Comparing Two Treatments via Simulation

Name: ________________ Date: ____________ Score: ______ / 33

Quick Review

Suppose you run an experiment with two treatments and find a difference in outcomes. How do you tell whether the difference is *real* or just a fluke of which subjects landed in which group?

The null hypothesis H_0 says the two treatments have *the same* effect. Any observed difference between groups is just noise from the random assignment.

The simulation idea (randomization test, also called a permutation test). (1) Pool all the outcomes together, ignoring which treatment each subject got. (2) Randomly re-deal the subjects into two groups of the same sizes. (3) Compute the simulated difference in group means. (4) Repeat many times to build a **null distribution** – the distribution of differences we'd expect if H_0 were true.

The p-value. The fraction of simulated differences as extreme as (or more extreme than) what we actually observed. A small p-value means our observation would be unusual under H_0, providing evidence against it.

Reading the result. If $p \leq 0.05$ (a common threshold), we say the result is *statistically significant* and we have evidence against H_0. If p is larger, we don't have evidence to reject H_0 – but that's not the same as proving H_0 true.

Why simulate? The randomization test mirrors the actual random-assignment mechanism of the experiment, so it makes very few assumptions. It works even for small samples and odd-shaped data.

Common slips. Confusing the null distribution (what we'd see under H_0) with the data distribution. Treating a non-significant result as proof of no effect (it's just *absence of evidence*). Reporting "the p-value is the probability H_0 is true" (wrong – p is computed *assuming* H_0, not about it).

PRACTICE

Identify the null/alternative, the null distribution, and how to read the simulation result.

1. State the null hypothesis when comparing two treatments. ________
2. What does a simulation-based test produce? ________
3. What's the alternative hypothesis H_a? ________
4. In a randomization test, what does each "shuffle" do? ________
5. The p-value is the chance of a result as extreme as ours, assuming what? ________
6. Small p-value (e.g., 0.02) means ________
7. From the simulation summary below, estimate the p-value and decide whether it is significant at $\alpha = 0.05$. ________

Observed diff.	Sims $\geq$ obs.	Total sims
2.8	60	1000

8. True or false: a p-value of 0.30 proves H_0. ________
9. Why use a randomization test instead of a formula-based test? ________
10. Scenario A in the table records a randomization test. Estimate its p-value. ________

Scenario	Observed diff.	Sims $\geq$ obs.	Total sims
A	3.5	20	1000
B	1.2	400	1000

Scan Me

11. Scenario B in the table above records a second test. Estimate its p-value. ________

Scenario	Observed diff.	Sims $\geq$ obs.	Total sims
A	3.5	20	1000
B	1.2	400	1000

12. To get a more reliable p estimate, run ________

13. What does the null distribution *not* depend on? ________

14. True or false: random assignment is what justifies the simulation. ________

15. After rejecting H_0, we conclude ________

16. True or false: the p-value tells you the probability the null is true. ________

17. What does "statistically significant at $\alpha = 0.05$" mean? ________

18. A larger observed difference $\Rightarrow$ smaller or larger p? ________

19. If two treatments truly have no difference, what should the long-run sample mean of simulated p-values be? ________

20. Why does a randomization test work even for small samples? ________

♦ Word Problems

21. A teacher splits 20 students at random into two groups of 10. Group A uses an old textbook; Group B uses a new one. After a unit, the mean score in A is 74 and in B is 80, a difference of 6 points. She runs a simulation under H_0 (no difference) and finds that 35 out of 1000 simulated differences were 6 or more. What is the estimated p-value, and what should she conclude at $\alpha = 0.05$? ____

22. A pharmaceutical researcher tests a new drug against a placebo. The observed difference in recovery rates is 4 percentage points. A simulation under H_0 shows that 250 of 1000 shuffles produce a difference of 4 percentage points or more. What's the estimated p-value and the right conclusion at $\alpha = 0.05$? ____

23. A coach randomly assigns 12 runners to two training plans, 6 in each. Plan X mean time: 58 seconds; Plan Y mean time: 55 seconds. Difference: 3 seconds. A randomization test finds the p-value is 0.42. Explain in plain language what this p-value means. ____

24. A researcher reports: "Treatment B produced a 2-point higher mean than Treatment A ($p = 0.003$)." A reporter writes: "There is a 0.3% chance that Treatments A and B are identical." Why is the reporter's interpretation wrong? ____

