Class 27

Daniel B. Rowe, Ph.D.

Department of Mathematics, Statistics, and Computer Science
Agenda:

Recap Chapter 10.1-10.3

Lecture Chapter 10.4-10.5
Recap Chapter 10.1-10.3
10: Inferences Involving Two Populations
10.2 Inference for Mean Difference Two Dependent Samples

Confidence Interval Procedure

Paired Difference
\[ d = x_1 - x_2 \]  \hspace{1cm} (10.1)

\[ \bar{d} = \frac{1}{n} \sum_{i=1}^{n} d_i \]
\[ s_d^2 = \frac{1}{n-1} \sum_{i=1}^{n} (d_i - \bar{d})^2 \]
\[ \mu_d = \mu_d \quad \sigma_d = \frac{s_d}{\sqrt{n}} \]

With \( \sigma_d \) unknown, a 1-\( \alpha \) confidence interval for \( \mu_d = (\mu_1 - \mu_2) \) is:

Confidence Interval for Mean Difference (Dependent Samples)
\[ \bar{d} - t(df, \alpha / 2) \frac{s_d}{\sqrt{n}} \] to \[ \bar{d} + t(df, \alpha / 2) \frac{s_d}{\sqrt{n}} \]
where \( df = n-1 \)  \hspace{1cm} (10.2)
10: Inferences Involving Two Populations

10.2 Inference for Mean Difference Two Dependent Samples

Example:
Construct a 95% CI for mean difference in Brand B – A tire wear.

\[d_i\text{'s: } 8, 1, 9, -1, 12, 9\]

\[n = 6\]
\[\bar{d} = 6.3\]
\[\alpha = 0.05\]
\[s_d = 5.1\]

\[\bar{d} = \frac{1}{n} \sum_{i=1}^{n} d_i\]

\[s_d^2 = \frac{1}{n-1} \sum_{i=1}^{n} (d_i - \bar{d})^2\]

\[\bar{d} \pm t(df, \alpha / 2) \frac{s_d}{\sqrt{n}} \rightarrow (0.090, 11.7)\]

Figure from Johnson & Kuby, 2012.

Rowe, D.B.
10: Inferences Involving Two Populations

10.2 Inference for Mean Difference Two Dependent Samples

Example:

Test mean difference of Brand B minus Brand A is zero.

Step 1 $H_0: \mu_d=0$ vs. $H_a: \mu_d \neq 0$

Step 2

$df = 5 \quad t^* = \frac{\bar{d} - \mu_{d0}}{s_d / \sqrt{n}}$

$\alpha = .05$

Step 3 $\bar{d} = 6.3 \quad t^* = \frac{6.3 - 0}{5.1 / \sqrt{6}} = 3.03$

Step 4 $t(df, \alpha / 2) = 2.57$

Step 5 Since $t^*>t(df, \alpha / 2)$, reject $H_0$

Conclusion: Significant difference in tread wear at .05 level.

Figures from Johnson & Kuby, 2012.
10: Inferences Involving Two Populations
10.3 Inference for Mean Difference Two Independent Samples

Confidence Interval Procedure

With $\sigma_1$ and $\sigma_2$ unknown, a $1 - \alpha$ confidence interval for $\mu_1 - \mu_2$ is:

$$\text{Confidence Interval for Mean Difference (Independent Samples)}$$

$$(\bar{x}_1 - \bar{x}_2) - t(df, \alpha / 2) \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} \text{ to } (\bar{x}_1 - \bar{x}_2) + t(df, \alpha / 2) \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

where $df$ is either calculated or smaller of $df_1$, or $df_2$  \hspace{1cm} (10.8)

Actually, this is for $\sigma_1 \neq \sigma_2$.

Next larger number than $df$

$$df = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2 + \left(\frac{s_1^2}{n_1} / n_1 - 1 \text{ and } \frac{s_2^2}{n_2} / n_2 - 1\right)^2}$$

If using a computer program.

