



Lesson 13: Summarizing Bivariate Categorical Data in a Two-Way Table

Student Outcomes

- Students organize bivariate categorical data into a two-way table.
- Students calculate row and column relative frequencies and interpret them in context.

Lesson Notes

In this lesson, students first organize data from a survey on a single categorical variable (i.e., a univariate categorical data) into a one-way frequency table. Some questions review content on random and representative samples that students first encountered in Grade 7. Then, they organize data on two categorical variables (i.e., bivariate categorical data) into two-way frequency tables. This lesson also introduces students to relative frequencies (e.g., row and column relative frequencies). Students then interpret relative frequencies in context.

Classwork

Exercises 1–5 (3–5 minutes)

Read the opening scenario to the class. Allow students a few minutes to choose their favorite ice cream flavors. Consider also asking students to raise their hands for each flavor preference and having them record the class data in the table provided for Exercise 1.

Exercises 1–8

On an upcoming field day at school, the principal wants to provide ice cream during lunch. She offers three flavors: chocolate, strawberry, and vanilla. She selected your class to complete a survey to help her determine how much of each flavor to buy.

- Answer the following question. Wait for your teacher to count how many students selected each flavor. Then, record the class totals for each flavor in the table below.

“Which of the following three ice cream flavors is your favorite: chocolate, strawberry, or vanilla?”

Answers will vary. One possibility is shown below.

Ice Cream Flavor	Chocolate	Strawberry	Vanilla	Total
Number of Students	17	4	7	28

- Which ice cream flavor do most students prefer?

Students should respond with the most-selected flavor. For the data set shown here, that is chocolate.

- Which ice cream flavor do the fewest students prefer?

Students should respond with the least-selected flavor. For the data set shown here, that is strawberry.

4. What percentage of students preferred each flavor? Round to the nearest tenth of a percent.

Answers will vary based on the data gathered in Exercise 1.

Chocolate: $\frac{17}{28} \approx 60.7\%$

Strawberry: $\frac{4}{28} \approx 14.3\%$

Vanilla: $\frac{7}{28} = 25\%$

5. Do the numbers in the table in Exercise 1 summarize data on a categorical variable or a numerical variable?

The numbers in this table summarize data on a categorical variable—the preferred flavor of ice cream.

Scaffolding:

Categorical variables are variables that represent categorical data. Data that represent specific descriptions or categories are called categorical data.

Exercises 6–8 (5 minutes)

Let students work with a partner to discuss and answer Exercises 6–8. These exercises review the concepts of random samples and representative samples from Grade 7. These exercises may also be used to structure a class discussion.

6. Do the students in your class represent a random sample of all the students in your school? Why or why not? Discuss this with your neighbor.

No, because there is no indication that the students were selected randomly.

7. Is your class representative of all the other classes at your school? Why or why not? Discuss this with your neighbor.

This class might be representative of the other eighth-grade classes but might not be representative of sixth- and seventh-grade classes.

8. Do you think the principal will get an accurate estimate of the proportion of students who prefer each ice cream flavor for the whole school using only your class? Why or why not? Discuss this with your neighbor.

It is unlikely to give a good estimate. It would depend on how representative the class is of all of the students at the school.

Example 1 (3–5 minutes)

In this example, be sure that students understand the vocabulary. *Univariate* means *one* variable. Thus, *univariate categorical data* means that there are data on one variable that are categorical, such as favorite ice cream flavor. A *one-way frequency table* is typically used to summarize values of univariate categorical data. When the data are categorical, it is customary to convert the table entries to *relative frequencies* instead of frequencies.

In other words, the fraction $\frac{\text{frequency}}{\text{total}}$ should be used, which is the relative frequency or proportion for each possible value of the categorical variable.

Scaffolding:

- Point out the prefix *uni-* means one. So, *univariate* means *one variable*.
- Some students may recognize the word *table* but may not yet know the mathematical meaning of the term. Point out that this lesson defines *table* as a tool for organizing data.

Students in another class were asked the same question about their favorite ice cream flavors. In this particular class of 25 students, 11 preferred chocolate, 4 preferred strawberry, and 10 preferred vanilla. Thus, the relative frequency for chocolate is $\frac{11}{25} = 0.44$. The interpretation of this value is “44% of the students in this class prefer chocolate ice cream.” Students often find writing interpretations to be difficult. Explain why this is not the case in this example.

