NONPARAMETRIC TESTS

-- Tests that do not assume that the data conforms to a particular probability distribution function (such as Gaussian or binomial distributions)

-- Useful for small sample sizes (where there is insufficient data to determine the pdf)

I. Wilcoxon Signed Rank Test

Purpose: to determine whether there is a statistically significant deviation from zero in the median of differences in related pairs of measurements

For example, to determine whether a blood pressure treatment works, we need to know whether the set of differences between the before-treatment and after-treatment blood pressures has a median that is significantly different from zero.

<table>
<thead>
<tr>
<th>Patient #</th>
<th>Before</th>
<th>After</th>
<th>Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>160</td>
<td>143</td>
<td>-17</td>
</tr>
<tr>
<td>2</td>
<td>152</td>
<td>162</td>
<td>+10</td>
</tr>
<tr>
<td>3</td>
<td>148</td>
<td>151</td>
<td>+3</td>
</tr>
<tr>
<td>4</td>
<td>170</td>
<td>169</td>
<td>-1</td>
</tr>
</tbody>
</table>

Although the sample median difference $\tilde{D} = 1$ in this case, we can’t infer anything about the population median difference $\mu_D$, because we don’t have a probability distribution function.

The purpose of this test is to decide if the data permits a statistically significant inference about $\mu_D$. Specifically, whether $\mu_D = 0$ can be rejected.

Requirements

- A data set consisting of a small ($n \leq 30$) list of differences in a particular a measurement (e.g. difference in systolic blood pressure of each subject before and after a treatment).
- A table or calculator of the critical values for this test (to use instead of $Z_{\alpha/2}$, $t_{\alpha/2}$, etc)

Procedure

1. Order the data.
   Order data, according to its absolute value, from lowest to highest.

   e.g. $(4, -2, -8, 5, 9, 0, 4, 11, -3) \rightarrow (0, -2, -3, 4, 4, 5, -8, 9, 11)$

2. Discard zeroes and rank the data.
   Discard zeroes for the data set, and assign a numerical rank to each measurement that corresponds to its position in the list.

<table>
<thead>
<tr>
<th>data</th>
<th>-2</th>
<th>-3</th>
<th>4</th>
<th>4</th>
<th>5</th>
<th>-8</th>
<th>9</th>
<th>11</th>
</tr>
</thead>
<tbody>
<tr>
<td>rank</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
<td>8</td>
</tr>
</tbody>
</table>
3. **Adjust ranks for ties.**
   In the case of ties, assign the each tied datum the average rank of the tied data.

   In this case, after adjusting for the ties, we have

<table>
<thead>
<tr>
<th>data</th>
<th>-2</th>
<th>-3</th>
<th>4</th>
<th>4</th>
<th>5</th>
<th>-8</th>
<th>9</th>
<th>11</th>
</tr>
</thead>
<tbody>
<tr>
<td>rank</td>
<td>1</td>
<td>2</td>
<td>3.5</td>
<td>3.5</td>
<td>5</td>
<td>6</td>
<td>7</td>
<td>8</td>
</tr>
</tbody>
</table>

4. **Separately sum the ranks of the negative and positive measurements.**

   Let $S_-$ = sum of the ranks of the negative measurements
   
   \[ S_- = 1 + 2 + 6 \]
   
   \[ S_- = 9 \]

   Let $S_+$ = sum of the ranks positive measurements

   \[ S_+ = 3.5 + 3.5 + 5 + 7 + 8 \]
   
   \[ S_+ = 27 \]

5. **Evaluate the test statistic, $W$.**

   $W$ = the smaller of $(S_-, S_+)$

   \[ W = 9 \text{ in this example.} \]

6. **Count the number of nonzero measurements**

   $R = $ number of nonzero measurements

   \[ R = \text{number of “signed ranks”} \]

   \[ R = \text{(number of plus signs in data set) + (number of minus signs in data set)} \]

   \[ R = 8 \text{ in this example} \]

7. **Choose the desired probability $\alpha$ for making a Type I error.**

   Let’s choose $\alpha = 0.05$ in this example.

8. **Obtain the critical value $C_{\alpha}(R)$**.

   Use this table to find the critical value that corresponds to $R$ and $\alpha$.