If not using a computer program.
10: Inferences Involving Two Populations

10.3 Inference Mean Difference

Confidence Interval

Example:
Interested in difference in mean heights between men and women. The heights of 20 females and 30 males is measured. Construct a 95% confidence interval for \( \mu_m - \mu_f \), \( \sigma_m \) & \( \sigma_f \) unknown

\[
(\bar{x}_m - \bar{x}_f) \pm t(df, \alpha / 2) \sqrt{\left(\frac{s_m^2}{n_m}\right) + \left(\frac{s_f^2}{n_f}\right)}
\]

\[
(69.8 - 63.8) \pm 2.09 \sqrt{\left(\frac{(1.92)^2}{30}\right) + \left(\frac{(2.18)^2}{20}\right)}
\]

therefore 4.75 to 7.25

\[
\alpha = 0.05 \quad t(19,.025) = 2.09
\]

Rowe, D.B.
10: Inferences Involving Two Populations

10.3 Inference for Mean Difference Two Independent Samples

Hypothesis Testing Procedure

Step 1

\( H_0: \mu_f = \mu_m \) vs. \( H_a: \mu_f \neq \mu_m \)

Step 2

\( t^* = \frac{(\bar{x}_m - \bar{x}_f) - (\mu_m - \mu_f)}{\sqrt{\left(\frac{s_m^2}{n_m}\right) + \left(\frac{s_f^2}{n_f}\right)}} \)

\( df = 26 \)

\( \alpha = 0.05 \)

Step 3

\( t^* = \frac{(71.7 - 64.4) - (0)}{\sqrt{\left(\frac{9.8}{27}\right) + \left(\frac{7.2}{53}\right)}} = 10.3 \)

Step 4

\( t(df, \alpha/2) = 2.06 \)

Step 5

Reject \( H_0 \) if \( 10.3 > 2.06 \), height males \( \neq \) height females

Rowe, D.B.
Chapter 10: Inferences Involving Two Populations

Questions?

Lecture Chapter 10.5-10.6
Chapter 10: Inference Involving Two Populations (continued)

Daniel B. Rowe, Ph.D.

Department of Mathematics, Statistics, and Computer Science
Inferences Involving One Population

9.2 Inference about the Binomial Probability of Success

We talked about a Binomial experiment with two outcomes. Each performance of the experiment is called a trial. Each trial is independent.

\[ P(x) = \frac{n!}{x!(n-x)!} p^x (1 - p)^{n-x} \]

\[ n = 1, 2, 3, \ldots \]
\[ 0 \leq p \leq 1 \]
\[ x = 0, 1, \ldots, n \]

\( n \) = number of trials or times we repeat the experiment.

\( x \) = the number of successes out of \( n \) trials.

\( p \) = the probability of success on an individual trial.
When we perform a binomial experiment we can estimate the probability of heads as:

$$p' = \frac{x}{n}$$  \hspace{1cm} (9.3)

where $x$ is the number of successes in $n$ trials.

This is a point estimate. Recall the rule for a CI is:

point estimate $\pm$ some amount
10: Inferences Involving Two Populations
10.4 Inference for Difference between Two Proportions

For Binomial, where $x$ is number of successes out of $n$ trials. We said that $\text{mean}(cx) = cnp$ and $\text{variance}(cx) = c^2npq$.

$\rightarrow \text{mean}(x / n) = p$ and $\text{variance}(x / n) = pq / n$. 

