Lesson 2: Calculating Probabilities of Events Using Two-Way Tables

Student Outcomes

- Students calculate probabilities given a two-way table of data.
- Students construct a *hypothetical 1000* two-way table given probability information.
- Students interpret probabilities in context.

Lesson Notes

In this lesson, students construct and interpret data in two-way tables. Questions are designed to help students understand what the tables are summarizing. Two-way frequency tables were introduced in Grade 8 (Module 6, Lesson 13) and revisited in Algebra I (Module 2, Lessons 9–11) as a way to organize and interpret bivariate categorical data. This lesson reviews and extends those concepts associated with two-way tables. Students also construct and interpret a hypothetical 1000 table. Research indicates that a hypothetical 1000 table makes understanding probabilities formed from a two-way table easier for students to understand. In Exercises 9–14, students create a hypothetical 1000 table (i.e., a two-way frequency table based on a population of $1,000$ students) to answer probability questions. The probabilities from this table are approximately equal to the probabilities from the actual frequencies, which students also interpret in the next lesson. The probability questions build on students' previous work with probabilities in Grade 7.

Lessons 2–4 work together to develop the standards for the cluster "Understand independence and conditional probability and use them to interpret data (S-CP)." Lesson 2 requires students to calculate several probabilities that are identified in Lesson 3 as conditional probabilities. However, for this lesson, students calculate the probabilities using the table, the question posed, and the context of the data. Lessons 3 and 4 offer a more formal description of a conditional probability and how it is used to determine if two events are independent.

Classwork

Example 1 (2 minutes): Building a New High School

Students encounter references to two-way tables often in this module. It may be important to discuss the word table with students, especially for students learning the language. Highlight that the word *table* is used in these lessons to describe a tool for organizing data. Data organized in rows and columns are referenced as two-way tables or two-way frequency tables.

A formal definition of tables is difficult for all students. Carefully point out the tables organized in these lessons, and point out the cells by describing what they represent.

Scaffolding:

For ELL students, provide visuals of the two different meanings for *table* and practice saying the word chorally.

Example 1: Building a New High School

The school board of Waldo, a rural town in the Midwest, is considering building a new high school primarily funded by local taxes. They decided to interview eligible voters to determine if the school board should build a new high school facility to replace the current high school building. There is only one high school in the town. Every registered voter in Waldo was interviewed. In addition to asking about support for a new high school, data on gender and age group were also recorded. The data from these interviews are summarized below.

	Should our town build a new high school?					
	Yes		No		No answer	
Age (in years)	Male	Female	Male	Female	Male	Female
18–25	29	32	8	6	0	0
26–40	53	60	40	44	2	4
41–65	30	36	44	35	2	2
66 and older	7	26	24	29	2	0

Scaffolding:

Consider using the following example to help students understand how to build and interpret a two-way table. If needed, Grade 8, Module 6, Lesson 13 can be used for remediation.

In a class of 20 students, 5 boys and 7 girls like chocolate ice cream. Six boys like vanilla ice cream.

- What are the variables?
- How could we organize the variables in the following table?

	Chocolate	*Vanilla*
Boys	5	6
Girls	7	

- What does the missing cell represent? How can you determine the missing value?

Exercises 1–8: Building a New High School (15–20 minutes)

MP.2 The following exercises ask students to interpret the data summarized in the table. The exercises allow students to reason abstractly by making summaries based on data. Discuss how the data summaries are being used to answer each question. Although the questions ask students to predict, the word *probability* is not always used in the questions. This is intentional, as it helps students to revisit their previous work with probability. As you work with students, emphasize the goal of predicting outcomes based on the data.

Allow students to work in small groups as they answer the following questions. Select questions to discuss as a whole group based on student work. Allow students to use a calculator. For these questions, a scientific calculator is sufficient.

Exercises 1–8

1. **Based on this survey, do you think the school board should recommend building a new high school? Explain your answer.**

   273 out of the 515 eligible voters (or approximately 53%) indicated "yes." As a result, I think the voters will recommend building a new high school.

