



Lesson 11: Conditional Relative Frequencies and Association

Student Outcomes

- Students calculate and interpret conditional relative frequencies from two-way frequency tables.
- Students evaluate conditional relative frequencies as an indication of possible association between two variables.
- Students explain why association does not imply causation.

Lesson Notes

Students continue the analysis of the bivariate categorical data that they began in Lessons 9 and 10. Lesson 9 summarized the data in a two-way frequency table. Relative frequencies were then introduced in Lesson 10. In each case, however, the question posed about whether there is a difference in the favorite superpower responses of males and females remains unclear. This lesson develops and interprets conditional relative frequencies to answer the statistical question posed in the previous lessons. Students familiar with two-way tables from Grade 8 will move through these questions within the suggested time frame. Students not as familiar with this work may need more time to complete these exercises.

The focus of this lesson is conditional relative frequencies and how they indicate a possible association. Definitions of conditional relative frequencies and association are provided in this lesson. Differences in conditional relative frequencies are used as evidence of possible association. Instructors should challenge students to think critically about the meaning of an association between two categorical variables and to be careful not to draw unwarranted conclusions about possible cause-and-effect relationships between two categorical variables. The last example and exercise discuss the issue of cause-and-effect.

Classwork

After further discussion, the students involved in designing the superhero comic strip decided that before any decision is made, a more careful look at the data on the special powers a superhero character could possess was needed. There is an association between gender and superpower response if the superpower responses of males are not the same as the superpower responses of females. Examining each row of the table can help determine whether or not there is an association.

Exploratory Challenge 1 (5 minutes): Conditional Relative Frequencies

Exploratory Challenge 1: Conditional Relative Frequencies

Recall the two-way table from the previous lesson.

	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

A *conditional relative frequency* compares a frequency count to the marginal total that represents the condition of interest. For example, the condition of interest in the first row is females. The row conditional relative frequency of females responding “invisibility” as the favorite superpower is $\frac{48}{228}$, or approximately 0.211. This conditional relative frequency indicates that approximately 21.1% of females prefer “invisibility” as their favorite superpower. Similarly, $\frac{27}{222}$, or approximately 0.122 or 12.2%, of males prefer “invisibility” as their favorite superpower.

Let students read through the introductory paragraph and Example 1.

Then, discuss:

- How is relative frequency calculated?
 - *It is calculated by dividing the frequency by the total observations.*
- How could we determine a relative frequency for only the female students?
 - *We could determine this by dividing the frequency by the total number of females.*
- Conditional relative frequencies are found by dividing the frequency by the marginal total.

Work with students to calculate the conditional relative frequency of females who chose “to fly” as a superpower:

$$\frac{49}{228} \approx 0.215, \text{ or } 21.5\%.$$

Then, discuss:

- How would we interpret the conditional relative frequency?
 - *21.5% of the females surveyed chose “to fly” as their favorite superpower.*
- How does this compare with the same cell in the relative frequency table from Example 1?
 - *10.9% of the students surveyed were females who chose “to fly” as their favorite superpower.*

Exercises 1–5 (10 minutes)

Let students continue to work in pairs on Exercises 1–5. Work with students more directly on the first set of questions.

Exercises 1–5

1. Use the frequency counts from the table in Exploratory Challenge 1 to calculate the missing row of conditional relative frequencies. Round the answers to the nearest thousandth.

	To Fly	Freeze Time	Invisibility	Super Strength	Telepathy	Total
Females	$\frac{49}{228} \approx 0.215$	$\frac{60}{228} \approx 0.263$	$\frac{48}{228} \approx 0.211$	$\frac{1}{228} \approx 0.004$	$\frac{70}{228} \approx 0.307$	$\frac{228}{228} = 1.000$
Males	$\frac{51}{222} \approx 0.230$	$\frac{71}{222} \approx 0.320$	$\frac{27}{222} \approx 0.122$	$\frac{25}{222} \approx 0.113$	$\frac{48}{222} \approx 0.216$	$\frac{222}{222} = 1.000$
Total	$\frac{100}{450} \approx 0.222$	$\frac{131}{450} \approx 0.291$	$\frac{75}{450} \approx 0.167$	$\frac{26}{450} \approx 0.058$	$\frac{118}{450} \approx 0.262$	$\frac{450}{450} = 1.000$

2. Suppose that a student is selected at random from those who completed the survey. What do you think is the gender of the student selected? What would you predict for this student's response to the superpower question?