We are often interested in comparisons between proportions $p_1 - p_2$. There is another rule that says that if $x_1$ and $x_2$ are random variables, then $\text{mean}(x_1 \pm x_2) = \text{mean}(x_1) \pm \text{mean}(x_2)$

further, $\text{mean}\left(\frac{x_1 \pm x_2}{n_1} \pm \frac{x_2}{n_2}\right) = \text{mean}\left(\frac{x_1}{n_1}\right) \pm \text{mean}\left(\frac{x_2}{n_2}\right)$

and $\text{variance}\left(\frac{x_1 \pm x_2}{n_1} \pm \frac{x_2}{n_2}\right) = \frac{p_1q_1}{n_1} + \frac{p_2q_2}{n_2}$. 

if $x_1$ & $x_2$ independent
10: Inferences Involving Two Populations
10.4 Inference for Difference between Two Proportions

That is where 1. and 2. in the green box below come from

If independent samples of size \(n_1\) and \(n_2\) are drawn … with \(p_1 = P_1\text{(success)}\) and \(p_2 = P_2\text{(success)}\), then the sampling distribution of \(p'_1 - p'_2\) has these properties:

1. mean \(\mu_{p'_1-p'_2} = p_1 - p_2\)

2. standard error \(\sigma_{p'_1-p'_2} = \sqrt{\frac{p_1q_1}{n_1} + \frac{p_2q_2}{n_2}}\)  (10.10)

3. approximately normal dist if \(n_1\) and \(n_2\) are sufficiently large.
   ie I \(n_1, n_2 > 20\) II \(n_1p_1, n_1q_1, n_2p_2, n_2q_2 > 5\) III sample<10% of pop
10: Inferences Involving Two Populations

10.4 Inference for Difference between Two Proportions

Confidence Interval Procedure

Assumptions for ... difference between two proportions:

\( p_1 - p_2 \): The \( n_1 \) ... and \( n_2 \) random observations ... are selected independently from two populations that are not changing.

Confidence Interval for the Difference between Two Proportions: \( p_1 - p_2 \)

\[
(p'_1 - p'_2) - z(\alpha / 2) \sqrt{\frac{p'_1 q'_1}{n_1} + \frac{p'_2 q'_2}{n_2}} \quad \text{to} \quad (p'_1 - p'_2) + z(\alpha / 2) \sqrt{\frac{p'_1 q'_1}{n_1} + \frac{p'_2 q'_2}{n_2}}
\]

where \( p'_1 = \frac{x_1}{n_1} \) and \( p'_2 = \frac{x_2}{n_2} \).  

(10.11)
Example:
Construct a 99% CI for proportion of female A’s minus male A’s
difference \( p_f - p_m \).

119 values

\[ z(\alpha / 2) = 2.58 \] \hspace{1cm} \( (p_f' - p_m') \pm z(\alpha / 2) \sqrt{\frac{p_f'q_f'}{n_f} + \frac{p_m'q_m'}{n_m}} \)

\( n_m = 45 \)
\( n_f = 74 \)
\( x_m = 17 \)
\( x_f = 30 \)

\[ p_f' = \frac{x_f}{n_f} = \frac{30}{74} = .41 \]
\[ p_m' = \frac{x_m}{n_m} = \frac{17}{45} = .38 \]

\( (.41 - .38) \pm 2.58 \sqrt{\frac{(.41)(.59)}{74} + \frac{(.38)(.62)}{45}} \)

\(-.210 \text{ to } .265\)
10: Inferences Involving Two Populations
10.4 Inference for Difference between Two Proportions

Hypothesis Testing Procedure

We can perform hypothesis tests on the proportion

\[ H_0: p_1 \geq p_2 \text{ vs. } H_a: p_1 < p_2 \]

\[ H_0: p_1 \leq p_2 \text{ vs. } H_a: p_1 > p_2 \]

\[ H_0: p_1 = p_2 \text{ vs. } H_a: p_1 \neq p_2 \]

Test Statistic for the Difference between two Proportions - Population Proportions Known

\[ z^* = \frac{(p_1' - p_2') - (p_{10} - p_{20})}{\sqrt{pq \left[ \frac{1}{n_1} + \frac{1}{n_2} \right]}} \]

\[ p_1' = \frac{x_1}{n_1} \quad p_2' = \frac{x_2}{n_2} \]

when \( p_1 = p_2 = p \).

