## EXERCISES 4.1 - 4.16

**4.1** ● The Insurance Institute for Highway Safety (www.iihs.org, June 11, 2009) published data on repair costs for cars involved in different types of accidents. In one study, seven different 2009 models of mini- and micro-cars were driven at 6 mph straight into a fixed barrier. The following table gives the cost of repairing damage to the bumper for each of the seven models.

| Model | Repair Cost |
|---|---|
| Smart Fortwo | $1,480 |
| Chevrolet Aveo | $1,071 |
| Mini Cooper | $2,291 |
| Toyota Yaris | $1,688 |
| Honda Fit | $1,124 |
| Hyundai Accent | $3,476 |
| Kia Rio | $3,701 |

Compute the values of the mean and median. Why are these values so different? Which of the two—mean or median—appears to be better as a description of a typical value for this data set?

**4.2** ● The article "Caffeinated Energy Drinks—A Growing Problem" (*Drug and Alcohol Dependence* [2009]: 1–10) gave the following data on caffeine concentration (mg/ounce) for eight top-selling energy drinks:

| Energy Drink | Caffeine Concentration (mg/oz) |
|---|---|
| Red Bull | 9.6 |
| Monster | 10.0 |
| Rockstar | 10.0 |
| Full Throttle | 9.0 |
| No Fear | 10.9 |
| Amp | 8.9 |
| SoBe Adrenaline Rush | 9.5 |
| Tab Energy | 9.1 |

a. What is the value of the mean caffeine concentration for this set of top-selling energy drinks?
b. Coca-Cola has 2.9 mg/ounce of caffeine and Pepsi Cola has 3.2 mg/ounce of caffeine. Write a sentence explaining how the caffeine concentration of top-selling energy drinks compares to that of these colas.

**4.3** ● Consumer Reports Health (www.consumerreports.org/health) reported the accompanying caffeine concentration (mg/cup) for 12 brands of coffee:

| Coffee Brand | Caffeine Concentration (mg/cup) |
|---|---|
| Eight O'Clock | 140 |
| Caribou | 195 |
| Kickapoo | 155 |
| Starbucks | 115 |
| Bucks Country Coffee Co. | 195 |
| Archer Farms | 180 |
| Gloria Jean's Coffees | 110 |
| Chock Full o'Nuts | 110 |
| Peet's Coffee | 130 |
| Maxwell House | 55 |
| Folgers | 60 |
| Millstone | 60 |

Use at least one measure of center to compare caffeine concentration for coffee with that of the energy drinks of the previous exercise. (Note: 1 cup = 8 ounces)

**4.4** ● Consumer Reports Health (www.consumerreports.org/health) reported the sodium content (mg) per 2 tablespoon serving for each of 11 different peanut butters:

120 50 140 120 150 150 150 65
170 250 110

a. Display these data using a dotplot. Comment on any unusual features of the plot.
b. Compute the mean and median sodium content for the peanut butters in this sample.
c. The values of the mean and the median for this data set are similar. What aspect of the distribution of sodium content—as pictured in the dotplot from Part (a)—provides an explanation for why the values of the mean and median are similar?

**4.5** In August 2009, Harris Interactive released the results of the "Great Schools" survey. In this survey, 1086 parents of children attending a public or private school were asked approximately how much time they spent volunteering at school per month over the last school year. For this sample, the mean number of hours per month was 5.6 hours and the median number of hours was 1.0. What does the large difference between the mean and median tell you about this data set?

The **population interquartile range** is the difference between the upper and lower population quartiles. If a histogram of the data set under consideration (whether a population or a sample) can be reasonably well approximated by a normal curve, then the relationship between the standard deviation (sd) and the interquartile range is roughly sd = iqr/1.35. A value of the standard deviation much larger than iqr/1.35 suggests a distribution with heavier (or longer) tails than a normal curve. For the degree data of Example 4.9, we had $s = 5.53$, whereas iqr/1.35 = 6/1.35 = 4.44. This suggests that the distribution of data values in Example 4.9 is indeed heavy-tailed compared to a normal curve. This can be seen in the stem-and-leaf display of Figure 4.7.

