Chapter 10: Analyzing the Association Between Categorical Variables

Section 10.3: How Strong is the Association?

Learning Objectives

1. Analyzing Contingency Tables
2. Measures of Association
3. Difference of Proportions
4. The Ratio of Proportions: Relative Risk
5. Properties of the Relative Risk
6. Large Chi-square Does Not Mean There’s a Strong Association

Learning Objective 1: Analyzing Contingency Tables

- ***Is there an association?***
  - **When the P-value is small, we infer that the variables are associated.**
  - **That is, when Chi-squared is large, there is evidence of the existence of an association between the variables.**

Learning Objective 1:
Analyzing Contingency Tables

- ***How do the cell counts differ from what independence predicts?***

- **To answer this question, we compare each observed cell count to the corresponding expected cell count**

Learning Objective 1: Analyzing Contingency Tables

- ***How strong is the association?***
- **Analyzing the *strength* of the association reveals whether the association is an important one, or if it is statistically significant but weak and unimportant in practical terms**

Learning Objective 2: Measures of Association

- **A *measure of association* is a statistic or a parameter that summarizes the strength of the dependence between two variables**
  - **a measure of association is useful for *comparing* associations**

Learning Objective 3:
Difference of Proportions

- **An easily interpretable measure of association is the *difference between the proportions* making a particular response**

	(A) Accept Credit Card			(B) Accept Credit Card		
INCOME	No	Yes	TOTAL	No	Yes	TOTAL
High	240 (60%)	160 (40%)	400 (100%)	0 (0%)	400 (100%)	400 (100%)
Low	360 (60%)	240 (40%)	600 (100%)	600 (100%)	0 (0%)	600 (100%)

Case (a) exhibits the weakest possible association – no association. The difference of proportions is 0

Case (b) exhibits the strongest possible association: The difference of proportions is 1

Learning Objective 3:
Difference of Proportions

- **In practice, we don't expect data to follow either extreme (0% difference or 100% difference), but the stronger the association, the larger the absolute value of the difference of proportions**

Learning Objective 3:
Difference of Proportions Example: Do Student Stress and Depression Depend on Gender?

TABLE 11.13: Conditional Distributions of Stress and Depression, by Gender

	Stress				Depression		
GENDER	Yes	No	TOTAL	GENDER	Yes	No	TOTAL
Female	35%	65%	100%	Female	8%	92%	100%
Male	16%	84%	100%	Male	6%	94%	100%

- **Which response variable, stress or depression, has the stronger sample association with gender?**
- **The difference of proportions between females and males was 0.35 – 0.16 = 0.19 for feeling stressed**
- **The difference of proportions between females and males was 0.08 – 0.06 = 0.02 for feeling depressed**

Learning Objective 3:
Difference of Proportions Example: Do Student Stress and Depression Depend on Gender?

- **In the sample, *stress* (with a difference of proportions = 0.19) has a *stronger association* with *gender* than *depression* has (with a difference of proportions = 0.02)**

Learning Objective 4:
The Ratio of Proportions: Relative Risk

- **Another measure of association, is the *ratio* of two proportions: p_1/p_2**
- **In medical applications in which the proportion refers to an adverse outcome, it is called the *relative risk***

Learning Objective 4:
Example: Relative Risk for Seat Belt Use and Outcome of Auto Accidents

- **Treating the auto accident outcome as the response variable, find and interpret the *relative risk***

TABLE 11.14: Outcome of Auto Accident by Whether or Not Subject Wore Seat Belt

	Outcome		
Wore Seat Belt	Survived	Died	**Total**
Yes	412,368	510	412,878
No	162,527	1601	164,128

Learning Objective 4:
Example: Relative Risk for Seat Belt Use and Outcome of Auto Accidents

- **The adverse outcome is death**
- **The relative risk is formed for that outcome**
- **For those who wore a seat belt, the proportion who died equaled 510/412,878 = 0.00124**
- **For those who did not wear a seat belt, the proportion who died equaled 1601/164,128 = 0.00975**

Learning Objective 4:
Example: Relative Risk for Seat Belt Use and Outcome of Auto Accidents

- **The relative risk is the ratio:**
  - **0.00124/0.00975 = 0.127**
  - **The proportion of subjects wearing a seat belt who died was 0.127 times the proportion of subjects not wearing a seat belt who died**

Learning Objective 4:
Example: Relative Risk for Seat Belt Use and Outcome of Auto Accidents

- **Many find it easier to interpret the relative risk but reordering the rows of data so that the relative risk has value above 1.0**

Learning Objective 4:
Example: Relative Risk for Seat Belt Use and Outcome of Auto Accidents

- **Reversing the order of the rows, we calculate the ratio:**
  - **0.00975/0.00124 = 7.9**
  - **The proportion of subjects not wearing a seat belt who died was 7.9 times the proportion of subjects wearing a seat belt who died**

Learning Objective 4:
Example: Relative Risk for Seat Belt Use and Outcome of Auto Accidents

- **A relative risk of 7.9 represents a strong association**
  - **This is far from the value of 1.0 that would occur if the proportion of deaths were the same for each group**
  - **Wearing a set belt has a practically significant effect in enhancing the chance of surviving an auto accident**

Learning Objective 5:
Properties of the Relative Risk

- **The relative risk can equal any nonnegative number**
- **When $p_1 = p_2$, the variables are independent and relative risk = 1.0**
- **Values farther from 1.0 (in either direction) represent stronger associations**
- **Rule of thumb: if the proportions are small, then a relative risk greater than 4 (or less than ¼) is considered strong.**

Learning Objective 6: Large χ^2 Does Not Mean There's a Strong Association

- **A large chi-squared value provides strong evidence that the variables are associated**
- **It does not imply that the variables have a strong association**
- **This statistic merely indicates (through its P-value) how certain we can be that the variables are associated, not how strong that association is**

EXAMPLE

In each case the difference between proportions is small, only 0.02.

	A.			B.			C.		
	Yes	No	*n*	Yes	No	*n*	Yes	No	*n*
Female	51%	49%	100	51%	49%	200	51%	49%	10,000
Male	49%	51%	100	49%	51%	200	49%	51%	10,000
Chi-squared = 0.08				Chi-squared = 0.16			Chi-squared = 8.0		
P-value = 0.78				P-value = 0.69			P-value = 0.005		