

Lesson 9: Summarizing Bivariate Categorical Data

Classwork

Recall from your work in Grade 6 and Grade 8 that categorical data are data that are not numbers. **Bivariate categorical data** results from collecting data on two categorical variables. In this lesson, you will see examples involving categorical data collected from two survey questions.

Exploratory Challenge 1: Superhero Powers

Superheroes have been popular characters in movies, television, books, and comics for many generations. Superman was one of the most popular series in the 1950s while Batman was a top rated series in the 1960s. 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 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.

Exercises 1–4

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 in Example 1 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 in Example 1 might have been selected? You can assume that there were 1,000 surveys to select from.

Exploratory Challenge 2: A Statistical Study Involving a Two-Way Frequency Table

The data in Example 1 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.

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.

12. 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						

Lesson Summary

- *Categorical data* are data that take on values that are categories rather than numbers. Examples include male or female for the categorical variable of gender or the five superpower categories for the categorical variable of superpower qualities.
- A *two-way frequency table* is used to summarize bivariate categorical data.
- The number in a two-way frequency table at the intersection of a row and column of the response to two categorical variables represents a *joint frequency*.
- The total number of responses for each value of a categorical variable in the table represents the *marginal frequency* for that value.

Problem Set

Several students at Rufus King High School were debating whether males or females were more involved in after-school activities. There are three organized activities in the after-school program—intramural basketball, chess club, and jazz band. Due to budget constraints, a student can only select one of these activities. The students were not able to ask every student in the school whether they participated in the after-school program or what activity they selected if they were involved.

1. Write questions that could be included in the survey to investigate the question the students are debating. Questions that could be used for this study include the following:
2. Rufus King High School has approximately 1,500 students. Sam suggested that the first 100 students entering the cafeteria for lunch would provide a random sample to analyze. Janet suggested that they pick 100 students based on a school identification number. Who has a better strategy for selecting a random sample? How do you think 100 students could be randomly selected to complete the survey?
3. Consider the following results from 100 randomly selected students:
 - Of the 60 female students selected, 20 of them played intramural basketball, 10 played chess, and 10 were in the jazz band. The rest of them did not participate in the after-school program.
 - Of the male students, 10 did not participate in the after-school program, 20 played intramural basketball, 8 played in the jazz band, and the rest played chess.

A two-way frequency table to summarize the survey data was started. What label is needed in the table cell identified with a “???”

	Intramural Basketball	Chess Club	Jazz Band	???	Total
Female					
Male					
Total					

4. Complete the above table for the 100 students who were surveyed.
5. The table shows the responses to the after-school activity question for males and females. Do you think there is a difference in the responses of males and females? Explain your answer.