2. **An eligible voter is picked at random. If this person is 21 years old, do you think he would indicate that the town should build a high school? Why or why not?**

   Most of the eligible voters ages 18–25 indicated "yes." 61 of the 75 eligible voters in this age group indicated "yes." I would predict a 21-year-old in this age group to have answered "yes."

3. **An eligible voter is picked at random. If this person is 55 years old, do you think she would indicate that the town should build a high school? Why or why not?**

   Most of the eligible voters ages 41–65 indicated "no" to the question about building a high school. I would predict that a person in this age group would indicate that the town should not build a high school.

4. **The school board wondered if the probability of recommending a new high school was different for different age categories. Why do you think the survey classified voters using the age categories 18–25 years old, 26–40 years old, 41–65 years old, and 66 years old and older?**

The age groups used in the table represent people with different interests or opinions regarding the building of a new high school. For example, people ages 26–40 are more likely to have children in school than people in the other age groups. The probability that a person in each of these age categories would recommend building a new high school might vary.

5. **It might be helpful to organize the data in a two-way frequency table. Use the given data to complete the following two-way frequency table. Note that the age categories are represented as rows, and the possible responses are represented as columns.**

	Yes	No	No Answer	Total
18–25 years old	61	14	0	75
26–40 years old	113	84	6	203
41–65 years old	66	79	4	149
66 years old and older	33	53	2	88
Total	273	230	12	515

Exercise 6 challenges students to think about a "headline summary" of data (something they may encounter in a newspaper, blog, or news report). Discuss how these types of short summaries of data need to be examined by readers. The exercise provides an opportunity for students to form their own conclusions from the data and to reconcile their conclusions with summaries expressed in the form of headlines. In this way, students are also critiquing the reasoning of others.

MP.3

6. **A local news service plans to write an article summarizing the survey results. Three possible headlines for this article are provided below. Is each headline accurate or inaccurate? Support your answer using probabilities calculated using the table above.**

Headline 1: ***Waldo Voters Likely to Support Building a New High School***

Yes, this is accurate. 273 out of 515, or approximately 0.530 (or 53.0%), support building a new school. The probability that an eligible voter would vote "yes" is greater than 0.50, so you would think it is likely that voters will support building a new high school.

Headline 2: ***Older Voters Less Likely to Support Building a New High School***

Yes, this headline is accurate. If you define older voters as 41 or older, then 132 out of 237 voters 41 or older, or approximately 0.557 (or 55.7%), indicated they would not support building a new high school.

Headline 3: ***Younger Voters Not Interested in Building a New High School***

This headline is not accurate. 61 out of 75 eligible voters, or approximately 0.813 (or 81.3%) representing the youngest eligible voters, indicated "yes" to building a new high school.

Exercise 7 is challenging. Students are expected to use both the probability of voting "yes" and the new information about the number of eligible voters expected to vote to answer this question. It may be necessary to help some students develop the final answer by working through a specific age group before they combine all age groups. Consider tackling this question as a whole group. This question also presents another opportunity to discuss the role of probability. Students previously indicated that most voters would vote "yes" to building a new high school. This question will raise the possibility that voters will vote "no" by adding information about past voter turnout.

7. **The school board decided to put the decision on whether or not to build the high school up for a referendum in the next election. At the last referendum regarding this issue, only 25 of the eligible voters ages 18–25 voted, 110 of the eligible voters ages 26–40 voted, 130 of the eligible voters ages 41–65 voted, and 80 of the eligible voters ages 66 and older voted. If the voters in the next election turnout in similar numbers, do you think this referendum will pass? Justify your answer.**

   Use the probabilities that a voter from an age group would vote "yes." Multiply the probabilities that a person would vote "yes" by the number of people estimated to vote in each age category:

   $25(0.813) + 110(0.557) + 130(0.442) + 80(0.375)$ *is approximately* 169.

   169 *out of the estimated* 345 *voters will vote "yes." Since this is less than half, we would predict that the vote will indicate the high school should not be built.*

8. **Is it possible that your prediction of the election outcome might be incorrect? Explain.**

   Yes. The above is only a prediction. The actual results could be different. It depends on the actual voter turnout and whether people actually vote as they indicated in the survey.