Scaffolding:
The word *relative* has multiple meanings, such as a family member. In this context, it refers to a measure that is compared to something else. Making this distinction clear aids in comprehension.

Example 1

Students in a different class were asked the same question about their favorite ice cream flavors. The table below shows the ice cream flavors and the number of students who chose each flavor for that particular class. This table is called a *one-way frequency table* because it shows the counts of a univariate categorical variable.

This is the univariate categorical variable. →

Ice Cream Flavor	Chocolate	Strawberry	Vanilla	Total
Number of Students	11	4	10	25

These are the counts for each category. →

We compute the relative frequency for each ice cream flavor by dividing the count by the total number of observations.

$$\text{relative frequency} = \frac{\text{count for a category}}{\text{total number of observations}}$$

Since 11 out of 25 students answered *chocolate*, the relative frequency would be $\frac{11}{25} = 0.44$. This relative frequency shows that 44% of the class prefers chocolate ice cream. In other words, the relative frequency is the proportional value that each category is of the whole.

Exercises 9–10 (3 minutes)

Let students work independently and confirm their answers with a neighbor.

Exercises 9–10

Use the table for the preferred ice cream flavors from the class in Example 1 to answer the following questions.

9. What is the relative frequency for the category *strawberry*?

Relative frequency = $\frac{4}{25} = 0.16$

10. Write a sentence interpreting the relative frequency value in the context of strawberry ice cream preference.

16% of the students in this class prefer strawberry ice cream.



Example 2 (3–5 minutes)

Read through the example as a class. In this example, the focus shifts to bivariate categorical data. The prefix *bi-* means two, so these data contain values for two variables that are both categorical, such as favorite ice cream flavor and gender.

Example 2

The principal also wondered if boys and girls have different favorite ice cream flavors. She decided to redo the survey by taking a random sample of students from the school and recording both their favorite ice cream flavors and their genders. She asked the following two questions:

- “Which of the following ice cream flavors is your favorite: chocolate, strawberry, or vanilla?”
- “What is your gender: male or female?”

The results of the survey are as follows:

- Of the 30 students who prefer chocolate ice cream, 22 are males.
- Of the 25 students who prefer strawberry ice cream, 15 are females.
- Of the 27 students who prefer vanilla ice cream, 13 are males.

The values of two variables, which were ice cream flavor and gender, were recorded in this survey. Since both of the variables are categorical, the data are bivariate categorical data.

Exercises 11–17 (10 minutes)

Present Exercises 11 and 12 to the class one at a time.

Exercises 11–17

11. Can we display these data in a one-way frequency table? Why or why not?

No, a one-way frequency table is for univariate data. Here we have bivariate data, so we would need to use a two-way table.

12. Summarize the results of the second survey of favorite ice cream flavors in the following table:

		Favorite Ice Cream Flavor			Total
		Chocolate	Strawberry	Vanilla	
Gender	Male	22	10	13	45
	Female	8	15	14	37
Total		30	25	27	82

Next, remind students how to calculate relative frequencies. Give students a few minutes to calculate the approximate relative frequencies and to write them in the table. A *cell relative frequency* is a cell frequency divided by the total number of observations. Let students work independently on Exercises 13–17. Discuss and confirm the answers to Exercises 16 and 17 as a class.

13. Calculate the relative frequencies of the data in the table in Exercise 12, and write them in the following table.

		Favorite Ice Cream Flavor			Total
		Chocolate	Strawberry	Vanilla	
Gender	Male	≈ 0.27	≈ 0.12	≈ 0.16	≈ 0.55
	Female	≈ 0.10	≈ 0.18	≈ 0.17	≈ 0.45
	Total	≈ 0.37	≈ 0.30	≈ 0.33	1.0

Use the relative frequency values in the table to answer the following questions:

14. What is the proportion of the students who prefer chocolate ice cream?

0.37

15. What is the proportion of students who are female and prefer vanilla ice cream?

0.17

16. Write a sentence explaining the meaning of the approximate relative frequency 0.55.

Approximately 55% of students responding to the survey are males.

17. Write a sentence explaining the meaning of the approximate relative frequency 0.10.

Approximately 10% of students responding to the survey are females who prefer chocolate ice cream.

