

Unit 1B Review — Solutions

AP Statistics - Mr. Levy

How to use this: every answer with the reasoning behind it. The italic line at the end of each is the mistake that most often produces a right-looking wrong answer — worth reading even on the ones you got right. Questions 3a, 11c, 13c, 14a and 19c have more than one good answer; what matters is whether you argued it.

1. **Both random → generalize to the population AND cause and effect. Random selection only → generalize, but association only. Random assignment only → cause and effect, but only for units like those in the study. Neither → neither.**

Random **selection** is what licenses the jump from sample to population. Random **assignment** is what licenses the jump from association to causation. They are separate mechanisms doing separate jobs, and a study can have one, both or neither.

| *The bottom-left box is the one people get wrong. A randomized experiment on volunteers does prove cause and effect — it just proves nothing about anyone outside the study.*

2. **I. E · II. O, R · III. E · IV. O, P**

I: the button colour is imposed by the researcher, so it is an experiment. II: units chosen now, data pulled from the past — retrospective. III: a checklist is *imposed* on some patients, so it is an experiment, even though “every third patient” is not random assignment. IV: nothing is imposed; 5,000 people are enrolled now and followed forward — prospective.

| *Randomization is not what makes something an experiment. Imposing a treatment is. III is an experiment with bad assignment, not an observational study.*

3. **a) Severity of the stroke (or the patient’s prior health). b) The study is observational, so the groups differed before treatment. c) Randomly assign patients to the two kinds of hospital.**

a) Both links: the most severe strokes are the ones transferred to or routed toward hospitals with a stroke unit; and severe strokes are more likely to end in death within 30 days. Severity therefore offers a complete alternative explanation for the gap. b) Hospitals were not assigned to patients by chance, so the sicker population at the stroke-unit hospitals is confounded with the care they received.

| *Naming a variable is not enough. Argue both links, or what you have named is an extraneous variable rather than a confounding one.*

4. **I. E (one mean, minutes per day) · II. T, 1 (one proportion; “fewer than”) · III. T, 2 (difference of two means; “differs”) · IV. E (difference of two proportions)**

The signal words do the work. “What is an estimate of...” and “what is the mean...” ask for an interval. “Is there convincing evidence that...” asks for a test, and the comparison word sets the direction: *fewer than* is one-sided, *differs from* is two-sided.

| *A difference of two proportions is still a single parameter. Name the parameter, with its units, every time.*

5. **a) 1. Whether a first-year teacher leaves the profession within five years (categorical), measured on first-year teachers in Florida. 2. Test a claim about one population proportion, one-sided. 3. Applies to all Florida first-year teachers; no cause and effect. b) i) missing all three components. ii) missing the third — it asks about the 50 batteries tested rather than about a population.**

b) i) “Do people like the new logo?” names no population (“people”), no measurable variable (what is “like?”), and no inference type. Rewrite: “What is an estimate of the proportion of all customers of this store who say they prefer the new logo to the old one?” ii) The batteries question names its variable (mean life, in hours) and its inference (test, one sided) perfectly well. What it gets wrong is the *population*: it asks about the 50 batteries already tested. Those are a sample — you do not need evidence about data in your hand, you just compute the mean. Rewrite: “...that the mean life of all batteries of this brand is greater than 900 hours?”

| *An investigative question is always about a population, never about the sample already in your hand.*

6. **I. T · II. F · III. F · IV. F**

I: true — an SRS does give every individual the same chance. II: false, and this is the trap. Equal chance for every *individual* is not the definition; an SRS requires every possible *sample* of size n to be equally likely. Line 20 students up and flip one coin to take all the odd positions or all the even ones: every individual has a 50% chance, and it is nowhere near an SRS. III: false for the same reason — a sample with 11 from one school is impossible, so not every sample of 30 can occur. IV: false; a census takes everybody, so it is not a sampling method at all.

| *The definition is about samples, not individuals. Every person having the same chance is not enough to make it an SRS.*

