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

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

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

Paired t-test

\[D_i = Y_i - X_i\]

\[D_1, \ldots, D_n \sim \text{iid normal}(\mu_B - \mu_A, \sigma_D)\]

sample mean \(\bar{D}\); sample SD \(s_D\)

\[T = \frac{\bar{D}}{(s_D/\sqrt{n})}\]

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

Suppose we are concerned about the normal assumption.

\((X_1, Y_1), \ldots, (X_n, Y_n)\) independent

Test \(H_0: X's \text{ and } Y's \text{ have the same distribution}\)

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

\((the \text{ number of pairs for which } X_i < Y_i)\)

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

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

\[\text{P-value} = 2 \times \Pr(S \geq S_{\text{obs}} \mid H_0) = 2 \times (1 - \text{pbinom}(S_{\text{obs}} - 1, n, 0.5))\]
Example

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

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

(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.

$$R = \text{sum of ranks of positive (or negative) values}$$

<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: \texttt{wilcox.test(d) $\rightarrow P = 0.054$}
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:**

\[
\begin{align*}
(117.3, 145.9) & \quad (100.1, 94.8) & \quad (94.5, 108.0) & \quad (135.5, 122.6) & \quad (92.9, 130.2) & \quad (118.9, 143.9) \\
(144.8, 149.9) & \quad (103.9, 138.5) & \quad (103.8, 91.7) & \quad (153.6, 162.6) & \quad (163.1, 202.5) & \quad T_{\text{obs}} = 2.50
\end{align*}
\]

**Example shuffled data:**

\[
\begin{align*}
(117.3, 145.9) & \quad (94.8, 100.1) & \quad (108.0, 94.5) & \quad (135.5, 122.6) & \quad (130.2, 92.9) & \quad (118.9, 143.9) \\
(144.8, 149.9) & \quad (138.5, 103.9) & \quad (103.8, 91.7) & \quad (162.6, 153.6) & \quad (163.1, 202.5) & \quad T^* = 0.19
\end{align*}
\]

**Permutation distribution**

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

**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

\[ X_1, \ldots, X_n \sim iid \ normal(\mu_A, \sigma) \quad Y_1, \ldots, Y_m \sim iid \ normal(\mu_B, \sigma) \]

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

\[ \bar{X} = 47.5 \quad s_A = 10.5 \quad n = 6 \]
\[ \bar{Y} = 74.3 \quad s_B = 20.6 \quad m = 9 \]
\[ s_p = 17.4 \quad T = -2.93 \]
\[ P = 2 \times \text{pt}(-2.93, 6+9-2) = 0.011 \]

Wilcoxon rank-sum test

Rank the X’s and Y’s from smallest to largest (1, 2, \ldots, n+m)

\[ R = \text{sum of ranks for X's} \]

(Also known as the Mann-Whitney Test)

\[
\begin{array}{ccc}
\text{X} & \text{Y} & \text{rank} \\
35.0 & & 1 \\
38.2 & & 2 \\
43.3 & & 3 \\
46.8 & & 4 \\
49.7 & & 5 \\
50.0 & & 6 \\
51.9 & & 7 \\
57.1 & & 8 \\
61.2 & & 9 \\
74.1 & & 10 \\
75.1 & & 11 \\
84.5 & & 12 \\
90.0 & & 13 \\
95.1 & & 14 \\
101.5 & & 15 \\
\end{array}
\]

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

P-value = 0.026

(\text{use} \ \text{wilcox.test}())

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

<table>
<thead>
<tr>
<th>X or Y group</th>
<th>X or Y group</th>
</tr>
</thead>
<tbody>
<tr>
<td>$X_1$</td>
<td>$X_1$</td>
</tr>
<tr>
<td>$X_2$</td>
<td>$X_2$</td>
</tr>
<tr>
<td>$\vdots$</td>
<td>$\vdots$</td>
</tr>
<tr>
<td>$X_n$</td>
<td>$X_n$</td>
</tr>
<tr>
<td>$Y_1$</td>
<td>$Y_1$</td>
</tr>
<tr>
<td>$Y_2$</td>
<td>$Y_2$</td>
</tr>
<tr>
<td>$\vdots$</td>
<td>$\vdots$</td>
</tr>
<tr>
<td>$Y_m$</td>
<td>$Y_m$</td>
</tr>
</tbody>
</table>

$\rightarrow T_{obs}$

<table>
<thead>
<tr>
<th>X or Y group</th>
<th>X or Y group</th>
</tr>
</thead>
<tbody>
<tr>
<td>$X_1$</td>
<td>$X_1$</td>
</tr>
<tr>
<td>$X_2$</td>
<td>$X_2$</td>
</tr>
<tr>
<td>$\vdots$</td>
<td>$\vdots$</td>
</tr>
<tr>
<td>$X_n$</td>
<td>$X_n$</td>
</tr>
<tr>
<td>$Y_1$</td>
<td>$Y_1$</td>
</tr>
<tr>
<td>$Y_2$</td>
<td>$Y_2$</td>
</tr>
<tr>
<td>$\vdots$</td>
<td>$\vdots$</td>
</tr>
<tr>
<td>$Y_m$</td>
<td>$Y_m$</td>
</tr>
</tbody>
</table>

Group status shuffled

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

Permutation distribution

![Histogram](https://via.placeholder.com/150)

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

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

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

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

If the “true” P-value $P = 5\%$ 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:**
  - 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.