This work is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike License. Your use of this material constitutes acceptance of that license and the conditions of use of materials on this site.
Section F (Optional)

Non-Parametric Analogue to the Two Sample t-test
Alternative to the Two Sample T-Test

- Nonparametric test for comparing two groups

- “Non-parametric” refers to a class of tests that do not assume anything about distribution of the data

- Nonparametric test for comparing two groups
  - Mann-Whitney Rank Sum Test (Wilcoxon Rank Sum Test)
  - Also called Mann-Whitney-Wilcoxon (a mouthful)

- Tries to answer the following question:
  - Are the two population distributions different?
Advantages

- Does not assume populations being compared are normally distributed
  - The two-sample t-test requires that assumption with very small samples sizes

- Uses only ranks

- Not sensitive to outliers
Disadvantage of the Nonparametric Test

- Nonparametric methods are often less sensitive (powerful) for finding true differences because they throw away information (they use only ranks)

- Need full data set, not just summary statistics

- Results do not include any confidence intervals quantifying range of possibility for true difference between populations
Example: Health Education Study

- Evaluate an intervention to educate high school students about health and lifestyle over a two-month period

- 10 students randomized to “intervention” or “control” group

- \[ x = \text{post test score} - \text{pre-test score} \] is outcome to compare between the intervention and control groups
Example: Health Education Study

- $x =$ post- pretest score for both groups

- Intervention (I)  5  0  7  2  19

- Control (C)  6  -5  -6  1  4
  - Only five individuals in each sample!!!
  - We want to compare the control and intervention groups to assess whether the “improvement” (post-pre) in scores are different, taking random sampling error into account.
With such a small sample size, we need to be sure score improvements are normally distributed if we want to use t-test (BIG assumption)

Possible approach:
- Mann-Whitney-Wilcoxon non-parametric test!
Example: Health Education Study

- First step—rank the pooled data (ignore groupings)

-6  -5  0  1  2  4  5  5  7  19

- Rank  1  2  3  4  5  6  7  8  9  10
Example: Health Education Study

- Second step—“reattach” group status

<table>
<thead>
<tr>
<th>Rank</th>
<th>Group</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>C</td>
</tr>
<tr>
<td>2</td>
<td>C</td>
</tr>
<tr>
<td>3</td>
<td>C</td>
</tr>
<tr>
<td>4</td>
<td>C</td>
</tr>
<tr>
<td>5</td>
<td>C</td>
</tr>
<tr>
<td>6</td>
<td>I</td>
</tr>
<tr>
<td>7</td>
<td>C</td>
</tr>
<tr>
<td>8</td>
<td>C</td>
</tr>
<tr>
<td>9</td>
<td>C</td>
</tr>
<tr>
<td>10</td>
<td>I</td>
</tr>
</tbody>
</table>

-6 -5 0 1 2 4 5 5 7 19
Example: Health Education Study

- Find the average rank in each of the two groups

- Intervention group average rank

\[
\frac{3 + 5 + 6 + 9 + 10}{5} = 6.8
\]

- Control group average rank

\[
\frac{1 + 2 + 4 + 6 + 8}{5} = 4.2
\]
Example: Health Education Study

- Statisticians have developed formulas and tables to determine the probability of observing such an extreme discrepancy in ranks (6.8 vs. 4.2) by chance alone
  - This is the p-value

- In the health education study, the p-value was .17
  - The interpretation is that the Mann-Whitney test did not show any significant difference in test score “improvement” between the intervention and control group (p = .17)
The two-sample t-test would give a different answer (p = .14)

Different statistical procedures can give different p-values

If the largest observation, 19, was changed, the p-value based on the Mann-Whitney test would not change but the two-sample t-test would change
The t-test or the nonparametric test?

- Statisticians will not always agree, but there are some guidelines
- Use non-parametric test if sample size is small and you have no reason to believe data is “well behaved” (normally distributed)
- Only “ranks” available
Using Stata to Perform Mann-Whitney-Wilcoxon

- Data, as entered

```
. list diff int_cntrl

+-----------------------+
<table>
<thead>
<tr>
<th>diff   int_cntrl</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
</tr>
<tr>
<td>2.</td>
</tr>
<tr>
<td>3.</td>
</tr>
<tr>
<td>4.</td>
</tr>
<tr>
<td>5.</td>
</tr>
<tr>
<td>-----------------------</td>
</tr>
<tr>
<td>6.</td>
</tr>
<tr>
<td>7.</td>
</tr>
<tr>
<td>8.</td>
</tr>
<tr>
<td>9.</td>
</tr>
<tr>
<td>10.</td>
</tr>
</tbody>
</table>
+-----------------------+
```
“ranksum” command
  Syntax:
  
  ranksum \textit{varname}, by(\textit{group\_var})

