Identifying Descriptive and Inferential Statistics

In a news report on the state of the media by Tom Rosenstiel and Amy Mitchell, they write the following:

"AOL had 900 journalists, 500 of them at its local Patch news operation.... By the end of 2011, Bloomberg expects to have 150 journalists and analysts for its new Washington operation, Bloomberg Government."

Source: Rosenstiel, Tom and Amy Mitchell. "Overview." The State of the News Media: An Annual Report on American Journalism. Pew Research Center's Project for Excellence in Journalism. 2011. http://stateofthemedia.org/2011/overview-2/ (12 Dec. 2011).

Identify the descriptive and inferential statistics used in this excerpt from their article.

Experimental Designs

Other than a census and sampling, another method for obtaining information is experimentation.

Observational Studies and Designed Experiments

Besides classifying statistical studies as either descriptive or inferential, we often need to classify them as either *observational studies* or *designed experiments*.

In an **observational study,** researchers simply observe characteristics and take measurements, as in a sample survey.

In a **designed experiment,** researchers impose treatments and controls and then observe characteristics and take measurements.

Observational studies can reveal only *association,* whereas designed experiments can help establish *causation.*

Vasectomies and Prostate Cancer

One study found 113 cases of prostate cancer among 22,000 men who had a vasectomy. This compares to a rate of 70 cases per 22,000 among men who didn't have a vasectomy." The study shows about a 60% elevated risk of prostate cancer for men who have had a vasectomy, thereby revealing an association between vasectomy and prostate cancer. But does it establish causation: that having a vasectomy causes an increased risk of prostate cancer?

The answer is no, because the study was observational. The researchers simply observed two groups of men, one with vasectomies and the other without. Thus, although an association was established between vasectomy and prostate cancer, the association might be due to other factors (e.g., temperament) that make some men more likely to have vasectomies and also put them at greater risk of prostate cancer.

Folic Acid and Birth Defects

For the study, the doctors enrolled 4753 women prior to conception and divided them randomly into two groups. One group took daily multivitamins containing 0.8 mg of folic acid, whereas the other group received only trace elements (minute amounts of copper, manganese, zinc, and vitamin C). A drastic reduction in the rate of major birth defects occurred among the women who took folic acid: 13 per 1000, as compared to 23 per 1000 for those women who did not take folic acid.

This is a designed experiment and does help establish causation. The researchers did not simply observe two groups of women but, instead, randomly assigned one group to take daily doses of folic acid and the other group to take only trace elements.

Terminology of Experimental Design

Response Variable, Factors, Levels, and Treatments

Response variable: The characteristic of the experimental outcome that is to be measured or observed.

Factor: A variable whose effect on the response variable is of interest in the experiment.

Levels: The possible values of a factor.

Treatment: Each experimental condition. For one-factor experiments, the treatments are the levels of the single factor. For multifactor experiments, each treatment is a combination of levels of the factors.

Principles of Experimental Design

Three basic principles of experimental design: **control, randomization,** and **replication.**

Control: Two or more treatments should be compared.

Randomization: The experimental units should be randomly divided into groups to avoid unintentional selection bias in constituting the groups.

Replication: A sufficient number of experimental units should be used to ensure that randomization creates groups that resemble each other closely and to increase the chances of detecting any differences among the treatments.

Principles of Experimental Design

Control: The doctors compared the rate of major birth defects for the women who took folic acid to that for the women who took only trace elements.

Randomization: The women were divided randomly into two groups to avoid unintentional selection bias.

Replication: A large number of women were recruited for the study to make it likely that the two groups created by randomization would be similar and also to increase the chances of detecting any effect due to the folic acid.

Definition 1.8

Randomized Block Design

In a **randomized block design,** the experimental units are assigned randomly among all the treatments separately within each block.

Although the completely randomized design is commonly used and simple, it is not always the best design. Several alternatives to that design exist. For instance, in a **randomized block design**, experimental units that are similar in ways that are expected to affect the response variable are grouped in **blocks**. Then the random assignment of experimental units to the treatments is made block by block.

Example 1.16 Statistical Designs

Golf Ball Driving Distances

Suppose we want to compare the driving distances for five different brands of golf ball. For 40 golfers, discuss a method of comparison based on

a. a completely randomized design.
b. a randomized block design.

