Unit 3 Lesson 3: Associations in Categorical Data

1 Cake or Pie (Warm up)

Student Task Statement

The table displays the dessert preference and dominant hand (left- or right-handed) for a sample of 300 people.

	prefers cake	prefers pie	total
left-handed	10	20	30
right-handed	90	180	270
total	100	200	300

For each of the calculations, describe the interpretation of the percentage in terms of the situation.

1. 10% from $\frac{10}{100} = 0.1$

2. 67% from $\frac{180}{270} \approx 0.67$

3. 30% from $\frac{90}{300} = 0.3$

2 Associations in Categorical Data

Student Task Statement

1. The two-way table displays data about 55 different locations. Scientists have a list of possible chemicals that may influence the health of the coral. They first look at how nitrate concentration might be related to coral health. The table displays the health of the coral (healthy or unhealthy) and nitrate concentration (low or high).

	low nitrate concentration	high nitrate concentration	total
healthy	20	5	25
unhealthy	8	22	30
total	28	27	55

   a. Complete the two-way relative frequency table for the data in the two-way table in which the relative frequencies are based on the total for each column.

	low nitrate concentration	high nitrate concentration
healthy		
unhealthy		
total	100%	100%

b. When there is a low nitrate concentration, which had a higher relative frequency, healthy or unhealthy coral?

c. When there is a high nitrate concentration, is there a higher relative frequency of healthy or unhealthy coral?

d. Based on this data, is there a possible **association** between coral health and the level of nitrate concentration? Explain your reasoning.

e. The scientists next look at how silicon dioxide concentration might be related to coral health. The relative frequencies based on the total for each column are shown in the table. Based on this data, is there a possible association between coral health and the level of silicon dioxide concentration? Explain your reasoning.

	low silicon dioxide concentration	high silicon dioxide concentration
healthy	44%	46%
unhealthy	56%	54%
total	100%	100%

2. Jada surveyed 300 people from various age groups about their shoe preference. The two-way table summarizes the results of the survey.

	prefers sneakers without laces	prefers sneakers with laces	prefers shoes that are not sneakers	total
4–10 years old	21	12	3	36
11–17 years old	21	48	39	108
18–24 years old	15	54	87	156
total	57	114	129	300

Jada concludes that there is a possible association between age and shoe preference. Is Jada's conclusion reasonable? Explain your reasoning.

3. The two-way table summarizes data on writing utensil preference and the dominant hand for a sample of 100 people.

	left-handed	right-handed	total
prefers pen	7	82	89
prefers pencil	6	5	11
total	13	87	100

Is there a possible association between dominant hand and writing utensil preference? Explain your reasoning.

3 Associating Your Own Variables

Student Task Statement

1. Work with your group to identify a pair of categorical variables you think might be associated and another pair you think would not be associated.

2. Imagine your group collected data for each pair of categorical variables. Create a two-way table that could represent each set of data. Invent some data with 100 total values to complete each table. Remember that one table shows a possible association, and the other table shows no association.

3. Explain or show why there appears to be an association for the first pair of variables and why there appears to be no association for the other pair of variables.

4. Prepare a display of your work to share.