   In this example $R = 8$, $\alpha = 0.05$. So $C_{0.05}(8) = 3$

9. **Apply the decision rule.**

   $\hat{\mu}_D$ = the median of the population consisting of difference measurements.

   [Each difference measurement is a difference in the measurements on the same sampling unit before and after a particular treatment.]

   Null hypothesis $H_0$: $\hat{\mu}_D = 0$.

   Decision rule:

   \[ \text{If } W \leq C_{\alpha}(R), \text{ Reject } H_0 \]
In this example $W = 9$, $C_a(R) = C_{0.05}(8) = 3$.

$9 \leq 3$ is false. So we do not reject $H_0$.

Note: The decision rule (reject $H_0$, if $W \leq C_a(R)$) is counterintuitive. Unlike all the other test statistics we've used, it's the *small* values of $W$ that are improbable when the null hypothesis is true.

**Example Problem**

An experimental weight loss drug is tested on 7 patients. Their weights before being treated with the drug and after are shown in the table below.

<table>
<thead>
<tr>
<th>Patient #</th>
<th>Weight Before</th>
<th>Weight After</th>
<th>Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>237</td>
<td>138</td>
<td>-99</td>
</tr>
<tr>
<td>2</td>
<td>532</td>
<td>600</td>
<td>68</td>
</tr>
<tr>
<td>3</td>
<td>605</td>
<td>400</td>
<td>-205</td>
</tr>
<tr>
<td>4</td>
<td>359</td>
<td>219</td>
<td>-140</td>
</tr>
<tr>
<td>5</td>
<td>422</td>
<td>299</td>
<td>-123</td>
</tr>
<tr>
<td>6</td>
<td>385</td>
<td>305</td>
<td>-80</td>
</tr>
<tr>
<td>7</td>
<td>274</td>
<td>128</td>
<td>-146</td>
</tr>
</tbody>
</table>

Can we assert with significance level 0.05 that this drug works? That is, can we reject the null hypothesis that the median weight loss (the median difference $\tilde{\mu}_D$) is zero with a 5% probability of being wrong.

**Step 1.** Order the data by absolute value

68, -80, -99, -123, -140, -146, -205

**Step 2.** Discard zeroes and rank data

<table>
<thead>
<tr>
<th>data</th>
<th>68</th>
<th>-80</th>
<th>-99</th>
<th>-123</th>
<th>-140</th>
<th>-146</th>
<th>-205</th>
</tr>
</thead>
<tbody>
<tr>
<td>rank</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
</tr>
</tbody>
</table>

**Step 3.** Adjust for ties. Not needed.

**Step 4.** Sum ranks of negative and positive data.

$S_+ = 1$

$S_- = 2 + 3 + 4 + 5 + 6 + 7 = 27$

**Step 5.** Evaluate $W$. $W = \text{the smaller of } (1, 27) = 1$.

**Step 6.** Count the nonzero measurements. $R = 7$.

**Step 7.** Choose $\alpha$. $\alpha = 0.05$

**Step 8.** Look up $C_a(R)$. $C_{0.05}(7) = 2$.

**Step 9.** Apply decision rule. Reject if $W \leq C_{0.05}(7)$.
Have $W = 1, \ C_{0.05}(7) = 2. \\
1 \leq 2 \\
So reject null hypothesis.

II. Sign Test

Purpose: Same as above (same as for Wilcoxon Signed Rank Test)

Requirements: Same as above (same as for Wilcoxon Signed Rank Test)

Procedure

1. **State the null hypothesis.**
   You have 3 choices:
   
   \[
   H_0: \bar{\mu}_D \leq d_0 \quad \text{One-tailed test} \\
   H_0: \bar{\mu}_D \geq d_0 \quad \text{One-tailed test} \\
   H_0: \bar{\mu}_D = d_0 \quad \text{Two-tailed test}
   \]
   
   where $\bar{\mu}_D$ is the population median difference, $d_0$ is the value to which you wish to compare $\bar{\mu}_D$.
   
