

Permutation Tests Explained

Click through a few fixed-seed label shuffles below to see exactly how a permutation test builds its comparison distribution from your own data rather than from a formula.



Permutation Tests Explained

Exported from statreason.com. Deterministic, browser-local content; not professional statistical or research advice.

Fixed data: Group A [5, 7, 6], Group B [9, 10, 8]. Observed difference in means = 3.00.

Shuffles drawn	0
Latest shuffled grouping	—
Latest shuffled difference in means	—
Shuffles at least as extreme as observed	—

Run a full permutation test on your own data: [Permutation Test Calculator](#)

What this answers

This page answers "what is a permutation test actually doing when it computes a p-value?" A permutation test builds its comparison distribution directly from your own data by repeatedly reshuffling group labels, rather than assuming a theoretical distribution like the t-distribution.

Exchangeability and shuffled labels

The logic of a permutation test relies on exchangeability: if the null hypothesis of no group difference is true, the group labels attached to each observation carry no real information, so reshuffling them randomly should produce differences no more extreme, on average, than what was actually observed. Each shuffle randomly reassigns which observations belong to which group, keeping the total pool of values fixed, and recalculates the statistic of interest for that shuffled grouping.

Building the null distribution and the p-value

Repeating this shuffle many times, often all possible shuffles for a small sample or a large random subset for a bigger one, builds an empirical null distribution of the statistic under the assumption of no real group difference. The permutation p-value is the proportion of shuffles that produce a statistic at least as extreme as the one actually observed in your real, unshuffled data, typically with a plus-one correction in the numerator and denominator to avoid a p-value of exactly zero.

Worked example

Click "Shuffle labels again" a few times above. Each shuffle randomly reassigns the six fixed values between two groups of three and reports the new difference in means. As you draw more shuffles, you can see directly how often a shuffled difference matches or exceeds the actual observed difference of 3.00, which is exactly what a permutation p-value counts.

Assumption audit

Calculated from your data: the observed statistic and, in the full calculator, thousands of shuffled statistics using a seeded generator, from which the exact permutation p-value is computed.

Evidence to review: whether your groups are exchangeable under the null model, meaning nothing besides random assignment should systematically distinguish them if there were truly no effect.

You must verify: that your data were collected in a way where label shuffling is a meaningful null model, which generally requires the original group assignment to have been random or exchangeable in the first place.

Source

This explanation follows the standard permutation-test methodology described in the NIST/SEMATECH e-Handbook of Statistical Methods and the shared statistical reasoning contract every StatReason engine is built against.

Limitations

This page uses a tiny six-value example with only a few manual shuffles to illustrate the mechanism; a real permutation test requires many more shuffles or the full enumeration of possibilities, which the linked calculator performs automatically.