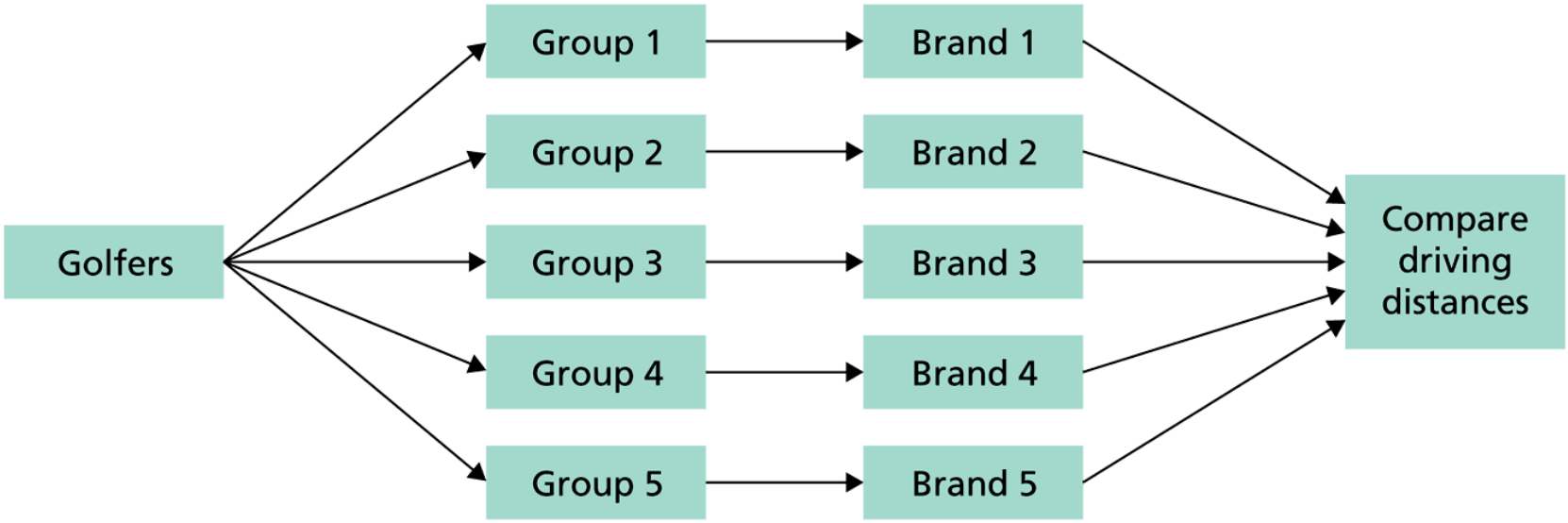
Solution

Here the experimental units are the golfers, the response variable is driving distance, the factor is brand of golf ball, and the levels (and treatments) are the five brands.

a. For a completely randomized design, we would randomly divide the 40 golfers into five groups of 8 golfers each and then randomly assign each group to drive a different brand of ball, as illustrated in Fig.1.5.

