

STAT540 — Computing in Statistics

Permutation Tests for Two-Sample Comparisons

Yen-Yi Ho

University of South Carolina

Section 1

Motivation

From Estimation to Testing

Resampling methods answer inference questions by re-using the observed data. Here the question is a **comparison** between two groups:

The two-sample question

Do physically active adults have a different **mean BMI** than inactive adults?

$$H_0 : \mu_{\text{Yes}} = \mu_{\text{No}} \quad \text{vs} \quad H_1 : \mu_{\text{Yes}} \neq \mu_{\text{No}}$$

The classical answer is the two-sample t -test, which needs:

- approximate **normality** of each group (or a large n leaning on the CLT), and
- a variance assumption (equal variances, or Welch's correction)

The permutation alternative

If the two groups really are the same, the group **labels** are arbitrary. So shuffle them, recompute the difference, and see how unusual the *observed* difference looks. No normality assumption, no variance assumption, no reference table.

Section 2

The Permutation Idea

The Null Hypothesis as Exchangeability

Write the pooled data as $n = n_1 + n_2$ observations with labels $g_1, \dots, g_n \in \{\text{Yes}, \text{No}\}$.

Exchangeability under H_0

If H_0 is true — the two groups come from the **same distribution** — then every way of assigning n_1 of the n observations to group 1 is **equally likely**. The labels carry no information.

This gives us the **exact null distribution** of any test statistic T :

$$\Pr(T \geq t \mid H_0) = \frac{\#\{\text{label arrangements with } T \geq t\}}{\binom{n}{n_1}}$$

- The null distribution is **built from the data**, not assumed
- The randomness comes from the **label assignment**, not from a population model
- Valid for *any* statistic: difference in means, in medians, in trimmed means, t , ...

The Permutation Algorithm

Given: values $\mathbf{x} = (x_1, \dots, x_n)$ and labels $\mathbf{g} = (g_1, \dots, g_n)$

① Compute the **observed statistic** $t = \frac{\bar{x}_{\text{group 1}} - \bar{x}_{\text{group 2}}}{\sqrt{\frac{s_{\text{group 1}}^2}{n_1} + \frac{s_{\text{group 2}}^2}{n_1}}}$

② **Repeat** M times ($b = 1, \dots, M$):

- ▶ **Shuffle** the labels: $\mathbf{g}^{*b}$ = a random permutation of $\mathbf{g}$ (*without replacement — the values stay put*)
- ▶ Recompute the statistic with the shuffled labels: t^{*b}

③ $\{t^{*1}, \dots, t^{*M}\}$ approximates the **null distribution** of t

④ Two-sided p -value: $p = \frac{\#\{|t^{*b}| \geq |t|\}}{M}$

Why Shuffling Works: A Tiny Example

Six observations, $n_1 = n_2 = 3$. Group A: 8, 6, 7
 $\hat{d} = 7 - 4 = 3$.

There are $\binom{6}{3} = 20$ ways to split the pooled values $\{3, 4, 5, 6, 7, 8\}$ into two groups of three:

Group A	Group B	d^*
6, 7, 8	3, 4, 5	+3.0
5, 7, 8	3, 4, 6	+2.3
5, 6, 8	3, 4, 7	+1.7
$\vdots$	$\vdots$	$\vdots$
3, 4, 6	5, 7, 8	-2.3
3, 4, 5	6, 7, 8	-3.0

Group B: 3, 5, 4. Observed difference

Only **2** of the 20 splits give $|d^*| \geq 3$ (the most extreme split and its mirror image), so the exact two-sided p -value is

$$p = \frac{2}{20} = 0.10.$$

Exact vs Monte Carlo

$\binom{n}{n_1}$ explodes: at $n_1 = n_2 = 20$ there are already 1.4×10^{11} splits. So we **sample** M random shuffles instead — a Monte Carlo approximation to the exact test.

Summary

Conceptual

- Permutation encodes H_0 as **exchangeable labels**
- The null distribution is **built from the data**
- p -value: proportion of shuffles at least as extreme
- No normality or variance assumption
- Valid for *any* statistic, standard or not

Programming

- Wrap logic in a **function** for reusability
- **Pre-allocate** storage before the loop
- Shuffle labels with `sample(g)` — *no replace* = TRUE
- Define the statistic **once** as a helper

When to reach for a permutation test

Small samples; skewed or heavy-tailed data; unequal group variances; or a statistic whose null distribution has no tractable form — a difference in medians, a trimmed mean, a ratio.