

What is the definition of self-selection bias and what are some examples of it?

Authored by
stats writer

April 26, 2024

RECOMMENDED CITATION

stats writer (2024). *What is the definition of self-selection bias and what are some examples of it?*. PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=139537>

Self-selection bias refers to the potential distortion in research or data analysis caused by individuals choosing to participate or be included in a study or sample based on their own characteristics or preferences. This can lead to a biased representation of the population being studied and can impact the accuracy and generalizability of the results.

For example, in a study examining the effectiveness of a weight loss program, individuals who volunteer to participate may be more motivated and committed to losing weight, leading to an overestimation of the program's success. Similarly, in surveys about satisfaction with a product, individuals who choose to respond may have strong positive or negative opinions, while those who do not respond may have a more neutral stance, resulting in a biased representation of overall satisfaction levels.

Self-Selection Bias: Definition & Examples

Self-selection bias occurs when individuals select themselves to be included in a survey.

For example, suppose a local government mails out a survey to all of its residents asking them whether or not they think a new intersection should be placed in the middle of the town.

Residents who drive through that particular part of town often and residents who have to deal with daily traffic are more likely to hold a strong opinion about a new intersection and are far more likely to actually respond to the survey.

On the other hand, residents who work from home or simply have no interest in the happenings around the

town are unlikely to take the time to respond to the survey.

Thus, the percentage of individuals in the survey who are in favor of the new intersection is unlikely to match the percentage of all residents in the town who are in favor of the new intersection.

Self-Selection Bias: When individuals select themselves to be included in a survey.

This results in a sample of individuals that is not representative of the overall population.

This makes it difficult to generalize the findings from the sample to the .

In other words, there is bias in our sample data. This makes it difficult to generalize the findings from the sample data to the overall population of interest.

Examples of Self-Selection Bias

The following examples illustrate a few scenarios where self-selection bias is likely to occur.

Example 1: Test Prep

Suppose a teacher wants to know if a new test prep course helps students improve test scores. She posts a sign-up sheet outside of her classroom and lets students decide if they'd like to participate in the course.

Self-selection bias is likely to occur because students who are more studious are more likely to sign up which means the sample of students who take the course aren't likely to match the overall population of students who could potentially take the course.

Example 2: Multiple Languages

Suppose a local government mails out a survey asking its residents if they should include other languages besides English on street signs to make it easier for people who speak other languages to navigate around town.

Self-selection bias is likely to occur because only residents who can actually read English will respond to the survey. This means the opinions of survey

respondents are unlikely to match the opinions of all residents in the town.

Example 3: Biology Research

Suppose a biologist is trying to estimate the average height of a certain species of deer so she places a certain deer feed in an open meadow and takes pictures of the deer that enter the meadow to eat the food.

Thus, it's unlikely that the average height of deer in this sample will match the average height of deer in the overall population.

Why Self-Selection Bias is a Problem

Self-selection bias is a problem because it causes the individuals in the sample to not be of the population.

Recall that the purpose of collecting sample data is to use it to draw conclusions about some population of interest. However, we can only draw valid conclusions if we use a representative sample.

Representative sample: A sample in which the characteristics of the individuals closely match the

characteristics of the overall population.

Ideally we would like the sample to be like a "mini version" of the population. This allows us to be confident about using the sample to draw conclusions about the population.

How to Reduce Self-Selection Bias

The obvious way to reduce self-selection bias is to not give individuals the ability to select themselves to be included in a survey.

Ideally, a should be used to obtain a sample.

Probability sampling method: A sampling method in which each member in a population has an equal probability of being selected to be in the sample.

Examples of probability sampling methods include:

1. Simple random sample: Randomly select individuals through the use of a random number generator or some means of random selection.

2. Systematic random sample: Put every member of a population into some order. Choose a random starting

point and select every n th member to be in the sample.

3. Stratified random sample: Split a population into groups. Randomly select some members from each group to be in the sample.

4. Cluster random sample: Split a population into clusters. Randomly select some clusters and use each individual in the chosen clusters to be in the sample.

Each of these methods is likely to produce samples that are representative of the population we're interested in, which allows us to generalize the findings from the sample data to the population.