



# Sampling Distributions and Sampling Variance

*Brady T. West*



# Lecture Overview

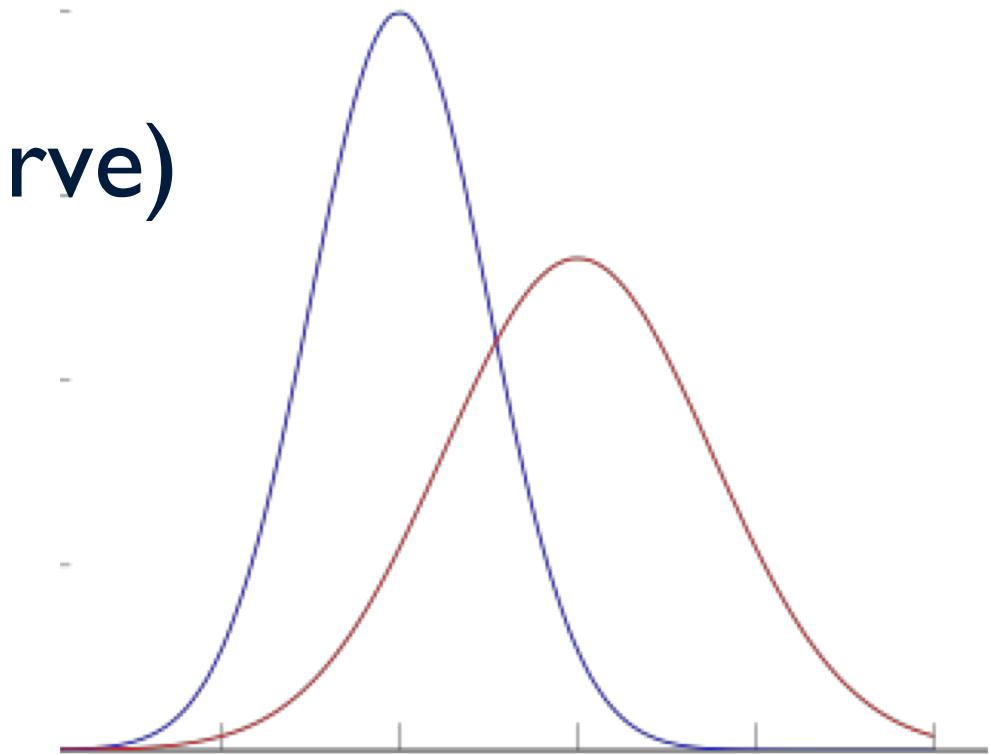
- **What is a sampling distribution?**
- **What is sampling variance?**
- **Why is sampling variance so important** for making population inferences based on probability samples?

# What is a Sampling Distribution?

- **Recall: Distribution of values on a variable of interest**

Example: **Normal distribution** (bell curve)

- **Assume** values on variable of interest would follow certain distribution *if we could measure entire population*

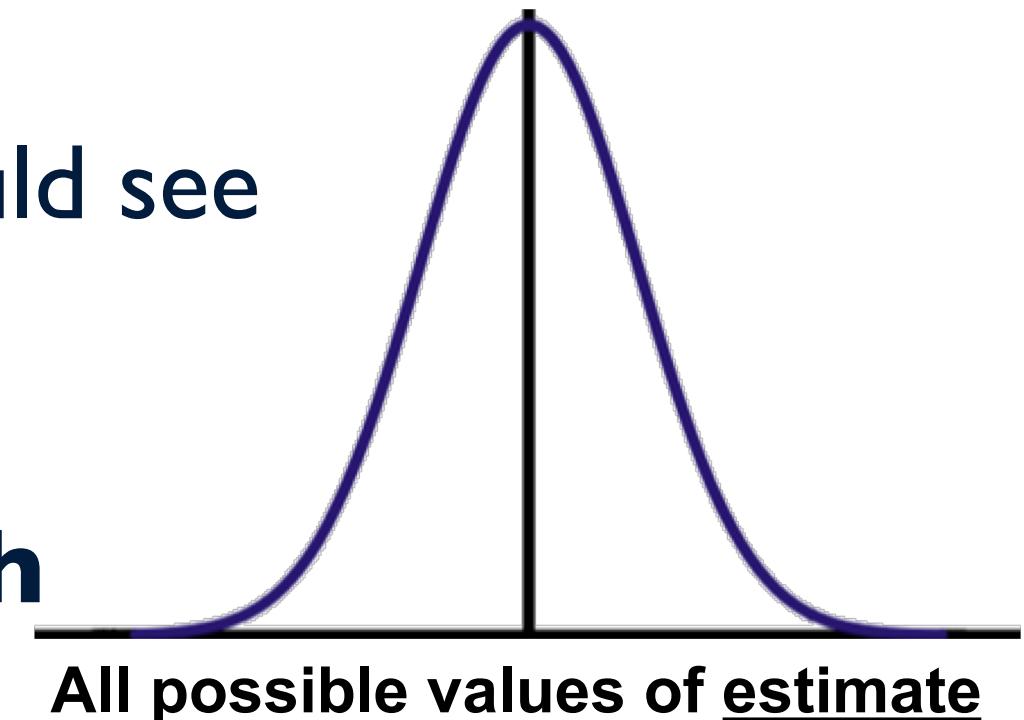


All possible values of variable of interest

[Bell curve by Jake CC-BY 2.0](#)

# What is a Sampling Distribution?

- **Recall:** When we select probability samples to make inferential statements about larger populations  
→ we refer to a **sampling distribution**
- **Sampling distribution** = distribution of **survey estimates** we would see *if* we selected many random samples using **same sampling design**, and **computed an estimate from each**



# What is a Sampling Distribution?

- **Key properties** of sampling distributions:
  - **Hypothetical!** What would happen if we had luxury of drawing thousands of probability samples and measuring each of them?
  - Generally very **different in appearance from distribution of values on a single variable of interest...**
  - With **large enough probability sample size**, sampling distribution of estimates will look like a **normal distribution**, regardless of what estimates are being computed! **Central Limit Theorem: CLT**

# What is Sampling Variance?

- **Sampling variance** = variability in the estimates described by the sampling distribution
- Because we select a **sample** (do not measure everyone in a **population**), a survey estimate based on a **single sample** **will not** be exactly equal to population quantity of interest  
(cases are randomly selected!)

