Class 28

Daniel B. Rowe, Ph.D.

Department of Mathematics, Statistics, and Computer Science
Agenda:

Recap Chapter 10.5 and 10.6

Lecture Chapter 11.1-11.2
Recap Chapter 10.5 and 10.6
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 
\( p_1 = P_1 \) (success) and \( p_2 = P_2 \) (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 \( \sigma_{p_1' - p_2'} = \sqrt{\frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}} \) (10.10)
3. approximately normal dist if \( n_1 \) and \( n_2 \) are sufficiently large.

ie \( \text{I } n_1, n_2 > 20 \text{ II } n_1 p_1, n_1 q_1, n_2 p_2, n_2 q_2 > 5 \text{ III sample<10% of pop} \)
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 \( p_1 - p_2 \)

\[
(p_1' - p_2') - z(\alpha / 2) \sqrt{\frac{p_1'q_1'}{n_1} + \frac{p_2'q_2'}{n_2}} ~ \text{to} ~ (p_1' - p_2') + z(\alpha / 2) \sqrt{\frac{p_1'q_1'}{n_1} + \frac{p_2'q_2'}{n_2}}
\]

where \( p_1' = \frac{x_1}{n_1} \) and \( p_2' = \frac{x_2}{n_2} \).

\( (10.11) \)
10: Inferences Involving Two Populations

10.4 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 \( p_f - p_m \).

40 values

\( n_m = 9 \)
\( n_f = 31 \)
\( x_m = 2 \)
\( x_f = 11 \)

\( z(\alpha / 2) = 2.58 \)

\( p_f' = \frac{x_f}{n_f} = \frac{11}{31} = .35 \)

\( p_m' = \frac{x_m}{n_m} = \frac{2}{9} = .22 \)

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

\( (.35 - .22) \pm 2.05 \sqrt{\frac{(.35)(.65)}{31} + \frac{(.22)(.78)}{9}} \)

\(-.287 \text{ to } .553 \)
10: Inferences Involving Two Populations

10.4 Inference for Difference between Two Proportions

Hypothesis Testing Procedure

We can perform hypothesis tests on the proportion

\( H_0: p_1 \geq p_2 \) vs. \( H_a: p_1 < p_2 \)

\( H_0: p_1 \leq p_2 \) vs. \( H_a: p_1 > p_2 \)

\( H_0: p_1 = p_2 \) vs. \( H_a: p_1 \neq p_2 \)

\[
\frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2} = pq \left[ \frac{1}{n_1} + \frac{1}{n_2} \right]
\]

when \( p_1 = p_2 = p \).

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

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

\[
p'_1 = \frac{x_1}{n_1} \quad p'_2 = \frac{x_2}{n_2}
\]
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 \( \hat{p}_p \) estimated

Test Statistic for the Difference between two Proportions UnKnown

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

\[
(10.15)
\]

\[
p_1' = \frac{x_1}{n_1} \quad p_2' = \frac{x_2}{n_2} \quad \frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2} = p q \left[ \frac{1}{n_1} + \frac{1}{n_2} \right] \quad \hat{p}_p = \frac{x_1 + x_2}{n_1 + n_2}
\]
10: Inferences Involving Two Populations

10.4 Inference for Difference between Two Proportions

Hypothesis Testing Procedure

Step 1

\[ H_0: p_s - p_c \leq 0 \text{ vs. } H_a: p_s - p_c > 0 \]

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 \( z < .05 \)

\[ .02 < p-value < .023 \text{ or } 2.04 > 1.65 \]

<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>

\[ p_s' = \frac{x_s}{n_s} = \frac{15}{150} \]

\[ p_c' = \frac{x_c}{n_c} = \frac{6}{150} \]

\[ p_p' = \frac{x_s + x_c}{n_s + n_c} = \frac{15 + 6}{150 + 150} \]

Figure from Johnson & Kuby, 2012.
We can perform hypothesis tests on two variances

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

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

\[ H_0 : \sigma_1^2 = \sigma_2^2 \] \quad 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.