Figure 1.5

Completely randomized design for golf ball experiment



ALWAYS LEARNING
PEARSON

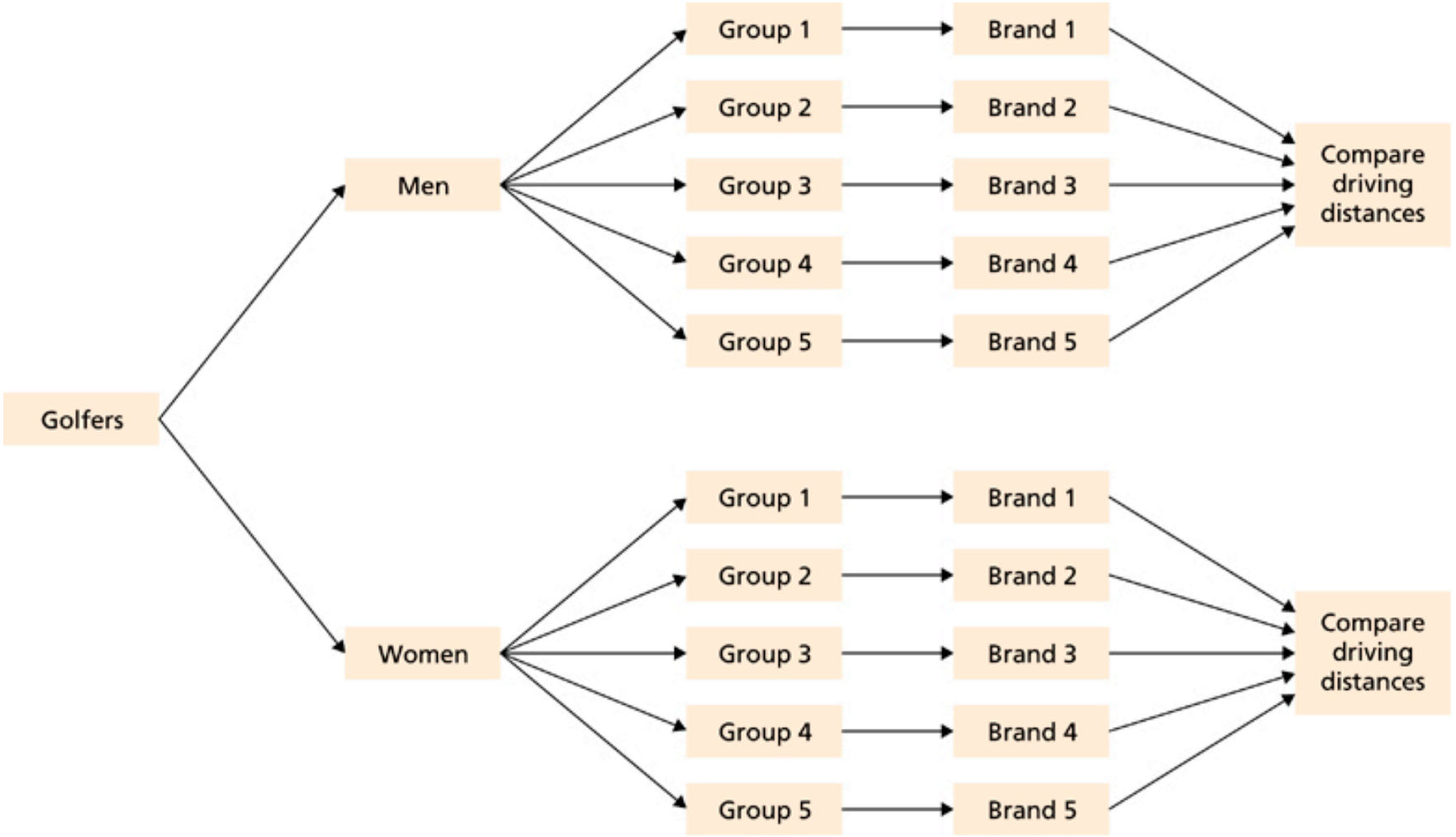
Example 1.16 Statistical Designs

Golf Ball Driving Distances

b. Because driving distance is affected by gender, using a randomized block design that blocks by gender is probably a better approach. We could do so by using 20 men golfers and 20 women golfers. We would randomly divide the 20 men into five groups of 4 men each and then randomly assign each group to drive a different brand of ball, as shown in Fig.1.6. Likewise, we would randomly divide the 20 women into five groups of 4 women each and then randomly assign each group to drive a different brand of ball, as also shown in Fig.1.6.

Figure 1.6

Randomized block design for golf ball experiment



ALWAYS LEARNING
PEARSON

By blocking, we can isolate and remove the variation in driving distances between men and women and thereby make it easier to detect any differences in driving distances among the five brands of golf ball. Additionally, blocking permits us to analyze separately the differences in driving distances among the five brands for men and women.

As illustrated in Example 1.16, blocking can isolate and remove systematic differences among blocks, thereby making any differences among treatments easier to detect. Blocking also makes possible the separate analysis of treatment effects on each block.

Chapter 2

Organizing Data

Definition 2.1

Variable: A characteristic that varies from one person or thing to another.

Qualitative variable: A nonumerically valued variable.

