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.

\[ X_i \sim \text{Normal}(\mu_A, \sigma_A) \quad Y_i \sim \text{Normal}(\mu_B, \sigma_B) \]

Test \(H_0 : \mu_A = \mu_B\) vs \(H_a : \mu_A \neq \mu_B\)

Paired t-test: \(D_i = Y_i - X_i\)
\[ \rightarrow D_1, \ldots, D_n \sim \text{iid Normal}(\mu_B - \mu_A, \sigma_D) \]

Sample mean \(\bar{D}\); sample SD \(s_D\)
\[ \rightarrow T = \frac{\bar{D}}{s_D/\sqrt{n}} \]

Compare to a t distribution with \(n - 1\) d.f.

Example

\(
\begin{array}{c}
\begin{array}{c}
X \\
Y
\end{array}
\end{array}
\)

\(
\begin{array}{c}
\begin{array}{c}
D \\
X
\end{array}
\end{array}
\)

\[ \bar{d} = 14.7 \quad s_D = 19.6 \quad n = 11 \]

\[ T = 2.50 \quad P = 2 \times (1 - pt(2.50, 10)) = 0.031 \]
**Wilcoxon signed rank test**

A “nonparametric” test.

Rank the differences according to their absolute values.

\[ R = \text{sum of ranks of positive (or negative) values.} \]

<table>
<thead>
<tr>
<th>D</th>
<th>28.6</th>
<th>−5.3</th>
<th>13.5</th>
<th>−12.9</th>
<th>37.3</th>
<th>25.0</th>
<th>5.1</th>
<th>34.6</th>
<th>−12.1</th>
<th>9.0</th>
<th>39.4</th>
</tr>
</thead>
<tbody>
<tr>
<td>rank</td>
<td>8</td>
<td>2</td>
<td>6</td>
<td>5</td>
<td>10</td>
<td>7</td>
<td>1</td>
<td>9</td>
<td>4</td>
<td>3</td>
<td>11</td>
</tr>
</tbody>
</table>

\[ R = 2 + 4 + 5 = 11 \]

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

In R: \[ \text{wilcox.test(d) } \rightarrow \text{ P = 0.054} \]

---

**Permutation test**

\((X_1, Y_1), \ldots, (X_n, Y_n) \rightarrow T_{\text{obs}}\)

- 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:**

\[ (117.3, 145.9) (100.1, 94.8) (94.5, 108.0) (135.5, 122.6) (92.9, 130.2) (118.9, 143.9) \]

\[ (144.8, 149.9) (103.9, 138.5) (103.8, 91.7) (153.6, 162.6) (163.1, 202.5) \rightarrow T_{\text{obs}} = 2.50 \]

**Example shuffled data:**

\[ (117.3, 145.9) (94.8, 100.1) (108.0, 94.5) (135.5, 122.6) (130.2, 92.9) (118.9, 143.9) \]

\[ (144.8, 149.9) (138.5, 103.9) (103.8, 91.7) (162.6, 153.6) (163.1, 202.5) \rightarrow T^* = 0.19 \]
**Permutation distribution**

P-value = Pr(|T^*| ≥ |T_{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 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(\( \mu_A, \sigma \)) \hspace{1cm} \( Y_1, \ldots, Y_m \) iid Normal(\( \mu_B, \sigma \))

Test \( H_0 : \mu_A = \mu_B \) vs \( H_a : \mu_A \neq \mu_B \)

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

\( \rightarrow \) Compare to the t distribution with \( n + m - 2 \) d.f.

Example

\[
\begin{align*}
\bar{x} &= 47.5 \hspace{1cm} s_A = 10.5 \hspace{1cm} n = 6 \\
\bar{y} &= 74.3 \hspace{1cm} s_B = 20.6 \hspace{1cm} m = 9 \\
s_p &= 17.4 \hspace{1cm} T = -2.93
\end{align*}
\]

\( \rightarrow \) \( P = 2 \times \text{pt} (-2.93, 6+9-2) = 0.011. \)
## 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)

<table>
<thead>
<tr>
<th>X</th>
<th>Y</th>
<th>rank</th>
</tr>
</thead>
<tbody>
<tr>
<td>35.0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>38.2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>43.3</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>46.8</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>49.7</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>50.0</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td>51.9</td>
<td>7</td>
<td></td>
</tr>
<tr>
<td>57.1</td>
<td>8</td>
<td></td>
</tr>
<tr>
<td>61.2</td>
<td>9</td>
<td></td>
</tr>
<tr>
<td>74.1</td>
<td>10</td>
<td></td>
</tr>
<tr>
<td>75.1</td>
<td>11</td>
<td></td>
</tr>
<tr>
<td>84.5</td>
<td>12</td>
<td></td>
</tr>
<tr>
<td>90.0</td>
<td>13</td>
<td></td>
</tr>
<tr>
<td>95.1</td>
<td>14</td>
<td></td>
</tr>
<tr>
<td>101.5</td>
<td>15</td>
<td></td>
</tr>
</tbody>
</table>

\[ R = 1 + 2 + 3 + 6 + 8 + 9 = 29 \]

P-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
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>( X_1 )</td>
<td>1</td>
</tr>
<tr>
<td>( X_2 )</td>
<td>1</td>
</tr>
<tr>
<td>( \vdots )</td>
<td>1</td>
</tr>
<tr>
<td>( X_n )</td>
<td>1</td>
</tr>
<tr>
<td>( Y_1 )</td>
<td>2</td>
</tr>
<tr>
<td>( Y_2 )</td>
<td>2</td>
</tr>
<tr>
<td>( \vdots )</td>
<td>2</td>
</tr>
<tr>
<td>( Y_m )</td>
<td>2</td>
</tr>
</tbody>
</table>

\( \rightarrow T_{obs} \)

X or Y | group
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>( X_1 )</td>
<td>2</td>
</tr>
<tr>
<td>( X_2 )</td>
<td>2</td>
</tr>
<tr>
<td>( \vdots )</td>
<td>1</td>
</tr>
<tr>
<td>( X_n )</td>
<td>2</td>
</tr>
<tr>
<td>( Y_1 )</td>
<td>1</td>
</tr>
<tr>
<td>( Y_2 )</td>
<td>2</td>
</tr>
<tr>
<td>( \vdots )</td>
<td>1</td>
</tr>
<tr>
<td>( Y_m )</td>
<td>1</td>
</tr>
</tbody>
</table>

\( \rightarrow T^* \)

Group status shuffled

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

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

- 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 \( \hat{P} \) 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.

\[ \hat{P} = \frac{X}{N} \quad \text{where} \quad X \sim \text{Binomial}(N,P) \]

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

If the “true” P-value was \( P = 5\% \), and we do \( N=1000 \) shuffles:
\( \text{SD}(\hat{P}) = 0.7\% \).
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.

→ 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.