Since there were almost the same number of males and females, the selected student could be male or female. If the responses of both males and females were combined, students would probably select "freeze time" because the relative frequency of "freeze time" is $\frac{131}{450}$ (or approximately 29%) of all students. This superpower was selected by more of the males and females than any other superpower.

3. Suppose that a student is selected at random from those who completed the survey. If the selected student is male, what do you think was his response to the selection of a favorite superpower? Explain your answer.

"Freeze time" was also the most popular selection for male students.

4. Suppose that a student is selected at random from those who completed the survey. If the selected student is female, what do you think was her response to the selection of a favorite superpower? Explain your answer.

"Telepathy" was the most popular selection for female students.

5. What superpower was selected by approximately one-third of the females? What superpower was selected by approximately one-third of the males? How did you determine each answer from the conditional relative frequency table?

"Telepathy" was selected by approximately one-third of the females. "Freeze time" was selected by approximately one-third of the males. These selections were based on row conditional relative frequencies approximately equal to 0.333 for females and for males.

Exploratory Challenge 2 (5 minutes): Possible Association Based on Conditional Relative Frequencies**Exploratory Challenge 2: Possible Association Based on Conditional Relative Frequencies**

Two categorical variables are associated if the row conditional relative frequencies (or column relative frequencies) are different for the rows (or columns) of the table. For example, if the selection of superpowers selected for females is different than the selection of superpowers for males, then gender and superpower favorites are associated. This difference indicates that knowing the gender of a person in the sample indicates something about their superpower preference.

The evidence of an association is strongest when the conditional relative frequencies are quite different. If the conditional relative frequencies are nearly equal for all categories, then there is probably not an association between variables.

Verbally summarize the paragraphs and how they relate to the theme of the lesson.

Association can be a challenging concept for students. Discuss additional examples of association involving the “condition” of gender:

- If you know that a person is female, could you predict that she spends more than thirty minutes getting ready for school?
 - *Maybe. Girls may tend to spend more time getting ready than boys, and, if so, there is an association between gender and whether a student spends more than thirty minutes getting ready for school.*
- If you know that a person is male, could you predict that country music is his favorite type of music?
 - *Probably not. There is no association between gender and music preference.*

Discuss other examples of questions that investigate association:

- If dogs are classified as large, medium, or small based on weight, are small dogs more likely to pass an obedience course?
- If users of a social network are classed as active, average, or inactive, is a person classified as an *active* user more likely to be a good writer than those classified in the other categories?

Then, discuss the evidence that supports association:

- There is strong evidence of association when there is a *noticeable* difference in conditional relative frequencies.
- What is a *noticeable* difference in conditional relative frequencies? This is subjective; students should use their best judgment at this time. Evaluating the differences more formally is discussed in Algebra II and Precalculus.

Exercises 6–10 (7 minutes)

Allow students to work independently on Exercises 6–10. Then, discuss and confirm answers as a class.

Exercises 6–10

Examine the conditional relative frequencies in the two-way table of conditional relative frequencies you created in Exercise 1. Note that for each superpower, the conditional relative frequencies are different for females and males.

6. For what superpowers would you say that the conditional relative frequencies for females and males are very different?

The most noticeable differences in conditional relative frequencies would be for “invisibility,” “super strength,” and “telepathy.” Students may also indicate a noticeable difference for “freeze time.”

7. For what superpowers are the conditional relative frequencies nearly equal for males and females?

The superpower “to fly” has nearly the same conditional relative frequencies for males and females.

8. Suppose a student is selected at random from the students who completed the survey. If you had to predict which superpower this student selected, would it be helpful to know the student’s gender? Explain your answer.

