

Introductory
STATISTICS

9TH EDITION



Neil
WEISS

Chapter 2

Organizing Data



Section 2.1

Variables and Data



Definition 2.1

Variables

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

Qualitative variable: A nonnumerically valued variable.

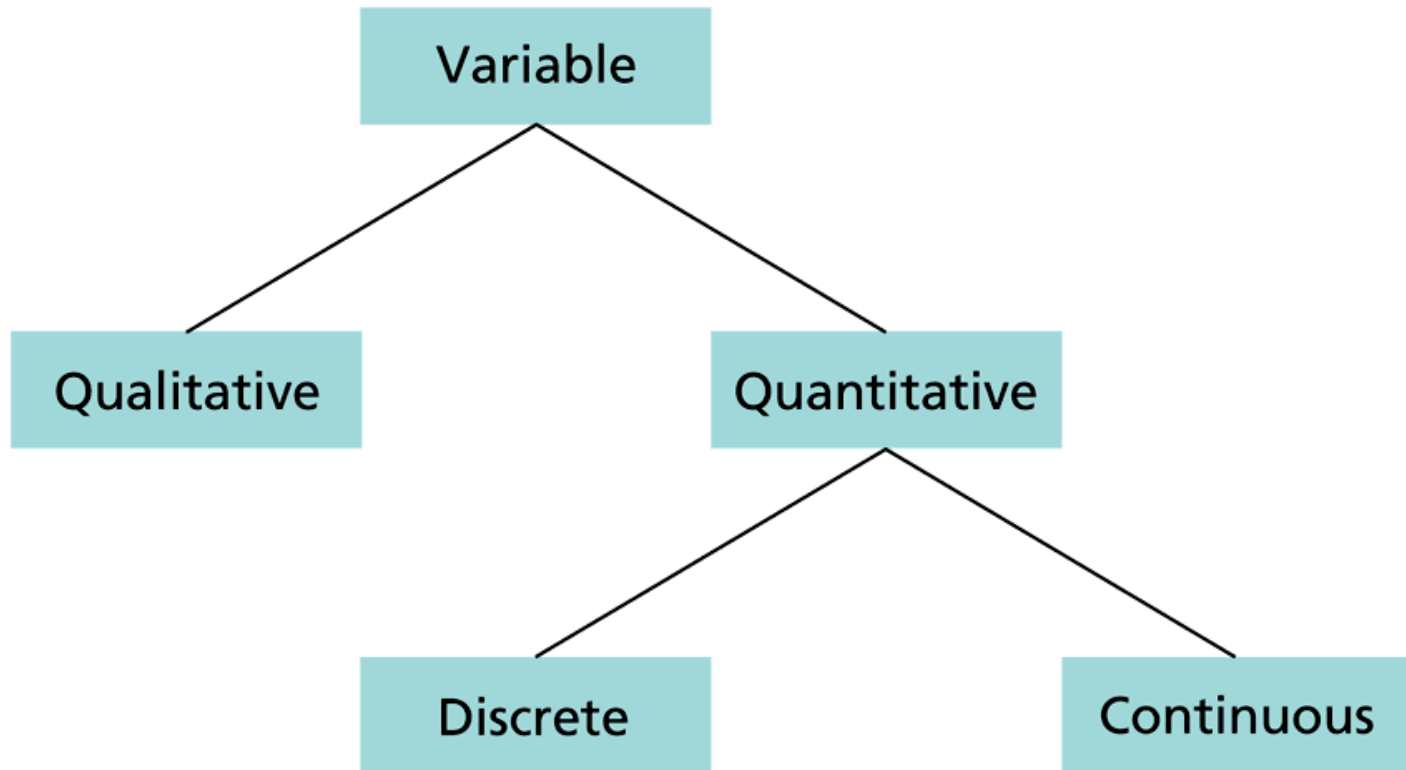
Quantitative variable: A numerically valued variable.

Discrete variable: A quantitative variable whose possible values can be listed.

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.

Section 2.2

Organizing Qualitative Data



Definition 2.3

Frequency Distribution of Qualitative Data

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

Table 2.1

Political party affiliations of the students in introductory statistics

D	R	O	R	R	R	R	R
D	O	R	D	O	O	R	D
D	R	O	D	R	R	O	R
D	O	D	D	D	R	O	D
O	R	D	R	R	R	R	D

Table 2.2

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

Party	Tally	Frequency
Democratic		13
Republican		18
Other		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.

