Lecture 33

Analysis of Bivariate Categorical Data

STAT 8010 Statistical Methods I
November 13, 2019

Whitney Huang
Clemson University

Example

A psychologist is interested in whether or not handedness is related to gender. She collected data on handedness for 100 individuals and the data set is summarized in the table below

	Right-handed	Left-handed	Total
Males	43	9	52
Females	44	4	48
Total	87	13	100

Example

A psychologist is interested in whether or not handedness is related to gender. She collected data on handedness for 100 individuals and the data set is summarized in the table below

	Right-handed	Left-handed	Total
Males	43	9	52
Females	44	4	48
Total	87	13	100

- Grand total: 100

Example

A psychologist is interested in whether or not handedness is related to gender. She collected data on handedness for 100 individuals and the data set is summarized in the table below

	Right-handed	Left-handed	Total
Males	43	9	52
Females	44	4	48
Total	87	13	100

- Grand total: 100
- Marginal total for males: 52

Example

A psychologist is interested in whether or not handedness is related to gender. She collected data on handedness for 100 individuals and the data set is summarized in the table below

	Right-handed	Left-handed	Total
Males	43	9	52
Females	44	4	48
Total	87	13	100

- Grand total: 100
- Marginal total for males: 52
- Marginal total for females: 48

Example

A psychologist is interested in whether or not handedness is related to gender. She collected data on handedness for 100 individuals and the data set is summarized in the table below

	Right-handed	Left-handed	Total
Males	43	9	52
Females	44	4	48
Total	87	13	100

- Grand total: 100
- Marginal total for males: 52
- Marginal total for females: 48
- Marginal total for right-handed: 87

Example

A psychologist is interested in whether or not handedness is related to gender. She collected data on handedness for 100 individuals and the data set is summarized in the table below

	Right-handed	Left-handed	Total
Males	43	9	52
Females	44	4	48
Total	87	13	100

- Grand total: 100
- Marginal total for males: 52
- Marginal total for females: 48
- Marginal total for right-handed: 87
- Marginal total for left-handed: 13

Example

A psychologist is interested in whether or not handedness is related to gender. She collected data on handedness for 100 individuals and the data set is summarized in the table below

	Right-handed	Left-handed	Total
Males	43	9	52
Females	44	4	48
Total	87	13	100

- Grand total: 100
- Marginal total for males: 52
- Marginal total for females: 48
- Marginal total for right-handed: 87
- Marginal total for left-handed: 13

Example

A psychologist is interested in whether or not handedness is related to gender. She collected data on handedness for 100 individuals and the data set is summarized in the table below

	Right-handed	Left-handed	Total
Males	43	9	52
Females	44	4	48
Total	87	13	100

- Grand total: 100
- Marginal total for males: 52
- Marginal total for females: 48
- Marginal total for right-handed: 87
- Marginal total for left-handed: 13

This is an example of a contingency table

Contingency Tables

- Bivariate categorical data is typically displayed in a contingency table
- The number in each cell is the frequency for each category level combination
- Contingency table for the previous example:

	Right-handed	Left-handed	Total
Males	43	9	52
Females	44	4	48
Total	87	13	100

For a given contingency table, we want to test **if two variables have a relationship or not?** $\Rightarrow \chi^2$-Test

χ^2-Test for Homogeneity (Independence Test)

1. Define the null and alternative 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 frequencies:

$$\text{Expected cell frequency} = \frac{\text{Row Total} \times \text{Column Total}}{\text{Grand Total}}$$

4. Calculate the partial χ^2 values (χ^2 value for each cell of the table):

$$\text{Partial } \chi^2 \text{ value} = \frac{(\text{observed - expected})^2}{\text{expected}}$$

χ^2-Test for Homogeneity Cont'd

5. Calculate the χ^2 statistic:

$$\chi^2_{obs} = \sum \text{partial } \chi^2 \text{ value}$$

6. Calculate the degrees of freedom (df)

$$df = (\#\text{of rows} - 1) \times (\#\text{of columns} - 1)$$

7. Find the χ^2 critical value with respect to α

8. Draw the conclusion:

Reject H_0 if χ^2_{obs} is bigger than the χ^2 critical value $\Rightarrow$ There is an statistical evidence that there is a relationship between the two variables at α level

Handedness/Gender Example Revisited

	Right-handed	Left-handed	Total
Males	43	9	52
Females	44	4	48
Total	87	13	100

Is the percentage left-handed men in the population different from the percentage of left-handed women?

Example

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 χ^2 test from beginning to end. Use $\alpha = .10$

(Observed)	Married	Divorced	Total
Married	581	487	
Divorced	455	477	
Total			

Example Cont'd

1. Define the Null and Alternative hypotheses:

   H_0 : there is no relationship between parents' marital status and childrens' marital status.