Due to the noticeable differences in the conditional relative frequencies of selecting a particular superpower based on gender, it would be helpful to know a student’s gender when predicting the selection of a superpower.

9. Is there evidence of an association between gender and a favorite superpower? Explain why or why not.

Based on the definition of association presented in this lesson, there appears to be an association between superpower selected and gender. Knowing the gender of a student helps predict the superpower response.

10. What superpower would you recommend the students at Rufus King High School select for their superhero character? Justify your choice.

Consider several answers. Students are expected to justify their answers based on the conditional relative frequencies, or the relative frequencies for all students.

Exploratory Challenge 3 (5 minutes): Association and Cause-and-Effect

Exploratory Challenge 3: Association and Cause-and-Effect

Students were given the opportunity to prepare for a college placement test in mathematics by taking a review course. Not all students took advantage of this opportunity. The following results were obtained from a random sample of students who took the placement test.

	Placed in Math 200	Placed in Math 100	Placed in Math 50	Total
Took Review Course	40	13	7	60
Did Not Take Review Course	10	15	15	40
Total	50	28	22	100

This example will introduce the important idea that you should not infer a cause-and-effect relationship from an association between two categorical variables.

Read through the example with students.

Pose the following questions to the class. Let students discuss their ideas.

- Do you think there is an association between taking the review course and a student’s placement in a math class?
- If you knew that a student took a review course, would it make a difference in what you predicted for which math course they were placed in?
- Do you think taking a course *caused* a student to place higher in a math placement?

Exercises 11–16 (6 minutes)

Let students work in pairs. Then, discuss and confirm answers to Exercises 11–13 and 14–16.

Exercises 11–16

11. Construct a row conditional relative frequency table of the above data.

	Placed in Math 200	Placed in Math 100	Placed in Math 50	Total
Took Review Course	$\frac{40}{60} \approx 0.667$	$\frac{13}{60} \approx 0.217$	$\frac{7}{60} \approx 0.117$	$\frac{60}{60} = 1.000$
Did Not Take Review Course	$\frac{10}{40} = 0.250$	$\frac{15}{40} = 0.375$	$\frac{15}{40} = 0.375$	$\frac{40}{40} = 1.000$
Total	$\frac{50}{100} = 0.500$	$\frac{28}{100} = 0.280$	$\frac{22}{100} = 0.220$	$\frac{100}{100} = 1.000$

12. Based on the conditional relative frequencies, is there evidence of an association between whether a student takes the review course and the math course in which the student was placed? Explain your answer.

There is evidence of association because the conditional relative frequencies are noticeably different for those students who took the course and for those students who did not take the course.

13. Looking at the conditional relative frequencies, the proportion of students who placed into Math 200 is much higher for those who took the review course than for those who did not. One possible explanation is that taking the review course caused improvement in placement test scores. What is another possible explanation?

Another possible explanation is that students who took the review course are more interested in mathematics or were already better prepared in mathematics and, therefore, performed better on the mathematics placement test.

- Do you think that this is an example of a cause-and-effect relationship? Be sure that they understand that even though there is an association, this does not mean that there is a cause and effect relationship.

Now consider the following statistical study:

Fifty students were selected at random from students at a large middle school. Each of these students was classified according to sugar consumption (high or low) and exercise level (high or low). The resulting data are summarized in the following frequency table.

		Exercise Level		Total
		High	Low	
Sugar Consumption	High	14	18	32
	Low	14	4	18
Total		28	22	50

14. Calculate the row conditional relative frequencies, and display them in a row conditional relative frequency table.

		Exercise Level		Total
		High	Low	
Sugar Consumption	High	$\frac{14}{32} = 0.4375$	$\frac{18}{32} = 0.5625$	$\frac{32}{32} = 1.000$
	Low	$\frac{14}{18} \approx 0.778$	$\frac{4}{18} \approx 0.222$	$\frac{18}{18} = 1.000$
Total		$\frac{28}{50} = 0.56$	$\frac{22}{50} = 0.44$	$\frac{50}{50} = 1.000$

15. Is there evidence of an association between sugar consumption category and exercise level? Support your answer using conditional relative frequencies.