Table 2.3

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

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

Figure 2.2

Pie chart of the political party affiliation data in Table 2.1

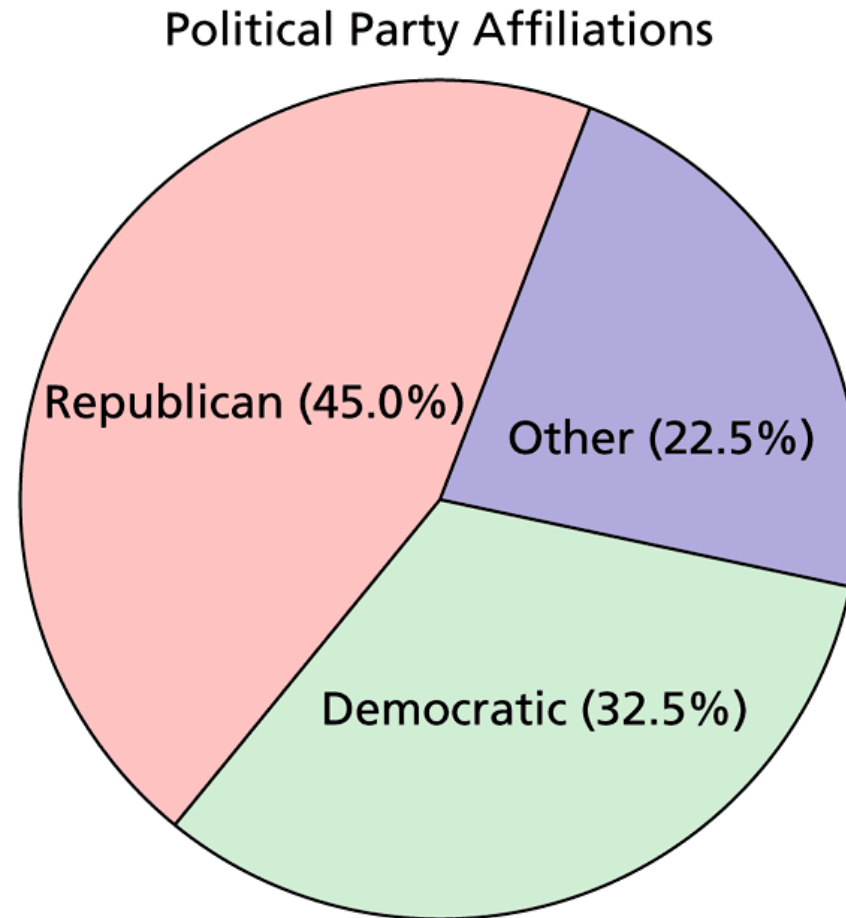
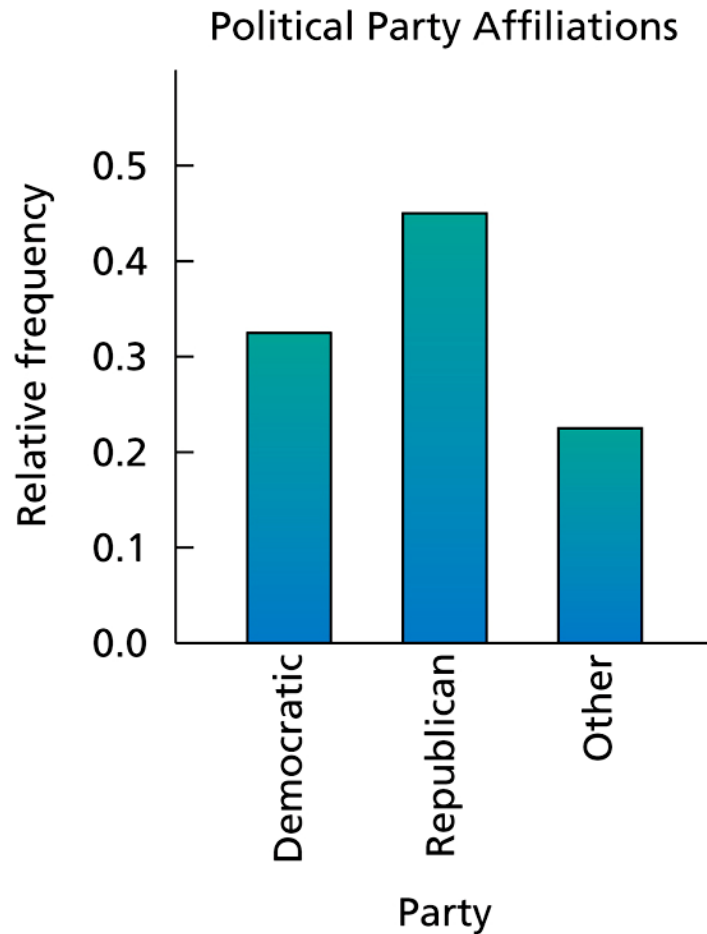


