



M5-05: Sampling

Part of the "Polling, Confidence Intervals, and the Normal Distribution" Learning Badge

Video Walkthrough: <https://discovery.cs.illinois.edu/m5-05/>

Sampling and Inference

In our discussion of random variables, we started with games of chance because we can easily think about all possible outcomes and the probabilities associated with them.

Now, we are going to see how random variables relate to gathering information about large populations from small samples. This is also known as *inference*.

We take a sample to find out about a larger population. We usually don't have the resources to gather information on everyone in the whole population so instead, we select a small sample and use it to make inferences about the larger population.

Definitions

- **Population:** the whole class of individuals about whom the investigator wants to generalize.
- **Sample:** the part of the population the investigator examines.
- **Inferences:** generalizations about the population that come from the sample.
- **Parameters:** numerical facts about the population.
- **Statistics:** estimates of the parameters computed from the sample.

Main Idea: The closer the sample is to the population, the more accurate our inferences will be.

Consider the following methods of sample selection. Which is the **best method** for drawing a sample as much like the population as possible?

- The researcher handpicks the sample to resemble the population on all relevant characteristics.
- The researcher publicly posts the survey (online usually) and allows anyone to respond.
- The researcher uses simple random sampling, meaning the sample is *randomly* selected and everyone in the population has an equal chance of being chosen.