

This work is licensed under a [Creative Commons Attribution-NonCommercial-ShareAlike License](https://creativecommons.org/licenses/by-nc-sa/4.0/). Your use of this material constitutes acceptance of that license and the conditions of use of materials on this site.



Copyright 2006, The Johns Hopkins University and Karl Broman. All rights reserved. Use of these materials permitted only in accordance with license rights granted. Materials provided "AS IS"; no representations or warranties provided. User assumes all responsibility for use, and all liability related thereto, and must independently review all materials for accuracy and efficacy. May contain materials owned by others. User is responsible for obtaining permissions for use from third parties as needed.

# 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  $(X_1, Y_1), \dots, (X_n, Y_n)$  independent

$$X_i \sim \text{normal}(\mu_A, \sigma_A) \quad Y_i \sim \text{normal}(\mu_B, \sigma_B)$$

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

Paired t-test

$$D_i = Y_i - X_i$$

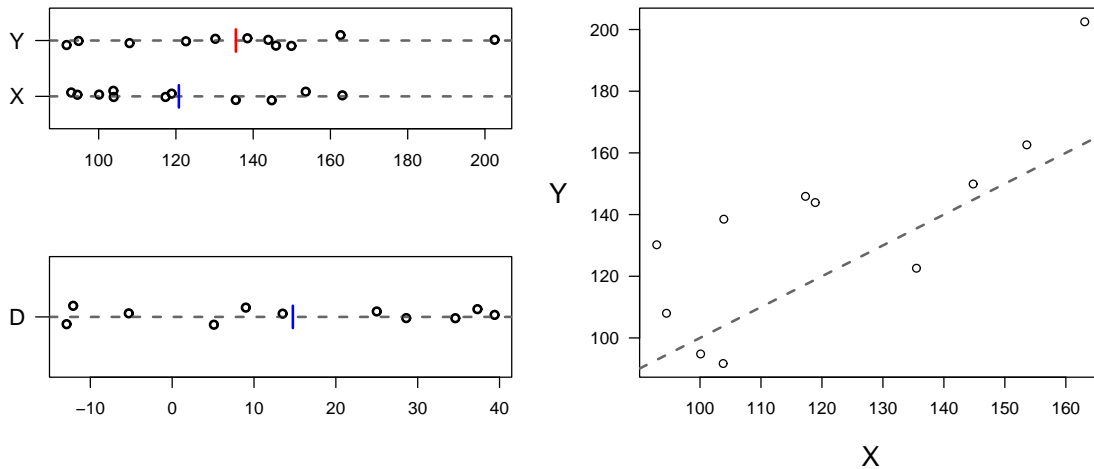
$$D_1, \dots, D_n \sim \text{iid normal}(\mu_B - \mu_A, \sigma_D)$$

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

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

Compare to t distribution with  $n - 1$  d.f.

# Example



$$\bar{D} = 14.7 \quad s_D = 19.6 \quad n = 11$$

$$T = 2.50 \quad P = 2 * (1 - \text{pt}(2.50, 10)) = 0.031$$

# Sign test

Suppose we are concerned about the normal assumption.

$(X_1, Y_1), \dots, (X_n, Y_n)$  independent

Test  $H_0$  : X's and Y's have the same distribution

Another statistic:  $S = \#\{i : X_i < Y_i\} = \#\{i : D_i > 0\}$

(the number of pairs for which  $X_i < Y_i$ )

Under  $H_0$ ,  $S \sim \text{binomial}(n, p=0.5)$

Suppose  $S_{\text{obs}} > n/2$ .

$$\text{P-value} = 2 \times \Pr(S \geq S_{\text{obs}} \mid H_0)$$

$$= 2 * (1 - \text{pbinom}(S_{\text{obs}} - 1, n, 0.5))$$

# Example

---

For our example, 8 out of 11 pairs had  $Y_i > X_i$ .

$$\text{P-value} = 2 * (1 - \text{pbinom}(7, 11, 0.5)) = 23\%$$

(Compare this to  $P = 3\%$  for the t-test.)

## Signed Rank test

---

Another “nonparametric” test. (This one is also called the Wilcoxon signed rank 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)` →  $P = 0.054$

# Permutation test

---

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

- 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:

