Selection and Deliberate Bias

by Sophia

WHAT'S COVERED

This tutorial will cover the topics of selection, deliberate, and unintentional bias. These may all impact the selection of the right group of people for your sample, so it's very important to be aware of them when attempting to generalize findings. Our discussion breaks down as follows:

1. Selection Bias

You may recall that sampling is like a pot of soup. Selecting a little bit of each ingredient for the soup is like obtaining a representative sample for an experiment. However, things can go wrong with the taste test, which may limit the ability to draw conclusions about the pot of soup as a whole.

Selection bias is also called undercoverage bias. It occurs when a significant subset of the population is left out of the sample. This is not necessarily intentional, but rather, occurs when they were systematically ignored by whoever was taking the sample.

IN CONTEXT

In 2008, almost every poll showed Barack Obama leading by at least five percentage points leading up to the New Hampshire presidential primary. All of these were based on random digit dialers calling a random sample of New Hampshire households. It was a well-done survey of all accounts.

Polling Data

NEW HAMPSHIRE DEMOCRATIC PARTY POLLING AND RESULT

Average Joe Poll	01/06	475	44	30	18	6	Obaama +8.0
WINNERS Polling	01/05	500	39	29	20	6	Obama + 8.5
SURE Stuff Poller	01/06	600	47	34	13	4	Obama +8.0
Randy on the Street	01/07	85	40	30	13	10	Obama +7.0
Giant Poll	01/05	1000	42	28	12	8	Obama +13.0
Random Poll	01/06	850	39	22	20	3	Obama +6.0
Mr. Ed Poll	01/05	700	40	20	18	2	Obama +9.0
FINAL RESULTS			39	30	16	7	Clinton +2.6

source: wikipedia

However, what happened was that Clinton gained some support in the last few days. Mainly, a lot of college students ended up coming out in support of Hillary Clinton in the last days when people were expecting all college students to come out in support of Obama.

Because a lot of the college students are from out of state, they aren't actually New Hampshire residents. For that reason, they were not counted and, as a result, the sample got every prediction wrong and Clinton ended up winning.

TERM TO KNOW

Selection Bias

A bias that results from systematically excluding certain subsets of the population from the sample. It is not necessarily intentional.

2. Random Digit Dialing

The New Hampshire primary used random digit dialers. **Random digit dialing** involves using a machine to select random phone numbers from within selected area codes. It doesn't randomly select the area code necessarily, but once it's in the area code, it can randomly select digits and dial that particular phone number after which the poll can be conducted.

The biggest advantage of using random digit dialers is that they can reach mobile phones and unlisted numbers that you wouldn't be able to obtain using a phone book. So, it evens the playing field a bit since anyone can be selected for that sample as long as the phone number is within that particular area code.

THINK ABOUT IT

How does selection bias affect what we think is in the soup? Imagine that certain ingredients were located only in certain locations in the pot. Maybe noodles sunk to the bottom. If you tasted only from the top, it doesn't matter how big that taste is. If you missed the noodles, you wouldn't even know they were there. That's the same as dealing with selection bias. Because you didn't select the representative group of

ingredients from the population, you don't get the right idea of what's going on. It limits your ability to generalize your findings to the general population.

TERM TO KNOW

Random Digit Dialing

A method of contacting people on the phone. Random numbers are dialed, so this allows researchers to sample people with unlisted phone numbers.

3. Deliberate Bias

Deliberate bias is exactly what it sounds like: it's a bias that's done on purpose. While deliberate bias doesn't happen very often, it can occur when there's a conflict of interest between the people performing research and the people funding--who are usually the ones benefiting from--that research.

Typically deliberate bias is motivated by an interest unrelated to the integrity of whatever you're researching. Most research is done with integrity, but when personal prestige, the advancement of some ideology, or money get in the way, it's harder to prove that intentions are pure. Politics can be an industry ripe for deliberate bias. Perhaps people call with a poll, but the survey includes a leading question to cause the person to respond in a certain way. When this is done it's called "push polling" and it's highly suspect.

IN CONTEXT

Deliberate bias can happen in other areas too--even the medical field. Suppose there are two drugs: Drug A and Drug B. The company for Drug B posed the following leading question:

"If Drug A was linked to cancer, would you be:

- more likely to choose Drug B?
- less likely to choose Drug B?
- equally likely to choose Drug B?"

Based on how this question was posed, Drug B would be more likely to be chosen.

But there's more. They've put a thought into the participant's head that Drug A is linked to cancer. Did they ever explicitly say that? No, they said *if* it was linked to cancer. However, now they've placed the association in the participant's mind. Subconsciously they're beginning to steer consumers away from Drug A and towards Drug B.

If a drug company funds a study to determine if it's latest drug is effective, the researchers stand to gain a lot of money and prestige for having tested the drug, if proven effective. For this reason, they might not be the best choice to test the drug.

IN CONTEXT

An environmental research group is hired by a real estate developer to investigate the effects of a new building. If the results are favorable, they might get another contract with that real estate developer. If the environmental research group doesn't come through with a favorable interpretation, another group will, and that group will get the next contract.

The environmental research group wants to be hired by the developer on another project, so there is a conflict of interest.

TERM TO KNOW

Deliberate Bias

The purposeful misrepresentation of data for the purpose of advancing an agenda.

4. Unintentional Bias

Unintentional bias occurs when there is simply an error in the design of the study. Two types of unintentional bias include:

- Response bias, which involves the wording of questions or refers to people feeling like they have to lie.
- Selection bias, which involves how the sample was selected, such as when people are not included in the selection process, even though they make up a portion of the population.

Both are simply errors with no hidden agenda. They're not intentional and are not meant to purposely steer the direction of the respondents.

TERM TO KNOW

Unintentional Bias

Bias that is not purposeful. It exists because of errors in the design of the study.

SUMMARY

Selection bias occurs when some subset of the population is left out. It might be intentional or unintentional. Since some section of the population is left out, the coverage is lacking, which is why selection bias is also known as "under-coverage". Random digit dialing is a great tool to use since it helps extend coverage to mobile phones and unlisted numbers. Most of the time, deliberate bias-- a bias that is done on purpose--is not typically a cause of concern. Sometimes, however, people with personal interests, like the advancement of an ideology or financial gain, steer results towards outcomes that are favorable to them. Most of the time, research is done with integrity. When bias does

occur, it is accidental, which is called unintentional bias.

Good luck!

TERMS TO KNOW

Deliberate Bias

The purposeful misrepresentation of data for the purpose of advancing an agenda.

Random Digit Dialing

A method of contacting people on the phone. Random numbers are dialed, so this allows researchers to sample people with unlisted phone numbers.

Selection Bias

A bias that results from systematically excluding certain subsets of the population from the sample. It is not necessarily intentional.

Unintentional Bias

Bias that is not purposeful. It exists because of errors in the design of the study.