Lesson

6-7B Contingency Tables

Vocabulary

categorical variable

contingency table, two-way table

BIG IDEA Considering all the possibilities, or contingencies, in a situation can help in making reasonable conclusions from data.

Frequencies in a Contingency Table

Suppose voters in a state were polled to determine if they are satisfied with their governor. The results were summarized in the table below.

Satisfaction Poll for Governor

Employment Status of Voter	Satisfied	Dissatisfied
Employed	412	129
Unemployed	31	28

In the table, the cells for the "Employment Status of Voter" ("Employed" or "Unemployed") are not filled with numbers. Employment status is called a **categorical variable** because its values represent characteristics rather than measures. Likewise, being "Satisfied" or "Dissatisfied" is not numerical. A table that divides outcomes among two or more categorical variables is called a **contingency table** or a **two-way table**.

In the contingency table above, the top left data cell indicates that 412 polled voters were employed *and* were satisfied with the way the governor was handling his job. Similarly, the bottom right cell indicates that 28 voters were unemployed *and* were dissatisfied with the way the governor was handling his job.

STOP QY1

Contingency tables can help analyze whether or not there is an association between the variables. For the data in the contingency table above, a question of interest may be, "Are voters who are employed more likely to be satisfied than those who are unemployed?" Essentially, we are asking if a voter's level of satisfaction *depends on* or *is contingent on* his or her employment status. Example 1 uses the contingency table above to examine this question.

QY1

a. How many polled voters were employed?

b. How many polled voters were unemployed?

c. How many voters were polled in all?

Example 1

Use the contingency table on the previous page. Is there an association between the employment status of a voter and his or her satisfaction with the governor's performance? If so, describe the association.

Solution

Step 1 Find the percent of polled voters who were satisfied.
Add the numbers in the Satisfied column: $412 + 31 = 443$.
600 voters were polled, so $\frac{443}{600} \approx 73.8\%$ were satisfied.

Step 2 Find the percent of employed voters in the poll who were satisfied.
$412 + 129 = 541$ polled voters were employed.
Of those voters, 412 were satisfied, so $\frac{412}{541} \approx 76.2\%$ of employed voters were satisfied.

Step 3 Find the percent of unemployed voters in the poll who were satisfied.
$31 + 28 = 59$ polled voters were unemployed.
Of those voters, 31 were satisfied, so $\frac{31}{59} \approx 52.5\%$ of unemployed voters were satisfied.

So there appears to be an association in the data between the employment status of a voter and his or her satisfaction with the governor's performance. A much higher percentage of employed voters than unemployed voters were satisfied with the governor's performance.

GUIDED

Example 2

Again consider the contingency table. Of the polled voters who are satisfied, what percent are employed? Round to the nearest tenth of a percent.

Solution Of the polled voters, there are 412 satisfied voters who are employed. The total number of satisfied voters is $412 + \underline{\ ?\ } = \underline{\ ?\ }$, so the percent of satisfied voters who are employed is $\frac{412}{?}$, or about $\underline{\ ?\ }$%.

The different answers found in Step 2 of Example 1 and Example 2 may be confusing because the questions seem so similar. However, the questions are different because they ask about different populations. Step 2 of Example 1 asks about the population of *employed voters*, of which there are 541. Example 2 asks about the population of *voters who are satisfied*, of which there are 443.

In forming a $\frac{part}{whole}$ ratio, you must be clear about the answer to the question "What is the whole?" For questions about ratios based on a contingency table, you must be careful to correctly identify the whole, that is, the population to which the question refers.

STOP **QY2**

▶ QY2

Out of all voters who are dissatisfied, what percent are unemployed?

Why Are Contingency Tables Important?

Some advertisements assert that a certain vitamin supplement will increase your energy, if you also eat well and get sleep. Let us suppose that 40 people were studied who took a particular vitamin supplement.

Regimen	Increased Energy	Did Not Increase Energy
Took vitamin supplement, Ate well, Got sleep	31	9

This looks like great evidence that the vitamin supplement works. $\frac{31}{40}$, or 77.5% of the people had increased energy.

The problem with making a conclusion that the supplement works is that all the contingencies have not been considered. It is often the case in medical studies that, instead of a real pill, a person will be given a fake pill called a *placebo* so that the person thinks he or she is getting the real pill. Let us suppose that in this study, there were 35 people given the placebo who also ate well and got sleep.

