Module 4: Describing Distributions

TOPIC 2: TWO-VARIABLE CATEGORICAL DATA

In this topic, students first learn how to create two-way frequency distributions to display two variables whose data can be grouped into categories and analyze the data for any possible trends or associations. However, because analyzing raw frequencies can lead to misinterpretations, students learn to display categorical data using a relative frequency distribution. Conditional relative frequency distributions organize categorical data so that students can determine the percent of occurrence of a category given the specific value of another category. Finally, students use all the tools that they have to analyze a set of categorical data and make recommendations based on their interpretation.

Where have we been?

Since middle school, students have collected, organized, and interpreted data using the statistical process. And in middle school, students were first introduced to bivariate categorical data. They have constructed and interpreted two-way frequency and relative frequency tables and used them to describe possible associations.

Where are we going?

As students increase their proficiency with the statistical process, they are able to perform more advanced analyses of complex data sets. Building on their deep understanding of how to use two-way tables to draw inferences about categorical data, they will be able to use the two-way table to decide if probabilistic events are independent and to approximate conditional probabilities.
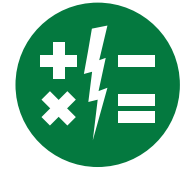
Conditional Relative Frequency

A conditional relative frequency distribution is the percent or ratio of occurrences of a category given the specific value of another category.

This conditional relative frequency distribution gives the percent of participants with each handedness, given the type of sports they play.

Sports Participation

Hand Favored		Individual	Team	Does Not Play
	Left	$\frac{3}{10} = 30\%$	$\frac{13}{39} = 33\%$	$\frac{8}{14} = 57\%$
	Right	$\frac{6}{10} = 60\%$	$\frac{23}{39} = 59\%$	$\frac{4}{14} = 29\%$
	Mixed	$\frac{1}{10} = 10\%$	$\frac{3}{39} = 8\%$	$\frac{2}{14} = 14\%$
	Total	$\frac{10}{10} = 100\%$	$\frac{39}{39} = 100\%$	$\frac{14}{14} = 100\%$

Blinded Me with Science!

In the Age of Enlightenment, which took place in the 17th and 18th centuries, there was rapid scientific advancement where scientists such as Descartes and Newton confirmed scientific thinking with experiments and mathematics. Today there are two major groups of sciences: natural sciences and social sciences. The natural sciences include topics such as astronomy, biology, chemistry, physics, and earth science. The social sciences include topics dealing with society and human behavior such as economics, linguistics, and psychology.

Both the natural sciences and the social sciences make heavy use of statistics and statistical analysis.

Talking Points

Categorical data is an important topic to know about for college admissions tests.

Here is a sample question:

At a diner, each meal comes with a free appetizer or dessert. Data were collected regarding the choice of the past 500 customers based on age. Out of a group of 10 high school students who just entered the diner, use the data to predict how many will order a free appetizer.

	Appetizer	Dessert
Ages 10–24	70	100
Ages 25–39	135	50
Ages 40–64	50	95

High school students would be in the 10−24 year age range. According to the data for that age range, $\frac{70}{170}$ or 41.1% chose an appetizer. So, for 10 students, 10 × 0.41, or about 4 of them would probably order a free appetizer.

Key Terms

categorical data
Categorical data are data that can be grouped into categories, unlike numerical data that can be placed on a numerical scale and compared.

marginal frequency distribution
A marginal frequency distribution displays the total of the frequencies of the rows or columns of a frequency distribution.

relative frequency distribution
A relative frequency distribution provides the ratio of occurrences in each category to the total number of occurrences.