   [For example, if you want to know whether patients' weight loss was above 0, below 0, or equal to 0, choose $d_0 = 0$. If you want to know if their weight loss was above 50 lbs, below 50 lbs, or equal to 50 lbs, choose $d_0 = 50$]

2. **Convert each sample difference $d_i$ into a “+”, a “-”, or “0”.**
   
   Use
   
   \[
   d_i = “+” \quad \text{if } d_i > d_0 \\
   d_i = “0” \quad \text{if } d_i = d_0 \\
   d_i = “-” \quad \text{if } d_i < d_0
   \]
   
   For example, if we choose $d_0 = 0$, and $(d_1, d_2, d_3, \ldots) = (4, -2, -8, 5, 9, 0, 4, 11, -3)
   
   $(4, -2, -8, 5, 9, 0, 4, 11, -3) \rightarrow (+, -, -, +, +, 0, +, +, -)$

3. **Evaluate the test statistic, $W$.**

   \[ W = \text{the number of plus signs in the list obtained in Step 3.} \]

   For the example in Step 2, $W = 5$.

4. **Count the number of nonzero measurements.**

   \[ n = (\text{number of plus signs}) + (\text{number of minus signs}) \]

   For example in Step 2, $n = 8$.

5. **Choose the maximum acceptable probability $\alpha$ for making a Type I error.**

   This is also called the maximum acceptable p-value of our test.
   
   Let's choose $\alpha = 0.05$
6. Obtain the critical values $C$ and $C'$.

Option 1: Use this table.
- Determine whether your test is one-tailed or two-tailed (see Step 1)
- Find the rows associated with the value of $n$ (obtained in Step 4)
- Select the row whose $\alpha$ value is about equal to or less than your chosen $\max \alpha$
- Read off $C$ and $C'$

For example, suppose that we had chosen a two-tailed test. For $n = 8$ and $\max \alpha = 0.05, \ C = 7$ and $C' = 1$.

<table>
<thead>
<tr>
<th>$n$</th>
<th>$c'$</th>
<th>$c$</th>
<th>One Tail</th>
<th>Two Tails</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>1</td>
<td>3</td>
<td>0.063</td>
<td>0.126</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>4</td>
<td>0.031</td>
<td>0.062</td>
</tr>
<tr>
<td>6</td>
<td>1</td>
<td>5</td>
<td>0.016</td>
<td>0.032</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>4</td>
<td>0.109</td>
<td>0.218</td>
</tr>
<tr>
<td>7</td>
<td>1</td>
<td>6</td>
<td>0.008</td>
<td>0.016</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>5</td>
<td>0.063</td>
<td>0.126</td>
</tr>
<tr>
<td>8</td>
<td>1</td>
<td>7</td>
<td>0.004</td>
<td>0.008</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>6</td>
<td>0.035</td>
<td>0.070</td>
</tr>
<tr>
<td>9</td>
<td>1</td>
<td>8</td>
<td>0.002</td>
<td>0.004</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>7</td>
<td>0.020</td>
<td>0.040</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>6</td>
<td>0.090</td>
<td>0.180</td>
</tr>
<tr>
<td>10</td>
<td>1</td>
<td>9</td>
<td>0.001</td>
<td>0.002</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>8</td>
<td>0.011</td>
<td>0.022</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>7</td>
<td>0.055</td>
<td>0.110</td>
</tr>
<tr>
<td>11</td>
<td>2</td>
<td>9</td>
<td>0.006</td>
<td>0.012</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>8</td>
<td>0.033</td>
<td>0.066</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>7</td>
<td>0.113</td>
<td>0.226</td>
</tr>
</tbody>
</table>

We could not select the row for which the two-tailed $\alpha = 0.07$, because that exceeds 0.05. [However, if instead of having specified the $\max \alpha$, we had specified the approximate $\alpha$, then we would have selected that row (0.07 is closer to 0.05 than is 0.008).]

Option 2: Use this calculator.

A. Enter 0.5 into the probabilities box.

B. Enter the value of $n$ into both the “successes” box and the “trials box”

C. Press the “Calculate button”.

D. Write down the appropriate p-value for your test. (0.0078 in this example).
E. Write down the critical values $C$ and $C'$.

\[ C = \text{(number that I've highlighted in yellow)} - 1 \]
\[ C' = \text{(number that I've highlighted in green)} + 1 \]

In this example, $C = 7$, $C' = 1$.

F. Click the "back" button of browser. Reduced the entry in the "successes" box by 1, leave other boxes, the same.

- Number of "successes" you observed = 7
- Number of trials or experiments = 8

Repeat Steps C and D.

