8 Comparing Two Independent Populations

We’ll study these methods for comparing two independent populations:

1. The Two-Sample T-Test (Normal with Equal Variances)
2. The Welch T-Test (Normal)
3. Bootstrap for Two Samples
4. Permutation Test
5. The Wilcoxon Rank Sum or Mann-Whitney Test
6. Comparing Two Population Proportions

8.1 The Two-Sample T-Test (Normal with Equal but Unknown Variances)

e.g. The horned lizard has spikes on its head that may protect against its primary predator, the loggerhead shrike. Researchers wanted to compare dead lizards killed by shrikes with live lizards from the same area. A SRS was taken from each population. The longest spike was measured, in mm. Is there a difference in longest spike length across the two populations? Here are the data:

Live: 23.76, 21.17, 26.13, 20.18, 23.01, 24.84, 19.34, 24.94, 27.14, 25.87, 18.95, 22.61

Start with graphical and numerical summaries:
Here are some numerical summaries:

<table>
<thead>
<tr>
<th>Group</th>
<th>Sample Size, n</th>
<th>Sample Mean</th>
<th>Sample SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dead</td>
<td>10</td>
<td>20.79</td>
<td>2.22</td>
</tr>
<tr>
<td>Live</td>
<td>12</td>
<td>23.16</td>
<td>2.76</td>
</tr>
</tbody>
</table>

The summaries all show 

Is it 

or just the result of 

? The shift of sample means matters only 

of the sample data.

We compare two population means, \( \mu_1 \) and \( \mu_2 \), by studying their 

Notation:

<table>
<thead>
<tr>
<th>Variable</th>
<th>Population 1</th>
<th>Population 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>( \mu_1 )</td>
<td>( \mu_2 )</td>
</tr>
<tr>
<td>Variance</td>
<td>( \sigma_1^2 )</td>
<td>( \sigma_2^2 )</td>
</tr>
<tr>
<td>Sample size</td>
<td>( n_1 )</td>
<td>( n_2 )</td>
</tr>
<tr>
<td>Sample mean</td>
<td>( \bar{X}_1 )</td>
<td>( \bar{X}_2 )</td>
</tr>
<tr>
<td>Sample variance</td>
<td>( s_1^2 )</td>
<td>( s_2^2 )</td>
</tr>
</tbody>
</table>

For inference about \( \mu_1 - \mu_2 \), use the statistic 

, and then

• test \( H_0 : \mu_1 - \mu_2 = \delta_0 \) \( (\delta_0 = 0 \implies \) \)

• find a confidence interval for \( \mu_1 - \mu_2 \)

To do this, we need the distribution of 

Recall, for independent \( X \) and \( Y \):

• \( E(X - Y) = \)

• \( \text{VAR}(X - Y) = \)

• \( E(\bar{X}) = \)

• \( \text{VAR}(\bar{X}) = \)

• If \( X \sim N(\mu_X, \sigma_X^2) \) and \( Y \sim N(\mu_Y, \sigma_Y^2) \), then \( X - Y \sim \)

• For normal \( X \), \( \bar{X} \sim \)

It follows that, for normal populations 1 and 2, \( \bar{X}_1 - \bar{X}_2 \sim N \left( \mu_1 - \mu_2, \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2} \right) \).
But we don’t know $\mu_1$ or $\mu_2$. Here we assume they are equal and calculate a variance estimate:

$$s^2_p = \frac{\sum_{i=1}^{n_1}(X_{1,i} - \bar{X}_1)^2 + \sum_{i=1}^{n_2}(X_{2,i} - \bar{X}_2)^2}{n_1 + n_2 - 2} = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$$

Now we can state a test and a confidence interval:

- Many hypothesis tests use test statistics of the form
  $$\frac{(\text{point estimate}) - (\text{parameter value})}{\text{(estimated or true) \hspace{1cm} of point estimate}}$$

  This ______ point estimate tells how far the estimate is from the parameter, in ______.

