



Lesson 9: Summarizing Bivariate Categorical Data

Student Outcomes

- Students distinguish between categorical data and numerical data.
- Students summarize data on two categorical variables collected from a sample using a two-way frequency table.

Lesson Notes

Categorical data are often summarized in the media, research studies, or general discussions. However, categorical data are summarized differently than numerical data. There is no mean or median that answers the question, “What is your favorite soft drink?” Methods for analyzing categorical data are developed in this lesson.

Categorical data was introduced to students in Grades 6 and 8. This lesson, along with Lessons 10 and 11, revisits and extends work students did in Grade 8 where data on two categorical variables were organized in a two-way table. Students also work with a random sample in this lesson and build on their understanding of a random sample developed in Grade 7. The data used in this lesson were obtained from the Census at School project. Census at School is further explained at the American Statistical Association website (www.amstat.org/censusatschool) and can be a source for data that might interest teenagers.

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 (9 minutes): Superhero Powers

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.

MP.1

Direct students to read through the example and examine the “data.” Then, discuss the following questions to determine their understanding of the example:

- What is the most popular superpower?
 - *Freeze time*
- What is the least popular?
 - *Super strength*
- Why would the survey include gender?
 - *The survey includes gender in order to investigate whether or not there is a difference in the selection of a superpower for males or females.*
- Do you think gender plays a role in superhero power preference?
 - *Answers will vary, but, in most cases, there are differences in the selection of a superpower by males or females.*

Exercises 1–4 (8 minutes)

Allow students to work independently or in small groups for two to three minutes on Exercises 1–4. Then, discuss and confirm answers as a class.

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.

- How many more females than males indicated their favorite power is “telepathy”?
22 more females indicated their favorite power was “telepathy.”
- How many more males than females indicated their favorite power was “to fly”?
2 more males indicated their favorite power was “to fly.”
- Write survey questions that you think might have been used to collect this data.
 - *What is your gender? ___ Male ___ Female*
 - *If you could possess just one superpower, what would it be?*
___ To fly ___ Freeze time ___ Invisibility ___ Super strength ___ Telepathy
- 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.
A process involving random selection is needed.

Exploratory Challenge 2 (5 minutes): A Statistical Study Involving a Two-Way Frequency Table

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.

Ask the students why the question posed is a statistical question. Remind students that a statistical question is a question that is answered by data and that it is anticipated that the data will vary. This question is a statistical question because you collect the responses (data), and you expect that the responses will vary (there are a variety of responses).

Explain the definition of categorical data and how it relates to Example 1.

Then, discuss:

- What is the difference between categorical and numerical data? Remind students that answers to questions such as the superhero powers involve responses that are categories.
- Ask students to provide examples of numerical data from past lessons to contrast with categorical data.

Have the students examine the two-way frequency table.

Then, discuss:

- Is it possible to switch the row and column categories?
 - Yes, male and female could be used as the column headings, and the superpowers could be represented in each row.

Exercises 5–12 (15 minutes)

Allow students to work in small groups on Exercises 5–12. If students struggle with these questions, discuss them as a whole group.

Exercises 5–12

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

Cell (a) represents the number of females who chose “to fly.”

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

Cell (j) represents the number of males who chose “super strength.”

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

Cell (l) represents the total number of males who completed the surveys in this sample.

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

Cell (n) represents the total number of students who chose “freeze time” as their favorite superpower.

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

Cell (r) represents the total number of students in the sample.

10. Cell (i) is the number of male students who selected “invisibility” as their favorite superpower. Using the information given in Example 1, what is the value of this number?

27 males selected “invisibility.”

11. Cell (d) is the number of females whose favorite superpower is “super strength.” Using the information given in Example 1, what is the value of this number?

1 female selected “super strength.”

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	49	60	48	1	70	228
Males	51	71	27	25	48	222
Total	100	131	75	26	118	450

Closing (3 minutes)

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.

Exit Ticket (5 minutes)

Name _____

Date _____

Lesson 9: Summarizing Bivariate Categorical Data

Exit Ticket

1. 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?

Exit Ticket Sample Solutions

1. 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?

The first question would result in numerical data; the second question would result in categorical data.

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).

Cell (c) has 15 students, (e) has 8 students, (j) has 4 students, and (q) has 11 students.

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

The cells (c), (e), and (j) are joint frequencies.

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

Cell (q) is a marginal frequency.

Problem Set Sample Solutions

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:

What is your gender? (Circle one) Female Male

Indicate if you participate in an after-school activity: (Circle one)

Intramural basketball Chess club Jazz band I do not participate in after-school activities.

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?

Sam's suggestion is the least likely to generate a random sample because it will be primarily a convenience sample based on factors that influence who has lunch first (for example, certain grade levels, certain classes). Selections that involve ID numbers from the entire school are more likely to result in a more random selection. Contacting the 100 students selected by their IDs and asking them to complete the survey is not necessarily an easy or even workable process. I think I would ask students to take the survey just as they are leaving school. Nearly everyone in school gathers outside for a few minutes. I would try to get a completed survey from one out of every 50 students so that I would not get surveys just from one group.

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 "???"

The "???" could be labeled, "Do not participate in after-school program."

	Intramural Basketball	Chess Club	Jazz Band	???	Total
Female	20	10	10	20	60
Male	20	2	8	10	40
Total	40	12	18	30	100

4. Complete the above table for the 100 students who were surveyed.

Answers are provided in the table.

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.

Yes, I think that there are differences in the responses for males and females. Quite a few more females selected chess club. However, more females were surveyed than males. Also, half of the males selected basketball, while only a third of the females selected basketball. (Allow students to indicate that they are not sure how to compare the frequencies. For students forming an answer, the frequencies in the table are compared to the number of males or the number of females. Use this question to point out that more needs to be considered before we can really answer the question of whether or not there is a difference in the responses for males or females. Also, point out that a strategy for answering this type of question is developed in the next two lessons.)