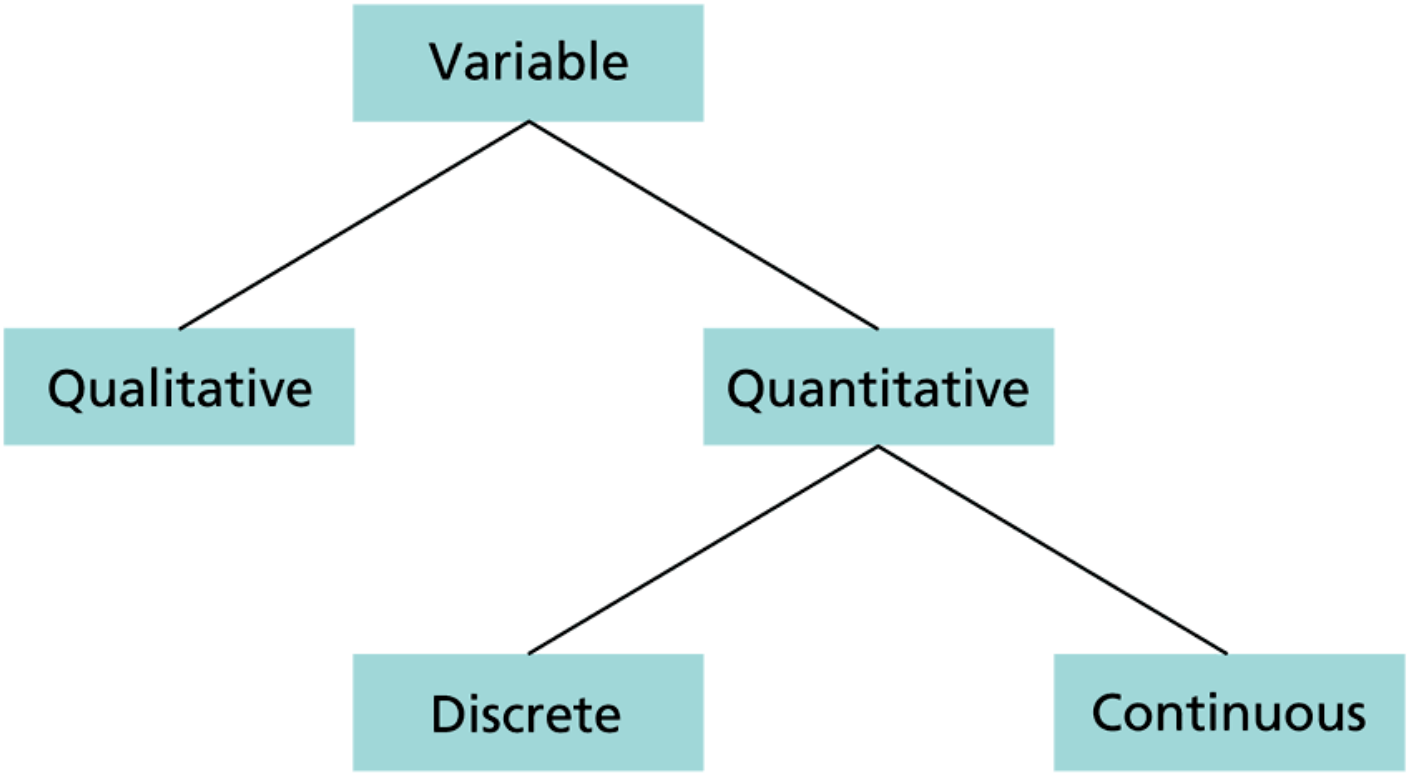
Quantitative variable: A numerically valued variable.

Discrete variable: A quantitative variable whose possible values can be listed. In particular, a quantitative variable with only a finite number of possible values is a discrete variable.

Continuous variable: A quantitative variable whose possible values form some interval of numbers.

Figure 2.1

Types of variables



Definition 2.2

Data

Data: Values of a variable.

Qualitative data: Values of a qualitative variable.

Quantitative data: Values of a quantitative variable.

Discrete data: Values of a discrete variable.

Continuous data: Values of a continuous variable.

Definition 2.3

Frequency Distribution of Qualitative Data

A **frequency distribution** of qualitative data is a listing of the distinct values and their frequencies.

Procedure 2.1

To Construct a Frequency Distribution of Qualitative Data

Step 1 List the distinct values of the observations in the data set in the first column of a table.

Step 2 For each observation, place a tally mark in the second column of the table in the row of the appropriate distinct value.

Step 3 Count the tallies for each distinct value and record the totals in the third column of the table.