  For our difference of two means, this is

  $$t = \frac{\bar{X}_1 - \bar{X}_2 - \delta_0}{\sqrt{s^2_p \frac{1}{n_1} + s^2_p \frac{1}{n_2}}} \sim t_{n_1 + n_2 - 2} \ (\approx)$$

- Recall that many confidence intervals have the form

  $$(\text{point estimate}) \pm (\text{margin of error})$$

  $= (\text{point estimate}) \pm (\_\_\_\_ \text{value for confidence}) \times [\_\_\_\_\_\_ \text{of point estimate}]$

  $= \hat{\theta} \pm (\text{table value for confidence}) \times \sigma_{\hat{\theta}}$

  Our $100\%(1 - \alpha)$ confidence interval for $\mu_1 - \mu_2$, assuming normal populations and equal population variances, is

  $$(\bar{X}_1 - \bar{X}_2) \pm t_{n_1 + n_2 - 2,\alpha/2} \sqrt{\frac{s^2_p}{n_1} + \frac{s^2_p}{n_2}}$$

  e.g. Test whether the lizard populations have the same population spike lengths and find a confidence interval for the difference in lengths.

  Check normality:
Check equal variances from original plots, above, or from a rule of thumb that it’s plausible that population variances are equal if the larger sample variance is less than twice the smaller.

e.g. Our sample variances are ______ and ______, so we can ________________.

e.g. Calculate a test and interval for the lizard spikes:

Summary:

Suppose we have independent simple random samples from normal populations with means \( \mu_1 \) and \( \mu_2 \) and variances \( \sigma^2_1 \) and \( \sigma^2_2 \), where \( \sigma^2_1 = \sigma^2_2 \).

• To test \( H_0 : \mu_1 - \mu_2 = \delta_0 \),

  1. State null and alternative hypotheses, \( H_0 \) and \( H_A \)
  2. Check assumptions (rule of thumb: \( \sigma^2_1 = \sigma^2_2 \) is plausible for sample variances within a factor of 2)
  3. Find the pooled variance estimate \( s^2_p = \frac{(n_1 - 1)s^2_1 + (n_2 - 1)s^2_2}{n_1 + n_2 - 2} \)
  4. Find the test statistic \( t = \frac{\bar{x}_1 - \bar{x}_2 - \delta_0}{\sqrt{\frac{s^2_p}{n_1} + \frac{s^2_p}{n_2}}} \)
  5. Find the \( P \)-value, which is an area under the \( t_{n_1+n_2-2} \) curve depending on \( H_A \):
     \( H_A : \mu_1 - \mu_2 > \delta_0 \implies P\text{-value} = P(T > t) \), the area right of \( t \)
     \( H_A : \mu_1 - \mu_2 < \delta_0 \implies P\text{-value} = P(T < t) \), the area left of \( t \)
     \( H_A : \mu_1 - \mu_2 \neq \delta_0 \implies P\text{-value} = P(|T| > |t|) \), the sum of the two tail areas
  6. Draw a conclusion

• A 100%(1 - \( \alpha \)) confidence interval for \( \mu_1 - \mu_2 \) is

\[
(\bar{X}_1 - \bar{X}_2) \pm t_{(n_1+n_2-2, \alpha/2)} \sqrt{\frac{s^2_p}{n_1} + \frac{s^2_p}{n_2}}
\]

Note: I recommend using Welch’s \( t \)-test, below, instead of the two-sample \( t \)-test, above. We introduced the equal-variances two-sample \( t \)-test to see its \( s^2_p \), which has the form sum of squared deviations from respective sample means degrees of freedom, which we’ll see again in soon in §10 on ANOVA.
8.2 The Welch T-Test (Normal without Assuming Equal Variances)

e.g. Concrete is often reinforced with steel “rebar” (“reinforcing bar”). Steel is strong, but tends
to corrode over time. An experiment tested two corrosion-resistant materials, one fiberglass and
the other carbon.

Eight concrete beams with fiberglass reinforcement, and 11 with carbon reinforcement, were poured.
Each was subjected to a load test, with the breaking force measured in kN (kiloNewtons):

Fiberglass: 38.3, 29.6, 33.4, 33.6, 30.7, 32.7, 34.6, 32.3
Carbon: 48.8, 38.0, 42.2, 45.1, 32.8, 47.2, 50.6, 44.0, 43.9, 40.4, 45.8

Is there a difference in the (population) mean strengths of the two types of beams? We test:

\[ H_0 : \mu_{\text{fiber}} - \mu_{\text{carbon}} = \]

vs.

\[ H_A : \mu_{\text{fiber}} - \mu_{\text{carbon}} \]

First, make graphical and numerical summaries.