## EXERCISES 4.17 - 4.31

4.17 ● The following data are cost (in cents) per ounce for nine different brands of sliced Swiss cheese (www.consumerreports.org):

29 62 37 41 70 82 47 52 49

**a.** Compute the variance and standard deviation for this data set.

**b.** If a very expensive cheese with a cost per slice of 150 cents was added to the data set, how would the values of the mean and standard deviation change?

4.18 ● Cost per serving (in cents) for six high-fiber cereals rated very good and for nine high-fiber cereals rated good by *Consumer Reports* are shown below. Write a few sentences describing how these two data sets differ with respect to center and variability. Use summary statistics to support your statements.

**Cereals Rated Very Good**

46 49 62 41 19 77

**Cereals Rated Good**

71 30 53 53 67 43 48 28 54

4.19 ● Combining the cost-per-serving data for high-fiber cereals rated very good and those rated good from the previous exercise gives the following data set:

46 49 62 41 19 77 71 30
53 53 67 43 48 28 54

**a.** Compute the quartiles and the interquartile range for this combined data set.

**b.** Compute the interquartile range for just the cereals rated good. Is this value greater than, less than, or about equal to the interquartile range computed in Part (a)?

4.20 ● The paper "Caffeinated Energy Drinks—A Growing Problem" (*Drug and Alcohol Dependence* [2009]: 1–10) gave the accompanying data on caffeine per ounce for eight top-selling energy drinks and for 11 high-caffeine energy drinks:

**Top-Selling Energy Drinks**

9.6 10.0 10.0 9.0 10.9 8.9 9.5 9.1

**High-Caffeine Energy Drinks**

21.0 25.0 15.0 21.5 35.7 15.0
33.3 11.9 16.3 31.3 30.0

The mean caffeine per ounce is clearly higher for the high-caffeine energy drinks, but which of the two groups of energy drinks (top-selling or high-caffeine) is the most variable with respect to caffeine per ounce? Justify your choice.

4.21 ● The Insurance Institute for Highway Safety (www.iihs.org, June 11, 2009) published data on repair costs for cars involved in different types of accidents. In one study, seven different 2009 models of mini- and micro-cars were driven at 6 mph straight into a fixed barrier. The following table gives the cost of repairing damage to the bumper for each of the seven models:

| Model | Repair Cost |
|---|---|
| Smart Fortwo | $1,480 |
| Chevrolet Aveo | $1,071 |
| Mini Cooper | $2,291 |
| Toyota Yaris | $1,688 |
| Honda Fit | $1,124 |
| Hyundai Accent | $3,476 |
| Kia Rio | $3,701 |

**a.** Compute the values of the variance and standard deviation. The standard deviation is fairly large. What does this tell you about the repair costs?

With two or more data sets consisting of observations on the same variable (for example, fuel efficiencies for four types of car or weight gains for a control group and a treatment group), **comparative boxplots** (more than one boxplot drawn using the same scale) can tell us a lot about similarities and differences between the data sets.

## EXAMPLE 4.13 NBA Salaries Revisited

The 2009–2010 salaries of NBA players published on the web site hoopshype.com were used to construct the comparative boxplot of the salary data for five teams shown in Figure 4.11.

