Lecture 28
Chi-Square Analysis

STAT 225 Introduction to Probability Models
April 23, 2014

Whitney Huang
Purdue University
**Chi-Square Analysis**

**$\chi^2$ test for two qualitative random variables**

- For a given contingency table, we want to test if two variables have a relationship or not.
- To answer this question, we need to perform a statistical hypothesis test.
- A statistical hypothesis test is a method of making decisions using data from a scientific study.
- A $\chi^2$ test is suitable for answering whether two qualitative variables have a relationship or not.
For a given contingency table, we want to test if two variables have a relationship or not.

To answer this question, we need to perform a statistical hypothesis test.

A statistical hypothesis test is a method of making decisions using data from a scientific study.

A $\chi^2$ test is suitable for answering whether two qualitative variables have a relationship or not.
Chi-Square Analysis

$\chi^2$ test for two qualitative random variables

- For a given contingency table, we want to test if two variables have a relationship or not.
- To answer this question, we need to perform a statistical hypothesis test.
- A statistical hypothesis test is a method of making decisions using data from a scientific study.
- A $\chi^2$ test is suitable for answering whether two qualitative variables have a relationship or not.
For a given contingency table, we want to test if two variables have a relationship or not.

To answer this question, we need to perform a statistical hypothesis test.

A statistical hypothesis test is a method of making decisions using data from a scientific study.

A $\chi^2$ test is suitable for answering whether two qualitative variables have a relationship or not.
The procedure of $\chi^2$ test for two qualitative random variables:

1. Define the Null ($H_0$) and Alternative ($H_A$) hypotheses
   - $H_0$: there is no relationship between the 2 variables
   - $H_A$: there is a relationship between the 2 variables

2. (If necessary) Calculate the marginal totals, and the grand total

3. Calculate the expected cell counts
   \[ \text{expected cell count} = \frac{\text{Row Total} \times \text{Column Total}}{\text{Grand Total}} \]

4. Calculate the partial $\chi^2$ values (a $\chi^2$ value for each cell of the table)
   \[ \text{partial } \chi^2 \text{ value} = \frac{(\text{observed} - \text{expected})^2}{\text{expected}} \]
The procedure of $\chi^2$ test for two qualitative random variables:

1. Define the Null ($H_0$) and Alternative ($H_A$) hypotheses
   - $H_0$ : there is no relationship between the 2 variables
   - $H_A$ : there is a relationship between the 2 variables

2. (If necessary) Calculate the marginal totals, and the grand total

3. Calculate the expected cell counts
   $$\text{expected cell count} = \frac{\text{Row Total} \times \text{Column Total}}{\text{Grand Total}}$$

4. Calculate the partial $\chi^2$ values (a $\chi^2$ value for each cell of the table)
   $$\text{partial } \chi^2 \text{ value} = \frac{(\text{observed} - \text{expected})^2}{\text{expected}}$$
The procedure of $\chi^2$ test for two qualitative random variables:

1. Define the Null ($H_0$) and Alternative ($H_A$) hypotheses
   - $H_0$: there is no relationship between the 2 variables
   - $H_A$: there is a relationship between the 2 variables

2. (If necessary) Calculate the marginal totals, and the grand total

3. Calculate the expected cell counts
   \[
   \text{expected cell count} = \frac{\text{Row Total} \times \text{Column Total}}{\text{Grand Total}}
   \]

4. Calculate the partial $\chi^2$ values (a $\chi^2$ value for each cell of the table)
   \[
   \text{partial } \chi^2 \text{ value} = \frac{(\text{observed} - \text{expected})^2}{\text{expected}}
   \]
The procedure of $\chi^2$ test for two qualitative random variables:

1. Define the Null ($H_0$) and Alternative ($H_A$) hypotheses
   - $H_0$: there is no relationship between the 2 variables
   - $H_A$: there is a relationship between the 2 variables

2. (If necessary) Calculate the marginal totals, and the grand total

3. Calculate the expected cell counts
   \[
   \text{expected cell count} = \frac{\text{Row Total} \times \text{Column Total}}{\text{Grand Total}}
   \]

