

Unit 1B Review

AP Statistics - Mr. Levy

Instructions: The following questions cover the main ideas of Unit 1B. One student will be chosen at random to answer each question out loud.

Quiz: Thursday, October 1.

1. Every study you meet this year sits in one of four boxes. Fill in all four: what you may conclude about the **population**, and whether you may say the explanatory variable **caused** the change.

	Treatments randomly assigned	Treatments <i>not</i> randomly assigned
Units randomly selected		
Units <i>not</i> randomly selected		

2. Write **O** for observational study or **E** for experiment. If it is observational, use the second blank for **R** (retrospective) or **P** (prospective).

_____ *I.* A streaming service shows half its users, chosen at random, a red play button and the other half a green one, then records hours watched.

_____ *II.* A researcher compares the lifespans of orchestra conductors with those of the general public, using sixty years of obituary records.

_____ *III.* A hospital gives every third patient admitted a new discharge checklist, then compares readmission rates.

_____ *IV.* A health service signs up 5,000 adults today and records their diet and their heart events every year for the next decade.

3. Patients treated at hospitals that have a dedicated stroke unit are **more** likely to die within 30 days of a stroke than patients treated at hospitals that have none.

a) Name one variable confounded with the kind of hospital here and explain **both** links.

b) A headline says the study shows stroke units are dangerous. What is wrong with that?

c) What one change to the design would let a researcher say *caused*?

4. Write **E** if the question asks you to estimate a parameter or **T** if it tests a claim about one. If it is a test, use the second blank for **1** (one-sided) or **2** (two-sided).

_____ *I.* What is the mean number of minutes per day that all students here spend on their phones?

_____ *II.* Is there convincing evidence that fewer than 30% of all adults in this county recycle?

_____ *III.* Is there convincing evidence that the mean commute time differs between the two districts?

_____ *IV.* What is an estimate of the difference in the proportions of boys and girls here who play a sport?

5. A finished investigative question does three jobs: it names the **variable(s)**, fixes the **type of inference**, and settles the **type of conclusion**.
- Name all three for: "Is there convincing evidence that the proportion of all first-year teachers in Florida who leave the profession within five years is greater than 0.40?"
 - Say what is missing from each question below, then rewrite it so it is finished.
 - "Do people like the new logo?"
 - "Is there convincing evidence that the 50 batteries we tested have a mean life greater than 900 hours?"
6. A district employs 900 teachers. Write **T** or **F** in each blank.
- _____ I. In a simple random sample of size 30, every teacher has the same chance of being chosen.
- _____ II. If every teacher has the same chance of being chosen, the sample must be an SRS.
- _____ III. A stratified random sample of 30, ten from each of three schools, is an SRS of size 30.
- _____ IV. A census is one of the four random sampling methods.
7. Name the sampling method, then give the feature of *that population* that makes it a sensible choice.
- A wildlife agency divides a lake into shallow, mid-depth and deep zones, then takes a random sample of 10 net hauls from each zone.
 - Loaves come off a bakery conveyor and there is no list of them. A worker picks a random number from 1 to 12, weighs that loaf and every 12th loaf after it.
 - An airline randomly selects 8 of the 400 flights it operates that day and surveys every passenger aboard those 8.
 - A museum numbers its 1,200 catalogued paintings and uses a random number generator to pick 25 different numbers.
8. A reef survey team must choose 5 of the 32 marked plots to dive this season.
- Describe how to choose a **simple random sample** of 5 plots.
 - Plots on the sheltered side and the exposed side of the reef are likely to differ a lot. Which method should the team use instead, and what advantage does it give over an SRS of the same size?
9. Fill in the blanks, then answer the question underneath.
- Strata are built so that units are _____ within a stratum and _____ between strata.
- Clusters are built so that each cluster is a _____ of the whole population.
- A school has a complete list of all 1,400 students and no travel cost. Why would cluster sampling be the wrong choice here?

10. Check every statement below that is **true**.

- ☐ Bias is a property of the sampling method, not of one unlucky sample.
- ☐ Taking a larger sample makes a biased method less biased.
- ☐ Response bias can show up even in a perfect simple random sample.
- ☐ Nonresponse bias and voluntary response bias are two names for the same problem.