Rowe, D.B.
10: Inferences Involving Two Populations

10.4 Inference for Difference between Two Proportions

Hypothesis Testing Procedure

where we assume $p_1 = p_2$ and use pooled estimate of proportion

\[
p'_p = \frac{x_1 + x_2}{n_1 + n_2}
\]

Test Statistic for the Difference between two Proportions - Population Proportions UnKnown

\[
z^* = \frac{(p'_1 - p'_2) - (p_{10} - p_{20})}{\sqrt{p'_p q'_p \left[ \frac{1}{n_1} + \frac{1}{n_2} \right]}}
\]

\[p'_p\] estimated

(10.15)
**10: Inferences Involving Two Populations**

**10.4 Inference for Difference between Two Proportions**

**Hypothesis Testing Procedure**

**Step 1**

$H_0: p_s - p_c \leq 0$ vs. $H_a: p_s - p_c > 0$

**Step 2**

$$z^* = \frac{(p'_s - p'_c) - (p_{0s} - p_{0c})}{\sqrt{p'_p q'_p \left[ \frac{1}{n_s} + \frac{1}{n_c} \right]}}$$

$\alpha = .05$

**Step 3**

$$z^* = \frac{(.10 - .04) - (0)}{\sqrt{(.07)(.93) \left[ \frac{1}{150} + \frac{1}{150} \right]}} = 2.04$$

**Step 4**

$z(\alpha) = 1.65$

**Step 5**

Reject $H_0$ if $z > 1.65$

$p_{value} < .02$ or $2.04 > 1.65$

<table>
<thead>
<tr>
<th>Product</th>
<th>Number Defective</th>
<th>Number Checked</th>
</tr>
</thead>
<tbody>
<tr>
<td>Salesperson’s</td>
<td>$x_s = 15$</td>
<td>$n_s = 150$</td>
</tr>
<tr>
<td>Competitor’s</td>
<td>$x_c = 6$</td>
<td>$n_c = 150$</td>
</tr>
</tbody>
</table>

$$p'_s = \frac{x_s}{n_s} = \frac{15}{150} \quad p'_c = \frac{x_c}{n_c} = \frac{6}{150}$$

$$p'_p = \frac{x_s + x_c}{n_s + n_c} = \frac{15 + 6}{150 + 150}$$

Figure from Johnson & Kuby, 2012.
10: Inferences Involving Two Populations
10.5 Inference for Ratio of Two Variances Two Ind. Samples
Hypothesis Testing Procedure

We can perform hypothesis tests on two variances

- $H_0: \sigma_1^2 \geq \sigma_2^2$ vs. $H_a: \sigma_1^2 < \sigma_2^2$
- $H_0: \sigma_1^2 \leq \sigma_2^2$ vs. $H_a: \sigma_1^2 > \sigma_2^2$
- $H_0: \sigma_1^2 = \sigma_2^2$ vs. $H_a: \sigma_1^2 \neq \sigma_2^2$

Assumptions: Independent samples from normal distribution

Test Statistic for Equality of Variances

\[ F^* = \frac{s_n^2}{s_d^2} \]

with $df_n = n_n - 1$ and $df_d = n_d - 1$.

\[ (10.16) \]

Actually ignore

\[ F^* = \left[ \frac{(n_n - 1)s_n^2 / \sigma^2}{(n_d - 1)s_d^2 / \sigma^2} \right] / (n_n - 1) / (n_d - 1) \]
10: Inferences Involving Two Populations

10.5 Inference for Ratio of Two Variances Two Ind. Samples

Properties of $F$ distribution

1. $F$ is non-negative
2. $F$ is nonsymmetrical
3. $F$ is a family of dists.

