and their goal is to survey a small number of people in order to accurately represent the adult population of a country. And they use this, often, with phone interviews. Now, with the phone interviews you need your sample to be representative, so they need to use the random digit dialing.

This has been your tutorial on selection bias.