Figure 2.3

Bar chart of the political party affiliation data in Table 2.1



Section 2.3

Organizing Quantitative Data



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

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

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		8	0.200
60–69		10	0.250
70–79		7	0.175
80–89		7	0.175
90–99		4	0.100
		40	1.000

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.

Definition 2.8

Terms Used in Cutpoint Grouping

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

Upper class cutpoint: The largest value that could go in the next-higher class (equivalent to the lower cutpoint of the next-higher class).

Class width: The difference between the cutpoints of a class.

Class midpoint: The average of the two cutpoints of a class.

Definition 2.9

Histogram

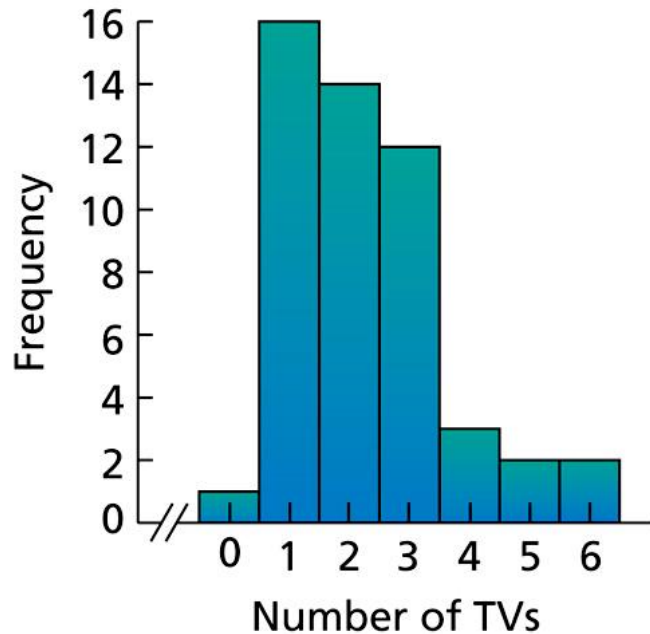
A **histogram** displays the classes of the quantitative data on a horizontal axis and the frequencies (relative frequencies, percents) of those classes on a vertical axis. The frequency (relative frequency, percent) of each class is represented by a vertical bar whose height is equal to the frequency (relative frequency, percent) of that class. The bars should be positioned so that they touch each other.

- For single-value grouping, we use the distinct values of the observations to label the bars, with each such value centered under its bar.
 - For limit grouping or cutpoint grouping, we use the lower class limits (or, equivalently, lower class cutpoints) to label the bars.
- Note: Some statisticians and technologies use class marks or class midpoints centered under the bars.

Figure 2.4

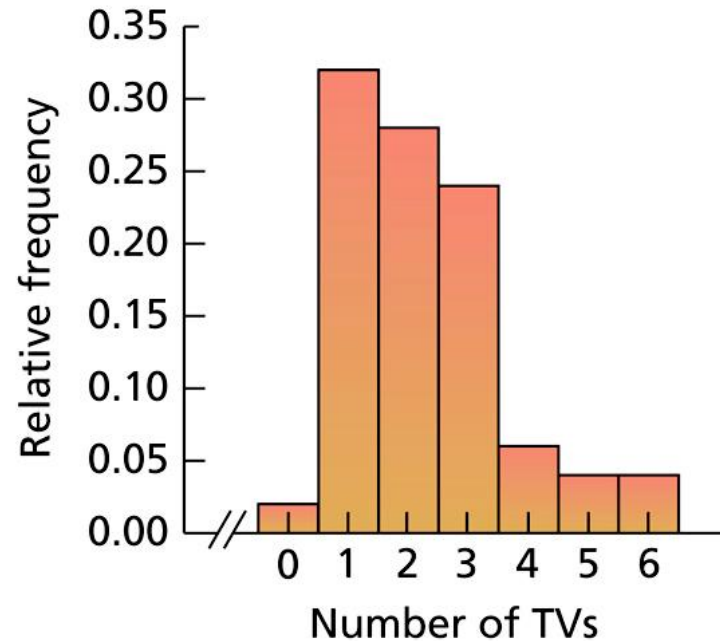
Single-value grouping. Number of TVs per household:
(a) frequency histogram; (b) relative-frequency histogram