Example 2 (2 minutes): Smoking and Asthma

Read through the example as a class.

Example 2: Smoking and Asthma

Health officials in Milwaukee, Wisconsin were concerned about teenagers with asthma. People with asthma often have difficulty with normal breathing. In a local research study, researchers collected data on the incidence of asthma among students enrolled in a Milwaukee public high school.

Students in the high school completed a survey that was used to begin this research. Based on this survey, the probability of a randomly selected student at this high school having asthma was found to be 0.193. Students were also asked if they had at least one family member living in their house who smoked. The probability of a randomly selected student having at least one member in his (or her) household who smoked was reported to be 0.421.

Exercises 9–14 (10–15 minutes)

Make sure students understand how the information will be organized in the given table. If necessary, point to a cell in the table and ask students to describe what this cell represents in the context of the data.

This scenario will be revisited in Lesson 4, where students will explore the relationship between having asthma and having at least one family member who smokes.

Allow students to continue to work in small groups as they answer the following questions.

Scaffolding:

Consider challenging advanced learners to build the hypothetical table independently or with a neighbor.

Exercises 9–14

It would be easy to calculate probabilities if the data for the students had been organized into a two-way table like the one used in Exercise 5. But there is no table here, only probability information. One way around this is to think about what the table might have been if there had been $1,000$ students at the school when the survey was given. This table is called a *hypothetical 1000 two-way table*.

What if the population of students at this high school was $1,000$? The population was probably not exactly $1,000$ students, but using an estimate of $1,000$ students provides an easier way to understand the given probabilities. Connecting these estimates to the actual population is completed in a later exercise. Place the value of $1,000$ in the cell representing the total population. Based on a hypothetical 1000 population, consider the following table.

	No household member smokes	At least one household member smokes	Total
Student indicates he (or she) has asthma	Cell 1	Cell 2	Cell 3
Student indicates he (or she) does not have asthma	Cell 4	Cell 5	Cell 6
Total	Cell 7	Cell 8	$1,000$

9. **The probability that a randomly selected student at this high school has asthma is 0.193. This probability can be used to calculate the value of one of the cells in the table above. Which cell is connected to this probability? Use this probability to calculate the value of that cell.**

   Cell 3. The value for this cell would be 193 students.

10. **The probability that a randomly selected student has at least 1 household member who smokes is 0.421. Which cell is connected to this probability? Use this probability to calculate the value of that cell.**

    Cell 8. The value for this cell would be 421 students.

11. **In addition to the previously given probabilities, the probability that a randomly selected student has at least one household member who smokes and has asthma is 0.120. Which cell is connected to this probability? Use this probability to calculate the value of that cell.**

    Cell 2. The value for this cell would be 120 students.

12. **Complete the two-way frequency table by calculating the values of the other cells in the table.**

	No household member smokes	At least one household member smokes	Total
Student indicates he (or she) has asthma	73	120	193
Student indicates he (or she) does not have asthma	506	301	807
Total	579	421	$1,000$

13. **Based on your completed two-way table, estimate the following probabilities as a fraction and also as a decimal (rounded to three decimal places):**

a. **A randomly selected student has asthma. What is the probability this student has at least 1 household member who smokes?**

The probability is $\frac{120}{193}$, *or approximately* 0.622.

b. **A randomly selected student does not have asthma. What is the probability this student has at least one household member who smokes?**

The probability is $\frac{301}{807}$, *or approximately* 0.373.

c. **A randomly selected student has at least one household member who smokes. What is the probability this student has asthma?**

The probability is $\frac{120}{421}$, *or approximately* 0.285.

14. **Do you think that whether or not a student has asthma is related to whether or not this student has at least one family member who smokes? Explain your answer.**

Yes. The probability a student with asthma has a household member who smokes is noticeably greater than the probability a student who does not have asthma has a household member who smokes.