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

10.5 Inference Ratio of Two Variances

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

\[ df_n = n_n - 1 \quad \text{and} \quad df_d = n_d - 1 \]

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>1</td>
<td>161.0</td>
<td>200.0</td>
<td>216.0</td>
<td>225.0</td>
<td>230.0</td>
<td>234.0</td>
<td>237.0</td>
<td>239.0</td>
<td>241.0</td>
<td>242.0</td>
</tr>
<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.
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”

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

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

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. $H_a: \sigma_1^2 \neq \sigma_2^2$ \quad \rightarrow \quad 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

10.5 Inference for Ratio of Two Variances Two Ind. Samples

Hypothesis Testing Procedure

Step 1

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

Step 2

\[ F^* = \frac{s_f^2}{s_m^2} \quad \text{df}_f = 23 \]
\[ F^* = \frac{s_f^2}{s_m^2} \quad \text{df}_f = 7 \]
\[ \alpha = .01 \]

Step 3

\[ F^* = \frac{8.2}{7.4} = 1.12 \]

Step 4

\[ F(23,7,.01) = 6.09 \]

Step 5

Fail to Reject \( H_0 \) \( 1.12 < 6.09 \)
Chapter 10: Inferences Involving Two Populations

Questions?

Lecture Chapter 11.1-11.3
Chapter 11: Applications of Chi-Square

Daniel B. Rowe, Ph.D.

Department of Mathematics, Statistics, and Computer Science
Data: The set of values collected from the variable from each of the elements that belong to the sample.
11: Applications of Chi-Square

11.1 Chi-Square Statistic Cooling a Great Hot Taste

Quite often we have qualitative data in categories.

**Example**: Cooling mouth after hot spicy food.

<table>
<thead>
<tr>
<th>Method</th>
<th>Water</th>
<th>Bread</th>
<th>Milk</th>
<th>Beer</th>
<th>Soda</th>
<th>Nothing</th>
<th>Other</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number</td>
<td>73</td>
<td>29</td>
<td>35</td>
<td>19</td>
<td>20</td>
<td>13</td>
<td>11</td>
</tr>
</tbody>
</table>
11: Applications of Chi-Square

11.1 Chi-Square Statistic

Data Setup

Example: Cooling mouth after hot spicy food.

<table>
<thead>
<tr>
<th>Method</th>
<th>Water</th>
<th>Bread</th>
<th>Milk</th>
<th>Beer</th>
<th>Soda</th>
<th>Nothing</th>
<th>Other</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number</td>
<td>73</td>
<td>29</td>
<td>35</td>
<td>19</td>
<td>20</td>
<td>13</td>
<td>11</td>
</tr>
</tbody>
</table>

Data set up: $k$ cells $C_1,\ldots,C_k$ that $n$ observations sorted into

- Observed frequencies in each cell $O_1,\ldots,O_k$.
- Expected frequencies in each cell $E_1,\ldots,E_k$.

<table>
<thead>
<tr>
<th>Cell</th>
<th>$C_1$</th>
<th>$C_2$</th>
<th>.</th>
<th>.</th>
<th>.</th>
<th>$C_k$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Observed</td>
<td>$O_1$</td>
<td>$O_2$</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>$O_k$</td>
</tr>
<tr>
<td>Expected</td>
<td>$E_1$</td>
<td>$E_2$</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>$E_k$</td>
</tr>
</tbody>
</table>
11: Applications of Chi-Square

11.1 Chi-Square Statistic

Outline of Test Procedure

When we have observed cell frequencies $O_1, \ldots, O_k$, we can test to see if they match with some expected cell frequencies $E_1, \ldots, E_k$.

Test Statistic for Chi-Square

$$\chi^2* = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i}$$

$$df = k - 1$$

If the $O_i$’s are different from $E_i$’s then $\chi^2*$ is “large.” Go through 5 hypothesis testing steps as before.

Figure from Johnson & Kuby, 2012.
11: Applications of Chi-Square

11.1 Chi-Square Statistic

Assumption for using the chi-square statistic to make inferences based upon enumerative data: … a random sample drawn from a population where each individual is classified according to the categories

\[
\chi^2* = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i}
\]

\[
df = k - 1
\]

observed cell frequencies \(O_1, \ldots, O_k\), expected cell frequencies \(E_1, \ldots, E_k\).
11: Applications of Chi-Square
11.2 Inferences Concerning Multinomial Experiments

Example: We work for Las Vegas Gaming Commission. We have received many complaints that dice at particular casino are loaded (weighted, not fair). We confiscate all dice and test one. We roll it \( n = 60 \) times. We get following data.