Television Sets per Household



(a)

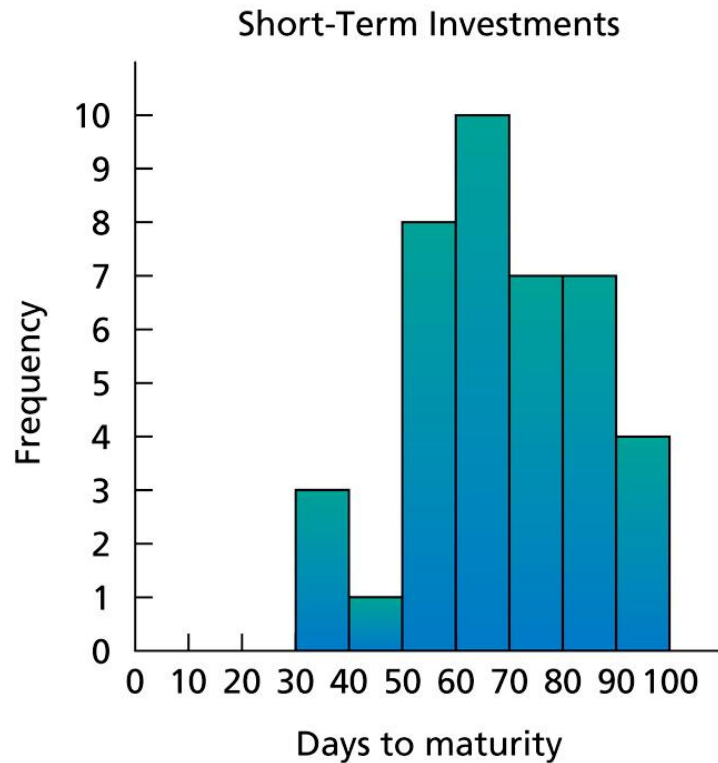
Television Sets per Household



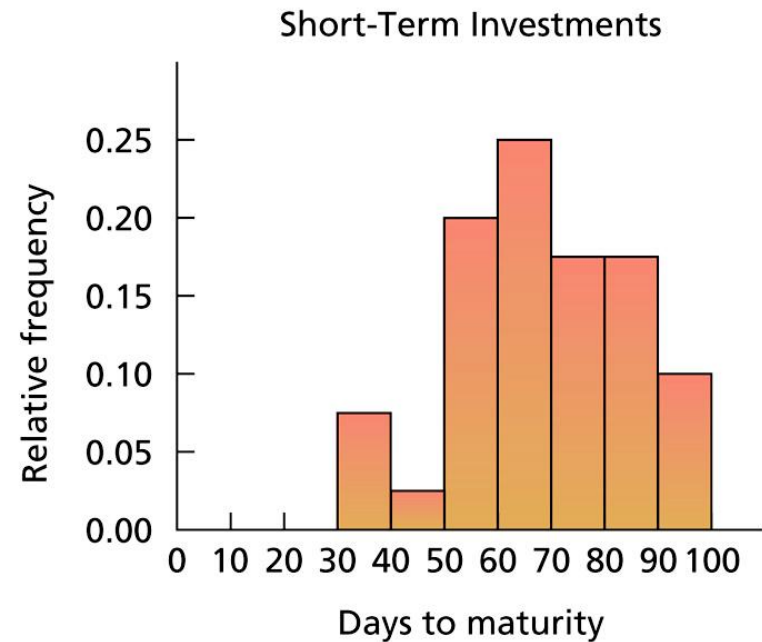
(b)

Figure 2.5

Limit grouping. Days to maturity: (a) frequency histogram; (b) relative-frequency histogram



(a)



(b)