4. Calculate the partial $\chi^2$ values (a $\chi^2$ value for each cell of the table)
   \[
   \text{partial } \chi^2 \text{ value} = \frac{(\text{observed} - \text{expected})^2}{\text{expected}}
   \]
**Chi-Square Analysis**

**Statistical analysis of two qualitative random variables**

χ² test cont’d

5. Calculate the χ² statistic
   \[ \chi^2 = \sum \text{partial } \chi^2 \text{ value} \]

6. Calculate the degrees of freedom (df)
   \[ df = (\# \text{of rows} - 1) \times (\# \text{of columns} - 1) \]

7. Find the χ² critical value with respect to α from the χ² table

8. Draw your conclusion:
   Reject \( H_0 \) if your χ² statistic is bigger than the χ² critical value
   ⇒ There is an statistical evidence that there is a relationship between the 2 variables at \( \alpha \) level
**Chi-Square Analysis**

**Statistical analysis of two qualitative random variables**

χ² test cont’d

5. Calculate the χ² statistic
   \[ \chi^2 = \sum \text{partial } \chi^2 \text{ value} \]

6. Calculate the degrees of freedom (df)
   \[ df = (\# \text{ of rows} - 1) \times (\# \text{ of columns} - 1) \]

7. Find the χ² critical value with respect to \( \alpha \) from the χ² table

8. Draw your conclusion:
   Reject \( H_0 \) if your \( \chi^2 \) statistic is bigger than the \( \chi^2 \) critical value
   \( \Rightarrow \) There is statistical evidence that there is a relationship between the 2 variables at \( \alpha \) level
Chi-Square Analysis

5. Calculate the \( \chi^2 \) statistic
\[ \chi^2 = \sum \text{partial } \chi^2 \text{ value} \]

6. Calculate the degrees of freedom (\( df \))
\[ df = (# \text{of rows} - 1) \times (# \text{of columns} - 1) \]

7. Find the \( \chi^2 \) critical value with respect to \( \alpha \) from the \( \chi^2 \) table

8. Draw your conclusion:
Reject \( H_0 \) if your \( \chi^2 \) statistic is bigger than the \( \chi^2 \) critical value \( \Rightarrow \) There is statistical evidence that there is a relationship between the 2 variables at \( \alpha \) level.
\( \chi^2 \) test cont’d

5. Calculate the \( \chi^2 \) statistic
   \[ \chi^2 = \sum \text{partial } \chi^2 \text{ value} \]

6. Calculate the degrees of freedom (\( df \))
   \[ df = (\# \text{ of rows} - 1) \times (\# \text{ of columns} - 1) \]

7. Find the \( \chi^2 \) critical value with respect to \( \alpha \) from the \( \chi^2 \) table

8. Draw your conclusion:
   Reject \( H_0 \) if your \( \chi^2 \) statistic is bigger than the \( \chi^2 \) critical value \( \Rightarrow \) There is an statistical evidence that there is a relationship between the 2 variables at \( \alpha \) level
Example 67

A 2011 study was conducted in Kalamazoo, Michigan. The objective was to determine if parents’ marital status affects children’s marital status later in their life. In total, 2,000 children were interviewed. The columns refer to the parents’ marital status. Use the contingency table below to conduct a $\chi^2$ test from beginning to end. Use $\alpha = .10$

<table>
<thead>
<tr>
<th>(Observed)</th>
<th>Married</th>
<th>Divorced</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Married</td>
<td>581</td>
<td>487</td>
<td></td>
</tr>
<tr>
<td>Divorced</td>
<td>455</td>
<td>477</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Example 67 cont’d

1. Define the Null and Alternative hypotheses:
   \( H_0 \): there is no relationship between parents’ marital status and children’s marital status
   \( H_A \): there is a relationship between parents’ marital status and children’s marital status

2. Calculate the marginal totals, and the grand total

<table>
<thead>
<tr>
<th>(Observed)</th>
<th>Married</th>
<th>Divorced</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Married</td>
<td>581</td>
<td>487</td>
<td>1068</td>
</tr>
<tr>
<td>Divorced</td>
<td>455</td>
<td>477</td>
<td>932</td>
</tr>
<tr>
<td>Total</td>
<td>1036</td>
<td>964</td>
<td>2000</td>
</tr>
</tbody>
</table>
Example 67 cont’d

