

## Chapter 3: Describing, Exploring, and Comparing Data

### Section 3-1: Measures of Center

1. The term *average* is not used in statistics. The term *mean* should be used for the result obtained by adding all of the sample values and dividing the total by the number of sample values.
2. No. The 50 amounts are all weighed equally in the calculation that yields the mean of 19.67, but some states have higher smoking rates, so the mean smoking rate should be calculated using a weighted mean that takes into account the smoking rates in the different states.
3. They use different approaches for providing a value (or values) of the center or middle of the sorted list of data.
4. Using the five values,  $\bar{x} = \frac{80+94+58+66+56}{4} = 70.8$  BPM and the median is 66.0 BPM.

Using the six values, including the outlier of 740,  $\bar{x} = \frac{80+94+58+66+56+740}{4} = 182.3$  BPM and the median is  $\frac{66+80}{2} = 73.0$  BPM.

The outlier caused the mean to change by a substantial amount, but the median changed very little. The median is resistant to the effect of the outlier, but the mean is not resistant.

5. The mean is  $\bar{x} = \frac{155,857 + \dots + 247,477 + 255,788 + \dots + 471,062}{10} = \$266,594.0$ .

The median is  $\frac{247,477 + 255,788}{2} = \$251,632.5$ .

There is no mode.

The midrange is  $\frac{155,857 + 471,062}{2} = \$313,459.5$ .

Apart from the fact that the other charges are lower than those given, nothing meaningful can be known about the population of all charges.

6. The mean is  $\bar{x} = \frac{26.8+27.0+27.1+27.3+27.4+27.7+27.9+28.1+28.2+28.3+28.3+28.4}{12} = 27.71$ .

The median is  $\frac{27.7+27.9}{2} = 27.80$ .

There is no mode.

The midrange is  $\frac{26.8+28.4}{2} = 27.60$

The measures of center do not provide any information about the change of MCAT scores over time.

7. The mean is  $\bar{x} = \frac{7+19+20+25+55+60+81+82+89+91+99}{11} = 57.1$ .

The median is 60.0.

There is no mode.

The midrange is  $\frac{7+99}{11} = 53.0$ .

The jersey numbers are nominal data that are just replacements for names, and they do not measure or count anything, so the resulting statistics are meaningless.

8. The mean is  $\bar{x} = \frac{189+190+190+195+202+225+235+252+254+305+305}{11} = 231.1$  lb.

The median is 225.0 lb.

The modes are 190 lb and 305 lb.

8. (continued)

The midrange is  $\frac{189+305}{2} = 247.0$  lb.

All of the weights are from players on only one of the 32 NFL teams, but it isn't likely that any of the teams has weights that are substantially different from those of any other team, so it is reasonable to conclude that the weights are somewhat representative of all NFL players.

9. The mean is  $\bar{x} = \frac{1+1+1+1+1+1+1+1+1+1+\cdots+2+2+2+3+3+3+3+3+3+4}{25} = 1.9$ .

The median is 2.0.

The mode is 1.

The midrange is  $\frac{1+4}{2} = 2.5$ .

The mode of 1 correctly indicates that the smooth-yellow peas occur more than any other phenotype, but the other measures of center do not make sense with these data at the nominal level of measurement.

10. The mean is  $\bar{x} = \frac{950+1150+1200+1400+\cdots+1600+1600+1750+1800}{12} = 1454.2$ , or \$1454.2.

The median is  $\frac{1500+1500}{2} = 1500.0$ , or \$1500.0.

The mode is \$1500.

The midrange is  $\frac{950+1800}{2} = 1375.0$ , or \$1375.0.

The sample consists of "best buy" TVs, so it is not a random sample and is not likely to be representative of the population. The lowest price is a relevant statistic for someone planning to buy one of the TVs.

11. The mean is  $\bar{x} = \frac{0.51+0.74+0.89+1.04+1.18+1.38+1.41+1.42+1.45+1.45+1.49}{11} = 1.178$  W/kg.

The median is 1.380 W/kg.

The mode is 1.45 W/kg.

The midrange is  $\frac{0.51+1.49}{2} = 1.000$  W/kg.

If concerned about radiation absorption, you might purchase the cell phone with the *lowest* absorption rate. All of the cell phones in the sample have absorption levels below the FCC maximum of 1.6 W/kg.

12. The mean is  $\bar{x} = \frac{0+0+0+\cdots+34+36+38+41+41+41+\cdots+53+54+55}{20} = 32.6$  mg.

The median is  $\frac{38+41}{2} = 39.5$  mg.