Table 2.11 & Figure 2.7

Prices, in dollars, of 16 DVD players

210	219	214	197
224	219	199	199
208	209	215	199
212	212	219	210



Table 2.12 & Figure 2.8

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

Constructing a stem-and-leaf diagram
for the days-to-maturity data

	Stems	Leaves
3	3	6 8 9
4	4	7
5	5	0 1 1 3 5 5 6 7
6	6	0 2 3 4 4 5 6 7 8 9
7	7	0 0 0 1 5 8 9
8	8	0 1 3 5 6 7 9
9	9	5 8 9 9

(a)

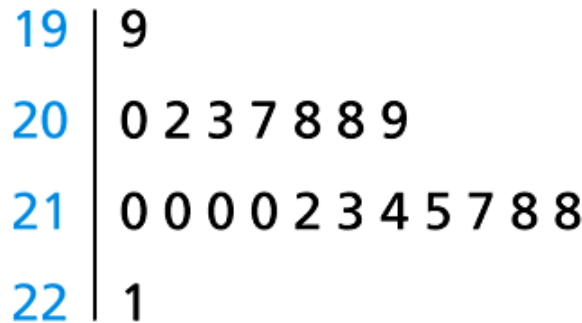
(b)

Table 2.13 & Figure 2.9

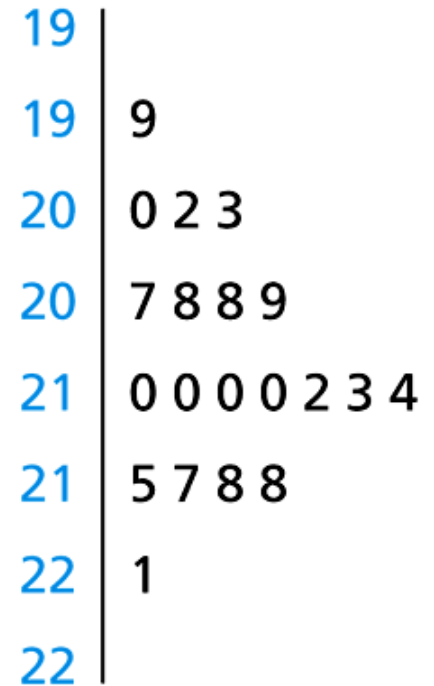
Cholesterol levels
for 20 high-level patients

210	209	212	208
217	207	210	203
208	210	210	199
215	221	213	218
202	218	200	214

Stem-and-leaf diagram for cholesterol levels: (a) one line per stem; (b) two lines per stem



(a)



(b)

Section 2.4

Distribution Shapes



Definition 2.10

Distribution of a Data Set

The **distribution of a data set** is a table, graph, or formula that provides the values of the observations and how often they occur.

Figure 2.10

Relative-frequency histogram and approximating smooth curve for the distribution of heights