$$df_n = \nu_n = n_n - 1, \quad df_d = \nu_d = n_d - 1.$$ 

$$\mu = \frac{\nu_d}{\nu_d - 2}, \quad \nu_d > 2$$

$$\sigma^2 = \frac{2\nu_d^2(\nu_n + \nu_d - 2)}{\nu_n(\nu_d - 2)^2(\nu_d - 4)}, \quad \nu_2 > 4$$

$$f(F | \nu_n, \nu_d) = \frac{\Gamma\left(\frac{\nu_n + \nu_d}{2}\right)}{\Gamma\left(\frac{\nu_n}{2}\right)\Gamma\left(\frac{\nu_d}{2}\right)} \frac{\nu_n^{\nu_n/2}}{\nu_d^{\nu_d/2}} \frac{1}{\left(1 + \frac{\nu_n}{\nu_d} F\right)^{(\nu_n + \nu_d)/2}}$$

Rowe, D.B.
10: Inferences Involving Two Populations
10.5 Inference for Ratio of Two Variances Two Ind. Samples
Hypothesis Testing Procedure

Test Statistic for Equality of Variances

\[ F^* = \frac{s^2_n}{s^2_d} \quad \text{with} \quad df_n = n_n - 1 \quad \text{and} \quad df_d = n_d - 1 . \tag{10.16} \]

Will also need critical values.

\[ P(F > F(df_n, df_d, \alpha)) = \alpha \]

Table 9
Appendix B
Page 722

Figure from Johnson & Kuby, 2012.
10: Inferences Involving Two Pops.
10.5 Inference Ratio of Two Variances

Example: Find $F(5,8,0.05)$.

Table 9, Appendix B, Page 722.

$\alpha = 0.05$

<table>
<thead>
<tr>
<th>$df_n$</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>18.5</td>
<td>19.0</td>
<td>19.2</td>
<td>19.2</td>
<td>19.3</td>
<td>19.3</td>
<td>19.4</td>
<td>19.4</td>
<td>19.4</td>
<td>19.4</td>
</tr>
<tr>
<td>3</td>
<td>10.1</td>
<td>9.55</td>
<td>9.28</td>
<td>9.12</td>
<td>9.01</td>
<td>8.94</td>
<td>8.89</td>
<td>8.85</td>
<td>8.81</td>
<td>8.79</td>
</tr>
<tr>
<td>4</td>
<td>7.71</td>
<td>6.94</td>
<td>6.59</td>
<td>6.39</td>
<td>6.26</td>
<td>6.16</td>
<td>6.09</td>
<td>6.04</td>
<td>6.00</td>
<td>5.96</td>
</tr>
<tr>
<td>5</td>
<td>6.61</td>
<td>5.79</td>
<td>5.41</td>
<td>5.19</td>
<td>5.05</td>
<td>4.95</td>
<td>4.88</td>
<td>4.82</td>
<td>4.77</td>
<td>4.74</td>
</tr>
<tr>
<td>6</td>
<td>5.99</td>
<td>5.14</td>
<td>4.76</td>
<td>4.53</td>
<td>4.39</td>
<td>4.28</td>
<td>4.21</td>
<td>4.15</td>
<td>4.10</td>
<td>4.06</td>
</tr>
<tr>
<td>7</td>
<td>5.59</td>
<td>4.74</td>
<td>4.35</td>
<td>4.12</td>
<td>3.97</td>
<td>3.87</td>
<td>3.79</td>
<td>3.73</td>
<td>3.68</td>
<td>3.64</td>
</tr>
<tr>
<td>8</td>
<td>5.32</td>
<td>4.46</td>
<td>4.07</td>
<td>3.84</td>
<td>3.69</td>
<td>3.58</td>
<td>3.50</td>
<td>3.44</td>
<td>3.39</td>
<td>3.35</td>
</tr>
<tr>
<td>9</td>
<td>5.12</td>
<td>4.26</td>
<td>3.86</td>
<td>3.63</td>
<td>3.48</td>
<td>3.37</td>
<td>3.29</td>
<td>3.23</td>
<td>3.18</td>
<td>3.14</td>
</tr>
<tr>
<td>10</td>
<td>4.96</td>
<td>4.10</td>
<td>3.71</td>
<td>3.48</td>
<td>3.33</td>
<td>3.22</td>
<td>3.14</td>
<td>3.07</td>
<td>3.02</td>
<td>2.98</td>
</tr>
</tbody>
</table>

Figures from Johnson & Kuby, 2012.