The mode is 0 mg.

The midrange is  $\frac{0+55}{2} = 27.5$  mg.

Americans consume some brands much more often than others, but the 20 brands are all weighted equally in the calculations, so the statistics are not necessarily representative of the population of all cans of the same 20 brands consumed by Americans.

13. The mean is  $\bar{x} = \frac{8+9+11+12+15+15+18+20+20+23+24+25+30+34}{14} = 18.9$  firefighters.

The median is  $\frac{18+20}{2} = 19.0$  firefighters.

The modes are 15 and 20 firefighters.

The midrange is  $\frac{8+34}{2} = 21.0$  firefighters.

13. (continued)

The data are time series data, but the measures of center do not reveal anything about a trend consisting of a pattern of change over time.

$$14. \text{ The mean is } \bar{x} = \frac{8.6+9.1+9.1+9.1+9.3+9.3+9.4+9.8+9.9+10+10.4}{11} = 9.45 \text{ in.}$$

The median is 9.30 in.

The mode is 9.1 in.

$$\text{The midrange is } \frac{8.6+10.4}{2} = 9.50 \text{ in.}$$

Because the measurements were made in 1988, they are not necessarily representative of the current population of all Army women.

$$15. \text{ The mean is } \bar{x} = \frac{53,323.00+\cdots+54,528.00+54,653.00+\cdots+57,261.00}{10} = \$54,862.0.$$

$$\text{The median is } \frac{54,528.00+54,653.00}{2} = \$54,590.5.$$

There is no mode.

$$\text{The midrange is } \frac{53,323.00+57,261.00}{2} = \$55,292.0.$$

Apart from the fact that all other colleges have tuition and fee amounts less than those listed, nothing meaningful can be known about the population.

$$16. \text{ The mean is } \bar{x} = \frac{0+0+0+0+0+0+\cdots+9+10+10+20+40+50}{50} = 2.8 \text{ cigarettes.}$$

The median is 0.0 cigarettes.

The mode is 0 cigarettes.

$$\text{The midrange is } \frac{0+50}{2} = 25.0 \text{ cigarettes.}$$

Because the selected subjects *report* the number of cigarettes smoked, it is very possible that the data are not at all accurate. And what about that person who smokes 50 cigarettes (or 2.5 packs) a day? What are they thinking?

17. Systolic:

$$\text{The mean is } \bar{x} = \frac{96+116+118+120+122+126+128+136+156+158}{10} = 127.6 \text{ mm Hg.}$$

$$\text{The median is } \frac{122+126}{2} = 124.0 \text{ mm Hg.}$$

Diastolic:

$$\text{The mean is } \bar{x} = \frac{52+58+64+72+74+76+80+82+88+90}{10} = 73.6 \text{ mm Hg.}$$

$$\text{The median is } \frac{74+76}{2} = 75.0 \text{ mm Hg.}$$

Given that systolic and diastolic blood pressures measure different characteristics, a comparison of the measures of center doesn't make sense. Because the data are matched, it would make more sense to investigate whether there is an association or *correlation* between systolic blood pressure measurements and diastolic blood pressure measurements.

18. White Blood Cell Count:

$$\text{The mean is } \bar{x} = \frac{4.1+4.9+5.6+5.7+6.1+6.9+6.9+7.5+8.0+8.1+8.3+8.7}{12} = 6.73 \text{ (1000 cells/}\mu\text{L).}$$

$$\text{The median is } \frac{6.9+6.9}{2} = 6.90 \text{ (1000 cells/}\mu\text{L).}$$

18. (continued)

Red Blood Cell Count:

$$\text{The mean is } \bar{x} = \frac{4.0 + 4.1 + 4.2 + 4.3 + 4.4 + 4.5 + 4.6 + 4.7 + 4.7 + 4.8 + 5.0 + 5.5}{12} = 4.57 \text{ (1000 cells/}\mu\text{L)}.$$

$$\text{The median is } \frac{4.5 + 4.6}{2} = 4.55 \text{ (1000 cells/}\mu\text{L)}.$$

Because the data are matched, it would make more sense to investigate whether there is an association or *correlation* between white and red blood cell counts.

