

Unit 3 Kickoff Packet

TEACHER ANSWER KEY

Topics 3.1, 3.2, and 3.3

FOR TEACHER USE ONLY — Do not distribute to students

Video 1: Topic 3.1 — Do the Data Tell the Truth?

The Abraham Wald Airplane Problem

1. Statistical Research Group; Allied
2. armor; one
3. came back / returned
4. Pause & Think: **Wings/body** (where bullet holes appear)
5. context
6. came back / returned; make it back / survive
7. tail (or engines/cockpit)

Key Takeaways from Video 1

8. how the data was collected
9. representative; population
10. Critical; Cautious; Compassionate; Bad Statistics

Video 2: Topic 3.2 — Planning a Study

Generalization Rules

1. randomly selected; population
2. cannot

The Racial Income Gap

3. \$12,000; \$538,000
4. confounding
5. observational study; cannot
6. current/past; future

The Résumé Experiment

7. randomly
8. callback
9. experiment; imposed
10. can

Observational Study vs. Experiment Table

	Observational Study	Experiment
Treatments imposed?	No	Yes
Can show causation?	No	Yes
Example from video	Income survey	Résumé experiment

Video 3: Topic 3.3 (Part 1) — Random Sampling Methods**Why Sample?**

1. all; hard/difficult
2. median
3. representative

Simple Random Sample (SRS)

4. group; equal
5. random number generator; without replacement

Cluster Random Sample

- 6. clusters
- 7. clusters; all
- 8. certain/specific

Stratified Random Sample

- 9. strata
- 10. within

Systematic Random Sample

- 11. start point; periodic
- 12. 20th
- 13. easy to collect

Video 4: Topic 3.3 (Part 2) — Evaluating Sampling Methods**Understanding Bias & Variability**

- 1. accuracy; true value
- 2. precision
- 3. low/no; low
- 4. 51,000

Dartboard Labels

Target A	Target B	Target C	Target D
Biased	Biased	Unbiased	Unbiased
High Variability	Low Variability	High Variability	Low Variability
<i>Target D is the GOAL (“gold standard”)</i>			

Simulation Results

5. Non-Random: **Biased**; bias
6. SRS: **Unbiased**; Moderate; explain; implement
7. Cluster: **Unbiased**; High/Very high; within; between; heterogeneous
8. Stratified: **Unbiased**; Low; mix; homogeneous; difficult

Sampling Methods Comparison Table

Method	Bias?	Variability	Best When...
Simple Random Sample	Unbiased	Moderate	Population is uniform
Cluster Sample	Unbiased	High	Clusters are heterogeneous
Stratified Sample	Unbiased	Low	Strata are homogeneous

Key Takeaways from Video 4

9. unbiased
10. heterogeneous; similar
11. homogeneous

Part 3: Check for Understanding — ANSWERS

Section A: Observational Studies vs. Experiments

A1. Answer: D

Explanation: Observational studies cannot establish causation because no treatments are imposed. Confounding variables may explain the association between sleep and job performance.

A2. Answer: B

Explanation: The experimental units are the 50 people in the study—they are the entities to which treatments were applied.

A3. Answer: E

Explanation: We can only generalize to the population from which the sample was drawn. The sample has three characteristics: (1) women, (2) students, (3) at Midwestern colleges. Therefore, results only generalize to women who are students at Midwestern colleges in the United States.

Section B: Sampling Methods

B1. Answer: D (Stratified random sample)

Explanation: The racks are strata (homogeneous within—similar prices; different between—prices vary by rack), and the consumer samples FROM EACH rack. This is the definition of stratified sampling.

B2. Answer: C

Explanation: Because seniors' plans vary greatly between schools, the 5 selected schools may not represent all seniors in the district. This leads to **high variability** in estimates—if the selected schools happen to have mostly college-bound students, the results won't reflect schools with more students entering the workforce.

B3. Answer: D (Random sampling with replacement)

Explanation: Margo returns the marble before the next draw ("puts it back"), which is sampling with replacement.

Section C: You Try It!**C1.**

- (a) **Stratified sampling**
- (b) *Reasoning:* Cities are **homogeneous within** (employees in the same city have similar commute times) but **different between** (commute times vary by city). Stratified sampling works best when strata are homogeneous. By sampling from EACH city, we ensure representation from all locations and reduce variability.
- (c) *Wrong method consequence:* If the researcher used **cluster sampling** and only selected 2 cities, the estimate could have **very high variability**. If those 2 cities happened to have unusually long or short commutes, the estimate would be way off from the true average.

C2.

- (a) *Reasoning:* Deer populations differ by habitat (forest vs. lake vs. town). A simple random sample might, by chance, place most observation stations in one area and miss others entirely. **Stratified sampling** ensures all three areas are represented, giving a more accurate overall estimate of the deer population.
- (b) *Problem:* Counting the same deer multiple times would **overestimate** the total population. Each deer counted twice is essentially being treated as two separate deer, inflating the count.

Exit Ticket: Sample Strong Response

Two most important ideas:

1. **Sampling:** Random sampling methods (SRS, stratified, cluster) produce unbiased estimates, but the best method depends on the population structure. Stratified works best when groups are homogeneous; cluster works best when groups are heterogeneous.
2. **Study Design:** Only experiments can establish cause-and-effect because treatments are imposed on subjects. Observational studies can show association but not causation due to confounding variables.