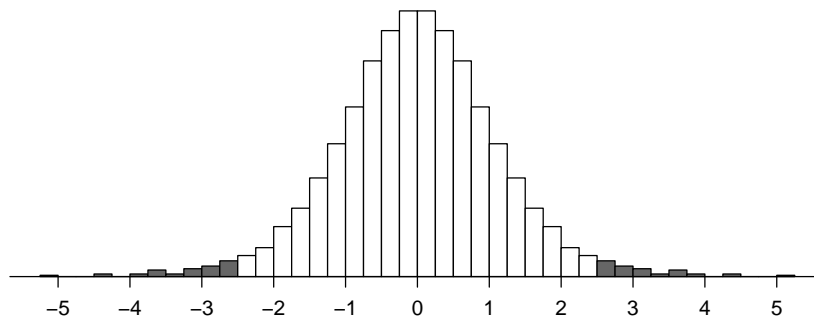
(117.3, 145.9) (100.1, 94.8) (94.5, 108.0) (135.5, 122.6) (92.9, 130.2) (118.9, 143.9)  
(144.8, 149.9) (103.9, 138.5) (103.8, 91.7) (153.6, 162.6) (163.1, 202.5)  $\longrightarrow T_{\text{obs}} = 2.50$

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)  $\longrightarrow T^* = 0.19$

## Permutation distribution

---



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

Small n: Look at all  $2^n$  possible flips

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 four choices:

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

Which to use?:

- Paired t-test depends on the normality assumption
- Sign test is pretty weak
- 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

---

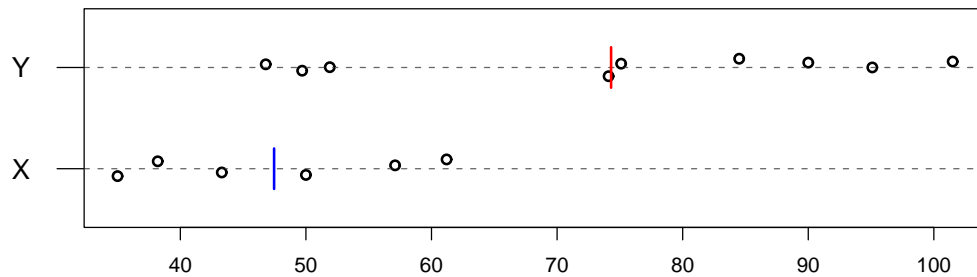
$X_1, \dots, X_n \sim \text{iid normal}(\mu_A, \sigma)$        $Y_1, \dots, Y_m \sim \text{iid normal}(\mu_B, \sigma)$

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

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

Compare to t distribution with  $n + m - 2$  degrees of freedom.

# Example



$$\bar{X} = 47.5 \quad s_A = 10.5 \quad n = 6$$

$$\bar{Y} = 74.3 \quad s_B = 20.6 \quad m = 9$$

$$s_p = 17.4 \quad T = -2.93$$

$$P = 2 * pt(-2.93, 6+9-2) = 0.011$$

## 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   |

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

$$P\text{-value} = 0.026$$

(use `wilcox.test()`)

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

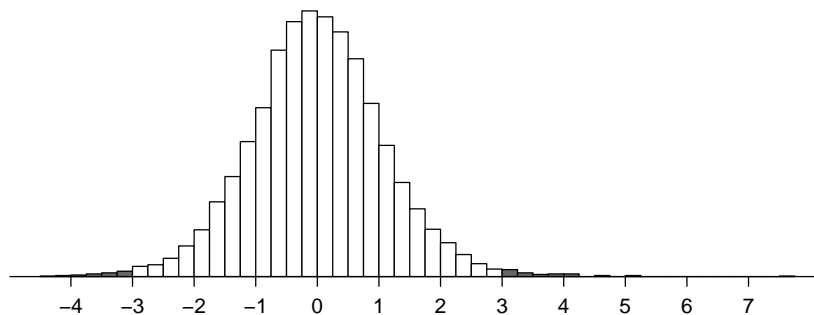
# Permutation test

| X or Y   | group |                              | X or Y   | group |                   |
|----------|-------|------------------------------|----------|-------|-------------------|
| $X_1$    | 1     |                              | $X_1$    | 2     |                   |
| $X_2$    | 1     |                              | $X_2$    | 2     |                   |
| $\vdots$ | 1     |                              | $\vdots$ | 1     |                   |
| $X_n$    | 1     | $\rightarrow T_{\text{obs}}$ | $X_n$    | 2     | $\rightarrow T^*$ |
| $Y_1$    | 2     |                              | $Y_1$    | 1     |                   |
| $Y_2$    | 2     |                              | $Y_2$    | 2     |                   |
| $\vdots$ | 2     |                              | $\vdots$ | 1     |                   |
| $Y_m$    | 2     |                              | $Y_m$    | 1     |                   |

Group status shuffled

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

## Permutation distribution



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

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  $P$  = true P-value (if we do all possible shuffles)

Do  $N$  shuffles, and let  $X$  = # times the statistic after shuffling  $\geq$  the observed statistic

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

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

If the “true” P-value  $P = 5\%$  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:
  - Sign test
  - Signed rank test
  - Permutation test
- Unpaired data:
  - Rank-sum test
  - Permutation test

**Crucial assumption: 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.