Closing (5 minutes)

- What was the role of probability in our two examples?
  - *To use data to predict outcomes of events, such as selecting a voter who will vote "yes" or selecting a student who has asthma and at least one household member who smokes.*
- As you used probabilities to make predictions, did you learn anything about the context of the data?
  - *Answers will vary. Anticipate that students indicate that younger voters were less likely to vote or that older voters in Waldo were less likely to support the new high school or that students with asthma were more likely to have a household family member who smokes. Emphasize that probabilities must be interpreted in context.*

Ask students to summarize the main ideas of the lesson in writing or with a neighbor. Use this as an opportunity to informally assess comprehension of the lesson. The Lesson Summary below offers some important ideas that should be included.

Lesson Summary

Data organized in a two-way frequency table can be used to calculate probabilities.

In certain problems, probabilities that are known can be used to create a hypothetical 1000 two-way table. The hypothetical population of $1,000$ can then be used to calculate probabilities.

Probabilities are always interpreted in context.

Exit Ticket (5 minutes)

Name ____ Date ____________________

Lesson 2: Calculating Probabilities of Events Using Two-Way Tables

Exit Ticket

Did males and females respond similarly to the survey question about building a new high school? Recall the original summary of the data:

	Should our town build a new high school?					
	Yes		No		No answer	
Age (in years)	Male	Female	Male	Female	Male	Female
18–25	29	32	8	6	0	0
26–40	53	60	40	44	2	4
41–65	30	36	44	35	2	2
66 and older	7	26	24	29	2	0

1. Complete the following two-way frequency table:

	Yes	No	No answer	Total
Male	119		6	
Female				
Total		230	12	515

2. Use the above two-way frequency table to answer the following questions:
   a. If a randomly selected eligible voter is a female, what is the probability she will vote to build a new high school?

   b. If a randomly selected eligible voter is male, what is the probability he will vote to build a new high school?

3. An automobile company has two factories assembling its luxury cars. The company is interested in whether consumers rate cars produced at one factory more highly than cars produced at the other factory. Factory A assembles 60% of the cars. A recent survey indicated that 70% of the cars made by this company (both factories combined) were highly rated. This same survey indicated that 10% of all cars made by this company were both made at Factory B and were not highly rated.

   a. Create a hypothetical 1000 two-way table based on the results of this survey by filling in the table below.

	Car was highly rated by consumers	Car was not highly rated by consumers	Total
Factory A			
Factory B			
Total			

   b. A randomly selected car was assembled in Factory B. What is the probability this car is highly rated?

Exit Ticket Sample Solutions

Did males and females respond similarly to the survey question about building a new high school? Recall the original summary of the data:

	Should our town build a new high school?					
	Yes		No		No answer	
Age (in years)	Male	Female	Male	Female	Male	Female
$18–25$	29	32	8	6	0	0
$26–40$	53	60	40	44	2	4
$41–65$	30	36	44	35	2	2
66 and older	7	26	24	29	2	0

1. **Complete the following two-way frequency table:**

	Yes	No	No answer	Total
Male	119	116	6	241
Female	154	114	6	274
Total	273	230	12	515

2. **Use the above two-way frequency table to answer the following questions:**

   a. **If a randomly selected eligible voter is a female, what is the probability she will vote to build a new high school?**

   $\frac{154}{274}$*, or approximately* 0.562*.*

   b. **If a randomly selected eligible voter is male, what is the probability he will vote to build a new high school?**

   $\frac{119}{241}$*, or approximately* 0.494*.*

3. **An automobile company has two factories assembling its luxury cars. The company is interested in whether consumers rate cars produced at one factory more highly than cars produced at the other factory. Factory A assembles 60% of the cars. A recent survey indicated that 70% of the cars made by this company (both factories combined) were highly rated. This same survey indicated that 10% of all cars made by this company were both made at Factory B and were not highly rated.**

   a. **Create a hypothetical 1000 two-way table based on the results of this survey by filling in the table below.**