. ranksum diff, by( int\_cntrl)

Two-sample Wilcoxon rank-sum (Mann-Whitney) test

<table>
<thead>
<tr>
<th></th>
<th>obs</th>
<th>rank sum</th>
<th>expected</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>5</td>
<td>21</td>
<td>27.5</td>
</tr>
<tr>
<td>1</td>
<td>5</td>
<td>34</td>
<td>27.5</td>
</tr>
<tr>
<td>combined</td>
<td>10</td>
<td>55</td>
<td>55</td>
</tr>
</tbody>
</table>

unadjusted variance  22.92
adjustment for ties  0.00
adjusted variance    22.92

Ho: \text{diff(int\_cntrl == 0) = diff(int\_cntrl == 1)}
\ z = -1.358
\ Prob > |z| = 0.1745
Using Stata to Perform Mann-Whitney-Wilcoxon

- “ranksum” command
  - Syntax:
    - ranksum varname, by(group_var)

```
. ranksum diff, by( int_cntrl)

Two-sample Wilcoxon rank-sum (Mann-Whitney) test

     int_cntrl |     obs  rank sum  expected
-------------------------+-----------+
          0  |    5    21    27.5
          1  |    5    34    27.5
-------------------------+
    combined  |   10    55    55

unadjusted variance     22.92
adjustment for ties     0.00
                        ------
adjusted variance       22.92

Ho: diff(int_cntrl==0) = diff(int_cntrl==1)
    z =  -1.358
    Prob > |z| =   0.1745
```
Using Stata to Perform t-test

- "ttest" command without "i" on end when data already in Stata
  - Syntax:

    ttest varname, by(group_var)

  . ttest diff, by( int_cntr1)

Two-sample t test with equal variances

<table>
<thead>
<tr>
<th>Group</th>
<th>Orig</th>
<th>Mean</th>
<th>Std. Err.</th>
<th>Std. Dev.</th>
<th>[95% Conf. Interval]</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>5</td>
<td>0</td>
<td>2.387467</td>
<td>5.338539</td>
<td>-6.628672 6.628672</td>
</tr>
<tr>
<td>1</td>
<td>5</td>
<td>6.6</td>
<td>3.325658</td>
<td>7.436397</td>
<td>-2.633506 15.83351</td>
</tr>
<tr>
<td>combined</td>
<td>10</td>
<td>3.3</td>
<td>2.221361</td>
<td>7.02456</td>
<td>-1.725068 8.325068</td>
</tr>
<tr>
<td>diff</td>
<td>-6.6</td>
<td>4.093898</td>
<td></td>
<td>-16.04055</td>
<td>2.840545</td>
</tr>
</tbody>
</table>

\[
\text{diff} = \text{mean}(0) - \text{mean}(1) \quad \quad t = -1.6122
\]

Ho: \( \text{diff} = 0 \) \quad degrees of freedom = 8

<table>
<thead>
<tr>
<th>Ha: ( \text{diff} &lt; 0 )</th>
<th>Ha: ( \text{diff} &gt; 0 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \Pr(T &lt; t) = 0.0728 )</td>
<td>( \Pr(T &gt; t) = 0.9272 )</td>
</tr>
<tr>
<td>( \Pr(</td>
<td>T</td>
</tr>
</tbody>
</table>
Summary: Educational Intervention Example

- **Statistical methods**
  - 10 high school students were randomized to either receive a two-month health and lifestyle education program (or no program)
  - Each student was administered a test regarding health and lifestyle issues prior to randomization (and after the two-month period)
Summary: Educational Intervention Example

- **Statistical methods**
  - Differences in the two test scores (after-before) were computed for each student
  - Mean and median test score changes were computed for each of the two study groups
  - A Mann-Whitney rank sum test was used to determine if there was a statistically significant difference in test score change between the intervention and control groups at the end of the two-month study period
Summary: Educational Intervention Example

Result

- Participants randomized to the educational intervention scored a median five points higher on the test given at the end of the two-month study period, as compared to the test administered prior to the intervention.
- Participants randomized to receive no educational intervention scored a median one point higher on the test given at the end of the two-month study period.
- The difference in test score improvements between the intervention and control groups was not statistically significant (p = .17).