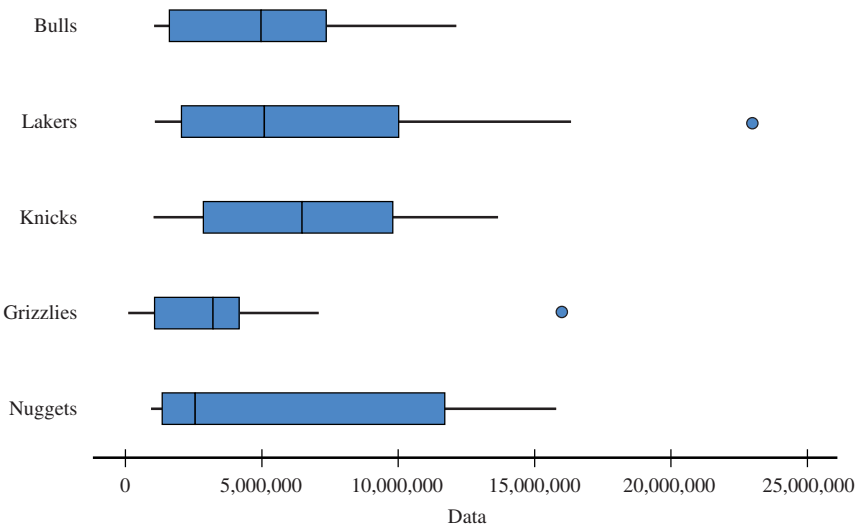


FIGURE 4.11
Comparative boxplot for salaries for five NBA teams.

The comparative boxplot reveals some interesting similarities and differences in the salary distributions of the five teams. The minimum salary is lower for the Grizzlies, but is about the same for the other four teams. The median salary was lowest for the Nuggets—in fact the median for the Nuggets is about the same as the lower quartile for the Knicks and the Lakers, indicating that half of the players on the Nuggets have salaries less than about $2.5 million, whereas only about 25% of the Knicks and the Lakers have salaries less than about $2.5 million. The Lakers had the player with by far the highest salary. The Grizzlies and the Lakers were the only teams that had any salary outliers. With the exception of one highly paid player, salaries for players on the Grizzlies team were noticeably lower than for the other four teams.

## EXERCISES 4.32 - 4.37

4.32 Based on a large national sample of working adults, the U.S. Census Bureau reports the following information on travel time to work for those who do not work at home:

lower quartile = 7 minutes
median = 18 minutes
upper quartile = 31 minutes

Bold exercises answered in back ● Data set available online ✦ Video Solution available

Also given was the mean travel time, which was reported as 22.4 minutes.

a. Is the travel time distribution more likely to be approximately symmetric, positively skewed, or negatively skewed? Explain your reasoning based on the given summary quantities.
b. Suppose that the minimum travel time was 1 minute and that the maximum travel time in the sample was 205 minutes. Construct a skeletal boxplot for the travel time data.
c. Were there any mild or extreme outliers in the data set? How can you tell?

4.33 ● The report "Who Moves? Who Stays Put? Where's Home?" (*Pew Social and Demographic Trends*, December 17, 2008) gave the accompanying data for the 50 U.S. states on the percentage of the population that was born in the state and is still living there. The data values have been arranged in order from largest to smallest.

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| 75.8 | 71.4 | 69.6 | 69.0 | 68.6 | 67.5 | 66.7 | 66.3 | 66.1 | 66.0 | 66.0 |
| 65.1 | 64.4 | 64.3 | 63.8 | 63.7 | 62.8 | 62.6 | 61.9 | 61.9 | 61.5 | 61.1 |
| 59.2 | 59.0 | 58.7 | 57.3 | 57.1 | 55.6 | 55.6 | 55.5 | 55.3 | 54.9 | 54.7 |
| 54.5 | 54.0 | 54.0 | 53.9 | 53.5 | 52.8 | 52.5 | 50.2 | 50.2 | 48.9 | 48.7 |
| 48.6 | 47.1 | 43.4 | 40.4 | 35.7 | 28.2 | | | | | |

**a.** Find the values of the median, the lower quartile, and the upper quartile.
**b.** The two smallest values in the data set are 28.2 (Alaska) and 35.7 (Wyoming). Are these two states outliers?
**c.** Construct a boxplot for this data set and comment on the interesting features of the plot.