7. **a) Stratified · b) Systematic · c) Cluster · d) Simple random sample**

a) Fish differ a lot by depth but are alike within a depth zone — similar within, different between. Stratifying guarantees every depth is represented and gives a more precise estimate than an SRS of the same size. b) There is no list and no way to build one, but the loaves come past in order, so a random start plus a fixed interval spreads the sample evenly across the run. c) Each flight is a mini-version of the passenger population, and surveying 8 whole flights is far cheaper than chasing individuals scattered across 400. You only need a list of flights, not of passengers. d) A complete list exists, there is no subgroup structure worth exploiting and no cost reason to do anything else.

| *Naming the method is only half an answer. The justification has to point at a feature of this population.*

8. **a) Label 01–32; generate random integers 1–32, ignoring repeats, until 5 different numbers appear; dive those 5 plots. b) Stratified random sampling, with sheltered and exposed as the strata. The advantage is precision.**

a) All three steps are needed: distinct labels, the generator with repeats discarded, and the sentence naming which plots end up in the sample. b) Take an SRS from each side and combine. Two gains: both sides are guaranteed to appear, and because plots within a side resemble each other, repeated stratified samples vary less than repeated SRSs of the same size — a tighter estimate for the same five dives.

| *Ignoring repeats is the step that gets left out. Most random number generators sample with replacement, so without that clause your procedure can hand you the same plot twice.*

9. **Strata: similar within, different between. Clusters: each one a miniature of the whole population. Cluster sampling is wrong here because both of its reasons for existing are absent.**

Cluster sampling is chosen for *practical* reasons — the population is spread out, travel is expensive, and you have no list of individuals. This school has a complete list of all 1,400 students and travel costs nothing. Clustering would only add variability, because students inside a natural cluster (a homeroom, a class) tend to resemble each other, so you learn less per student sampled. An SRS or a stratified sample is better here.

| *Strata are alike inside; clusters are mixed inside. Stratifying buys precision, clustering buys convenience and costs precision.*

10. **Check the first and the third.**

1 is true: bias lives in the method, not in one sample — it is the systematic tendency to miss in the same direction every time. 2 is false: a bigger sample gives a more precise wrong answer; nothing about enlarging a flawed method removes the flaw. 3 is true: randomizing fixes *who* you ask, not *how* you ask — a leading question or a non-anonymous survey distorts a perfect SRS. 4 is false: nonresponse means you selected them and they did not answer; voluntary response means they invited themselves in.

| *Statement 2 is the one to think hardest about — question 12 is built on it.*

11. **a) Undercoverage — overestimate. b) Response bias — underestimate. c) Voluntary response and nonresponse.**

a) Members who did not renew had no chance of being in the sample at all, and they are exactly the ones who rarely came. The mean visit count of the survivors runs high. b) Raising a hand in front of a teacher and thirty classmates is not anonymous, so students who vape will not admit it — the proportion comes out too low. c) Two distinct failures: the 1,800 chose themselves by replying, and the other 18,200 were invited and did not answer.

| *Naming the bias without saying which way the estimate is off is half an answer.*

12. **a) Precision improves; accuracy is unchanged. b) Two errors: precision is not accuracy, and quadrupling the sample halves the standard error rather than quartering it. c) Change the method, not the size.**

a) More respondents means repeated samples cluster more tightly, so the estimate is more precise. But the panel is still voluntary, so it still misses in the same direction by the same amount — a more precise wrong answer. b) The colleague is confusing precision with accuracy, and has the arithmetic wrong too: quadrupling the sample size halves the standard error ($\sqrt{4} = 2$), it does not quarter it. c) Replace the voluntary panel with a random sample drawn from a real sampling frame of adults, and chase the non-respondents. Only the method can remove the bias.

| *Accuracy does not improve even slightly. 2,400 people who selected themselves are self-selected in exactly the way 600 were.*

13. **a) Two factors: light colour (4 levels), watering schedule (3 levels). b) $4 \times 3 = 12$ treatments, e.g. red light with the middle watering schedule. c) Units: the 120 basil seedlings. Response: grams of leaf harvested after six weeks.**

A treatment is one combination of one level of every factor, which is why the count is a product and not a sum. Twelve treatments across 120 seedlings leaves 10 seedlings per treatment.

| *Ask what chance actually assigned. If the light colour goes on a whole shelf and the water on a whole tray, the experimental unit is not the individual seedling.*

