



M5-05: Bias

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

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

RANDOM SAMPLES are best for surveys for exactly the same reasons that randomized controls are best for experiments:

- **Random selection is most likely to make the sample as like the population as possible because it eliminates selection bias.** With enough subjects, **random differences average out**, not only on the characteristics that the researcher has identified as relevant but on *all* characteristics, including hidden ones that the researcher might not realize are important.
- **Random samples can be translated into probability models so that we can use statistics to measure the accuracy of our estimates.**

Random sampling eliminates selection bias but there are still other possible sources of bias. The 2 main ones are:

Non-response bias – Not everyone who is chosen to participate in a survey responds. The people who respond may be systematically different from those who don't.

Response bias – The way a question is phrased, who asks it, or the images associated with it can influence responses.

With sampling, the goal is for our sample statistics to be as close to the population parameters as possible. This means we want to minimize bias and error.
--

Sampling Visual: