Class 26

Daniel B. Rowe, Ph.D.

Department of Mathematics, Statistics, and Computer Science
Agenda:

Recap Chapter 10.3 and 10.4

Lecture Chapter 10.5-10.6
Recap Chapter 10.3-10.4
10: Inferences Involving Two Populations

10.3 Inference for Mean Difference Two Dependent Samples

Confidence Interval Procedure

Paired Difference

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

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

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

Confidence Interval for Mean Difference (Dependent Samples)

\[
\bar{d} - t(df, \alpha / 2) \frac{s_d}{\sqrt{n}} \quad \text{to} \quad \bar{d} + t(df, \alpha / 2) \frac{s_d}{\sqrt{n}} \quad \text{where} \quad df=n-1
\]  \hspace{1cm} (10.2)
### 10: Inferences Involving Two Populations

#### 10.3 Inference for Mean Difference Two Dependent Samples

#### Example:

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

<table>
<thead>
<tr>
<th>Car</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>Brand A</td>
<td>125</td>
<td>64</td>
<td>94</td>
<td>38</td>
<td>90</td>
<td>106</td>
</tr>
<tr>
<td>Brand B</td>
<td>133</td>
<td>65</td>
<td>103</td>
<td>37</td>
<td>102</td>
<td>115</td>
</tr>
</tbody>
</table>

8, 1, 9, –1, 12, 9

\[
\bar{d} = \frac{1}{n} \sum_{i=1}^{n} d_i \\
\sigma_d = 6.3 \quad \alpha = 0.05 \\
\bar{d} = 5.1 \\
\bar{d} \pm t(df, \alpha / 2) \frac{s_d}{\sqrt{n}} \rightarrow (0.090, 11.7)
\]

Figure from Johnson & Kuby, 2008.
10: Inferences Involving Two Populations

10.3 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  
\[
\bar{d} = 6.3 \quad \alpha = .05
\]

Step 3  
\[
t^* = \frac{\bar{d} - \mu_{d0}}{s_d / \sqrt{n}}
\]

Step 4  
\[
t^* = \frac{6.3 - 0}{5.1/\sqrt{6}} = 3.03
\]

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

---

Rowe, D.B.

Figures from Johnson & Kuby, 2008.
10: Inferences Involving Two Populations

10.4 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:

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

where $df$ is either calculated or smaller of $df_1$, or $df_2$ (10.8)

Next larger number than

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

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

If using a computer program.

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

10.4 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

Figure from Johnson & Kuby, 2008.
10: Inferences Involving Two Populations

10.4 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

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

\( df = 37 \)

\( \alpha = .05 \)

Step 3

\[
t^* = \frac{(69.8 - 64.9) - (0)}{\sqrt{\left(\frac{15.8}{38}\right) + \left(\frac{9.2}{52}\right)}} = 6.42
\]

Step 4

\( t(df, \alpha / 2) = 2.03 \)

Step 5

Reject \( H_0 \) if \( t^* > t(df, \alpha / 2) \)

MW

90 values

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
9: Inferences Involving One Population
9.3 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.

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

- \( n = \text{number of trials or times we repeat the experiment.} \)
- \( x = \text{the number of successes out of } n \text{ trials.} \)
- \( p = \text{the probability of success on an individual trial.} \)
9: Inferences Involving One Population

9.3 Inference about the Binomial Probability of Success

When we perform a binomial experiment we can estimate the probability of heads as

\[ p' = \frac{x}{n} \]

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
10.5 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 \( \text{variance}(cx) = c^2npq \).
\[ \rightarrow \text{mean}(x/n) = p \text{ and variance}(x/n) = pq/n. \]

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

if \( x_1 \) & \( x_2 \) independent

Rowe, D.B.
10: Inferences Involving Two Populations
10.5 Inference for Difference between Two Proportions

That is where 1. and 2. in the orange 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
10.5 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)
10: Inferences Involving Two Populations
10.5 Inference for Difference between Two Proportions

Confidence Interval Procedure

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

116 values

$n_m = 48$

$n_f = 68$

$x_m = 20$

$x_f = 42$

$p_f' = \frac{x_f}{n_f} = \frac{42}{68} = .62$

$p_m' = \frac{x_m}{n_m} = \frac{20}{42} = .48$

$z(\alpha / 2) = 2.05$

$(p_f' - p_m') \pm z(\alpha / 2) \sqrt{\frac{p_f'q_f'}{n_f} + \frac{p_m'q_m'}{n_m}}$

$(.62)(.48) (.46)(.62) 2.05 \sqrt{\frac{(.62)(.48)}{68} + \frac{(.46)(.62)}{48}}$

$.009$ to $.431$
10: Inferences Involving Two Populations

10.5 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 \( p_1 = p_2 = p \).
10: Inferences Involving Two Populations
10.5 Inference for Difference between Two Proportions

Hypothesis Testing Procedure

where we assume \( p_1 = p_2 \) and use pooled estimate of proportion \( \hat{p} \) estimated