Table 2.1

Political party affiliations of the students in introductory statistics

Democratic	Other	Democratic	Other	Democratic
Republican	Republican	Other	Other	Republican
Republican	Republican	Republican	Democratic	Republican
Republican	Democratic	Democratic	Other	Republican
Democratic	Democratic	Republican	Democratic	Democratic
Republican	Republican	Other	Other	Democratic
Republican	Democratic	Republican	Other	Other
Republican	Republican	Republican	Democratic	Republican

Table 2.2

Table for constructing a frequency distribution for the political party affiliation data in Table 2.1

Party	Tally	Frequency
Democratic	𝍸 𝍸 |||	13
Republican	𝍸 𝍸 𝍸 |||	18
Other	𝍸 ||||	9
		40

Definition 2.4

Relative-Frequency Distribution of Qualitative Data

A **relative-frequency distribution** of qualitative data is a listing of the distinct values and their relative frequencies.

Procedure 2.2

To Construct a Relative-Frequency Distribution of Qualitative Data

Step 1 Obtain a frequency distribution of the data.

Step 2 Divide each frequency by the total number of observations.

Table 2.3

Relative-frequency distribution for the political party affiliation data in Table 2.1

Party	Relative frequency	
Democratic	0.325	← *13/40*
Republican	0.450	← *18/40*
Other	0.225	← *9/40*
	1.000	

Definition 2.5

Pie Chart

A **pie chart** is a disk divided into wedge-shaped pieces proportional to the relative frequencies of the qualitative data.

Procedure 2.3

To Construct a Pie Chart

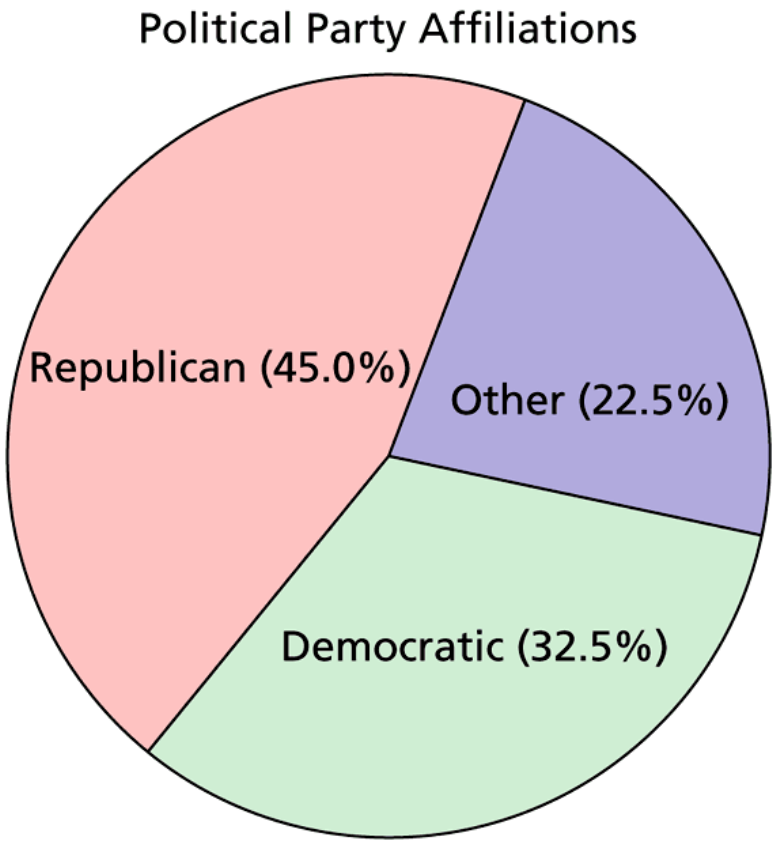
Step 1 Obtain a relative-frequency distribution of the data by applying Procedure 2.2.

Step 2 Divide a disk into wedge-shaped pieces proportional to the relative frequencies.

Step 3 Label the slices with the distinct values and their relative frequencies.

Figure 2.2

Pie chart of the political party affiliation data in Table 2.1



Definition 2.6

Bar Chart

A **bar chart** displays the distinct values of the qualitative data on a horizontal axis and the relative frequencies (or frequencies or percents) of those values on a vertical axis. The relative frequency of each distinct value is represented by a vertical bar whose height is equal to the relative frequency of that value. The bars should be positioned so that they do not touch each other.