<table>
<thead>
<tr>
<th>Beam Type</th>
<th>Sample Size</th>
<th>Mean</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fiber</td>
<td>8</td>
<td>33.15</td>
<td>2.63</td>
</tr>
<tr>
<td>Carbon</td>
<td>11</td>
<td>43.53</td>
<td>5.06</td>
</tr>
</tbody>
</table>

These summaries suggest ________________. Let’s test this. Is it plausible the two
populations are ______________? Here are QQ plots:
We’ll assume population. The first graph, and our rule of thumb, suggest in the Carbon group, so we assume equal variances.

Suppose, then, that we have independent simple random samples from two normal populations with means $\mu_1$ and $\mu_2$ and variances $\sigma_1^2$ and $\sigma_2^2$, which be equal. Recall (from §8.1, above) that

$$\bar{X}_1 - \bar{X}_2 \sim N\left(\mu_1 - \mu_2, \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}\right)$$

We could standardize $\bar{X}_1 - \bar{X}_2$ as $Z = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \sim \ldots$, but we don’t know or . We approximate them with and , and then get a distribution instead of a .

(Recall that the $t_\nu$ distributions look like , but are with ).

Experts say $T = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2/n_1}{n_1} + \frac{s_2^2/n_2}{n_2}}} \sim t_\nu (\approx)$, where $\nu = \frac{(s_1^2/n_1 + s_2^2/n_2)^2}{\frac{(s_1^2/n_1)^2}{n_1-1} + \frac{(s_2^2/n_2)^2}{n_2-1}}$, rounded .

Now we can state a test and interval if we recall the common test statistic form,

$$\frac{\text{(estimated or true) of point estimate} - \text{(estimated or true) of point estimate}}{\text{(estimated or true) \text{ confidence interval form}},}$$

which tells how far the estimate is from the parameter, in standard deviations, and the common confidence interval form,

$$(\text{point estimate}) \pm (\text{margin of error}) = \hat{\theta} \pm (\text{table value for confidence}) \times \sigma_{\hat{\theta}}$$
Suppose we have independent simple random samples from normal populations with means $\mu_1$ and $\mu_2$ and variances $\sigma^2_1$ and $\sigma^2_2$.

- To test $H_0 : \mu_1 - \mu_2 = \delta_0$,

  1. State null and alternative hypotheses, $H_0$ and $H_A$
  2. Check assumptions
  3. Find the test statistic $t = \frac{(\bar{x}_1 - \bar{x}_2) - \delta_0}{\sqrt{\frac{s^2_1}{n_1} + \frac{s^2_2}{n_2}}}$
  4. Find the degrees of freedom, $\nu = \frac{(\frac{s^2_1}{n_1})^2 + (\frac{s^2_2}{n_2})^2}{\frac{(s^2_1/n_1)^2}{n_1-1} + \frac{(s^2_2/n_2)^2}{n_2-1}}$, rounded down
  5. Find the $p$-value, which is an area under the $t_\nu$ curve depending on $H_A$:
     - $H_A : \mu_1 - \mu_2 > \delta_0 \implies p$-value $= P(T_\nu > t)$, the area right of $t$
     - $H_A : \mu_1 - \mu_2 < \delta_0 \implies p$-value $= P(T_\nu < t)$, the area left of $t$
     - $H_A : \mu_1 - \mu_2 \neq \delta_0 \implies p$-value $= P(|T_\nu| > |t|)$, the sum of the two tail areas
  6. Draw a conclusion

- $(\bar{X}_1 - \bar{X}_2) \pm t_{\nu, \alpha/2} \sqrt{\frac{s^2_1}{n_1} + \frac{s^2_2}{n_2}}$ contains $\mu_1 - \mu_2$ for a proportion $1 - \alpha$ of samples.

Note that these formulas are like the §8.1 formulas, except that the estimated __________ and __________ changed.

e.g. Test $H_0 : \mu_{\text{carbon}} - \mu_{\text{fiber}} = _____$ vs. $H_A : \mu_{\text{carbon}} - \mu_{\text{fiber}}$____________.

t = __________

$\nu = __________$

$p$-value $= __________$

conclusion: __________

95% interval for $\mu_{\text{carbon}} - \mu_{\text{fiber}}$:

Compare two-sided test and interval: __________
To decide between the §8.1 two-sample t-test and this §8.2 Welch’s t-test, consider

- If population variances are equal, but are not assumed to be equal (so Welch’s test is used), the test loses a little ____________, but is still a good test.

