Name: _____ Date: _____

UNDERSTANDING ASSOCIATIONS IN CATEGORICAL DATA
N-GEN MATH® 8

Statistics is about **understanding** and **quantifying** the **variation** in the data used to answer a statistical question. We can do this for categorical data by further examining the **conditional relative frequencies** associated with the data.

***Exercise* #1:** For his math project, Julian is trying to understand if there is a difference between whether people buy snacks at the movie theater based on whether they are a child or an adult. He asks 40 subjects exiting a movie and finds the data in the table below.

	Snack	No Snack	Total
Child	18	7	25
Adult	9	6	15
Total	27	13	40

(a) For each **row**, set up the row conditional relative frequency by dividing each entry in the table by its row total. Show the division (fraction). Express each decimal to the nearest hundredth.

	Snack	No Snack	Total
Child			
Adult			
Total			

Row Conditional Relative Frequencies

(b) What is the overall relative frequency of buying a snack? Give an interpretation that involves percent.

(c) What is the percent of children who buy a snack?

(d) What is the percentage of adults that buy a snack?

(e) Is there an association between a person's age and whether they buy a snack at the movies? Explain.

When we divide each entry by either a **row total** or a **column total** we are finding **conditional relative frequencies**. We can interpret differences in these to determine if there is any association between categories.

***Exercise* #2:** Harmony is trying to see if there is an association between the middle school grade a student is in and whether they prefer vanilla or chocolate ice cream. She asks a total of 50 6^{th} and 7^{th} grade students their preference. Her results are shown below.

	Vanilla	Chocolate	Total
6^{th} Grade	9	13	22
7^{th} Grade	12	16	28
Total	21	29	50

(a) Like in *Exercise* #1, create a table containing the row conditional relative frequencies. Show your division and express each decimal to two decimal place accuracy.

	Vanilla	Chocolate	Total
6^{th} Grade			
7^{th} Grade			
Total			

(b) What percentage of 6^{th} grade students preferred vanilla ice cream?

(c) What percentage of 7^{th} grade students preferred vanilla ice cream?

(d) What do your answers to (b) and (c) tell you about the association between ice cream preference and grade level?

(e) What would dividing by the column totals for this situation tell you?

Name: _____ Date: _____

UNDERSTANDING ASSOCIATIONS IN CATEGORICAL DATA
N-GEN MATH® 8 HOMEWORK

USING YOUR MATH

1. Minji believes there is likely an association between whether a student plays a sport or not and whether the student walks home or gets a ride home from school. She asks 80 random students if they play sports and how they get home after school. The results are shown below.

	Walk	Ride	Total
Sport	7	25	32
No Sport	21	27	48
Total	28	52	80

(a) Set up the row conditional relative frequency by dividing each entry in the table by its row total. Show the division (fraction). Express each decimal to the nearest hundredth.

	Walk	Ride	Total
Sport			
No Sport			
Total			

(b) What percentage of students who play a sport walk home?

(c) What percentage of students who do not play a sport walk home?

(d) Explain how your answers to (b) and (c) help support Minji's belief that an association exits between the two categories.

(e) Do you think it is a strong association or a weak association? Explain.

2. Luca is trying to determine if there is an association between animal type in an animal shelter and whether the animal is male or female. He visits a local animal shelter and finds the following data for a particular day.

	Cat	Dog	Total
Male	23	13	37
Female	19	10	28
Total	42	23	65

(a) Set up the **column** conditional relative frequency by dividing each entry in the table by the **column total**. Show the division. Express each decimal to the nearest hundredth.

	Cat	Dog	Total
Male			
Female			
Total			

(b) What percentage of cats were male?

(c) What percentage of dogs were male?

(d) Do the results in (b) and (c) support the idea that there is an association between the categories? Explain.

(e) When we divided by the column totals, we found the percentage of cats that were male or female and the percentage of dogs that were male or female.

What would have we found if we divided by the row totals instead?