   H_a : there is a relationship between parents' marital status and childrens' marital status

2. Calculate the marginal totals, and the grand total

(Observed)	Married	Divorced	Total
Married	581	487	1068
Divorced	455	477	932
Total	1036	964	2000

Example Cont'd

3. Calculate the expected cell counts

(Expected)	Married	Divorced
Married	$\frac{1068\times1036}{2000} = 553.224$	$\frac{1068\times964}{2000} = 514.776$
Divorced	$\frac{932\times1036}{2000} = 482.776$	$\frac{932\times964}{2000} = 449.224$

Example Cont'd

3. Calculate the expected cell counts

(Expected)	Married	Divorced
Married	$\frac{1068\times1036}{2000} = 553.224$	$\frac{1068\times964}{2000} = 514.776$
Divorced	$\frac{932\times1036}{2000} = 482.776$	$\frac{932\times964}{2000} = 449.224$

4. Calculate the partial χ^2 values

partial χ^2	Married	Divorced
Married	$\frac{(581-553.224)^2}{553.224} = 1.39$	$\frac{(487-514.776)^2}{514.776} = 1.50$
Divorced	$\frac{(455-482.776)^2}{482.776} = 1.60$	$\frac{(477-449.224)^2}{449.224} = 1.72$

Example Cont'd

3. Calculate the expected cell counts

(Expected)	Married	Divorced
Married	$\frac{1068\times1036}{2000} = 553.224$	$\frac{1068\times964}{2000} = 514.776$
Divorced	$\frac{932\times1036}{2000} = 482.776$	$\frac{932\times964}{2000} = 449.224$

4. Calculate the partial χ^2 values

partial χ^2	Married	Divorced
Married	$\frac{(581-553.224)^2}{553.224} = 1.39$	$\frac{(487-514.776)^2}{514.776} = 1.50$
Divorced	$\frac{(455-482.776)^2}{482.776} = 1.60$	$\frac{(477-449.224)^2}{449.224} = 1.72$

5. Calculate the χ^2 statistic

$\chi^2 = 1.39 + 1.50 + 1.60 + 1.72 = 6.21$

Example Cont'd

3. Calculate the expected cell counts

(Expected)	Married	Divorced
Married	$\frac{1068\times1036}{2000} = 553.224$	$\frac{1068\times964}{2000} = 514.776$
Divorced	$\frac{932\times1036}{2000} = 482.776$	$\frac{932\times964}{2000} = 449.224$

4. Calculate the partial χ^2 values

partial χ^2	Married	Divorced
Married	$\frac{(581-553.224)^2}{553.224} = 1.39$	$\frac{(487-514.776)^2}{514.776} = 1.50$
Divorced	$\frac{(455-482.776)^2}{482.776} = 1.60$	$\frac{(477-449.224)^2}{449.224} = 1.72$

5. Calculate the χ^2 statistic

   $\chi^2 = 1.39 + 1.50 + 1.60 + 1.72 = 6.21$

6. Calculate the degrees of freedom (df)

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

Example Cont'd

3. Calculate the expected cell counts

(Expected)	Married	Divorced
Married	$\frac{1068\times1036}{2000} = 553.224$	$\frac{1068\times964}{2000} = 514.776$
Divorced	$\frac{932\times1036}{2000} = 482.776$	$\frac{932\times964}{2000} = 449.224$

4. Calculate the partial χ^2 values

partial χ^2	Married	Divorced
Married	$\frac{(581-553.224)^2}{553.224} = 1.39$	$\frac{(487-514.776)^2}{514.776} = 1.50$
Divorced	$\frac{(455-482.776)^2}{482.776} = 1.60$	$\frac{(477-449.224)^2}{449.224} = 1.72$

5. Calculate the χ^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 χ^2 critical value with respect to α from the χ^2 table

   The $\chi^2_{\alpha=0.1,df=1} = 2.71$

Example Cont'd

3. Calculate the expected cell counts

(Expected)	Married	Divorced
Married	$\frac{1068\times1036}{2000} = 553.224$	$\frac{1068\times964}{2000} = 514.776$
Divorced	$\frac{932\times1036}{2000} = 482.776$	$\frac{932\times964}{2000} = 449.224$

4. Calculate the partial χ^2 values

partial χ^2	Married	Divorced
Married	$\frac{(581-553.224)^2}{553.224} = 1.39$	$\frac{(487-514.776)^2}{514.776} = 1.50$
Divorced	$\frac{(455-482.776)^2}{482.776} = 1.60$	$\frac{(477-449.224)^2}{449.224} = 1.72$

5. Calculate the χ^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 χ^2 critical value with respect to α from the χ^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.

Another Example

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 χ^2 test from beginning to end. Use $\alpha = .05$

(Observed)	Female	Male	Total
Liberal Arts	378	262	640
Science	99	175	274
Engineering	104	510	614
Total	581	947	1528