CHAPTER 6: Two-Way Tables

The Basic Practice of Statistics

6th Edition

Moore / Notz / Fligner

Lecture PowerPoint Slides

Chapter 6 Concepts

- Two-Way Tables
- Row and Column Variables
- Marginal Distributions
- Conditional Distributions
- Simpson’s Paradox

Chapter 6 Objectives

- Construct and interpret two-way tables
- Construct and interpret marginal distributions
- Construct and interpret conditional distributions
- Examine the effects of Simpson’s Paradox

Categorical Variables

- **Review:** Categorical Variables place individuals into one of several groups or categories.
  - The values of a categorical variable are labels for the different categories.
  - The distribution of a categorical variable lists the count or percent of individuals who fall into each category.

When a dataset involves two categorical variables, we begin by examining the counts or percents in various categories for *one* of the variables.

Two-way Table – describes two categorical variables, organizing counts according to a ***row variable*** and a ***column variable***.

Two-Way Table

Young adults by gender and chance of getting rich			
	Female	Male	Total
Almost no chance	96	98	194
Some chance, but probably not	426	286	712
A 50-50 chance	696	720	1416
A good chance	663	758	1421
Almost certain	486	597	1083
Total	2367	2459	4826

What are the variables described by this two-way table?

How many young adults were surveyed?

Marginal Distribution

The **Marginal Distribution** of one of the categorical variables in a two-way table of counts is the distribution of values of that variable among all individuals described by the table.

Note: Percents are often more informative than counts, especially when comparing groups of different sizes.

To examine a marginal distribution:

1. Use the data in the table to calculate the marginal distribution (in percents) of the row or column totals.

2. Make a graph to display the marginal distribution.

Marginal Distribution

Young adults by gender and chance of getti			
	Female	Male	Total
Almost no chance	96	98	194
Some chance, but probably not	426	286	712
A 50-50 chance	696	720	1416
A good chance	663	758	1421
Almost certain	486	597	1083
Total	2367	2459	4826

Examine the **marginal distribution** of chance of getting rich.

Response	Percent
Almost no chance	194/4826 = 4.0%
Some chance	712/4826 = 14.8%
A 50-50 chance	1416/4826 = 29.3%
A good chance	1421/4826 = 29.4%
Almost certain	1083/4826 = 22.4%

Conditional Distribution

- Marginal distributions tell us nothing about the relationship between two variables.

A **Conditional Distribution** of a variable describes the values of that variable among individuals who have a specific value of another variable.

To examine or compare conditional distributions:

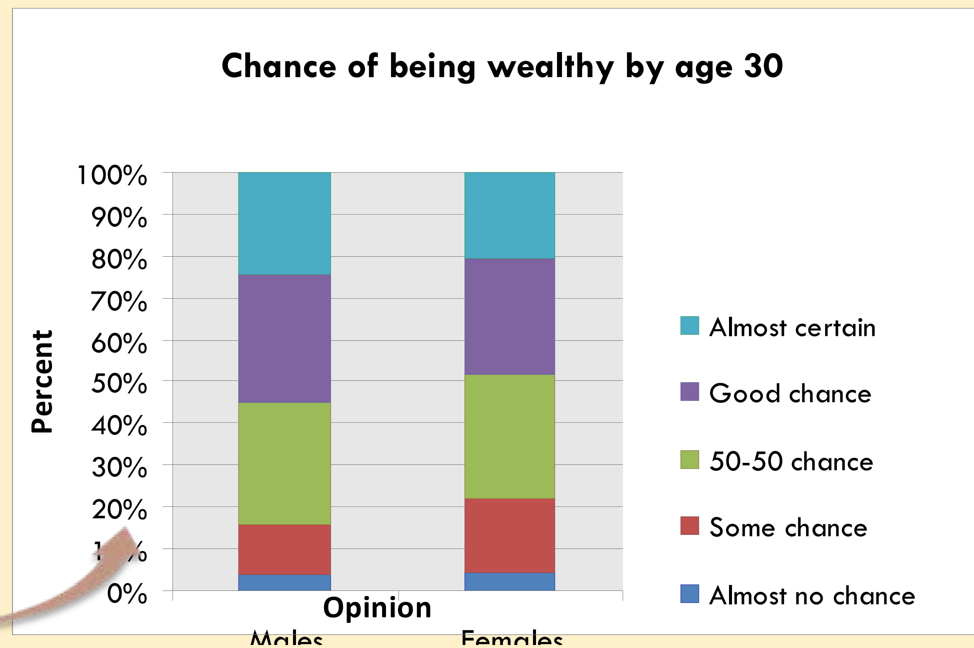
1. Select the row(s) or column(s) of interest.
2. Use the data in the table to calculate the conditional distribution (in percents) of the row(s) or column(s).
3. Make a graph to display the conditional distribution.
   - Use a **side-by-side bar graph** or **segmented bar graph** to compare distributions.

Conditional Distribution

Young adults by gender and chance of getting rich			
	Female	Male	Total
Almost no chance	96	98	194
Some chance, but probably not	426	286	712
A 50-50 chance	696	720	1416
A good chance	663	758	1421
Almost certain	486	597	1083
Total	2367	2459	4826

Calculate the conditional distribution of opinion among males.
Examine the relationship between gender and opinion.

Response	Male	Female
Almost no chance	98/2459 = 4.0%	96/2367 = 4.1%
Some chance	286/2459 = 11.6%	426/2367 = 18.0%
A 50-50 chance	720/2459 = 29.3%	696/2367 = 29.4%
A good chance	758/2459 = 30.8%	663/2367 = 28.0%
Almost certain	597/2459 = 24.3%	486/2367 = 20.5%



Simpson's Paradox

- When studying the relationship between two variables, there may exist a **lurking variable** that creates a reversal in the direction of the relationship when the lurking variable is ignored as opposed to the direction of the relationship when the lurking variable is considered.
- The lurking variable creates subgroups, and failure to take these subgroups into consideration can lead to misleading conclusions regarding the association between the two variables.

> An association or comparison that holds for all of several groups can reverse direction when the data are combined to form a single group. This reversal is called **Simpson's paradox.**

Simpson's Paradox

Consider the acceptance rates for the following groups of men and women who applied to college.

Counts	Accepted	Not accepted	Total
Men	198	162	360
Women	88	112	200
Total	286	274	560

Percents	Accepted	Not accepted
Men	**55%**	45%
Women	44%	56%

A higher percentage of **men** were accepted: Is there evidence of discrimination?

Simpson's Paradox

Consider the acceptance rates when broken down by type of school.

BUSINESS SCHOOL

Counts	Accepted	Not accepted	Total
Men	18	102	120
Women	24	96	120
Total	42	198	240

Percents	Accepted	Not accepted
Men	15%	85%
Women	**20%**	80%

ART SCHOOL

Counts	Accepted	Not accepted	Total
Men	180	60	240
Women	64	16	80
Total	244	76	320

Percents	Accepted	Not accepted
Men	75%	25%
Women	**80%**	20%

Simpson's Paradox

- **Lurking variable:** Applications were split between the Business School (240) and the Art School (320).

Within each school a higher percentage of women were accepted than men.

There is not any discrimination against women!!!

This is an example of **Simpson's Paradox.**

When the lurking variable (Type of School: Business or Art) is ignored the data seem to suggest discrimination against women.

However, when the type of school is not considered, the association is reversed and suggests discrimination against men.

Chapter 6 Objectives Review

- Construct and interpret two-way tables
- Construct and interpret marginal distributions
- Construct and interpret conditional distributions
- Examine the effects of Simpson's Paradox