3.2 Sampling Good and Bad Part 1: How to Sample Poorly with Convenience Samples

Many statistical studies use information from a sample to make a conclusion about an entire population. To ensure that these conclusions are accurate, we must be mindful of how the sample is selected.

Suppose we want to know how much time students at a large high school spent doing homework last week. We might go to the school library and ask the first 30 students we see about their homework time. The sample we get is called a **convenience sample**.

Choosing individuals from the population who are easy to reach results in a **convenience sample**.

Convenience sampling often produces unrepresentative data. Students who hang out in the library probably spend more time doing homework than a typical student. Our estimate for the average amount of time spend doing homework will be too high.

In fact, if we were to repeat this sampling process again and again, we would almost always overestimate the average time spent doing homework in the population of all students.

This predictable overestimation is due to **bias** in the sampling method. The design of a statistical study shows **bias** if it would consistently underestimate  consistently overestimate the value you want to know when the study is repeated many times.

Bias is not just bad luck in one sample. It's the result of a bad study design that will consistently miss the truth about the population in the same direction.

Example: A farmer brings a juice company many crates of oranges each week. A company inspector looks at 10 oranges from the top of each crate before deciding whether to buy all the oranges. Explain why this sampling method is biased. Is the population of damaged oranges in the sample likely greater than or less than the population of all oranges in the crate that are damaged?

① The oranges at the top should not predict how the rest of the oranges are.

② Less than