

STAT — Computing in Statistics

Permutation Test Quiz

Yen-Yi Ho

University of South Carolina

2026-09-16

Section 1

The Questions

Q1. What Does H_0 Actually Assume?

Under the null hypothesis, a two-sample permutation test treats the data as if:

- A. Both groups were drawn from normal distributions with the same variance
- B. The two group means are equal, but the shapes may differ freely
- C. The group **labels are arbitrary** — every way of assigning them to the observed values is equally likely
- D. Each observation was drawn independently from its own group's distribution

Q1. What Does H_0 Actually Assume?

Under the null hypothesis, a two-sample permutation test treats the data as if:

- A. Both groups were drawn from normal distributions with the same variance
- B. The two group means are equal, but the shapes may differ freely
- C. The group **labels are arbitrary** — every way of assigning them to the observed values is equally likely
- D. Each observation was drawn independently from its own group's distribution

Answer: C — exchangeability

This is the whole engine of the test. If the labels carry no information, the arrangement we observed is just one of many equally likely arrangements — so we generate the others by shuffling and see how unusual ours looks. Normality and equal variances are assumptions of the t -test, not of this one.

Q2. Why Is the Null Distribution Centred at Zero?

Why is the permutation null distribution of a difference in means centred near zero?

- A. Because shuffling the labels destroys any systematic difference, so two arbitrarily labelled groups differ by nothing on average
- B. Because the observed difference is subtracted from every replicate
- C. Because the Central Limit Theorem forces any difference to be centred at zero
- D. Because the statistic is divided by its standard error before plotting

Q2. Why Is the Null Distribution Centred at Zero?

Why is the permutation null distribution of a difference in means centred near zero?

- A. Because shuffling the labels destroys any systematic difference, so two arbitrarily labelled groups differ by nothing on average
- B. Because the observed difference is subtracted from every replicate
- C. Because the Central Limit Theorem forces any difference to be centred at zero
- D. Because the statistic is divided by its standard error before plotting

Answer: A

Each replicate compares two groups formed *at random* from the same pooled values, so neither has any reason to be systematically larger. The arithmetic does not impose zero — it emerges because of the random shuffling.

Q3. Reading the p -value

A permutation test returns $p = 0.03$. The correct reading is:

- A. There is a 3% probability that the null hypothesis is true
- B. The two groups differ by 3%
- C. If the study were repeated, the result would replicate 97% of the time
- D. About 3% of random relabellings produced a difference **at least as extreme** as the one observed

Q3. Reading the p -value

A permutation test returns $p = 0.03$. The correct reading is:

- A. There is a 3% probability that the null hypothesis is true
- B. The two groups differ by 3%
- C. If the study were repeated, the result would replicate 97% of the time
- D. About 3% of random relabellings produced a difference **at least as extreme** as the one observed

Answer: D

A p -value describes the data under an assumed null — not the probability that the null is true, which would require a prior. The 0.03 is the **proportion of relabellings whose test statistic value was at least as extreme as the observed value** — something we can point at in the tails of the histogram and count.

Q4. Assumption-Light, Not Assumption-Free

Twenty students had their pulse measured **before** and **after** exercise. Someone pools all 40 values and shuffles the before/after labels freely across them. Why is that wrong?

- A. The measurements are **paired**: two readings from the same student are dependent, so free shuffling breaks the exchangeability the test assumes
- B. The sample size is too small for a permutation test to be valid
- C. Pulse measurements are not normally distributed
- D. “Before” and “after” are ordered categories and cannot be permuted

Q4. Assumption-Light, Not Assumption-Free

Twenty students had their pulse measured **before** and **after** exercise. Someone pools all 40 values and shuffles the before/after labels freely across them. Why is that wrong?

- A. The measurements are **paired**: two readings from the same student are dependent, so free shuffling breaks the exchangeability the test assumes
- B. The sample size is too small for a permutation test to be valid
- C. Pulse measurements are not normally distributed
- D. “Before” and “after” are ordered categories and cannot be permuted

Answer: A

The shuffling must match the study design. Only the two labels *within* each student are exchangeable, so the correct test flips before/after independently per student — 2^{20} relabellings, not $\binom{40}{20}$. Free shuffling ignores the pairing, inflates the null spread, and costs power.

Q5. So What *Is* the Right Test?

Same 20 students, pulse measured before and after exercise. Which procedure gives a **valid** permutation test of “exercise has no effect”?

- A. Pool all 40 values and shuffle the labels, keeping 20 in each group
- B. For each student, randomly flip the before/after label. Then calculate the difference.
- C. Bootstrap all 40 values with replacement and take the middle 95%
- D. Randomly split the 20 students into two groups of 10 and compare their pulses

Q5. So What *Is* the Right Test?

Same 20 students, pulse measured before and after exercise. Which procedure gives a **valid** permutation test of “exercise has no effect”?

- A. Pool all 40 values and shuffle the labels, keeping 20 in each group
- B. For each student, randomly flip the before/after label. Then calculate the difference.
- C. Bootstrap all 40 values with replacement and take the middle 95%
- D. Randomly split the 20 students into two groups of 10 and compare their pulses

Answer: B — a sign-flip test

If exercise does nothing, each d_i is equally likely to have come out $+$ or $-$, so flipping signs is exactly the relabelling the *design* permits — and it is the same thing as swapping before/after *within* a student. That gives $2^{20} \approx 1.05$ million possible outcomes. The analysis collapses to **one number per student**, which is what paired designs buy you: between-student variation cancels.