

1.1 Analyzing Categorical Data (Two-Way Tables)

Learning Targets

1. Calculate and display the marginal distribution of a categorical variable from a two-way table.
2. Calculate and display the conditional distribution of a categorical variable for a particular value of the other categorical variable in a two-way table.
3. Describe the association between two categorical variables by comparing appropriate conditional distributions.

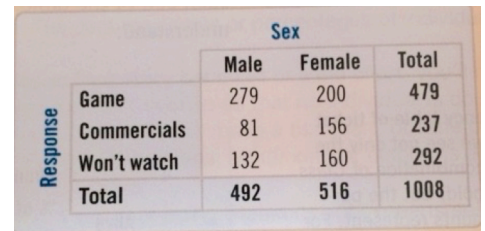
Vocabulary: marginal distributions. Conditional distribution, side-by-side bar graph, segmented bar graph, association

Read 11–18

What is a two-way table? What is a marginal distribution?

Example: “Super Bowl Survey” - A recent Gallup poll asked 1008 Americans age 18 and over whether they planned to watch the upcoming Super Bowl. The pollsters asked those who planned to watch whether they were looking forward to seeing the football game or the commercials. The results are summarized in the table below.

- a. What are the two categorical variables displayed on this table?



	Sex		Total
	Male	Female	
Game	279	200	479
Commercials	81	156	237
Won't watch	132	160	292
Total	492	516	1008

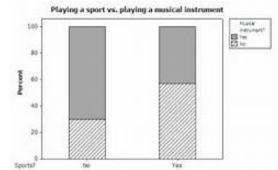
- b. What does the “160” value in the table represent?

- c. Calculate the marginal distribution (in percents) of responses.

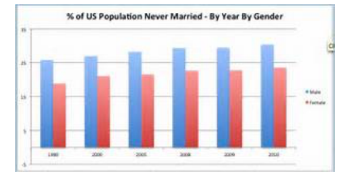
- d. Make a graph to display the marginal distribution. Describe what you see. *Always label axis and title of graph.

What is a conditional distribution? How do we know which variable to condition on?

What is a segmented bar graph (aka stacked bar graph)? Why are they good to use?



What is a side-by-side graph? What are they good for?



Going back to the “Super Bowl” example, calculate the conditional distribution of responses among males and calculate the conditional distribution of responses among the females. Create a graph that compares the responses amongst males and females.

	Sex		Total
	Male	Female	
Game	279	200	479
Commercials	81	156	237
Won't watch	132	160	292
Total	492	516	1008

What does it mean for two variables to have an association? How can you tell by looking at a graph?

Example: Cell Phones

The Pew Research Center asked a random sample of 2024 adult cell phone owners from the United States which type of cell phone they own: iPhone, Android, or other (including non-smart phones). Here are the results, broken down by age category

	18–34	35–54	55+	Total
iPhone	169	171	127	467
Android	214	189	100	503
Other	134	277	643	1054
Total	517	637	870	2024

(a) Explain what it would mean if there was no association between age and cell phone type.

(b) Based on this data, can we conclude there is an association between age and cell phone type? Justify.

Activities: Check Your Understanding (p. 18) & A Titanic Disaster (p. 19)

HW: page 22 (20, 22, 23, 25, 27–34) & work through the “I’m Gonna Be Rich” running example on pp. 12-15, 17, Read 1.2