Topic 3.3: Random Sampling and Data Collection

Evaluating Sampling Methods: Bias & Variability

Video Follow-Along Worksheet — Daily Video 2

Learning Objective (DAT-2.D): Explain why a particular sampling method is or is not appropriate for a given situation.

Essential Knowledge: There are advantages and disadvantages for each sampling method depending upon the question that is to be answered and the population from which the sample will be drawn.

Key Vocabulary

Bias	A measure of **accuracy**; whether estimates are centered at the true population value
Variability	A measure of **precision**; how spread out estimates are from each other
Unbiased	Estimates are centered at the true value (accurate)
Low Variability	Estimates are close together (precise)

Part 1: Introduction & Review [0:00–1:42]

1. In this video, we will learn what criteria we can use to _____ sampling methods.

2. **San Antonio Example Review:** San Antonio has about _____ households. We wanted to randomly sample _____ of them (2% of the population).

3. Fill in the sample median estimates from each method:

Sampling Method	Sample Median
Simple Random Sample (SRS)	$_____
Cluster Random Sample	$_____
Stratified Random Sample	$_____

4. The question we need to answer: Which of these estimates should we _____ the most?

Name: _____ Period: _____

Part 2: Understanding Bias & Variability [1:42–2:52]

5. **Bias** is a measure of _____. It asks: Are you centered at the _____ you want to be estimating?

6. **Variability** is about _____. It asks: How much distance is there between the different _____ that you might have gotten?

7. Label each dartboard with the correct description:

Target A (off-center, spread)	**Target B** (off-center, clustered)	**Target C** (centered, spread)	**Target D** (centered, clustered)
_____ _____	_____ _____	_____ _____	_____ _____

8. The "gold standard" (#Goals) is _____ and _____.

9. The true median household income in San Antonio (from the census) is $_____.

Part 3: Simulation Results [2:52–4:50]

10. When we simulate sampling many times, we can see how our estimates compare to the _____.

11. **Non-Random Sample** (students picking "representative" areas):
    - Result: The center of sample medians was _____ (systematically overestimated).
    - This happens because non-random samples can lead to _____.

12. **Simple Random Sample (SRS):**
    - Bias: _____ (centered at true median)
    - Variability: _____

13. SRS Advantages:
    - It is _____
    - It is easy to _____ to people

14. SRS Disadvantages:
    - In certain scenarios, it is difficult to _____
    - May not be as _____ as other methods

Name: _____ Period: _____

Part 4: Cluster Sampling Analysis [4:50–6:43]

15. When simulating cluster samples many times, the estimates are spread _____ across the number line.

16. **Why does this happen?** Income is consistent _____ regions, but it varies a lot _____ regions.

17. If we happen to select north side clusters: median estimate is very _____.
If we happen to select lower-income clusters: median estimate is very _____.

18. **Cluster Sample Results:**

- Bias: _____ (because it uses random selection)
- Variability: _____ (not precise)

19. Cluster Sample Advantages:

- It is _____
- Can be _____ (only need to visit certain areas)

20. Cluster Sample Disadvantages:

- If clusters are _____ but very different from one another, you can have very _____.

21. **Key insight:** For cluster sampling to work well, you want clusters that are _____ within groups, and the groups should be _____ to one another.

Part 5: Stratified Sampling Analysis [6:43–7:42]

22. When simulating stratified samples many times, the estimates are clustered _____ around the true median.

23. **Stratified Sample Results:**

- Bias: _____
- Variability: _____ (very precise!)

24. **Why does this work so well?** Because income is _____ within regions, each sample from a stratified method will have a similar _____ of incomes.

25. Stratified Sample Advantages:

- It is _____

Name: _____ Period: _____

- When strata are homogeneous, tends to have _____

26. Stratified Sample Disadvantage:
    - Can be very _____ (complicated procedure, must visit all areas)

Sampling Methods Comparison

Method	Bias?	Variability	Best When...
Simple Random Sample			
Cluster Sample			
Stratified Sample			

Part 6: Key Takeaways

[7:42–7:58]

Complete the main takeaways from the video:

27. Random sampling tends to provide _____ estimates.

28. Cluster random sampling is effective when clusters are _____ and _____ to one another.

29. Stratified random sampling is most effective when strata are _____.

30. As always when analyzing data: Be _____. Be _____. Be _____. And avoid BS—_____!

Part 7: Post-Video Reflection

31. **Connecting to the Big Idea:** In your own words, explain why random sampling methods tend to be unbiased while non-random methods often produce biased estimates.

Name: _____ Period: _____

32. **Identifying Key Elements:** For the San Antonio example:
    - What was the **population**? _____
    - What was the **parameter of interest**? _____
    - Why was income **segregation** important to consider? _____

33. **Real-World Application:** A researcher wants to estimate the average commute time for employees at a large company with offices in 5 different cities. The commute times are similar within each city but vary greatly between cities. Should the researcher use cluster or stratified sampling? Explain your reasoning.

34. **Looking Ahead:** Even well-designed random samples can have problems. What are some potential sources of bias that could affect any sampling method? (Hint: Think about who might not respond or be included.)

Exit Ticket

In 1–2 sentences, explain what makes a sampling method "good" and which method performed best for the San Antonio income example.

Name: _____ Period: _____