- If population variances are different, but are assumed equal (so the two-sample t-test is used), the test can make ____________ conclusions.

### 8.3 Bootstrap for Two Samples

e.g. When sage crickets mate, the male allows the female to eat part of his hind wings. Does female hunger influence desire to mate? An experiment randomly assigned 24 females to two groups. One group of 11 was starved for two days, while the other group of 13 was fed normally. Each female was presented with a male and the time to mating (in hours) was recorded. Do starved females have a different mean time to mating than normally fed females? Here are the data:

Starved: 1.9, 2.1, 3.8, 9.0, 9.6, 13.0, 14.7, 17.9, 21.7, 29.0, 72.3
Fed: 1.5, 1.7, 2.4, 3.6, 5.7, 22.6, 22.8, 39.0, 54.4, 72.1, 73.6, 79.5, 88.9

We test: \( H_0 : \mu_{\text{starved}} - \mu_{\text{fed}} = \) ____________ vs. \( H_A : \mu_{\text{starved}} - \mu_{\text{fed}} \) ____________

<table>
<thead>
<tr>
<th>Group</th>
<th>Sample Size</th>
<th>Mean</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Starved</td>
<td>11</td>
<td>17.73</td>
<td>19.96</td>
</tr>
<tr>
<td>Fed</td>
<td>13</td>
<td>35.98</td>
<td>33.63</td>
</tr>
</tbody>
</table>

(So far, we might consider a ____________, as the variances seem ____________.)

Note that the fed times ____________, while the starved times include ____________.
We cannot use the Welch’s T-test because ______________. We can again use a __________ method.

To do a bootstrap test for $H_0: \mu_1 - \mu_2 = 0$,

1. Draw simple random samples $x_{1,1}, x_{1,2}, \ldots, x_{1,n_1}$ of size $n_1$ from the first population and $x_{2,1}, x_{2,2}, \ldots, x_{2,n_2}$ of size $n_2$ from the second. Compute $\bar{x}_1$, $s_1^2$, $\bar{x}_2$, and $s_2^2$. Find $t_{obs} = \frac{(\bar{x}_1 - \bar{x}_2) - 0}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$.

2. Draw simple random samples with replacement, $x^*_{1,1}, \ldots, x^*_{1,n_1}$, from the first sample and $x^*_{2,1}, \ldots, x^*_{2,n_2}$ from the second.

3. Compute the means and variances of the resampled data for each group. Call these $\bar{x}^*_1$ and $s^*_1$, and $\bar{x}^*_2$ and $s^*_2$.

4. Compute the statistic $\hat{t} = \frac{(\bar{x}^*_1 - \bar{x}^*_2) - (\bar{x}_1 - \bar{x}_2)}{\sqrt{\frac{s^*_1}{n_1} + \frac{s^*_2}{n_2}}}$.

5. Repeat steps 2-4 a large number, $B$, times to get a collection of $\hat{t}$ values that approximate the sampling distribution of $t$.

6. Find the $p$-value, an area under the approximate sampling distribution given by __________, where $m$ depends on $H_A$:

   $H_A : \mu_1 - \mu_2 > 0 \implies m$ is the number of values of $\hat{t}$ for which $\hat{t} \quad t_{obs}$

   $H_A : \mu_1 - \mu_2 \leq 0 \implies m$ is the number of values of $\hat{t}$ for which $\hat{t} < t_{obs}$

   $H_A : \mu_1 - \mu_2 \neq 0 \implies m$ is the number of values of $\hat{t}$ for which __________