MP.6

Example 3 (3–5 minutes)

In this example, students learn that they can also use row and column totals to calculate relative frequencies. This concept provides a foundation for future work with conditional relative frequencies in Algebra I.

Point out that students need to carefully decide which total (i.e., table total, row total, or column total) they should use.

Scaffolding:

- English language learners may need a reminder about the difference between columns and rows.
- A column refers to a vertical arrangement, and a row refers to a horizontal arrangement in the table.
- Keeping a visual aid posted that labels these parts aids in comprehension.

Example 3

In the previous exercises, you used the total number of students to calculate relative frequencies. These relative frequencies were the proportion of the whole group who answered the survey a certain way. Sometimes we use row or column totals to calculate relative frequencies. We call these *row relative frequencies* or *column relative frequencies*.

Below is the two-way frequency table for your reference. To calculate “the proportion of male students who prefer chocolate ice cream,” divide the 22 male students who preferred chocolate ice cream by the total of 45 male students. This proportion is $\frac{22}{45} \approx 0.49$. Notice that you used the row total to make this calculation. This is a row relative frequency.

		Favorite Ice Cream Flavor			Total
		Chocolate	Strawberry	Vanilla	
Gender	Male	22	10	13	45
	Female	8	15	14	37
Total		30	25	27	82

Exercises 18–22 (8–10 minutes)

Discuss Exercise 18 as a class. When explaining the problem, try covering the unused part of the table with paper to focus attention on the query at hand.

Exercises 18–22

In Exercise 13, you used the total number of students to calculate relative frequencies. These relative frequencies were the proportion of the whole group who answered the survey a certain way.

18. Suppose you are interested in the proportion of male students who prefer chocolate ice cream. How is this value different from “the proportion of students who are male and prefer chocolate ice cream”? Discuss this with your neighbor.

The proportion of students who are male and prefer chocolate ice cream is $\frac{22}{82} \approx 0.27$. This proportion uses all 82 students. The proportion of male students who prefer chocolate ice cream is $\frac{22}{45} \approx 0.49$. This proportion uses only the 45 male students as its total.

Now, allow students time to answer Exercises 19–22. Discuss student answers stressing which *total* was used in the calculation.

19. Use the table provided in Example 3 to calculate the following relative frequencies.

- a. What proportion of students who prefer vanilla ice cream are female?

$\frac{14}{27} \approx 0.52$

- b. What proportion of male students prefer strawberry ice cream? Write a sentence explaining the meaning of this proportion in the context of this problem.

$\frac{10}{45} \approx 0.22$ *Twenty-two percent of male students in this survey prefer strawberry ice cream.*

- c. What proportion of female students prefer strawberry ice cream?

$$\frac{15}{37} \approx 0.41$$

- d. What proportion of students who prefer strawberry ice cream are female?

$$\frac{15}{25} \approx 0.60$$

20. A student is selected at random from this school. What would you predict this student's favorite ice cream to be? Explain why you chose this flavor.

I would predict that the student's favorite flavor is chocolate because more students chose chocolate in the survey.

21. Suppose the randomly selected student is male. What would you predict his favorite flavor of ice cream to be? Explain why you chose this flavor.

I would predict his favorite flavor to be chocolate because more male students chose chocolate in the survey.

22. Suppose the randomly selected student is female. What would you predict her favorite flavor of ice cream to be? Explain why you chose this flavor.

I would predict her favorite flavor to be strawberry because more female students chose strawberry in the survey.

Closing (2 minutes)

Review the Lesson Summary with students.

Lesson Summary

- Univariate categorical data are displayed in a one-way frequency table.
- Bivariate categorical data are displayed in a two-way frequency table.
- *Relative frequency* is the frequency divided by a total ().
- A *cell relative frequency* is a cell frequency divided by the total number of observations.
- A *row relative frequency* is a cell frequency divided by the row total.
- A *column relative frequency* is a cell frequency divided by the column total.

Exit Ticket (5 minutes)



3. A random group of students are polled about how they get to school. The results are summarized in the table below.

		School Transportation Survey			Total
		Walk	Ride Bus	Carpool	
Gender	Male	9	26	9	44
	Female	8	26	24	58
Total		17	52	33	102

- Calculate the relative frequencies for the table above. Write them as a percent in each cell of the table. Round to the nearest tenth of a percent.
- What is the relative frequency for the Carpool category? Write a sentence interpreting this value in the context of school transportation.
- What is the proportion of students who are female and walk to school? Write a sentence interpreting this value in the context of school transportation.
- A student is selected at random from this school. What would you predict this student's mode of school transportation to be? Explain.