1. Define the Null and Alternative hypotheses:
   \( H_0 \): there is no relationship between parents’ marital status and children’s marital status
   \( H_A \): there is a relationship between parents’ marital status and children’s marital status

2. Calculate the marginal totals, and the grand total

<table>
<thead>
<tr>
<th></th>
<th>Married</th>
<th>Divorced</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Married</td>
<td>581</td>
<td>487</td>
<td>1068</td>
</tr>
<tr>
<td>Divorced</td>
<td>455</td>
<td>477</td>
<td>932</td>
</tr>
<tr>
<td>Total</td>
<td>1036</td>
<td>964</td>
<td>2000</td>
</tr>
</tbody>
</table>
Example 67 cont’d

1. Calculate the expected cell counts

<table>
<thead>
<tr>
<th>(Expected)</th>
<th>Married</th>
<th>Divorced</th>
</tr>
</thead>
<tbody>
<tr>
<td>Married</td>
<td>$\frac{1068 \times 1036}{2000} = 553.224$</td>
<td>$\frac{1068 \times 964}{2000} = 514.776$</td>
</tr>
<tr>
<td>Divorced</td>
<td>$\frac{932 \times 1036}{2000} = 482.776$</td>
<td>$\frac{932 \times 964}{2000} = 449.224$</td>
</tr>
</tbody>
</table>

2. Calculate the partial $\chi^2$ values

<table>
<thead>
<tr>
<th>partial $\chi^2$</th>
<th>Married</th>
<th>Divorced</th>
</tr>
</thead>
<tbody>
<tr>
<td>Married</td>
<td>$\frac{(581-553.224)^2}{553.224} = 1.39$</td>
<td>$\frac{(487-514.776)^2}{514.776} = 1.50$</td>
</tr>
<tr>
<td>Divorced</td>
<td>$\frac{(455-482.776)^2}{482.776} = 1.60$</td>
<td>$\frac{(477-449.224)^2}{449.224} = 1.72$</td>
</tr>
</tbody>
</table>
Example 67 cont’d

3 Calculate the expected cell counts

<table>
<thead>
<tr>
<th>(Expected)</th>
<th>Married</th>
<th>Divorced</th>
</tr>
</thead>
<tbody>
<tr>
<td>Married</td>
<td>$\frac{1068 \times 1036}{2000} = 553.224$</td>
<td>$\frac{1068 \times 964}{2000} = 514.776$</td>
</tr>
<tr>
<td>Divorced</td>
<td>$\frac{932 \times 1036}{2000} = 482.776$</td>
<td>$\frac{932 \times 964}{2000} = 449.224$</td>
</tr>
</tbody>
</table>

4 Calculate the partial $\chi^2$ values

<table>
<thead>
<tr>
<th>partial $\chi^2$</th>
<th>Married</th>
<th>Divorced</th>
</tr>
</thead>
<tbody>
<tr>
<td>Married</td>
<td>$\frac{(581 - 553.224)^2}{553.224} = 1.39$</td>
<td>$\frac{(487 - 514.776)^2}{514.776} = 1.50$</td>
</tr>
<tr>
<td>Divorced</td>
<td>$\frac{(455 - 482.776)^2}{482.776} = 1.60$</td>
<td>$\frac{(477 - 449.224)^2}{449.224} = 1.72$</td>
</tr>
</tbody>
</table>
Example 67 cont’d

5. Calculate the $\chi^2$ statistic
$\chi^2 = 1.39 + 1.50 + 1.60 + 1.72 = 6.21$

6. Calculate the degrees of freedom ($df$)
The $df$ is $(2 - 1) \times (2 - 1) = 1$

7. Find the $\chi^2$ critical value with respect to $\alpha$ from the $\chi^2$ table
The $\chi^2_{\alpha=0.1, df=1} = 2.71$

8. Draw your conclusion:
We reject $H_0$ and conclude that there is a relationship between parents’ marital status and children’s marital status.
Example 67 cont’d

5. Calculate the $\chi^2$ statistic
\[ \chi^2 = 1.39 + 1.50 + 1.60 + 1.72 = 6.21 \]

6. Calculate the degrees of freedom ($df$)
The $df$ is $(2 - 1) \times (2 - 1) = 1$

7. Find the $\chi^2$ critical value with respect to $\alpha$ from the $\chi^2$ table
The $\chi^2_{\alpha=0.1, df=1} = 2.71$