7. Draw a conclusion: $\left\{\begin{array}{ll}
p$-value $\leq \alpha$ (where $\alpha$ is the level, .05 by default) $\implies$ reject $H_0$ 
p$-value > \alpha \implies$ retain $H_0$ as plausible
# Here's one way to do the bootstrap for a difference of two means in R:

```r
starved = c(1.9, 2.1, 3.8, 9.0, 9.6, 13.0, 14.7, 17.9, 21.7, 29.0, 72.3)
fed = c(1.5, 1.7, 2.4, 3.6, 5.7, 22.6, 22.8, 39.0, 54.4, 72.1, 73.6, 79.5, 88.9)
summary(starved) # numerical summaries
sd(starved)
sd(fed)
```

```r
# install.packages("lattice") # Run once to download R code to your computer.
require("lattice")

all = c(starved, fed) # graphs
group = c(rep("Starved", 11), rep("Fed", 13))
dotplot(~all|group, layout = c(1,2), as.table = T, xlab = 'Time (hrs)')
histogram(~all|group, layout = c(1,2), as.table = T, xlab = "Time (hrs)"
qqnorm(starved, main = "QQ Plot of Starved")
qqnorm(fed, main = "QQ Plot of Fed")
```

```r
# dat1 and dat2 are data from the two groups. nboot is the number of resamples.
boottwo = function(dat1, dat2, nboot) {
  bootstat = numeric(nboot)
  truediff = mean(dat1) - mean(dat2)
  n1 = length(dat1)
  n2 = length(dat2)
  for(i in 1:nboot) {
    samp1 = sample(dat1, size = n1, replace = T)
    samp2 = sample(dat2, size = n2, replace = T)
    bootmean1 = mean(samp1)
    bootmean2 = mean(samp2)
    bootvar1 = var(samp1)
    bootvar2 = var(samp2)
    bootstat[i] = (bootmean1 - bootmean2 - truediff)/sqrt((bootvar1/n1) + (bootvar2/n2))
  }
  return(bootstat)
}

B = 5000

```r
```
```
```r
cricketboot = boottwo(starved, fed, B)
t.obs = (mean(starved) - mean(fed)) /
        sqrt(var(starved) / length(starved) + var(fed) / length(fed))
low = sum(cricketboot < -abs(t.obs))
high = sum(cricketboot > abs(t.obs))
p.val = (low + high) / B
```
e.g. For the starved/fed cricket data, we find $t_{\text{obs}} =$

I used R to run $B = 5000$ resamples and found ______ $t$ values less than $t_{\text{obs}}$ and ______ greater, for a $p$-value of ______

We conclude ____________

8.4 The Wilcoxon Rank Sum or Mann-Whitney Test

One more test of location for two populations that may not be normal is the Wilcoxon Rank Sum Test or Mann-Whitney Test. e.g. Consider again the cricket data:

starved: 1.9, 2.1, 3.8, 9.0, 9.6, 13.0, 14.7, 17.9, 21.7, 29.0, 72.3 ($n_{\text{starved}} = 11$)
fed: 1.5, 1.7, 2.4, 3.6, 5.7, 22.6, 22.8, 39.0, 54.4, 72.1, 73.6, 79.5, 88.9 ($n_{\text{fed}} =$)

Our hypotheses are not in terms of the two population _______ (not _______):

$H_0$: The distributions of the two groups are identical vs.

$H_A$: The distributions of the two groups ____________ but one is ________ relative to the other.

We assume independence of sample data between and within groups and that the distributions of the two groups have the same shape.

The test statistic is related to ____________ of the samples, so we rank the data without regard for sample, while retaining sample labels. Then we find:

- $R =$ sum of sample 1 ranks, $R_{\text{min}} = \frac{n_1(n_1+1)}{2} =$ minimum possible sum, and $U = R - R_{\text{min}}$

- $p$-value: \[
\begin{align*}
H_A: \text{population 1 is shifted left of 2} \implies p\text{-value} = P(U \leq U_{\text{obs}}) \\
H_A: \text{population 1 is shifted right of 2} \implies p\text{-value} = P(U \geq U_{\text{obs}}) \\
H_A: \text{population 1 is shifted from 2} \implies p\text{-value} = 2 \cdot \min[P(U \leq U_{\text{obs}}), P(U \geq U_{\text{obs}}), \frac{1}{2}]
\end{align*}
\]

e.g. Here’s a simpler example for which it is not hard to calculate the $p$-value by hand. Suppose sample A is 4.8, 2.2 and sample B is 3.0, 1.5, 3.5.