<table>
<thead>
<tr>
<th>Cell, ( i )</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>Observed, ( O_i )</td>
<td>7</td>
<td>12</td>
<td>10</td>
<td>12</td>
<td>8</td>
<td>11</td>
</tr>
<tr>
<td>Expected, ( E_i )</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
</tr>
</tbody>
</table>

Expected Value for Multinomial Experiment:

\[
E_i = np_i
\]  

(11.3)
11: Applications of Chi-Square
11.2 Inferences Concerning Multinomial Experiments

Example: We roll it $n=60$ times. We get following data.

<table>
<thead>
<tr>
<th>Cell, $i$</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>Observed, $O_i$</td>
<td>7</td>
<td>12</td>
<td>10</td>
<td>12</td>
<td>8</td>
<td>11</td>
</tr>
<tr>
<td>Expected, $E_i$</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
</tr>
</tbody>
</table>

$E_i = 60(1/6)$

Is the die fair? Need to go through the hypothesis testing procedure to determine if it is fair.
### 11: Applications of Chi-Square

#### 11.2 Inferences Concerning Multinomial Experiments

**Example:** Is the die fair?

Calculating $\chi^2$

<table>
<thead>
<tr>
<th>Number</th>
<th>Observed ($O$)</th>
<th>Expected ($E$)</th>
<th>$O - E$</th>
<th>$(O - E)^2$</th>
<th>$\frac{(O - E)^2}{E}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>7</td>
<td>10</td>
<td>-3</td>
<td>9</td>
<td>0.9</td>
</tr>
<tr>
<td>2</td>
<td>12</td>
<td>10</td>
<td>2</td>
<td>4</td>
<td>0.4</td>
</tr>
<tr>
<td>3</td>
<td>10</td>
<td>10</td>
<td>0</td>
<td>0</td>
<td>0.0</td>
</tr>
<tr>
<td>4</td>
<td>12</td>
<td>10</td>
<td>2</td>
<td>4</td>
<td>0.4</td>
</tr>
<tr>
<td>5</td>
<td>8</td>
<td>10</td>
<td>-2</td>
<td>4</td>
<td>0.4</td>
</tr>
<tr>
<td>6</td>
<td>11</td>
<td>10</td>
<td>1</td>
<td>1</td>
<td>0.1</td>
</tr>
</tbody>
</table>

**Total** | **60** | **60** | **0** | **0** | **0** |

$O_1 + \ldots + O_k = n$  \hspace{1cm} $E_1 + \ldots + E_k = n$

\[ \chi^2* = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i} \]

$D$ of $F$ for Mult: $df = k - 1$ \hspace{1cm} (11.2)

Figure from Johnson & Kuby, 2012.
11: Applications of Chi-Square

11.2 Inference for Mean Difference Two Dependent Samples

<table>
<thead>
<tr>
<th>Cell</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>Observed</td>
<td>7</td>
<td>12</td>
<td>10</td>
<td>12</td>
<td>8</td>
<td>11</td>
</tr>
<tr>
<td>Expected</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
</tr>
</tbody>
</table>

Observed different than expected?

Step 1

\[ H_0: \text{Die fair } p_i's = \frac{1}{6} \quad H_a: \text{Die not fair } p_i's \neq \frac{1}{6} \]

Step 2

\[ \chi^2* = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i} \quad df = k - 1 \]

Step 3

\[ \chi^2* = 2.2 \quad \chi^2(df, \alpha) = 11.1 \]

Step 5

Since \( .05 < p-value = .82 \) or because \( \chi^2* < \chi^2(df, \alpha) \), fail to reject \( H_0 \)

Figures from Johnson & Kuby, 2012.

Rowe, D.B.
Chapter 11: Applications of Chi-Square

Questions?

Homework: Chapter 11 # 3, 5, 11, 15, 21, 49, 53