8. Draw your conclusion:
We reject $H_0$ and conclude that there is a relationship between parents’ marital status and childrens’ marital status.
Example 67 cont’d

5. Calculate the $\chi^2$ statistic
   $\chi^2 = 1.39 + 1.50 + 1.60 + 1.72 = 6.21$

6. Calculate the degrees of freedom ($df$)
The $df$ is $(2 - 1) \times (2 - 1) = 1$

7. Find the $\chi^2$ critical value with respect to $\alpha$ from the $\chi^2$ table
   The $\chi^2_{\alpha=0.1, df=1} = 2.71$

8. Draw your conclusion:
   We reject $H_0$ and conclude that there is a relationship between parents’ marital status and childrens’ marital status.
Example 67 cont’d

5. Calculate the $\chi^2$ statistic
   
   $\chi^2 = 1.39 + 1.50 + 1.60 + 1.72 = 6.21$

6. Calculate the degrees of freedom ($df$)
   
   The $df$ is $(2 - 1) \times (2 - 1) = 1$

7. Find the $\chi^2$ critical value with respect to $\alpha$ from the $\chi^2$ table
   
   The $\chi^2_{\alpha=0.1, df=1} = 2.71$

8. Draw your conclusion:
   
   We reject $H_0$ and conclude that there is a relationship between parents’ marital status and childrens’ marital status.
Example 68

The following contingency table contains enrollment data for a random sample of students from several colleges at Purdue University during the 2006-2007 academic year. The table lists the number of male and female students enrolled in each college. Use the two-way table to conduct a $\chi^2$ test from beginning to end. Use $\alpha = .01$

<table>
<thead>
<tr>
<th></th>
<th>Female</th>
<th>Male</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Liberal Arts</td>
<td>378</td>
<td>262</td>
<td>640</td>
</tr>
<tr>
<td>Science</td>
<td>99</td>
<td>175</td>
<td>274</td>
</tr>
<tr>
<td>Engineering</td>
<td>104</td>
<td>510</td>
<td>614</td>
</tr>
<tr>
<td>Total</td>
<td>581</td>
<td>947</td>
<td>1528</td>
</tr>
</tbody>
</table>
### Example 68 cont’d

<table>
<thead>
<tr>
<th></th>
<th>Female</th>
<th>Male</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Expected)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Liberal Arts</td>
<td>$\frac{640 \times 581}{1528} = 243.35$</td>
<td>$\frac{640 \times 947}{1528} = 396.65$</td>
</tr>
<tr>
<td>Science</td>
<td>$\frac{274 \times 581}{1528} = 104.18$</td>
<td>$\frac{274 \times 947}{1528} = 169.82$</td>
</tr>
<tr>
<td>Engineering</td>
<td>$\frac{614 \times 581}{1528} = 233.46$</td>
<td>$\frac{614 \times 947}{1528} = 380.54$</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Female</th>
<th>Male</th>
</tr>
</thead>
<tbody>
<tr>
<td>partial $\chi^2$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lib Arts</td>
<td>$\frac{(378 - 243.35)^2}{243.35} = 74.50$</td>
<td>$\frac{(262 - 396.65)^2}{396.65} = 45.71$</td>
</tr>
<tr>
<td>Sci</td>
<td>$\frac{(99 - 104.18)^2}{104.18} = 0.26$</td>
<td>$\frac{(175 - 169.82)^2}{169.82} = 0.16$</td>
</tr>
<tr>
<td>Eng</td>
<td>$\frac{(104 - 233.46)^2}{233.46} = 71.79$</td>
<td>$\frac{(510 - 380.54)^2}{380.54} = 44.05$</td>
</tr>
</tbody>
</table>

$$\chi^2 = 74.50 + 45.71 + 0.26 + 0.16 + 71.79 + 44.05 = 236.47$$

The $df = (3 - 1) \times (2 - 1) = 2$

The $\chi^2_{0.01, df=2} = 9.21$

We reject $H_0$ and conclude that there is a relationship between gender and major.