Permutation & Non-Parametric Tests

Statistical tests

- Gather data to assess some hypothesis (e.g., does this treatment have an effect on this outcome?)
- Form a test statistic for which large values indicate a departure from the hypothesis.
- Compare the observed value of the statistic to its distribution under the null hypothesis.

Paired t-test

Pairs $(X_1, Y_1), \ldots, (X_n, Y_n)$ independent.

$X_i \sim \text{Normal}(\mu_A, \sigma_A)$ $\qquad$ $Y_i \sim \text{Normal}(\mu_B, \sigma_B)$

Test $H_0 : \mu_A = \mu_B$ vs $H_a : \mu_A \neq \mu_B$

Paired t-test: $D_i = Y_i - X_i$

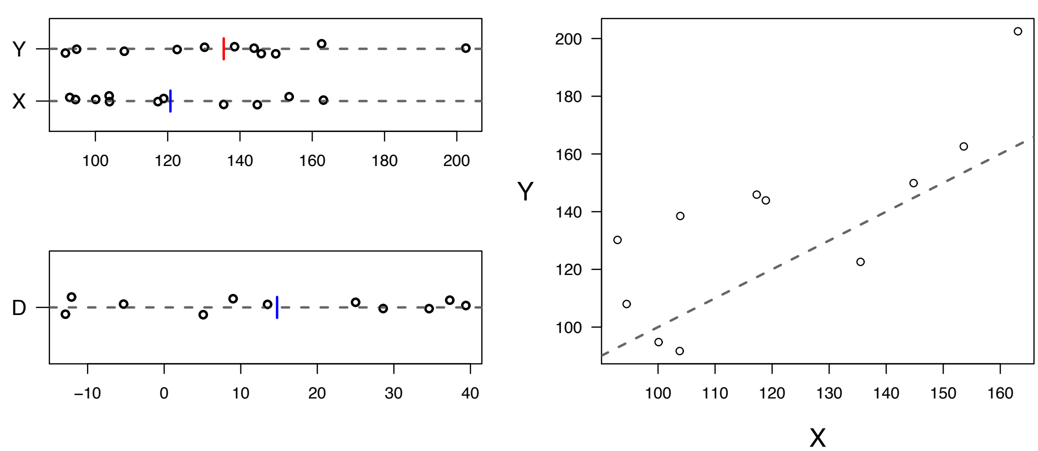
$\longrightarrow$ $D_1, \ldots, D_n \sim$ iid Normal$(\mu_B - \mu_A, \sigma_D)$

Sample mean $\bar{D}$; sample SD s_D

$\longrightarrow$ $T = \bar{D}/(s_D/\sqrt{n})$

Compare to a t distribution with n – 1 d.f.

Example



$\bar{d} = 14.7$ $\qquad$ $s_D = 19.6$ $\qquad$ $n = 11$

$T = 2.50$ $\qquad$ P = `2*(1-pt(2.50,10))` = 0.031

Wilcoxon signed rank test

A "nonparametric" test.

Rank the differences according to their absolute values.

R = sum of ranks of positive (or negative) values.

D	28.6	–5.3	13.5	–12.9	37.3	25.0	5.1	34.6	–12.1	9.0	39.4
rank	8	2	6	5	10	7	1	9	4	3	11

R = 2 + 4 + 5 = 11

Compare this to the distribution of R when each rank has an equal chance of being positive or negative.

In R: `wilcox.test(d)` $\longrightarrow$ P = 0.054

Permutation test

$(X_1, Y_1), \ldots, (X_n, Y_n) \longrightarrow T_{obs}$

- Randomly flip the pairs. (For each pair, toss a fair coin. If heads, switch X and Y; if tails, do not switch.)
- Compare the observed T statistic to the distribution of the T-statistic when the pairs are flipped at random.
- If the observed statistic is extreme relative to this permutation/randomization distribution, then reject the null hypothesis (that the X's and Y's have the same distribution).

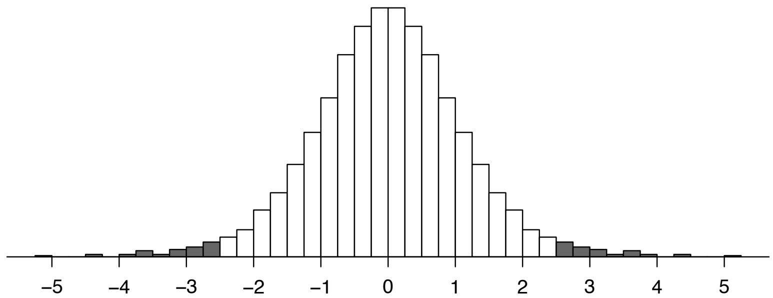
Actual data:

(117.3,145.9) (100.1,94.8) (94.5,108.0) (135.5,122.6) (92.9,130.2) (118.9,143.9)
(144.8,149.9) (103.9,138.5) (103.8,91.7) (153.6,162.6) (163.1,202.5) $\longrightarrow T_{obs} = 2.50$

Example shuffled data:

(117.3,145.9) (94.8,100.1) (108.0,94.5) (135.5,122.6) (130.2,92.9) (118.9,143.9)
(144.8,149.9) (138.5,103.9) (103.8,91.7) (162.6,153.6) (163.1,202.5) $\longrightarrow T^\star = 0.19$

Permutation distribution



P-value = $\Pr(|T^\star| \geq |T_{obs}|)$

$\longrightarrow$ Small n: Look at all 2^n possible flips.

$\longrightarrow$ Large n: Look at a sample (w/ repl) of 1000 such flips.