Additional Practice

25. Find the mean of $4, 6, 8, 10$. ________

26. Find the median of $3, 9, 4, 10, 7$. ________

27. Find the range of $12, 5, 9, 20$. ________

28. Find the mode of $2, 3, 3, 5, 8$. ________

29. Find z for $x = 72$, mean 60, standard deviation 6. ________

30. Interpret $z = -1.5$. ________

31. Predicted y for $\hat{y} = 2x + 5$ at $x = 6$. ________

Scan Me

32. Residual if actual $= 20$ and predicted $= 17$. ________

33. Positive association: slope sign? ________

Answer Keys

1. the two treatments have the same effect
2. a null distribution of test statistics
3. the treatments have different effects
4. re-deals subjects into the two groups
5. H_0 is true
6. evidence against H_0
7. $p \approx 0.06;$ not significant
8. False
9. fewer assumptions; mirrors the actual assignment
10. 0.02
11. 0.40
12. more simulations
13. the actual treatment labels
14. True
15. the treatments likely differ
16. False
17. $p \le 0.05$
18. smaller
19. uniformly distributed on $[0, 1]$
20. it doesn't rely on the CLT
21. $p \approx 0.035;$ reject H_0
22. $p \approx 0.25;$ do not reject H_0
23. $p = 0.42;$ not statistically significant
24. p is not the probability that H_0 is true

Additional Practice Answers

25. 7
26. 7
27. 15
28. 3
29. 2
30. 1.5 SD below mean
31. 17
32. 3
33. positive

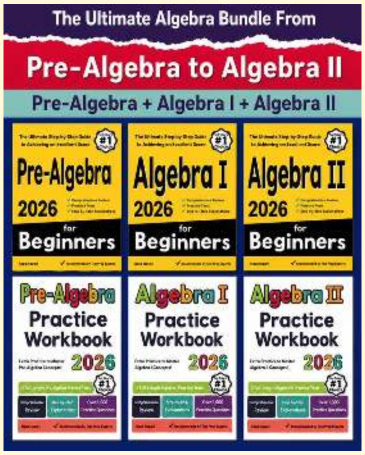
Additional Practice: Answers for all numbered items, including the added practice, are shown in the grid above.

Step-by-Step Explanations

1. A careful way to see it: H_0: no real treatment effect. Any observed difference is just from random assignment. That gives a quick check on the answer.
2. The simulated distribution of what the test statistic would look like if the null were true.
3. Some real treatment difference exists. Could be directional (greater than / less than) or two-sided (any difference).
4. Pool all outcomes, then re-assign them to treatment groups at random. Compute the resulting difference each time.
5. A careful way to see it: p is computed *under the null*. It's not the probability the null is true. That gives a quick check on the answer.
6. Our observation would be unusual if H_0 held. We interpret that as evidence the null is wrong.
7. One steady path is: $p \approx \frac{60}{1000} = 0.06$. Since $0.06 > 0.05$, it just misses the threshold – not statistically significant at $\alpha = 0.05$. That gives a quick check on the answer.
8. A large p just means we don't have evidence against H_0. Absence of evidence isn't evidence of absence.
9. Randomization tests work for small samples, weird distributions, and unusual statistics where the formula-based approach would be shaky.
10. Scenario A: $p \approx \frac{20}{1000} = 0.02$. Below 0.05, so we'd reject H_0 at the 5% level.
11. One steady path is: Scenario B: $p \approx \frac{400}{1000} = 0.40$. Way above 0.05 – no evidence against H_0. That gives a quick check on the answer.
12. More iterations smooth out the simulated p. $10{,}000$ or more is common for reliable estimates near the 0.05 boundary.
13. Once we pool the outcomes, the original labels are discarded. The null distribution comes from random reshufflings.
14. The simulation mimics the actual random-assignment mechanism. Without random assignment, the test's logic is shaky.
15. Statistically significant difference $\Rightarrow$ evidence of a real treatment effect. Doesn't tell us the size of the effect.
16. Start with the key idea: p is computed *assuming* H_0. It's $P(\text{data as extreme} \mid H_0)$, not $P(H_0 \mid \text{data})$. That gives a quick check on the answer.
17. Means the p-value is at or below the threshold α – evidence strong enough to reject H_0 by our chosen rule.
18. More extreme observations are less likely under H_0, so they get smaller p-values.
19. Under H_0, p is uniformly distributed on $[0, 1]$ – so any value is equally likely.
20. Randomization tests use the actual sampled values – no normality or large-n assumptions are needed.
21. A careful way to see it: $p \approx 35/1000 = 0.035$. That's below $\alpha = 0.05$, so she rejects H_0 and concludes there's statistically significant evidence that the new textbook produces higher mean scores than the old one. (Note: the practical significance – 6 points – is also worth thinking about. Statistical significance doesn't automatically mean practical importance.) That gives a quick check on the answer.
22. Keep the rule visible: $p \approx 250/1000 = 0.25$. That's far above $\alpha = 0.05$, so we fail to reject H_0 – there isn't statistically significant evidence that the new drug works better. Important caveat: this is *not* the same as proving the drug is ineffective. The sample might be too small to detect a real but modest effect. The next step would be a larger study. That gives a quick check on the answer.
23. Under the null hypothesis (the two plans are equally effective), 42% of random shuffles of the 12 runners would produce a difference at least as large as the observed 3 seconds. That's very common, so a 3-second gap is well within what chance can produce. The coach can't claim Plan Y is better based on this study – the effect isn't distinguishable from noise. A larger sample might tell a different story.
24. The p-value is computed *assuming* H_0 is true and measures how surprising the observed difference would be under that assumption. It is $P(\text{data as extreme} \mid H_0)$, *not* $P(H_0 \mid \text{data})$. The reporter is flipping the conditional. A correct rewrite: "If the two treatments were truly identical, a difference as large as the one observed would happen only 0.3% of the time – so the observed difference is strong evidence the treatments do differ."