19. Males:

$$\text{The mean is } \bar{x} = \frac{4.1 + 4.9 + 5.1 + 5.4 + 5.6 + 5.7 + 5.7 + 6.1 + 6.1 + 7.5 + 8.3 + 9.5}{12} = 6.17 \text{ (1000 cells/}\mu\text{L)}.$$

$$\text{The median is } \frac{5.7 + 5.7}{2} = 5.70 \text{ (1000 cells/}\mu\text{L)}.$$

Females:

$$\text{The mean is } \bar{x} = \frac{4.8 + 5.9 + 6.3 + 6.4 + 6.9 + 6.9 + 7.2 + 8.0 + 8.1 + 8.1 + 8.7 + 10.9}{12} = 7.35 \text{ (1000 cells/}\mu\text{L)}.$$

$$\text{The median is } \frac{6.9 + 7.2}{2} = 7.05 \text{ (1000 cells/}\mu\text{L)}.$$

Females appear to have higher white blood cell counts.

20. Single Line:

$$\text{The mean is } \bar{x} = \frac{390 + 396 + 402 + 408 + 426 + 438 + 444 + 462 + 462 + 462}{10} = 429.0 \text{ seconds.}$$

$$\text{The median is } \frac{426 + 438}{2} = 432.0 \text{ seconds.}$$

Individual Lines:

$$\text{The mean is } \bar{x} = \frac{252 + 324 + 348 + 372 + 402 + 462 + 462 + 510 + 558 + 600}{10} = 429.0 \text{ seconds.}$$

$$\text{The median is } \frac{402 + 462}{2} = 432.0 \text{ seconds.}$$

Individual lines: same results as with a single line. Although the mean and median are the same, the times with individual lines are much more *varied* than those with a single line.

21. Using all values, the mean is  $\bar{x} = 53.7$  mg/dL and the median is 52.0 mg/dL. The highest value of 138 mg/dL appears to be an outlier. Excluding 138 mg/dL, the mean is  $\bar{x} = 53.4$  mg/dL and the median is 52.0 mg/dL. Excluding the outlier does not cause much of a change in the mean, and the median remains the same.

22. Using all values, the mean is  $\bar{x} = 113.7$  mg/dL and the median is 113.0 mg/dL. The highest value of 251 mg/dL appears to be an outlier. Excluding 251 mg/dL, the mean is  $\bar{x} = 113.3$  mg/dL and the median is 113.0 mg/dL. Excluding the outlier does not cause much of a change in the mean, and the median remains the same.

23. The mean is  $\bar{x} = 98.20^\circ\text{F}$  and the median is  $98.40^\circ\text{F}$ . These results suggest that the mean is less than  $98.6^\circ\text{F}$ .

24. The mean is  $\bar{x} = 3152.0$  g and the median is 3300 g. All of the weights end in 00, so they are all rounded to the nearest 100 grams. This suggests that the results should be rounded as follows:  $\bar{x} = 3150.0$  g and the median is 3300 g.

25. The mean is  $\bar{x} = \frac{1(49.5) + 51(149.5) + 90(249.5) + 10(349.5) + 0(449.5) + 0(549.5) + 1(649.5)}{1 + 51 + 90 + 10 + 0 + 0 + 1}$   
 $= 224.0$  (1000 cells /  $\mu\text{L}$ ). The mean from the frequency distribution is quite close to the mean of  
 $= 224.3$  (1000 cells /  $\mu\text{L}$ ), obtained by using the original list of values.
26. The mean is  $\bar{x} = \frac{25(149.5) + 92(249.5) + 28(349.5) + 0(449.5) + 2(549.5)}{25 + 92 + 28 + 0 + 2} = 255.6$  (1000 cells /  $\mu\text{L}$ ). The mean  
from the frequency distribution is quite close to the mean of  $= 225.1$  (1000 cells /  $\mu\text{L}$ ), obtained by using the  
original list of values.
27. The mean is  $\bar{x} = \frac{3(4) + 3(2) + 3(3) + 4(4) + 1(1)}{3 + 3 + 3 + 4 + 1} = 3.14$ , so the student made the dean's list.
28. The mean is  $\bar{x} = 0.60 \left( \frac{63 + 91 + 88 + 84 + 79}{5} \right) + 0.10(86) + 0.15(90) + 0.15(70) = 81.2$ , so the student earned a B.
29. a. The missing value is  $5(78.0) - 82 - 78 - 56 - 84 = 90$  beats per minute.  
b.  $n - 1$
30. The mean ignoring the presidents who are still alive is 15.0 years. The mean including the presidents who are  
still alive is at least 15.7 years. The results do not differ by a considerable amount.
31. Mean: 113.7 mg/dL; 10% trimmed mean: 112.6 mg/dL; 20% trimmed mean: 112.2 mg/dL. The 10% trimmed  
mean and 20% trimmed mean are both fairly close, but the untrimmed mean of 113.7 mg/dL differs from them  
because it is more strongly affected by the outliers.