Test Statistic for the Difference between two Proportions UnKnown

\[
\begin{align*}
Z^* &= \frac{(p'_1 - p'_2) - (p_{10} - p_{20})}{\sqrt{p'_p q'_p \left[ \frac{1}{n_1} + \frac{1}{n_2} \right]}} \\
&= \frac{0}{\sqrt{p'_p q'_p \left[ \frac{1}{n_1} + \frac{1}{n_2} \right]}}
\end{align*}
\] (10.15)

where \( p'_1 = \frac{x_1}{n_1} \), \( p'_2 = \frac{x_2}{n_2} \), \( p'_p = \frac{x_1 + x_2}{n_1 + n_2} \)

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

10.5 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

\[
 z^* = \frac{(p_s' - p_c') - (p_{0s} - p_{0c})}{\sqrt{p_p'q_p' \left[ \frac{1}{n_s} + \frac{1}{n_c} \right]}}
\]

\( p_s' = \frac{x_s}{n_s} = \frac{15}{150} \quad \text{and} \quad p_c' = \frac{x_c}{n_c} = \frac{6}{150} \)

\( p_p' = \frac{x_s + x_c}{n_s + n_c} = \frac{15 + 6}{150 + 150} \)

\( \alpha = .05 \)

Step 3

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

Step 4

\( z(\alpha) = 1.65 \)

Step 5

Reject \( H_0 \) if \( z > 1.65 \) or \( p-value < .05 \)

\( .02 < p-value < .023 \) or \( 2.04 > 1.65 \)

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

We can perform hypothesis tests on two variances

\[ H_0: \sigma_1^2 \geq \sigma_2^2 \quad \text{vs.} \quad H_a: \sigma_1^2 < \sigma_2^2 \]

\[ H_0: \sigma_1^2 \leq \sigma_2^2 \quad \text{vs.} \quad H_a: \sigma_1^2 > \sigma_2^2 \]

\[ H_0: \sigma_1^2 = \sigma_2^2 \quad \text{vs.} \quad H_a: \sigma_1^2 \neq \sigma_2^2 \]

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 \( df_d = n_d - 1 \).

\[ (10.16) \]

Use new table to find areas for new statistic.
Properties of F distribution

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

$$df_n = \nu_n = n_n - 1, \quad df_d = \nu_d = n_d - 1.$$
10: Inferences Involving Two Populations
10.6 Inference for Ratio of Two Variances Two Ind. Samples

Hypothesis Testing Procedure

Will also need critical values.

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

Table 9
Appendix B
Page 668

Figure from Johnson & Kuby, 2008.
10: Inferences Involving Two Pops.

10.6 Inference Ratio of Two Variances

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

$$df_n = n_n - 1 \quad df_d = n_d - 1$$

Table 9, Appendix B, Page 668.

<table>
<thead>
<tr>
<th>$df_d$</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>

$\alpha = 0.05$

Figures from Johnson & Kuby, 2008.

Rowe, D.B.
10: Inferences Involving Two Populations
10.6 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”

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

$H_0: \sigma_1^2 \leq \sigma_2^2$ vs. $H_a: \sigma_1^2 > \sigma_2^2$ $\rightarrow$ $H_0: \sigma_1^2 / \sigma_2^2 \leq 1$ vs. $H_a: \sigma_1^2 / \sigma_2^2 > 1$  \[ F^* = \frac{s_1^2}{s_2^2} \]

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

$H_0: \sigma_1^2 = \sigma_2^2$ vs. $H_a: \sigma_1^2 \neq \sigma_2^2$ $\rightarrow$ $H_0: \sigma_n^2 / \sigma_d^2 = 1$ vs. $H_a: \sigma_n^2 / \sigma_d^2 \neq 1$

$\sigma_n^2 = \sigma_1^2$ if $s_1^2 > s_2^2$, $\sigma_n^2 = \sigma_2^2$ if $s_2^2 > s_1^2$

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

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

10.6 Inference for Ratio of Two Variances Two Ind. Samples

Hypothesis Testing Procedure

Step 1

\[ H_0: \sigma_m^2 \geq \sigma_f^2 \quad \text{vs.} \quad H_a: \sigma_m^2 < \sigma_f^2 \]

\[ H_0: \sigma_m^2 \leq \sigma_f^2 \quad \text{vs.} \quad H_a: \sigma_m^2 > \sigma_f^2 \]

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

Step 2

\[ F^* = \frac{s_m^2}{s_f^2} \quad \text{df}_m = 37 \]

\[ \text{df}_f = 51 \]

\[ \alpha = .01 \]

Step 3

\[ F^* = 15.8 / 9.2 = 1.72 \]

Step 4

\[ F(37,51,.01) = 2.02 \]

Step 5

Fail to Reject \( H_0 \) \( 1.72 < 2.02 \)
Chapter 10: Inferences Involving Two Populations

Questions?

Homework: Chapter 10 # 5, 7, 9, 11,
13, 17, 23, 25, 31, 35
49, 57, 65, 71
81, 83, 87, 91, 93, 99
109, 111, 113, 115, 119