	Car was highly rated by consumers	Car was not highly rated by consumers	Total
Factory A	400	200	600
Factory B	300	100	400
Total	700	300	$1,000$

   b. **A randomly selected car was assembled in Factory B. What is the probability this car is highly rated?**

   The probability a car from Factory B is highly rated is $\frac{300}{400}$*, or* 0.750*.*

Problem Set Sample Solutions

1. **The Waldo School Board asked eligible voters to evaluate the town's library service. Data are summarized in the following table.**

	How would you rate our town's library services?							
	Good		Average		Poor		Do not use library	
Age (in years)	Male	Female	Male	Female	Male	Female	Male	Female
18–25	10	8	5	7	5	5	17	18
26–40	30	28	25	30	20	30	20	20
41–65	30	32	26	21	15	10	5	10
66 and older	21	25	8	15	2	10	2	5

   a. **What is the probability that a randomly selected person who completed the survey rated the library as "good?"**

   $\frac{184}{515}$, *or* 0.357

   b. **Imagine talking to a randomly selected male who had completed the survey. How do you think this person rated the library services? Explain your answer.**

   Answers will vary. A general look at the table indicates that most males rated the library as "good." As a result, I would predict that this person would rate the library as "good."

   c. **Use the given data to construct a two-way table that summarizes the responses on gender and rating of the library services. Use the following template as your guide.**

	Good	Average	Poor	Do Not Use	Total
Male	91	64	42	44	241
Female	93	73	55	53	274
Total	184	137	97	97	515

   d. **Based on your table, answer the following:**

      i. **A randomly selected person who completed the survey is male. What is the probability he rates the library services as "good?"**

      $\frac{91}{241}$, *or approximately* 0.378

      ii. **A randomly selected person who completed the survey is female. What is the probability she rates the library services as "good?"**

      $\frac{93}{274}$, *or approximately* 0.339

   e. **Based on your table, answer the following:**

      i. **A randomly selected person who completed the survey rated the library services as "good." What is the probability this person is a male?**

      $\frac{91}{184}$, *or approximately* 0.495

ii. **A randomly selected person who completed the survey rated the library services as "good." What is the probability this person is a female?**

$\frac{93}{274}$, or approximately 0.339.

f. **Do you think there is a difference in how males and females rated library services? Explain your answer.**

Answers will vary. Yes. For example, 37.9% of the males rated the library services as good, but only 33.9% of the females rated the services this way. There are also differences between males and females for the other rating categories.

2. **Obedience School for Dogs is a small franchise that offers obedience classes for dogs. Some people think that larger dogs are easier to train and, therefore, should not be charged as much for the classes. To investigate this claim, dogs enrolled in the classes were classified as large (30 pounds or more) or small (under 30 pounds). The dogs were also classified by whether or not they passed the obedience class offered by the franchise. 45% of the dogs involved in the classes were large. 60% of the dogs passed the class. Records indicate that 40% of the dogs in the classes were small and passed the course.**

a. **Complete the following hypothetical 1000 two-way table.**

	Passed the course	Did not pass the course	Total
Large Dogs	200	250	450
Small Dogs	400	150	550
Total	600	400	$1,000$

b. **Estimate the probability that a dog selected at random from those enrolled in the classes passed the course.**

$\frac{600}{1000}$, or 0.600, meaning 60% passed the course.

c. **A dog was randomly selected from the dogs that completed the class. If the selected dog was a large dog, what is the probability this dog passed the course?**

$\frac{200}{450}$, or approximately 0.444, meaning that approximately 44.4% of large dogs passed the course.

d. **A dog was randomly selected from the dogs that completed the class. If the selected dog is a small dog, what is the probability this dog passed the course?**

$\frac{400}{550}$, or approximately 0.727, meaning that approximately 72.7% of small dogs passed the course.

e. **Do you think dog size and whether or not a dog passes the course are related?**

Answers will vary. Yes, there is a noticeably greater probability that a dog passed the obedience class if a dog is small than if the dog is large.

f. **Do you think large dogs should get a discount? Explain your answer.**

Answers will vary. No, large dogs are not as likely to have passed the obedience class as small dogs.