There is a noticeable difference in the conditional relative frequencies based on whether a person selected had high or low sugar consumption. The differences suggest an association between sugar consumption and exercise level.

16. Do you think it is reasonable to conclude that high sugar consumption is the cause of the observed differences in the conditional relative frequencies? What other explanations could explain a difference in the conditional relative frequencies? Explain your answer.

Students are encouraged to think about their responses to this exercise based on their understanding that the results should not be interpreted as a cause-and-effect relationship. Other factors such as eating habits and lifestyle could be mentioned by students.

If time permits, discuss the following with your students:

- It is possible that students in the above study who are more health conscious tend to be in the low sugar consumption category and also tend to be in the high exercise level category?
- It is not possible to determine if the difference in the conditional relative frequencies is due to a cause-and-effect relationship?
- The data summarized in this study were collected in an observational study. In an observational study, any observed differences in conditional relative frequencies might be explained by some factor other than the variables examined in the study. With an observational study, evidence of an association may exist, but it is not possible to imply that there is a cause-and-effect relationship.

Closing (2 minutes)

Discuss with students the Lesson Summary.

Lesson Summary

- A conditional relative frequency compares a frequency count to the marginal total that represents the *condition* of interest.
- The differences in conditional relative frequencies are used to assess whether or not there is an association between two categorical variables.
- The greater the differences in the conditional relative frequencies, the stronger the evidence that an association exists.
- An observed association between two variables does not necessarily mean that there is a cause-and-effect relationship between the two variables.

Exit Ticket (5 minutes)

Name _____

Date _____

Lesson 11: Conditional Relative Frequencies and Association

Exit Ticket

Juniors and seniors were asked if they plan to attend college immediately after graduation, seek full-time employment, or choose some other option. A random sample of 100 students was selected from those who completed the survey. Scott started to calculate the row conditional relative frequencies to the nearest thousandth.

	Plan to Attend College	Plan to Seek Full-Time Employment	Other Options	Totals
Seniors	$\frac{25}{55} \approx 0.455$	$\frac{10}{55} \approx 0.182$	$\frac{20}{55} \approx ???$	$\frac{55}{55} = 1.000$
Juniors	$\frac{35}{45} \approx ???$	$\frac{5}{45} \approx ???$	$\frac{5}{45} \approx 0.111$	$\frac{45}{45} = 1.000$
Totals	$\frac{60}{100} = 0.600$	$\frac{15}{100} = 0.150$	$\frac{25}{100} = 0.250$	$\frac{100}{100} = 1.000$

- Complete the calculations of the row conditional relative frequencies. Round your answers to the nearest thousandth.
- Are the row conditional relative frequencies for juniors and seniors similar, or are they very different?
- Do you think there is a possible association between grade level (junior or senior) and after high school plans? Explain your answer.

Exit Ticket Sample Solutions

Juniors and seniors were asked if they plan to attend college immediately after graduation, seek full-time employment, or choose some other option. A random sample of 100 students was selected from those who completed the survey. Scott started to calculate the row conditional relative frequencies to the nearest thousandth.

1. Complete the calculations of the row conditional relative frequencies. Round your answers to the nearest thousandth.

	Plan to Attend College	Plan to Seek Full-Time Employment	Other Options	Totals
Seniors	$\frac{25}{55} \approx 0.455$	$\frac{10}{55} \approx 0.182$	$\frac{20}{55} \approx 0.364$	$\frac{55}{55} = 1.000$
Juniors	$\frac{35}{45} \approx 0.778$	$\frac{5}{45} \approx 0.111$	$\frac{5}{45} \approx 0.111$	$\frac{45}{45} = 1.000$
Totals	$\frac{60}{100} = 0.600$	$\frac{15}{100} = 0.150$	$\frac{25}{100} = 0.250$	$\frac{100}{100} = 1.000$

2. Are the row conditional relative frequencies for juniors and seniors similar, or are they very different?

The conditional relative frequencies are noticeably different for juniors and seniors.

3. Do you think there is a possible association between grade level (junior or senior) and after high school plans? Explain your answer.

