|  |
| --- |
|  |

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 flavor. You can also ask students to raise hands for each flavor preference and have 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 will offer 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.

1. 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 chart 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** |  |  |  |  |

1. 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.

1. 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.

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

Answers will vary based on data gathered in Exercise 1.

Chocolate:

Strawberry:

Vanilla:

*Scaffolding:*

C*ategorical variables* are variables that represent categorical data.  Data that represent specific descriptions or categories are called categorical data.

1. Do the numbers in the chart above summarize data on a categorical variable or a numerical variable?

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

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. You may also use these exercises to structure a class discussion.

1. Do the students in your class represent a random sample of all 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.

1. **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.

1. Do you think the principal will get an accurate estimate of the proportion of students that 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 you a good estimate. It would depend on how representative your class is of all of the students at your school.

Example 1 (3–5 minutes)

*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.

In this example, be sure that students understand the vocabulary. *Univariate* means *one* variable. Thus, *univariate categorical data* means that you have data on one variable that is 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 is categorical, it is customary to convert the table entries to *relative frequencies* instead of frequencies. In other words, you should use the fraction, , which is the relative frequency or proportion for each possible value of the categorical variable.

Students in another class were asked the same question about their favorite ice cream flavor.  In this particular class of students, preferred chocolate, preferred strawberry, and preferred vanilla. Thus, the relative frequency for chocolate is . The interpretation of this value is “ 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 will aid in comprehension.

Example 1

Students in a different class were asked the same question about their favorite ice cream flavor. 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.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Ice Cream Flavor** | **Chocolate** | **Strawberry** | **Vanilla** | **Total** |
| **Number of Students** |  |  |  |  |

This is the univariate categorical
 variable.

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.

Since out of students answered chocolate, the relative frequency would be . This relative frequency shows that 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.

1. What is the relative frequency for the category strawberry?

Relative frequency

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

 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 this data will 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 flavor and their gender. 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 students who prefer chocolate ice cream, are males.
* Of the students who prefer strawberry ice cream, are females.
* Of the students who prefer vanilla ice cream, 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

1. 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.

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

|  |  |  |  |
| --- | --- | --- | --- |
|  |  | **Favorite Ice Cream Flavor** |  |
|  |  | **Chocolate** | **Strawberry** | **Vanilla** | **Total** |
| **Gender** | **Male** |  |  |  |  |
| **Female** |  |  |  |  |
|  | **Total** |  |  |  |  |

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 16–17 as a class.

1. Calculate the relative frequencies for the table above and write them in the table.

|  |  |  |  |
| --- | --- | --- | --- |
|  |  | **Favorite Ice Cream Flavor** |  |
|  |  | **Chocolate** | **Strawberry** | **Vanilla** | **Total** |
| **Gender** | **Male** |  |  |  |  |
| **Female** |  |  |  |  |
|  | **Total** |  |  |  |  |

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

1. What is the proportion of the students that prefer chocolate ice cream?
2. What is the proportion of students that are female and prefer vanilla ice cream?
3. Write a sentence explaining the meaning of the approximate relative frequency .

Approximately of students responding to the survey are males.

**MP.6**

1. Write a sentence explaining the meaning of the approximate relative frequency .

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

**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.

*Scaffolding:*

* ELL students 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 will aid in comprehension.

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

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 that prefer chocolate ice cream,” divide the male students who preferred chocolate ice cream by the total of male students. This proportion is . Notice that you used the row total to make this calculation. This is a row relative frequency.

|  |  |  |  |
| --- | --- | --- | --- |
|  |  | **Favorite Ice Cream Flavor** |  |
|  |  | **Chocolate** | **Strawberry** | **Vanilla** | **Total** |
| **Gender** | **Male** |  |  |  |  |
| **Female** |  |  |  |  |
|  | **Total** |  |  |  |  |

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.

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

The proportion of students that are male and prefer chocolate ice cream is . This proportion uses all students. The proportion of male students that prefer chocolate ice cream is This proportion uses only the 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.

1. Use the table provided in Example 3 to calculate the following relative frequencies.
	1. What proportion of students that prefer vanilla ice cream is female?
	2. What proportion of male students prefers strawberry ice cream? Write a sentence explaining the meaning of this proportion in context of this problem.

 Twenty-two percent of male students in this survey prefer strawberry ice cream.

* 1. What proportion of female students prefers strawberry ice cream?
	2. What proportion of students who prefer strawberry ice cream is female?
1. A student is selected at random from this school. What would you predict this student’s favorite ice cream to be? Explain why you choose this flavor.

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

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

I would predict that his favorite flavor is chocolate because more male students choose chocolate in the survey.

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

I would predict that her favorite flavor is strawberry because more female students choose 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 (frequency/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)

Name \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ Date\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

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

Exit Ticket

1. Explain what the term *bivariate categorical data* means.
2. Explain how to calculate relative frequency. What is another word for *relative frequency*?
3. A random group of students is polled about how they get to school. The results are summarized in the table below.

|  |  |  |  |
| --- | --- | --- | --- |
|  |  | **School Transportation Survey** |  |
|  |  | Walk | Ride Bus | Carpool | Total |
| **Gender** | Male |  |  |  |  |
| Female |  |  |  |  |
|  | Total |  |  |  |  |

* 1. 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.
	2. What is the relative frequency for the Carpool category? Write a sentence interpreting this value in the context of school transportation.
	3. What is the proportion of students that are female and walk to school? Write a sentence interpreting this value in the context of school transportation.
	4. 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.

1. 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.

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

|  |  |  |  |
| --- | --- | --- | --- |
|  |  | School Transportation Survey |  |
|  |  | Walk | Ride Bus | Carpool | Total |
| Gender | Male |  |  |  |  |
| Female |  |  |  |  |
|  | Total |  |  |  |  |

* 1. 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.

* 1. 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 , or . Approximately of the students surveyed use a carpool to get to school.

* 1. What is the proportion of students that are female and walk to school? Write a sentence interpreting this value in the context of school transportation.

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

* 1. 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 choose 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 eighth-grade students at the school.

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

* Of the students enrolled in band, are male.
* Of the students enrolled in art, are female.
* Of the students enrolled in choir, are male.
* Of the students enrolled in sports, are female.
1. Complete the table below.

|  |  |  |  |
| --- | --- | --- | --- |
|  |  | **Extracurricular Activities** |  |
|  |  | **Band** | **Choir** | **Sports** | **Art** | **Total** |
| **Gender** | **Female** |  |  |  |  |  |
| **Male** |  |  |  |  |  |
|  | **Total** |  |  |  |  |  |

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

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

1. What proportion of students is male and enrolled in choir?
2. What proportion of students is enrolled in a musical extracurricular activity (i.e., band or choir)?
3. What proportion of male students is enrolled in sports?
4. What proportion of students enrolled in sports is male?

A pregnant woman will often undergo ultrasound tests to monitor her baby’s health. These tests can also be used to predict the gender of the baby, but these predictions are not always accurate. Data on the gender predicted by ultrasound and the actual gender of the baby for babies are summarized in the two-way table below.

|  |  |  |
| --- | --- | --- |
|  |  | **Predicted Gender** |
|  |  | **Female** | **Male** |
| **Actual Gender** | **Female** |  |  |
| **Male** |  |  |

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

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

1. What is the proportion of babies predicted to be male but were actually female?
2. What is the proportion of incorrect ultrasound gender predictions?
3. For babies predicted to be female, what proportion of the predictions was correct?
4. For babies predicted to be male, what proportion of the predictions was correct?