**Sampling  
Error**

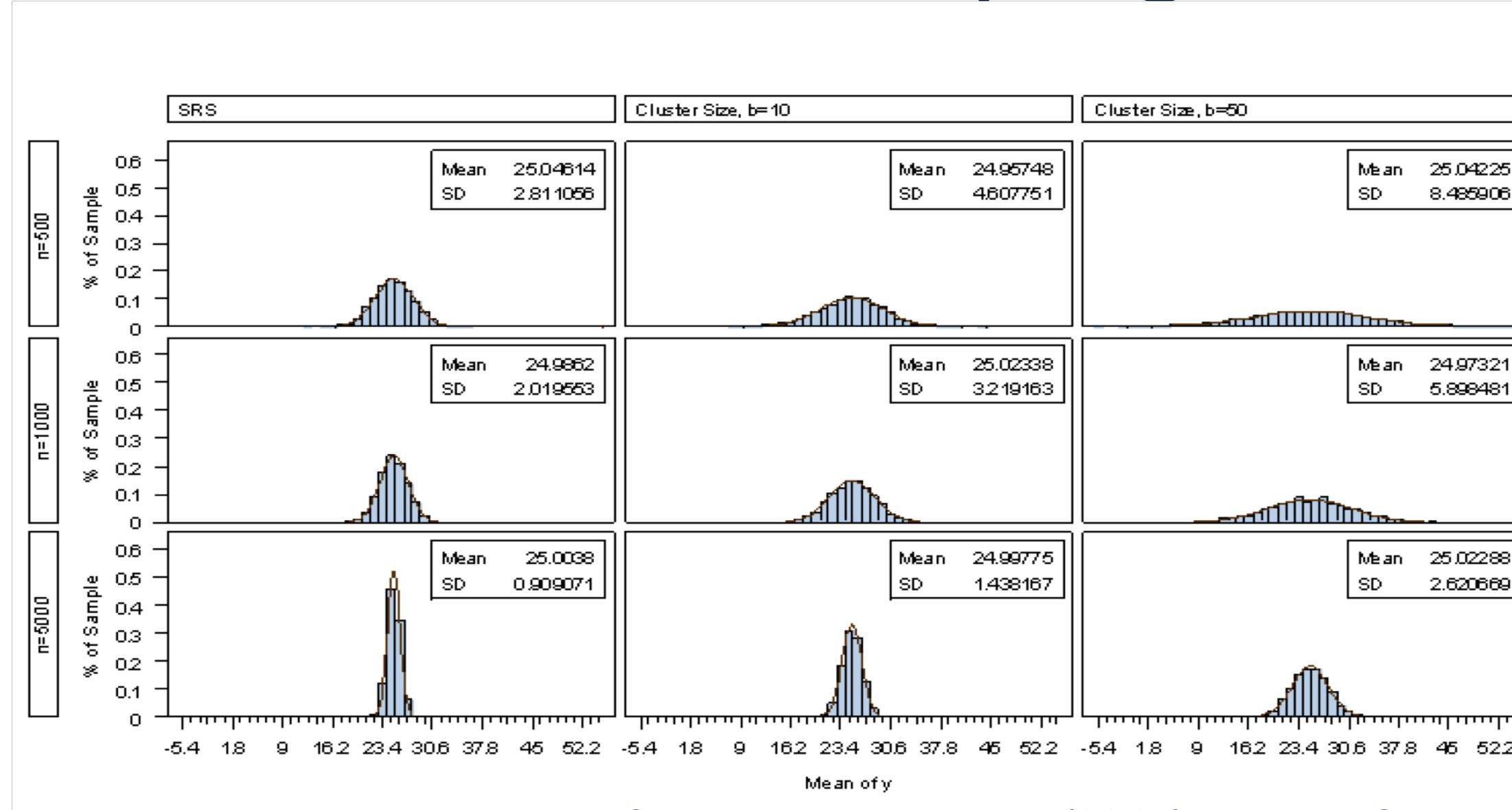
# What is Sampling Variance?

- Across hypothetical repeated samples, **these sampling errors will randomly vary** (some positive, some negative...)
- **Variability of these sampling errors** describes the variance of the sampling distribution
- **If every sample estimate was equal to population quantity of interest** (e.g., in the case of a Census), there would be **no** sampling error, and **no** sampling variance!

# What is Sampling Variance?

- With a **larger probability sample size**, sampling more from a given population → in theory there will be less sampling error, and sampling errors will be less variable
- **Larger samples** → **Less sampling variance!**  
**More precise** estimates, **more confidence** in inferential statements (**but** more costly!)
- **Spread** of sampling distribution becomes **smaller as sample size become larger**

# Simulated Sampling Distributions



As sample size increases (across rows), sampling distributions shrink (**less variance**)

With cluster sampling, (2nd and 3rd columns) distributions spread out (**more variance**)

Credit: Heeringa et al. (2017), Applied Survey Data Analysis, Second Edition