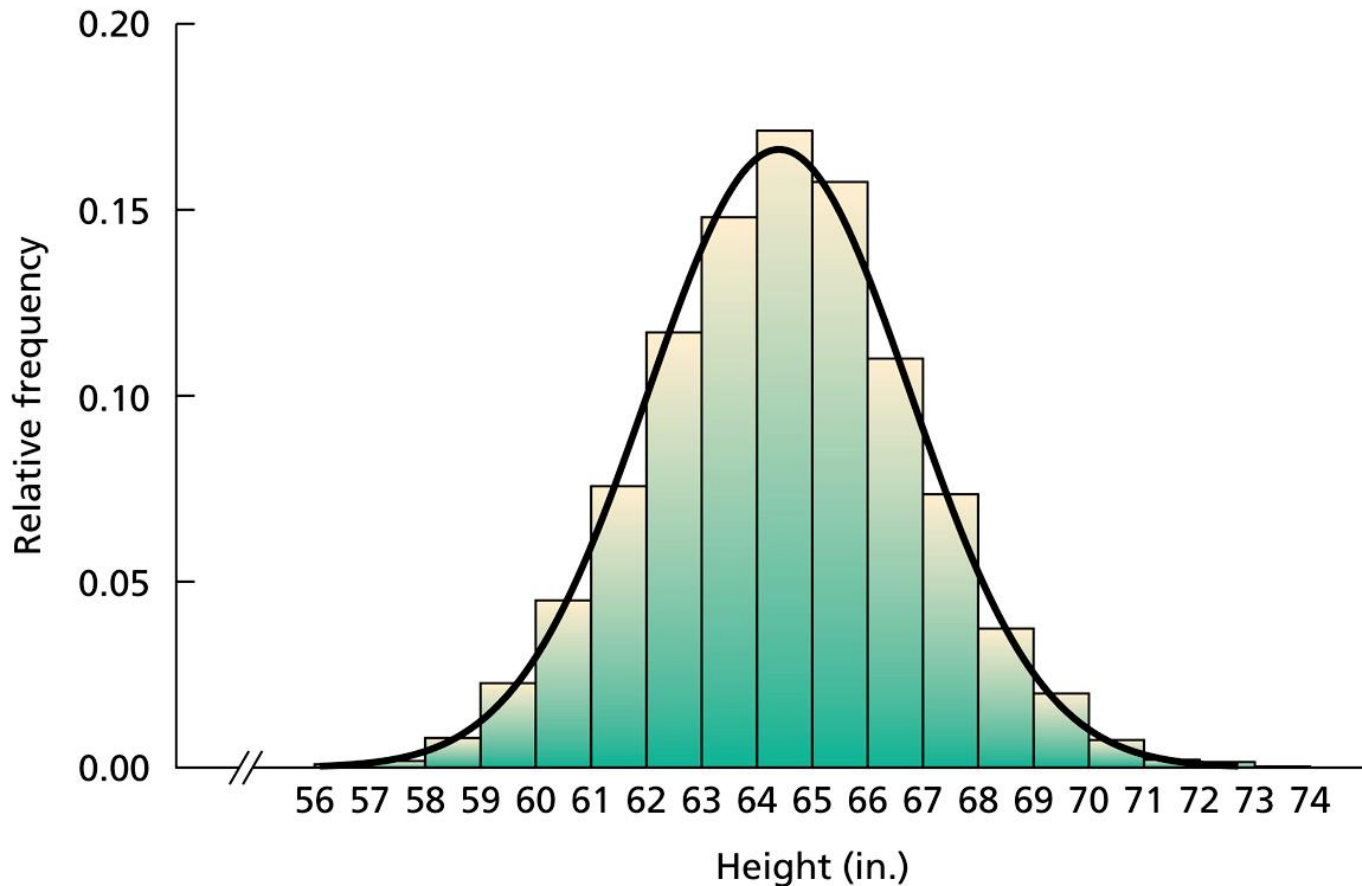
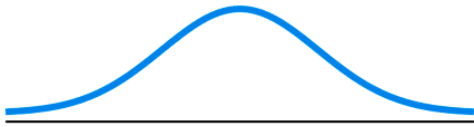
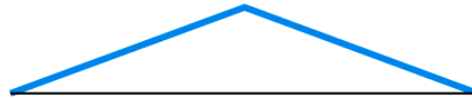


Figure 2.11

Common distribution shapes



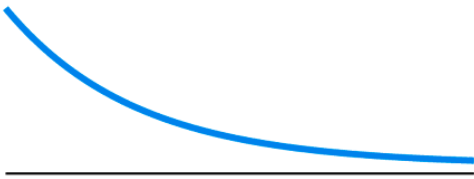
(a) Bell shaped



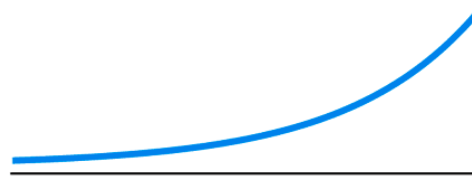
(b) Triangular



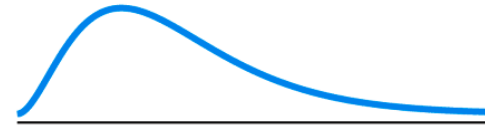
(c) Uniform (or rectangular)