Scan Me

Build Algebra Confidence From Pre-Algebra Through Algebra II



The Complete Algebra Success Bundle

Pre-Algebra, Algebra I, and Algebra II in one clear path

Friendly lessons, focused practice, and full-review support for every stage.

Scan Me

Scan for the Bundle

6 Books
3 Courses
1 Path

Bundle Value: Six coordinated books help students review missing skills, learn new algebra topics, and practice until the steps feel natural.

Complete Course Path

- ✓ Starts with Pre-Algebra foundations
- ✓ Moves smoothly into Algebra I skills
- ✓ Extends learning through Algebra II topics
- ✓ Great for review, tutoring, and summer study

One bundle, one steady path.

Step-by-Step Lessons

- ✓ Plain-English explanations students can follow
- ✓ Worked examples that show every important step
- ✓ Common mistakes called out before they stick
- ✓ Skill-building practice after each lesson
- ✓ Helpful for independent study or class support

Less guessing. More understanding.

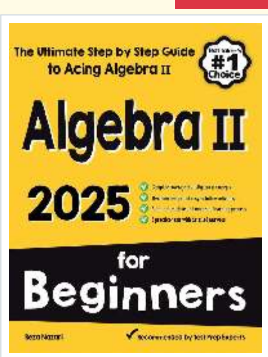
Practice That Sticks

- ✓ Matching practice workbooks for extra repetition
- ✓ Review sets to keep older skills fresh
- ✓ Answer explanations for checking thinking
- ✓ Strong support before tests and final exams
- ✓ Designed to build fluency and confidence

Practice today. Remember tomorrow.

STUDENT FAVORITE • Master Algebra II From the Ground Up

Algebra II for Beginners



✓ 100% Guaranteed
✓ Lifetime Support
✓ Trusted by Teachers

Start Your Algebra Journey Today! →

Written by a top math teacher & aligned with national and state Algebra II courses. From polynomial functions to logarithms, trigonometry, and rational expressions — explained the easy way.

- ✓ **Complete coverage** of every Algebra II concept — perfect companion to these worksheets
- ✓ **Step-by-step explanations** with worked examples on every topic
- ✓ **QR codes in every chapter** for free video lessons & bonus practice
- ✓ **2 full-length practice tests** with detailed answer keys

⋆ STUDENT'S #1 CHOICE ⋆

Teacher-recommended • 12,000+ Happy Students

↓ PDF EDITION

Scan Me

Instant download • any device

PAPERBACK

Scan Me

Paperback on Amazon

Hold it in your hands

Pair these free worksheets with *Algebra II for Beginners* and you have a complete self-paced course — concept lessons, daily practice, and full exam-style reviews, all in one path. → EffortlessMath.com/product/algebra-ii-for-beginners