The differences in the conditional relative frequencies suggest that there is an association.

Problem Set Sample Solutions

Consider again the summary of data from the 100 randomly selected students in the Rufus King High School investigation of after-school activities and gender.

	Intramural Basketball	Chess Club	Jazz Band	Not Involved	Total
Females	20	10	10	20	60
Males	20	2	8	10	40
Total	40	12	18	30	100

1. Construct a row conditional relative frequency table for this data. Decimal values are given to the nearest thousandth.

	Intramural Basketball	Chess Club	Jazz Band	Not Involved	Total
Females	$\frac{20}{60} \approx 0.333$	$\frac{10}{60} \approx 0.167$	$\frac{10}{60} \approx 0.167$	$\frac{20}{60} \approx 0.333$	$\frac{60}{60} = 1.000$
Males	$\frac{20}{40} = 0.500$	$\frac{2}{40} = 0.050$	$\frac{8}{40} = 0.200$	$\frac{10}{40} = 0.250$	$\frac{40}{40} = 1.000$
Total	$\frac{40}{100} = 0.400$	$\frac{12}{100} = 0.120$	$\frac{18}{100} = 0.180$	$\frac{30}{100} = 0.300$	$\frac{100}{100} = 1.000$

2. For what after-school activities do you think the row conditional relative frequencies for females and males are very different? What might explain why males or females select different activities?

There are noticeable differences in several of the after-school activity conditional relative frequencies for the two genders, suggesting that there is an association. (The most noticeable differences are in intramural basketball and chess club.) They could be different based on a student's interest in being with friends, having a specific coach, following a popular leader, or participating in a more popular tradition established at the school.

3. If John, a male student at Rufus King High School, completed the after-school survey, what would you predict was his response? Explain your answer.

50%, or half, of the males indicated they participated in intramural basketball. Since this was the greatest conditional relative frequency for males, you would anticipate that John, a male student, would select this response.

4. If Beth, a female student at Rufus King High School, completed the after-school survey, what would you predict was her response? Explain your answer.

Females selected "intramural basketball" or "not involved" equally often. You would anticipate that Beth, a female student, would select one of these two responses.

5. Notice that 20 female students participate in intramural basketball and that 20 male students participate in intramural basketball. Is it accurate to say that females and males are equally involved in intramural basketball? Explain your answer.

It is not accurate to say they were equally involved because there were more females in the sample. Thus, the 20 female students who participated in intramural basketball represent one-third of the females, while the 20 male students represent one-half of the males.

6. Do you think there is an association between gender and choice of after-school program? Explain.

There is a difference in the conditional relative frequencies, but the differences are not as obvious as in the other examples. (This is a problem in which students would be expected to justify how they would indicate association between gender and the after-school selections.)

Column conditional relative frequencies can also be computed by dividing each frequency in a frequency table by the corresponding column total to create a column conditional relative frequency table. Column conditional relative frequencies indicate the proportions, or relative frequencies, based on the column totals.

7. If you wanted to know the relative frequency of females surveyed who participated in chess club, would you use a row conditional relative frequency or a column conditional relative frequency?

You would use a row conditional relative frequency. Of the 60 females surveyed, 10 participated in the chess club. (The given condition is that the student is female.) The relative frequency of $\frac{10}{60}$ would be the conditional relative frequency of females who participated in chess club.

8. If you wanted to know the relative frequency of band members surveyed who were female, would you use a row conditional relative frequency or a column conditional relative frequency?

You would use a column conditional relative frequency. Of the 18 band members, 10 were females. (The given condition is that the student is in band.) The conditional relative frequency of $\frac{10}{18}$ would be the conditional relative frequency of band members who were females.

9. For the superpower survey data, write a question that would be answered using a row conditional relative frequency.

An example of a row conditional relative frequency question would be “What proportion of females selected ‘telepathy’ as their favorite superpower?”

10. For the superpower survey data, write a question that would be answered using a column conditional relative frequency.

An example of a column conditional relative frequency would be “What proportion of those who selected ‘telepathy’ are female?”