4.34 ● The National Climate Data Center gave the accompanying annual rainfall (in inches) for Medford, Oregon, from 1950 to 2008 (www.ncdc.noaa.gov/oa/climate/research/cag3/city.html):

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| 28.84 | 20.15 | 18.88 | 25.72 | 16.42 | 20.18 | 28.96 | 20.72 | 23.58 |
| 10.62 | 20.85 | 19.86 | 23.34 | 19.08 | 29.23 | 18.32 | 21.27 | 18.93 |
| 15.47 | 20.68 | 23.43 | 19.55 | 20.82 | 19.04 | 18.77 | 19.63 | 12.39 |
| 22.39 | 15.95 | 20.46 | 16.05 | 22.08 | 19.44 | 30.38 | 18.79 | 10.89 |
| 17.25 | 14.95 | 13.86 | 15.30 | 13.71 | 14.68 | 15.16 | 16.77 | 12.33 |
| 21.93 | 31.57 | 18.13 | 28.87 | 16.69 | 18.81 | 15.15 | 18.16 | 19.99 |
| 19.00 | 23.97 | 21.99 | 17.25 | 14.07 | | | | |

**a.** Compute the quartiles and the interquartile range.
**b.** Are there outliers in this data set? If so, which observations are mild outliers? Which are extreme outliers?
**c.** Draw a boxplot for this data set that shows outliers.

4.35 ● The accompanying data on annual maximum wind speed (in meters per second) in Hong Kong for each year in a 45-year period were given in an article that appeared in the journal *Renewable Energy* (March 2007). Use the annual maximum wind speed data to construct a boxplot. Is the boxplot approximately symmetric?

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| 30.3 | 39.0 | 33.9 | 38.6 | 44.6 | 31.4 | 26.7 | 51.9 | 31.9 |
| 27.2 | 52.9 | 45.8 | 63.3 | 36.0 | 64.0 | 31.4 | 42.2 | 41.1 |
| 37.0 | 34.4 | 35.5 | 62.2 | 30.3 | 40.0 | 36.0 | 39.4 | 34.4 |
| 28.3 | 39.1 | 55.0 | 35.0 | 28.8 | 25.7 | 62.7 | 32.4 | 31.9 |
| 37.5 | 31.5 | 32.0 | 35.5 | 37.5 | 41.0 | 37.5 | 48.6 | 28.1 |

4.36 ● Fiber content (in grams per serving) and sugar content (in grams per serving) for 18 high fiber cereals (www.consumerreports.com) are shown below.

**Fiber Content**

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| 7 | 10 | 10 | 7 | 8 | 7 | 12 | 12 | 8 |
| 13 | 10 | 8 | 12 | 7 | 14 | 7 | 8 | 8 |

**Sugar Content**

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| 11 | 6 | 14 | 13 | 0 | 18 | 9 | 10 | 19 |
| 6 | 10 | 17 | 10 | 10 | 0 | 9 | 5 | 11 |

a. Find the median, quartiles, and interquartile range for the fiber content data set.
b. Find the median, quartiles, and interquartile range for the sugar content data set.
c. Are there any outliers in the sugar content data set?
d. Explain why the minimum value for the fiber content data set and the lower quartile for the fiber content data set are equal.
e. Construct a comparative boxplot and use it to comment on the differences and similarities in the fiber and sugar distributions.

4.37 ● Shown here are the number of auto accidents per year for every 1000 people in each of 40 occupations (*Knight Ridder Tribune*, June 19, 2004):

| Occupation | Accidents per 1000 | Occupation | Accidents per 1000 |
|---|---|---|---|
| Student | 152 | Social worker | 98 |
| Physician | 109 | Manual laborer | 96 |
| Lawyer | 106 | Analyst | 95 |
| Architect | 105 | Engineer | 94 |
| Real estate broker | 102 | Consultant | 94 |
| Enlisted military | 99 | Sales | 93 |

*(continued)*