Example data:

All 2^{11} permutations: P = 0.037; sample of 1000: P = 0.040.

Paired comparisons

At least three choices:

- Paired t-test.
- Signed rank test.
- Permutation test with the t-statistic.

Which to use?

- Paired t-test depends on the normality assumption.
- Signed rank test ignores some information.
- Permutation test is recommended.

The fact that the permutation distribution of the t-statistic is generally well-approximated by a t distribution recommends the ordinary t-test. But if you can estimate the permutation distribution, do it.

2-sample t-test

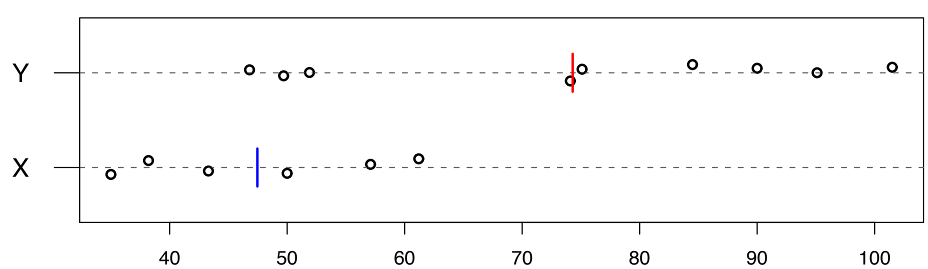
$X_1, \ldots, X_n$ iid Normal(μ_A, σ) $Y_1, \ldots, Y_m$ iid Normal(μ_B, σ)

Test $H_0 : \mu_A = \mu_B$ vs $H_a : \mu_A \neq \mu_B$

Test statistic: $T = \dfrac{\overline{X} - \overline{Y}}{s_p \sqrt{\frac{1}{n} + \frac{1}{m}}}$ where $s_p = \sqrt{\frac{s_A^2(n-1)+s_B^2(m-1)}{n+m-2}}$

$\longrightarrow$ Compare to the t distribution with n + m – 2 d.f.

Example



$\bar{x} = 47.5$ $s_A = 10.5$ $n = 6$

$\bar{y} = 74.3$ $s_B = 20.6$ $m = 9$

$s_p = 17.4$ $T = -2.93$

$\longrightarrow$ P = `2*pt(-2.93,6+9-2)` = 0.011.

Wilcoxon rank-sum test

Rank the X's and Y's from smallest to largest (1, 2, ..., n+m)

R = sum of ranks for X's.

(Also known as the Mann-Whitney Test)

X	Y	rank
35.0		1
38.2		2
43.3		3
	46.8	4
	49.7	5
50.0		6
	51.9	7
57.1		8
61.2		9
	74.1	10
	75.1	11
	84.5	12
	90.0	13
	95.1	14
	101.5	15

R = 1 + 2 + 3 + 6 + 8 + 9 = 29

P-value = 0.026

$\longrightarrow$ use `wilcox.test()`

Note: The distribution of R (given that X's and Y's have the same dist'n) is calculated numerically.

Permutation test

X or Y	group
X_1	1
X_2	1
$\vdots$	1
X_n	1
Y_1	2
Y_2	2
$\vdots$	2
Y_m	2

$\rightarrow T_{obs}$

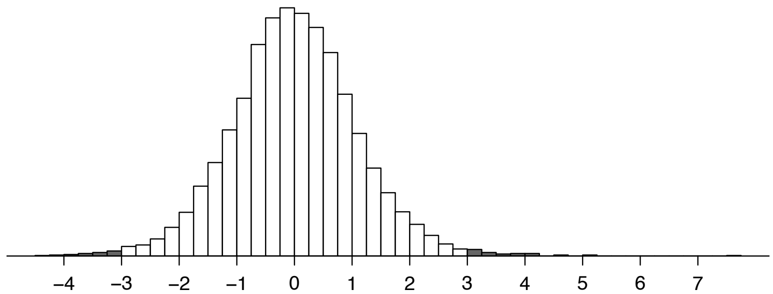
X or Y	group
X_1	2
X_2	2
$\vdots$	1
X_n	2
Y_1	1
Y_2	2
$\vdots$	1
Y_m	1

$\rightarrow T^{\star}$

Group status shuffled

Compare the observed t-statistic to the distribution obtained by randomly shuffling the group status of the measurements.

Permutation distribution



P-value = $\Pr(|T^\star| \geq |T_{obs}|)$

⟶ Small n & m: Look at all $\binom{n+m}{n}$ possible shuffles.

⟶ Large n & m: Look at a sample (w/repl) of 1000 such shuffles.

Example data:

All 5005 permutations: P = 0.015; sample of 1000: P = 0.013.

Estimating the permutation P-value

Let P be the true P-value (if we do *all possible* shuffles).

Do N shuffles, and let X be the number of times the statistic after shuffling is bigger or equal to the observed statistic.

⟶ $\hat{P} = \frac{X}{N}$ where $X \sim \text{Binomial}(N, P)$

⟶ $E(\hat{P}) = P$ $SD(\hat{P}) = \sqrt{\frac{P(1-P)}{N}}$

If the “true” P-value was P = 5%, and we do N=1000 shuffles:

$SD(\hat{P})$ = 0.7%.

Summary

The t-test relies on a normality assumption.

If this is a worry, consider:

- Paired data:
  - Signed rank test.
  - Permutation test.
- Unpaired data:
  - Rank-sum test.
  - Permutation test.

$\longrightarrow$ The crucial assumption is independence!

The fact that the permutation distribution of the t-statistic is often closely approximated by a t distribution is good support for just doing t-tests.