Sampling Distributions and Sampling Variance, Part 2

Brady T. West

Research Associate Professor, Survey Research Center, Institute for Social Research

Why is Sampling Variance Important?

- In practice, we only have the resources to select **one sample!**
- Important sampling theory (developed in early 1900s) allows us to **estimate features of sampling distribution** (including variance!) based on **one sample**

"Magic" of probability sampling:

Can select one probability sample and features of that design tell us what we need to know about the expected sampling distribution

The magic of probability sampling that we're talking about here is that we can

Why is Sampling Variance Important?

- **Because we can estimate variance** of sampling distribution based only one sample, **we can make inferential statements** about where most estimates based on a particular sample design will fall

  → **Can make statements about likely values of population parameters** that account for variability in sampling errors that arises from random probability sampling

So, we can make statements about likely values of population parameters

You speculate that the mean test performance in an undergraduate psychology class is 80 / 100. Based on a single sample of 30 students, you estimate the mean test score to be 90. The estimated sampling variance based on that one sample suggests that most estimates from repeated hypothetical samples of size 30 will lie between 85 and 95. How confident are you about your speculation?

- (●) We have fairly strong evidence against our speculation: a mean of 80 seems unrealistic.

Correct
The estimating sampling variance suggests that most estimates of the mean test performance will lie between 85 and 95, which makes 80 seem like an unrealistic value for the mean. We do not need to draw several samples of size 30, because sampling theory allows us to estimate the variance of sampling distribution based on one sample only!

- ( ) We have evidence in support of our speculation: 80 seems like a plausible mean value.
- ( ) We need to draw several more samples of size 30 before we can make a conclusion.

What's Next?

- Work with a **Web App** to **visualize sampling distributions** when selecting random samples from a population with certain features:

https://markkurzejaumich.shinyapps.io/multiple_population_bias/

- See how **random sampling** generally produces sampling distributions with means close to true population quantity of interest, and how larger samples produce sampling distributions with less variance
- See how **biased, non-representative samples** can produce sampling distributions that paint **misleading** pictures of the larger population

sampling distributions that paint misleading pictures of the larger population.