Procedure 2.4

To Construct a Bar Chart

Step 1 Obtain a relative-frequency distribution of the data by applying Procedure 2.2.

Step 2 Draw a horizontal axis on which to place the bars and a vertical axis on which to display the relative frequencies.

Step 3 For each distinct value, construct a vertical bar whose height equals the relative frequency of that value.

Step 4 Label the bars with the distinct values, the horizontal axis with the name of the variable, and the vertical axis with "Relative frequency."

Figure 2.3

Bar chart of the political party affiliation data in Table 2.1

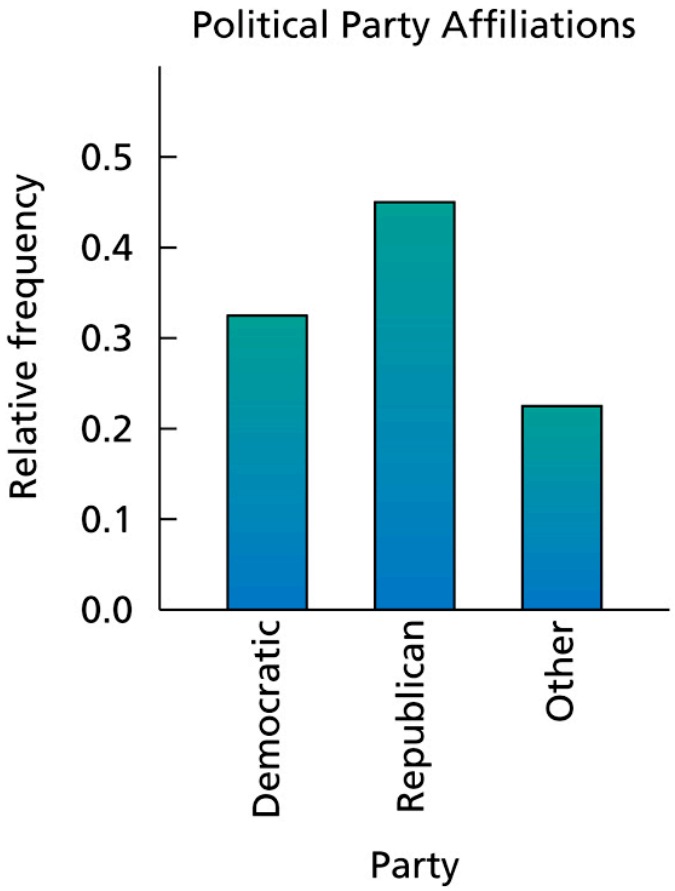


Table 2.4

Number of TV sets in each of 50 randomly selected households.

1	1	1	2	6	3	3	4	2	4
3	2	1	5	2	1	3	6	2	2
3	1	1	4	3	2	2	2	2	3
0	3	1	2	1	2	3	1	1	3
3	2	1	2	1	1	3	1	5	1

Organizing Quantitative Data

Table 2.5

Frequency and relative-frequency distributions, using **single-value grouping,** for the number-of-TVs data in Table 2.4

Number of TVs	Frequency	Relative frequency
0	1	0.02
1	16	0.32
2	14	0.28
3	12	0.24
4	3	0.06
5	2	0.04
6	2	0.04
	50	1.00

Definition 2.7

Terms Used in Limit Grouping

Lower class limit: The smallest value that could go in a class.

Upper class limit: The largest value that could go in a class.

Class width: The difference between the lower limit of a class and the lower limit of the next-higher class.

Class mark: The average of the two class limits of a class.

Table 2.6

Days to maturity for 40 short-term investments

70	64	99	55	64	89	87	65
62	38	67	70	60	69	78	39
75	56	71	51	99	68	95	86
57	53	47	50	55	81	80	98
51	36	63	66	85	79	83	70

Table 2.7

Frequency and relative-frequency distributions, using **limit grouping**, for the days-to-maturity data in Table 2.6

Days to maturity	Tally	Frequency	Relative frequency
30–39	|||	3	0.075
40–49	|	1	0.025
50–59	𝍸 |||	8	0.200
60–69	𝍸 𝍸	10	0.250
70–79	𝍸 ||	7	0.175
80–89	𝍸 ||	7	0.175
90–99	||||	4	0.100
		40	1.000