Rowe, D.B.
10: Inferences Involving Two Populations
10.5 Inference for Ratio of Two Variances Two Ind. Samples

Hypothesis Testing Procedure

One tailed tests: Arrange $H_0$ & $H_a$ so $H_a$ is always “greater than”

$H_0: \sigma_1^2 \geq \sigma_2^2 \hspace{1em} \text{vs.} \hspace{1em} H_a: \sigma_1^2 < \sigma_2^2 \rightarrow H_0: \frac{\sigma_2^2}{\sigma_1^2} \leq 1 \hspace{1em} \text{vs.} \hspace{1em} H_a: \frac{\sigma_2^2}{\sigma_1^2} > 1 \hspace{1em} F^* = \frac{s_2^2}{s_1^2}$

$H_0: \sigma_1^2 \leq \sigma_2^2 \hspace{1em} \text{vs.} \hspace{1em} H_a: \sigma_1^2 > \sigma_2^2 \rightarrow H_0: \frac{\sigma_2^2}{\sigma_1^2} \leq 1 \hspace{1em} \text{vs.} \hspace{1em} H_a: \frac{\sigma_2^2}{\sigma_1^2} > 1 \hspace{1em} F^* = \frac{s_1^2}{s_2^2}$

Reject $H_0$ if $F^* = \frac{s_n^2}{s_d^2} > F(df_n, df_d, \alpha)$.

Two tailed tests: put larger sample variance $s^2$ in numerator

$H_0: \sigma_1^2 = \sigma_2^2 \hspace{1em} \text{vs.} \hspace{1em} H_a: \sigma_1^2 \neq \sigma_2^2 \rightarrow H_0: \frac{\sigma_n^2}{\sigma_d^2} = 1 \hspace{1em} \text{vs.} \hspace{1em} H_a: \frac{\sigma_n^2}{\sigma_d^2} \neq 1$ $\sigma_n^2 = \sigma_1^2 \hspace{1em} \text{if} \hspace{1em} s_1^2 > s_2^2$ $\sigma_n^2 = \sigma_2^2 \hspace{1em} \text{if} \hspace{1em} s_2^2 > s_1^2$

Reject $H_0$ if $F^* = \frac{s_n^2}{s_d^2} > F(df_n, df_d, \alpha/2)$.
10: Inferences Involving Two Populations

10.5 Inference for Ratio of Two Variances Two Ind. Samples

Hypothesis Testing Procedure

**Step 1**

\[ H_0: \sigma_m^2 \geq \sigma_f^2 \text{ vs. } H_a: \sigma_m^2 < \sigma_f^2 \]
\[ H_0: \sigma_m^2 \leq \sigma_f^2 \text{ vs. } H_a: \sigma_m^2 > \sigma_f^2 \]
\[ H_0: \sigma_m^2 / \sigma_f^2 \leq 1 \text{ vs. } H_a: \sigma_m^2 / \sigma_f^2 > 1 \]

**Step 2**

\[ F^* = \frac{s_m^2}{s_f^2} \quad df_n = 26 \quad df_f = 52 \]
\[ \alpha = .01 \]

**Step 3**

\[ F^* = \frac{9.8}{7.2} = 1.36 \]

**Step 4**

\[ F(26,52,.01) = 2.14 \]

**Step 5**

Fail to Reject \( H_0 \) 1.36 < 2.14
Chapter 10: Inferences Involving Two Populations

Questions?