If the p-value exceeds your specified maximum (0.05 in this example), stop. Use the last critical values ($C$, $C'$) that you obtained.

If the p-value does not exceed your specified maximum, repeat Steps E and F.

In this example, we stop now (because $0.0703 > 0.05$).

In this example, the relevant output of this calculator is: $C = 7$, $C' = 1$.

[Had we specified an approximate $\alpha$, instead of the $\text{max } \alpha$, we stop after we've obtained the $C$ and $C'$ corresponding to the p-value that is closest to our specified $\alpha$.]
7. Use the following Rejection Table

<table>
<thead>
<tr>
<th>$H_0$</th>
<th>Reject $H_0$ if</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\bar{\mu}_D \leq d_0$</td>
<td>$W &gt; C$</td>
</tr>
<tr>
<td>$\bar{\mu}_D \geq d_0$</td>
<td>$W &lt; C'$</td>
</tr>
<tr>
<td>$\bar{\mu}_D = d_0$</td>
<td>$W &lt; C'$ or $W &gt; C$</td>
</tr>
</tbody>
</table>

In this example, we do not reject $H_0$, because $W=5$, $C=7$, $C'=1$, and

\[ 5 < 1 \text{ is false and } 5 > 7 \text{ is false.} \]

**Example Problem** (same problem used with Wilcoxon Signed-Rank test)

An experimental weight loss drug is tested on 7 patients. Their weights before being treated with the drug and after are shown in the table below.

<table>
<thead>
<tr>
<th>Patient #</th>
<th>Weight Before</th>
<th>Weight After</th>
<th>Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>237</td>
<td>138</td>
<td>-99</td>
</tr>
<tr>
<td>2</td>
<td>532</td>
<td>600</td>
<td>68</td>
</tr>
<tr>
<td>3</td>
<td>605</td>
<td>400</td>
<td>-205</td>
</tr>
<tr>
<td>4</td>
<td>359</td>
<td>219</td>
<td>-140</td>
</tr>
<tr>
<td>5</td>
<td>422</td>
<td>299</td>
<td>-123</td>
</tr>
<tr>
<td>6</td>
<td>385</td>
<td>305</td>
<td>-80</td>
</tr>
<tr>
<td>7</td>
<td>274</td>
<td>128</td>
<td>-146</td>
</tr>
</tbody>
</table>

Can we assert with significance level 0.05 that this drug works? That is, can we reject the null hypothesis that the median weight loss (the median difference $\bar{\mu}_D$) is zero with a 5% probability of being wrong.

**Step 1:** State the null hypothesis. $H_0: \bar{\mu}_D = d_0$ (Two-tailed test)

**Step 2:** Convert data to a list of pluses, minuses, and zeroes.

<table>
<thead>
<tr>
<th>data</th>
<th>-99</th>
<th>68</th>
<th>-205</th>
<th>-140</th>
<th>-123</th>
<th>-80</th>
<th>-146</th>
</tr>
</thead>
<tbody>
<tr>
<td>sign</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

**Step 3:** Evaluate test statistic.

$W =$ number of plus signs $= 1$

**Step 4:** Count nonzero measurements.

$n = 7$

**Step 5:** Choose $\alpha$.

Let’s read the problem to mean that we want the max probability of being wrong to be 5%, i.e. $\max \alpha = 0.05$. 

Step 6: Obtain critical values C and C’.

Using the table with \( n = 7 \), we see that \( C’ = 1 \) and \( C = 6 \). [We did not use the other row for \( n=7 \), because its \( \alpha (0.126) \) exceeds our specified max \( \alpha \) of 0.05.]

Step 7: Apply rejection condition.

We must reject \( H_0 \) if \( W < C’ \) or \( W > C \).

\[
\begin{align*}
W &= 1, \quad C’=1, \quad C=6 \\
1 < 1 \text{ is false} \quad \text{and} \quad 1 > 6 \text{ is false}
\end{align*}
\]

So we cannot reject \( H_0 \).

Reflect: Notice that in order to reject \( H_0 \) with the Sign Test, we would have to accept a p-value of 12.6% (see table). The Wilcoxon Signed-Rank Test was able to reject \( H_0 \) with a p-value of 5%. This shows that the Sign Test is cruder than the Wilcoxon Signed Rank Test.