11. Name the type of bias and say whether the estimate is an **overestimate** or an **underestimate**.

- a) A gym estimates the mean number of visits per member per month from the card swipes of the members who renewed this year.
- b) A school estimates the proportion of its students who vape by asking students to raise their hands in homeroom.
- c) A university emails a course-evaluation link to all 20,000 students; 1,800 reply and the averages rank professors. Name **two** sources.

12. A polling firm estimates the proportion of adults who support a ballot measure using a voluntary web panel. It grows the panel from 600 respondents to 2,400.

- a) What happens to the **precision** of the estimate? What happens to its **accuracy**?
- b) A colleague says the bigger panel makes the estimate four times as trustworthy. Respond.
- c) What would actually fix the problem?

13. A greenhouse tests **4 light colours** and **3 watering schedules** on 120 basil seedlings, recording the grams of leaf harvested after six weeks.

- a) List the factors and the number of levels of each.
- b) How many treatments? Show the multiplication, then describe one in words.
- c) What are the experimental units, and what is the response variable?

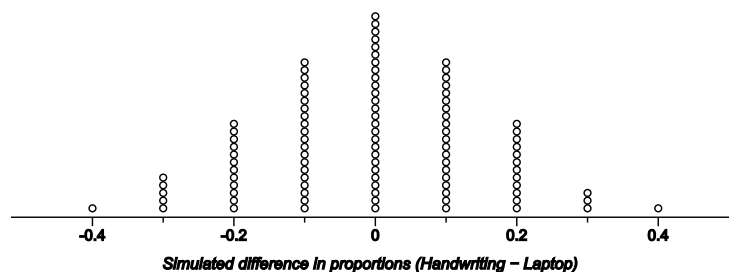
14. A coach has his 18 sprinters wear new carbon-plated spikes in a May time trial, compares their times with the March trial, and concludes that the spikes make runners faster.

- a) Faster compared to *what*? Name two variables confounded with the spikes here.
- b) Redesign the study so that it has all four elements of a well-designed experiment.

15. A drug company tests a new migraine tablet against a placebo on 300 volunteers who get frequent migraines.

- a) What is a placebo, and why does *this* study in particular need one?
- b) What would single-blind and double-blind each mean here?
- c) The nurse recording how quickly each headache ends knows which tablet each patient got. Name a specific way that could distort the results.

16. A researcher compares two fertilisers on tomato yield in one field. For each variable below, say whether you would **hold it constant**, **block** on it, or **let random assignment handle it** — and why.
- The amount of water each plant receives.
 - Soil drainage, far better at the north end of the field than the south end.
 - Tiny genetic differences between individual seedlings that nobody can measure.
17. A dairy tests **three feed supplements** on milk yield using 30 cows. The cows differ enormously to begin with — some already give nearly twice as much milk as others.
- What should they block on, and why?
 - Describe the random assignment within each block.
 - With only **two** supplements, and each cow receiving both in turn — what is that design called, and what still has to be randomized?
18. Forty volunteers were randomly assigned, 20 per group, to take lecture notes by hand or on a laptop. Of the 20 who wrote **by hand**, 13 recalled at least 15 of 20 words; of the 20 who used a **laptop**, 7 did. That is $0.65 - 0.35 = 0.30$ higher for handwriting. A computer then re-shuffled the 40 students into two groups of 20 a hundred times, assuming the method makes no difference. Each dot below is one of those re-shuffles.



100 simulated random assignments, assuming no treatment effect

- How many simulated differences are at least as large as 0.30? Is the result statistically significant, and what does that phrase mean here?
 - Write the conclusion, including who it applies to.
19. A random sample of 1,200 U.S. adults was asked about screen use before bed. Among those who used a screen in the hour before sleeping, 38% reported poor sleep quality; among those who did not, 22% did.
- What is the largest population these results apply to, and what in the design tells you that?
 - Can the survey support the claim that screen use *causes* poorer sleep? Why or why not?
 - Name a variable that could be confounded with screen use here, and explain both links.