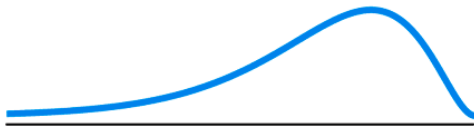
(d) Reverse J shaped



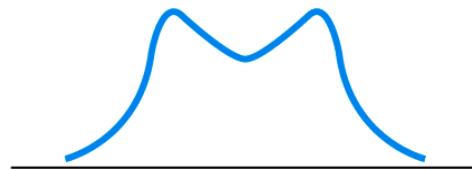
(e) J shaped



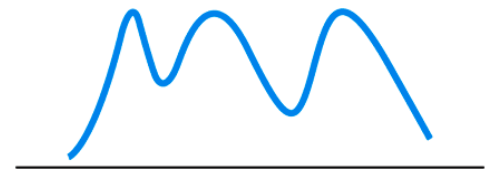
(f) Right skewed



(g) Left skewed



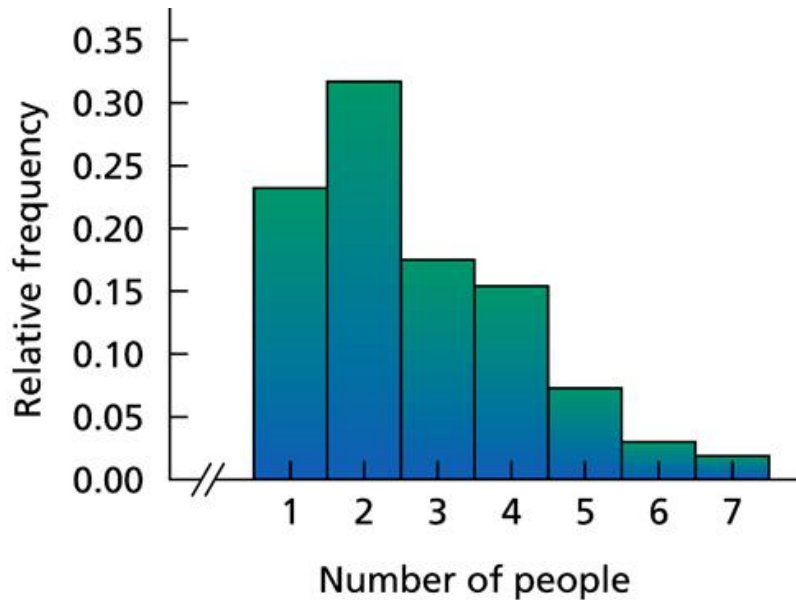
(h) Bimodal



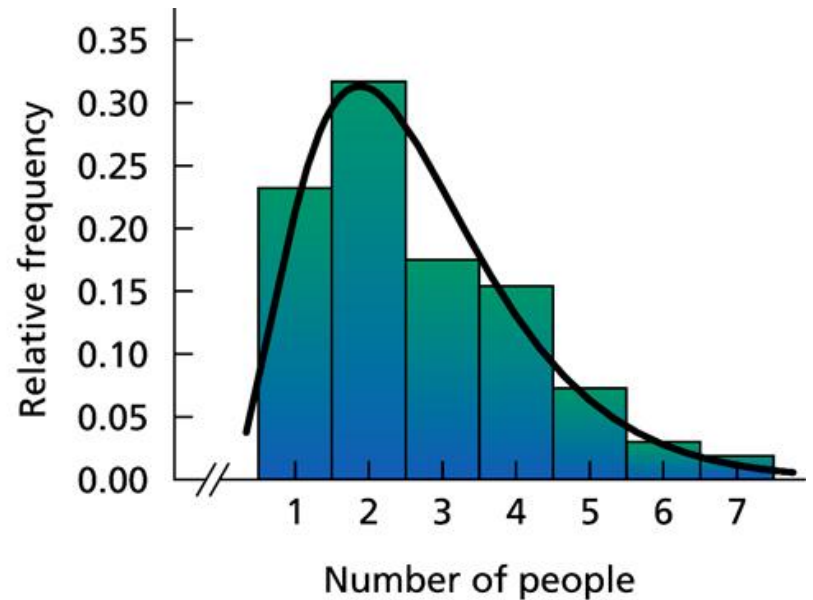
(i) Multimodal

Figure 2.12

Relative-frequency histogram for household size



(a)



