Lesson 2.1.1: Summarizing Bivariate Categorical Data

Targets:

1. I can distinguish between categorical data and numerical data.
2. I can summarize data using a two-way frequency table.

Background Information

Superheroes have been popular characters in movies, television, books, and comics for many generations. Superman was one of the most popular series in the 1950's while Batman was a top rated series in the 1960's. Each of these characters was also popular in movies released from 1990 to 2013. Other notable characters portrayed in movies over the last several decades include Captain America, She-Ra, and the Fantastic Four. What is special about a superhero? Is there a special superhero power that makes these characters particularly popular?

High school students in the United States were invited to complete an online survey in 2010. Part of the survey included questions about superhero powers. More than 1,000 students responded to this survey that included a question about a student's most favorite superhero power. 450 of the completed surveys were randomly selected. A rather confusing breakdown of the data by gender was compiled from the 450 surveys:

- 100 students indicated their favorite power was "to fly." 49 of those students were females.
- 131 students selected the power to "freeze time" as their favorite power. 71 of those students were males.
- 75 students selected "invisibility" as their favorite power. 48 of those students were females.
- 26 students indicated "super strength" as their favorite power. 25 of those students were males.
- And finally, 118 students indicated "telepathy" as their favorite power. 70 of those students were females.

Warm Up

Several superheroes portrayed in movies and television series had at least one extraordinary power. Some superheroes had more than one special power. Was Superman's power "to fly" the favorite power of his fans, or was it his "super strength"? Would females view the power "to fly" differently than males or in the same way? Use the survey information given above to answer the following questions.

1. How many more females than males indicated their favorite power is "telepathy?"

2. How many more males than females indicated their favorite power was "to fly?"

3. Write survey questions that you think might have been used to collect this data.

4. How do you think the 450 surveys used above might have been selected? You can assume that there were 1,000 surveys to select from.

Vocabulary

Read the following description below. Underline what you think is most important. Then watch the video and take notes.

The data in the warm up prompted students in a mathematics class to pose the statistical question, "Do high school males have different preferences for superhero powers than high school females?" Answering this statistical question involves collecting data as well as anticipating variability in the data collected.

The data consist of two responses from each student completing a survey. The first response indicates a student's gender, and the second response indicates the student's favorite superpower. For example, data collected from one student was "male" and "to fly." The data are **bivariate categorical data** (bi means two and there were two different categories of data).

The first step in analyzing the statistical question posed by the students in their mathematics class is to organize this data in a two-way frequency table.

A two-way frequency table that can be used to organize the categorical data is shown below. The letters below represent the frequency counts of the cells of the table.

	To Fly	Freeze time	Invisibility	Super Strength	Telepathy	Total
Females	(a)	(b)	(c)	(d)	(e)	(f)
Males	(g)	(h)	(i)	(j)	(k)	(l)
Total	(m)	(n)	(o)	(p)	(q)	(r)

- The shaded cells are called **marginal frequencies**. They are located around the "margins" of the table and represent the totals of the rows or columns of the table.
- The non-shaded cells *within* the table are called **joint frequencies**. Each joint cell is the frequency count of responses from the two categorical variables located by the intersection of a row and column.

Copy the notes from the Vocabulary Video here:

Numerical Data:

Categorical Data:

Bivariate Categorical Data:

Two-Way Frequency Table:

Practice 1

1. Describe the data that would be counted in cell (a).

2. Describe the data that would be counted in cell (j).

3. Describe the data that would be counted in cell (l).

4. Describe the data that would be counted in cell (n).

5. Describe the data that would be counted in cell (r).

6. Cell (i) is the number of male students who selected "invisibility" as their favorite superpower. Using the information given in the warm up, what is the value of this number?

7. Cell (d) is the number of females whose favorite superpower is "super strength." Using the information given in the warm up, what is the value of this number?

8. Complete the table below by determining a frequency count for each cell based on the summarized data.

	To Fly	Freeze time	Invisibility	Super Strength	Telepathy	Total
Females						
Males						
Total						

Exit Ticket

- A survey asked the question "How tall are you to the nearest inch?" A second question on this survey asked, "What sports do you play?" Indicate what type of data, numerical or categorical, would be collected from the first question? What type of data would be collected from the second question?
- Another random sample of 100 surveys was selected. Jill had a copy of the frequency table that summarized these 100 surveys. Unfortunately, she spilled part of her lunch on the copy. The following summaries were still readable:

	To Fly	**Freeze time**	**Invisibility**	**Super Strength**	**Telepathy**	**Total**
Females	12	15	(c)*	5	(e)*	55
Males	12	16	10	(j)*	3	45
Total	24	31	25	9	(q)*	100

2. Help Jill recreate the table by determining the frequencies for cells (c), (e), (j), and (q).

3. Of the cells (c), (e), (j), and (q), which cells represent joint frequencies?

4. Of the cells (c), (e), (j), and (q), which cells represent marginal frequencies?