Exit Ticket Sample Solutions

1. Explain what the term *bivariate categorical data* means.

Bivariate categorical data means that the data set comprises data on two variables that are both categorical.

2. Explain how to calculate relative frequency. What is another word for *relative frequency*?

Relative frequency is calculated by dividing a frequency by the total number of observations. Another word for relative frequency is proportion.

3. A random group of students are polled about how they get to school. The results are summarized in the table below.

		School Transportation Survey			Total
		Walk	Ride Bus	Carpool	
Gender	Male	9 ≈ 8.8%	26 ≈ 25.5%	9 ≈ 8.8%	44 ≈ 43.1%
	Female	7 ≈ 6.9%	26 ≈ 25.5%	25 ≈ 24.5%	58 ≈ 56.9%
Total		16 ≈ 15.7%	52 ≈ 51.0%	34 ≈ 33.3%	102 100.0%

- a. Calculate the relative frequencies for the table above. Write them as a percent in each cell of the table. Round to the nearest tenth of a percent.

See the completed table above.

- b. What is the relative frequency for the Carpool category? Write a sentence interpreting this value in the context of school transportation.

The relative frequency is 0.333, or 33.3%. Approximately 33.3% of the students surveyed use a carpool to get to school.

- c. What is the proportion of students who are female and walk to school? Write a sentence interpreting this value in the context of school transportation.

The proportion is 0.069, or 6.9%. Approximately 6.9% of the students surveyed are female and walk to school.

- d. A student is selected at random from this school. What would you predict this student’s mode of school transportation to be? Explain.

I would predict the student would ride the bus because more students in the survey chose this mode of transportation.

Problem Set Sample Solutions

Every student at Abigail Douglas Middle School is enrolled in exactly one extracurricular activity. The school counselor recorded data on extracurricular activity and gender for all 254 eighth-grade students at the school.

The counselor's findings for the 254 eighth-grade students are the following:

- Of the 80 students enrolled in band, 42 are male.
- Of the 65 students enrolled in choir, 20 are male.
- Of the 88 students enrolled in sports, 30 are female.
- Of the 21 students enrolled in art, 9 are female.

1. Complete the table below.

		Extracurricular Activities				Total
		Band	Choir	Sports	Art	
Gender	Female	38	45	30	9	122
	Male	42	20	58	12	132
	Total	80	65	88	21	254

2. Write a sentence explaining the meaning of the frequency 38 in this table.

The frequency of 38 represents the number of eighth-grade students who are enrolled in band and are female.

Use the table provided above to calculate the following relative frequencies.

3. What proportion of students are male and enrolled in choir?

$$\frac{20}{254} \approx 0.08$$

4. What proportion of students are enrolled in a musical extracurricular activity (i.e., band or choir)?

$$\frac{80 + 65}{254} \approx 0.57$$

5. What proportion of male students are enrolled in sports?

$$\frac{58}{132} \approx 0.44$$

6. What proportion of students enrolled in sports are male?

$$\frac{58}{88} \approx 0.66$$

Pregnant women often undergo ultrasound tests to monitor their babies' health. These tests can also be used to predict the gender of the babies, but these predictions are not always accurate. Data on the gender predicted by ultrasound and the actual gender of the baby for 1,000 babies are summarized in the two-way table below.

		Predicted Gender	
		Female	Male
Actual Gender	Female	432	48
	Male	130	390

7. Write a sentence explaining the meaning of the frequency 130 in this table.

The frequency of 130 represents the number of babies who were predicted to be female but were actually male (i.e., the ultrasound prediction was not correct for these babies).

Use the table provided above to calculate the following relative frequencies.

8. What is the proportion of babies who were predicted to be male but were actually female?

$$\frac{48}{1000} = 0.048$$

9. What is the proportion of incorrect ultrasound gender predictions?

$$\frac{130 + 48}{1000} = 0.178$$

10. For babies predicted to be female, what proportion of the predictions were correct?

$$\frac{432}{562} \approx 0.769$$

11. For babies predicted to be male, what proportion of the predictions were correct?

$$\frac{390}{438} \approx 0.890$$