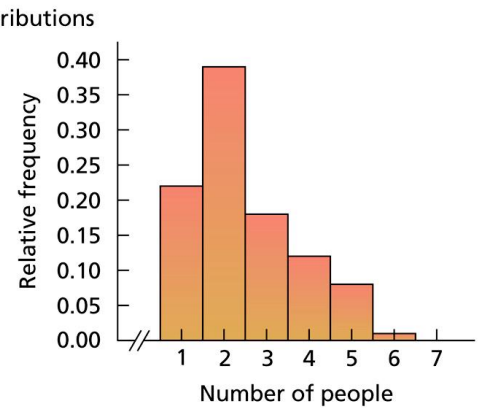
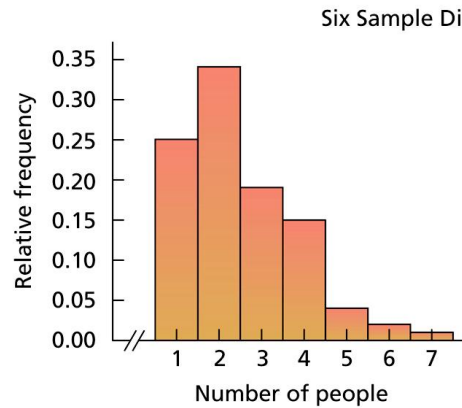
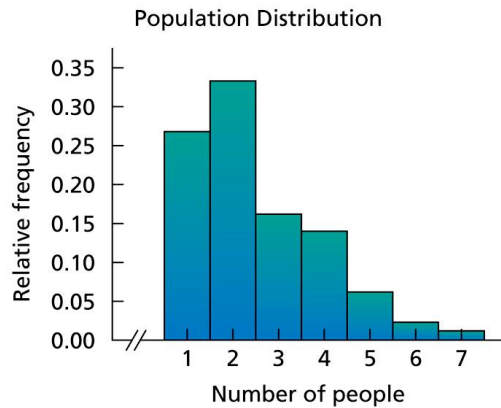
(b)

Definition 2.12

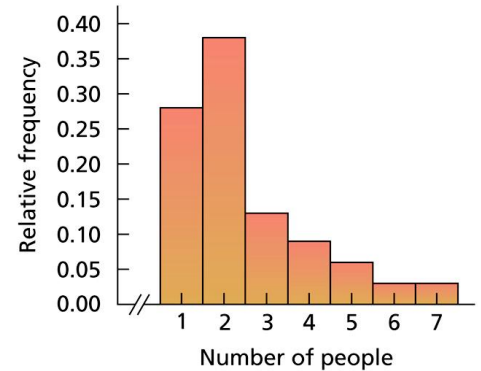
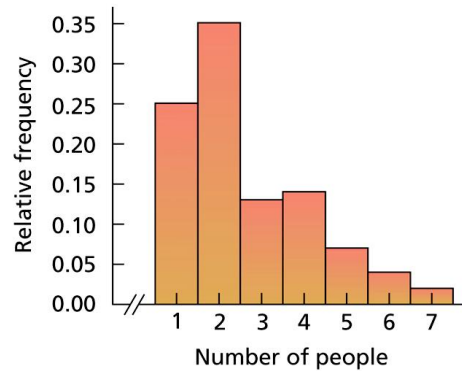
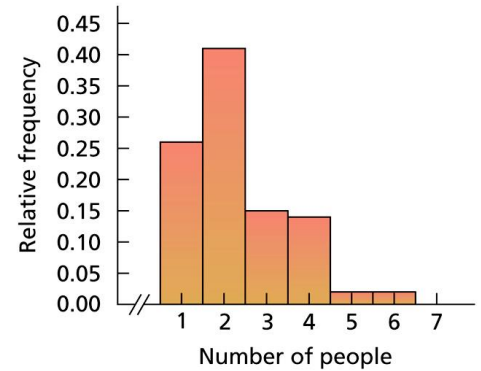
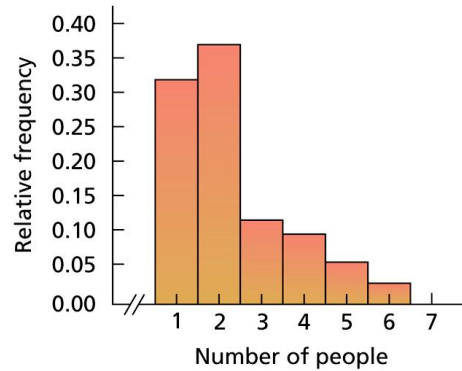
Population and Sample Distributions; Distribution of a Variable

The distribution of population data is called the **population distribution**, or the **distribution of the variable**.

The distribution of sample data is called a **sample distribution**.



(a)



(b)

Figure 2.13

Population distribution and six sample distributions for household size

Key Fact 2.1

Population and Sample Distributions

For a simple random sample, the sample distribution approximates the population distribution (i.e., the distribution of the variable under consideration). The larger the sample size, the better the approximation tends to be.