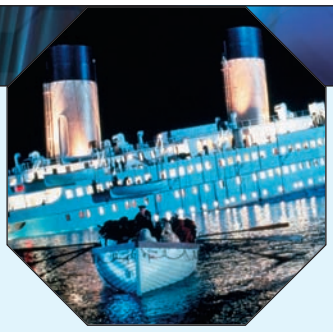

WHAT HAVE WE LEARNED?

We've learned that we can summarize categorical data by counting the number of cases in each category, sometimes expressing the resulting distribution as percents. We can display the distribution in a bar chart or a pie chart. When we want to see how two categorical variables are related, we put the counts (and/or percentages) in a two-way table called a contingency table.

- We look at the marginal distribution of each variable (found in the margins of the table).
- We also look at the conditional distribution of a variable within each category of the other variable.
- We can display these conditional and marginal distributions by using bar charts or pie charts.
- If the conditional distributions of one variable are (roughly) the same for every category of the other, the variables are independent.

Terms

Frequency table (Relative frequency table)
21. A frequency table lists the categories in a categorical variable and gives the count (or percentage of observations for each category.

Distribution
22. The distribution of a variable gives
- the possible values of the variable and
- the relative frequency of each value.

Area principle
22. In a statistical display, each data value should be represented by the same amount of area.

Bar chart (Relative frequency bar chart)
22. Bar charts show a bar whose area represents the count (or percentage) of observations for each category of a categorical variable.

Pie chart
23. Pie charts show how a "whole" divides into categories by showing a wedge of a circle whose area corresponds to the proportion in each category.

Categorical data condition
24. The methods in this chapter are appropriate for displaying and describing categorical data. Be careful not to use them with quantitative data.

Contingency table
24. A contingency table displays counts and, sometimes, percentages of individuals falling into named categories on two or more variables. The table categorizes the individuals on all variables at once to reveal possible patterns in one variable that may be contingent on the category of the other.

Marginal distribution
24. In a contingency table, the distribution of either variable alone is called the marginal distribution. The counts or percentages are the totals found in the margins (last row or column) of the table.

Conditional distribution
26. The distribution of a variable restricting the *Who* to consider only a smaller group of individuals is called a conditional distribution.

Independence
29. Variables are said to be independent if the conditional distribution of one variable is the same for each category of the other. We'll show how to check for independence in a later chapter.

Segmented bar chart
30. A segmented bar chart displays the conditional distribution of a categorical variable within each category of another variable.

Simpson's paradox
34. When averages are taken across different groups, they can appear to contradict the overall averages. This is known as "Simpson's paradox."

Skills

THINK
- Be able to recognize when a variable is categorical and choose an appropriate display for it.
- Understand how to examine the association between categorical variables by comparing conditional and marginal percentages.

SHOW
- Be able to summarize the distribution of a categorical variable with a frequency table.
- Be able to display the distribution of a categorical variable with a bar chart or pie chart.
- Know how to make and examine a contingency table.