Sample A’s ranks are ______ and ______, $R =$ _______, $R_{\text{min}} =$ _______, and $U =$ ______. Under $H_0$, ranks are randomly assigned to the two samples from \{1, 2, 3, 4, 5\}. Here are the possible sample A ranks and the statistics we get from them:

<table>
<thead>
<tr>
<th>Sample A ranks</th>
<th>1, 2</th>
<th>1, 3</th>
<th>1, 4</th>
<th>1, 5</th>
<th>2, 3</th>
<th>2, 4</th>
<th>2, 5 (observed)</th>
<th>3, 4</th>
<th>3, 5</th>
<th>4, 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>$R$</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>5</td>
<td>6</td>
<td></td>
<td>7</td>
<td>8</td>
<td>9</td>
<td></td>
</tr>
<tr>
<td>$R_{\text{min}} =$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>7</td>
<td>8</td>
<td>9</td>
<td></td>
</tr>
<tr>
<td>$U$</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>2</td>
<td>3</td>
<td></td>
<td>4</td>
<td>5</td>
<td>6</td>
<td></td>
</tr>
</tbody>
</table>

The $p$-value is ________.
e.g. Here are the cricket data again:

<table>
<thead>
<tr>
<th>rank</th>
<th>time</th>
<th>sample</th>
<th>starved ranks</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1.5</td>
<td>fed</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1.7</td>
<td>fed</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>2.1</td>
<td>starved</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>2.4</td>
<td>fed</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>3.6</td>
<td>fed</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>3.8</td>
<td>starved</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>5.7</td>
<td>fed</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>9.0</td>
<td>starved</td>
<td>9</td>
</tr>
<tr>
<td>10</td>
<td>9.6</td>
<td>starved</td>
<td>10</td>
</tr>
<tr>
<td>11</td>
<td>13.0</td>
<td>starved</td>
<td>11</td>
</tr>
<tr>
<td>12</td>
<td>14.7</td>
<td>starved</td>
<td>12</td>
</tr>
<tr>
<td>13</td>
<td>17.9</td>
<td>starved</td>
<td>13</td>
</tr>
<tr>
<td>14</td>
<td>21.7</td>
<td>starved</td>
<td>14</td>
</tr>
<tr>
<td>15</td>
<td>22.6</td>
<td>fed</td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>22.8</td>
<td>fed</td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>29.0</td>
<td>starved</td>
<td>17</td>
</tr>
<tr>
<td>18</td>
<td>39.0</td>
<td>fed</td>
<td></td>
</tr>
<tr>
<td>19</td>
<td>54.4</td>
<td>fed</td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>72.1</td>
<td>fed</td>
<td></td>
</tr>
<tr>
<td>21</td>
<td>72.3</td>
<td>starved</td>
<td>21</td>
</tr>
<tr>
<td>22</td>
<td>73.6</td>
<td>fed</td>
<td></td>
</tr>
<tr>
<td>23</td>
<td>79.5</td>
<td>fed</td>
<td></td>
</tr>
<tr>
<td>24</td>
<td>88.9</td>
<td>fed</td>
<td></td>
</tr>
</tbody>
</table>

$R = \phantom{0}$$

$R_{\text{min}} = \phantom{0}$$

$U = \phantom{0}$$

(For \phantom{0} observations, use \phantom{0} ranks. e.g. If samples had two 1.2s, they'd be #1 and #2 or #2 and #1, so each would get rank \phantom{0}.)

Here's one way to do this with R:

```
starved = c(1.9, 2.1, 3.8, 9.0, 9.6, 13.0, 14.7, 17.9, 21.7, 29.0, 72.3)
fed = c(1.5, 1.7, 2.4, 3.6, 5.7, 22.6, 22.8, 39.0, 54.4, 72.1, 73.6, 79.5, 88.9)
wilcox.test(starved, fed)
```

For the cricket data, R gives p-value \phantom{0} , so we \phantom{0}.
8.5 Comparing Two Population Proportions

e.g. Does handedness differ by sex? A SRS of $n_M = 54$ males and $n_F = 21$ females was taken. Each person indicated his or her dominant hand:

Female: 12 left, 9 right
Male: 23 left, 31 right

Let $\pi_{FL} =$ proportion of left-handed females and $\pi_{ML} =$ proportion of left-handed males in the population. We test:

$H_0 : \pi_{FL} - \pi_{ML} = 0$
$H_A : \pi_{FL} - \pi_{ML} \neq 0$

A natural point estimate for the population difference of proportions is \( \hat{\pi}_{FL} - \hat{\pi}_{ML} \). If $\pi_{FL}n_F$, $(1 - \pi_{FL})n_F$, $\pi_{ML}n_M$, and $(1 - \pi_{ML})n_M$ are all greater than _______, we can use the CLT to say:

$$\hat{\pi}_{FL} - \hat{\pi}_{ML} \sim N \left( \pi_{FL} - \pi_{ML}, \frac{\pi_{FL}(1 - \pi_{FL})}{n_F} + \frac{\pi_{ML}(1 - \pi_{ML})}{n_M} \right)$$

But we don’t know _____ and _______. Under $H_0$, they are _______ : $\pi_{FL} = \pi_{ML} = \pi_L$, and the distribution becomes:

$$\hat{\pi}_{FL} - \hat{\pi}_{ML} \sim N \left( 0, \pi_L(1 - \pi_L) \left( \frac{1}{n_F} + \frac{1}{n_M} \right) \right)$$

We don’t know the common proportion ________, but we estimate it with a weighted average of the sample proportions:

$$\hat{\pi}_L = \frac{\hat{\pi}_{FL}n_F + \hat{\pi}_{ML}n_M}{n_F + n_M} = \text{number of ________ in both samples combined}$$

______ our point estimate to get a test statistic:

$$Z = \frac{\hat{\pi}_{FL} - \hat{\pi}_{ML}}{\sqrt{\hat{\pi}_L(1 - \hat{\pi}_L) \left( \frac{1}{n_F} + \frac{1}{n_M} \right)}} \sim N(0, 1)$$

e.g. For the handedness data, we have:

$$\hat{\pi}_{FL} = \text{______}$$
$$\hat{\pi}_{ML} = \text{______}$$
$$\hat{\pi}_L = \text{________}$$
The (approximate) expected numbers of successes and failures are ________________.

\[ z = \] ________________

p-value = ________________

conclusion: ________________

We can also make a CI. However, it does not come with a ________________, so we use the more general form of the variance. An approximate $100(1-\alpha)\%$ CI for $\pi_{FL} - \pi_{ML}$ is:

\[
\hat{\pi}_{FL} - \hat{\pi}_{ML} \pm z_{\alpha/2} \sqrt{\frac{\hat{\pi}_{FL}(1-\hat{\pi}_{FL})}{n_F} + \frac{\hat{\pi}_{ML}(1-\hat{\pi}_{ML})}{n_M}}
\]

For our data, a 95% interval works out to:

Here is a summary:

Suppose $X \sim \text{Bin}(n_X, \pi_X)$ and $Y \sim \text{Bin}(n_Y, \pi_Y)$ are independent, with $n_X \pi_X, n_X(1-\pi_X), n_Y \pi_Y,$ and $n_Y(1-\pi_Y)$ all $> 10$.

- To test $H_0 : \pi_X - \pi_Y = 0$:
  1. State null and alternative hypotheses, $H_0$ and $H_A$
  2. Check assumptions
  3. Find $\hat{\pi}_X = \frac{X}{n_X}, \hat{\pi}_Y = \frac{Y}{n_Y}$, and pooled $\hat{\pi} = \frac{X+Y}{n_X+n_Y}$
  4. Find the test statistic, $z = \frac{(\hat{\pi}_X - \hat{\pi}_Y) - 0}{\sqrt{\hat{\pi}(1-\hat{\pi})(1/n_X + 1/n_Y)}}$
  5. Find the $P$-value, which is an area under the $N(0,1)$ curve depending on $H_1$:
     \[ H_A : \pi_X - \pi_Y > 0 \implies P\text{-value} = P(Z > z), \text{ the area right of } z \]
     \[ H_A : \pi_X - \pi_Y < 0 \implies P\text{-value} = P(Z < z), \text{ the area left of } z \]
     \[ H_A : \pi_X - \pi_Y \neq 0 \implies P\text{-value} = P(|Z| > |z|), \text{ the sum of the two tail areas} \]
  6. Draw a conclusion

- A $(100\%)(1-\alpha)$ confidence interval for $\pi_X - \pi_Y$ is

\[
(\hat{\pi}_X - \hat{\pi}_Y) \pm z_{\alpha/2} \sqrt{\frac{\hat{\pi}_X(1-\hat{\pi}_X)}{\hat{n}_X} + \frac{\hat{\pi}_Y(1-\hat{\pi}_Y)}{\hat{n}_Y}}
\]

In the next section, we compare two means when the populations are ________________.