Regimen	Increased Energy	Did Not Increase Energy
Took placebo, Ate well, Got sleep	29	6

The percents are close enough between the placebo and the vitamin supplement that we cannot conclude that the supplement worked. This illustrates the importance of considering all the contingencies before making a conclusion.

STOP **QY3**

▶ QY3

What percent of the people taking the placebo had increased energy?

Example 3

"Home field advantage" is an important idea in many sports. In Major League Baseball, teams play as many home games as road games. In 2009, the Philadelphia Phillies' record in home games was 45-36. Tell whether each conclusion is reasonable.

a. **The Philadelphia Phillies won about 56% of the games they played in 2009.**

b. **The Philadelphia Phillies had a higher percentage of games won at home than games won on the road.**

Solutions

a. The conclusion is not reasonable. They did win about 56% of the home games but we are not told how many games they won on the road.

b. The conclusion is not reasonable for the same reason as in Part a.

In fact, the Phillies won 48 road games and lost 33 in 2009. They won a higher percentage of games on the road than at home.

Questions

COVERING THE IDEAS

In 1–3, refer to the contingency table of satisfaction poll results in this lesson.

1. Of all voters in the poll, what percent are dissatisfied?
2. Of the satisfied voters in the poll, what percent are unemployed?
3. Of the dissatisfied voters in the poll, what percent are employed?
4. Use the data given in Example 3 and its solution.
   a. Construct a contingency table for the home games and road games won and lost by the Philadelphia Phillies in 2009.
   b. What was the total win-loss record of the Philadelphia Phillies in 2009?
   c. What percent of road games did the Phillies win?
   d. Of the games the Phillies won, what percent were road games?

In 5 and 6, identify a contingency that is not being considered.

5. A candidate states: If you vote for me, you will have 4 years of prosperity.
6. A store advertises: Come in before the end of the month and we will take an additional 10% off the price of any item you purchase.

APPLYING THE MATHEMATICS

In 7–11, use the following information. The Quality Widget Co. produces widgets at two plants, Plant A and Plant B. When a defective widget is sold, it is returned to the company. The contingency table at the right shows how many defective widgets and not defective widgets were recently produced at each plant.

Production at the Quality Widget Co.

	Defective	Not Defective
Plant A	39	561
Plant B	32	368

7. **Multiple Choice** Which of the following is *not* true?
   - **A** A total of 71 defective widgets were produced at the two plants.
   - **B** A total of 1,000 widgets were produced at the two plants.
   - **C** About 93% of the widgets produced at the two plants were not defective.
   - **D** About 45% of the defective widgets were produced at Plant A.
8. What percent of all of the widgets produced came from Plant A?
9. What percent of the widgets produced at Plant A were defective?
10. If any one of the defective widgets is selected, what is the probability that it came from Plant A?
11. Which plant was more likely to produce a defective widget? Explain your answer.

In 12–14, use the following information. In a talent contest involving dancing and singing, votes were cast for the best performer in each category. The results are given in the contingency table below.

Talent Contest Votes

	Dancing	Singing
Performer A	21	17
	35%	28%
Performer B	4	36
	7%	60%
Performer C	35	7
	58%	12%

12. Are the percents given in the table based on columns or on rows? How do you know?
13. What percent of Performer B's votes came from singing?
14. Who received the most votes total? What percent of all votes cast did that performer receive?

15. A biologist is breeding some rabbits where the parents and offspring have either white or gray fur. The results of the experiment are given in the contingency table below. In this table, percents are given for each row.

Rabbit Experiment

	Both Parents White Fur	Both Parents Gray Fur	One Parent White; One Parent Gray
White Offspring	98	0	52
% of row	65%	?	?
Gray Offspring	8	36	14
% of row	14%	?	?

a. Copy and complete the above table to show the percents for each row. Round each result to the nearest percent.

b. What is the sum of the percents in each row of the table?

c. What percent of the white offspring have one parent that has white fur and one parent that has gray?

d. When both parents have gray fur, what percent of the offspring have white fur?

e. When both parents have white fur, what percent of the offspring have white fur?

f. What percent of all of the offspring have gray fur?

g. If any gray offspring is selected, what is the probability that its parents both have gray fur?

h. What can you conclude about the offspring if one parent has white fur and the other has gray?

QY ANSWERS

1. a. 541

b. 59

c. 600

2. about 17.8%

3. about 82.9%