14. **a) Compared to their own March times — there is no control group. Confounded: two more months of training, and different track and weather conditions. b) Comparison, random assignment, replication and control — e.g. both shoe types run on the same day, with 9 runners drawn for each.**

a) Also accept: knowing they are being tested in new shoes, a different point in the competitive season, different pacing or competition. Any of these changed between March and May alongside the spikes. b) A redesign that earns all four: run both shoe types on the **same day** at the same track (control); write the 18 names on identical slips and draw 9 for the new spikes and 9 for their usual spikes (random assignment); 9 runners per group is the replication; and the usual-spike group is the comparison. Stronger still is matched pairs — every runner runs both, with the order decided by chance for each runner.

| *The March trial cannot serve as the control group, because everything else changed along with the shoes.*

15. **a) A dummy tablet with no active ingredient, identical in every other way; needed because migraine relief is self-reported and headaches end on their own. b) Single-blind: one side is in the dark. Double-blind: both are. c) Knowing the assignment, the nurse can record borderline cases as ending sooner for the real tablet.**

a) Without a placebo you cannot separate the drug from the expectation of being treated — the placebo effect is the difference between the average response to the dummy and the average response to no treatment at all. b) Single-blind: the volunteers do not know which tablet they got, or the researchers interacting with them do not — but not both. Double-blind: neither the volunteers nor the researchers who deal with them and measure the response know. c) “When did the headache end?” is a judgement call. A nurse who knows who got the real tablet may record borderline cases as resolving sooner, prompt those patients differently, or encourage them — bias built directly into the response measurement.

| *Blinding is not only about the patients. Whoever records the outcome can bias a study just as thoroughly as whoever receives the treatment.*

16. **a) Hold it constant. b) Block on it. c) Let random assignment handle it.**

a) Water is known to affect yield and you can physically give every plant the same amount. Controlling it removes it as an explanation and shrinks the variability in the response. b) Drainage is known to affect yield but cannot be changed. Make a north block and a south block and randomly assign both fertilisers *within each block*, so the drainage difference cannot be mistaken for a fertiliser difference. c) You cannot measure it, so you can neither control nor block on it. Randomization balances it out between the groups on average — which is the whole reason chance is used.

| *Randomization is not a formality. It is the only defence you have against variables nobody has thought of.*

17. **a) Block on each cow’s current milk yield. b) Rank the 30 cows by yield, form 10 blocks of 3, and randomly assign the three supplements within each block. c) Matched pairs; the order still has to be randomized.**

a) Baseline yield is the dominant source of variation in the response — a high-yield cow out-produces a low-yield cow whatever she is fed, and that noise would swamp any real supplement difference. Blocking takes it out of the comparison. b) Inside a block of three similar cows, put the three supplement names on identical slips, shuffle, and draw one per cow. Repeat separately in every block. c) Two treatments with blocks of size one — each cow is her own block — so chance decides which supplement she gets first, with a washout period between.

| *The randomization happens separately inside each block. And in matched pairs the order still has to be randomized — something is always left to chance.*

18. **a) 4 of the 100 — yes, statistically significant. b) Cause and effect, but only for people like these volunteers.**

The observed gap is $13/20 = 0.65$ against $7/20 = 0.35$, so 0.30. a) Three dots sit at 0.30 and one at 0.40, so 4 of the 100 re-shuffles produced a gap at least as large as the real one. Statistically significant means exactly that: a difference this big is too rare under “the method makes no difference” to be written off as chance variation in the random assignment. b) The students *were* randomly assigned, so handwriting caused the higher recall rate — but they were volunteers, not a random sample, so it cannot be extended to all students.

| *The dots sit 0.10 apart because the groups stay at 20 and 20: moving one success across changes the difference by 2/20. An odd number of successes would leave no dot at 0 at all.*

19. **a) All U.S. adults, because the 1,200 were a random sample of them. b) No — it is an observational study. c) Any variable with both links argued — shift work, for instance.**

a) Random selection is the mechanism that licenses generalizing from the sample to the population. b) Nobody was assigned to use or avoid screens; people chose for themselves, so the two groups may differ systematically in other ways and the association cannot be read as causation. c) Shift or late work, for instance: people working late are more likely to be on a screen in the hour before bed, and irregular late schedules produce poorer sleep on their own. Stress, age, caffeine and young children all work too.

| *Random selection but no random assignment: you get a claim about all U.S. adults, and association only.*