### Section 3-2: Measures of Variation

1.  $s \approx \frac{1439 - 963}{4} = 119.0 \text{ cm}^3$ , which is quite close to the exact value of the standard deviation of  $124.9 \text{ cm}^3$ .
2. Significantly low values are less than or equal to  $1126.0 - 2(124.9) = 876.2 \text{ cm}^3$  and significantly high values  
are greater than or equal to  $1126.0 + 2(124.9) = 1375.8 \text{ cm}^3$ . A brain volume of  $1440 \text{ cm}^3$  is significantly high.
3.  $(20.0414 \text{ kg})^2 = 401.6577 \text{ kg}^2$
4.  $s, \sigma, s^2, \sigma^2$
5. The range is  $471,062 - 155,857 = \$315,205.0$ .  
The variance is  $s^2 = \frac{(155,857 - 266,594)^2 + \cdots + (471,062 - 266,594)^2}{10 - 1} = 8,502,938,525.9 \text{ dollars}^2$ .  
The standard deviation is  $s = \sqrt{8,502,938,525.9} = \$92,211.4$ .  
(Many technologies will not provide all of the digits shown for the variance  $s^2$ , so any result close to the value  
shown here is acceptable.) Because only the 10 highest sample values are used, nothing much can be known about  
the population of all such charges.
6. The range is  $28.4 - 26.8 = 1.60$ .  
The variance is  $s^2 = \frac{(26.8 - 27.2)^2 + \cdots + (28.4 - 27.2)^2}{12 - 1} = 0.324$ .  
The standard deviation is  $s = \sqrt{0.324} = 0.57$ .  
The results do not provide any information about the change of MCAT scores over time.

7. The range is
- $99 - 7 = 92.0$
- .

$$\text{The variance is } s^2 = \frac{(7 - 57.1)^2 + \cdots + (99 - 57.1)^2}{11 - 1} = 1149.5.$$

$$\text{The standard deviation is } s = \sqrt{1149.5} = 33.9.$$

The jersey numbers are nominal data that are just replacements for names, and they do not measure or count anything, so the resulting statistics are meaningless.

8. The range is
- $305 - 189 = 116.0$
- lb.

$$\text{The variance is } s^2 = \frac{(189 - 231.1)^2 + \cdots + (305 - 231.1)^2}{11 - 1} = 1923.7 \text{ lb}^2.$$

$$\text{The standard deviation is } s = \sqrt{1923.7} = 43.9 \text{ lb.}$$

All of the weights are from players on only 1 of the 32 teams, but it isn't very likely that any of the teams has weights that are substantially different from those of any other team, so it is reasonable to conclude that the measures of variation are typical of NFL players.

9. The range is
- $4 - 1 = 3.0$
- .

$$\text{The variance is } s^2 = \frac{25(109) - (47)^2}{25(25 - 1)} = 0.9.$$

$$\text{The standard deviation is } s = \sqrt{0.9} = 0.9.$$

The measures of variation can be found, but they make no sense because the data don't measure or count anything. They are nominal data.

10. The range is
- $1800 - 950 = \$850.0$
- .

$$\text{The variance is } s^2 = \frac{12(26,047,500) - (17,450)^2}{12(12 - 1)} = 61,117.4 \text{ (dollars)}^2.$$

$$\text{The standard deviation is } s = \sqrt{61,117.4} = \$247.2.$$

The sample consists of "best buy" TVs, so it is not a random sample and is not likely to be representative of the population. The measures of variation are not likely to be typical of all TVs that are 60 inches or larger.

11. The range is
- $1.49 - 0.51 = 0.980$
- W/kg.

$$\text{The variance is } s^2 = \frac{11(16.4078) - (12.96)^2}{11(12 - 1)} = 0.114 \text{ (W/kg)}^2.$$

$$\text{The standard deviation is } s = \sqrt{0.114} = 0.337 \text{ W/kg.}$$

No. Some models of cell phones have a much larger market share than others, so the measures from the different models should be weighted according to their size in the population.

12. The range is
- $55 - 0 = 55.0$
- mg.

$$\text{The variance is } s^2 = \frac{20(29,045) - (651)^2}{20(20 - 1)} = 413.4 \text{ mg}^2.$$

$$\text{The standard deviation is } s = \sqrt{413.4} = 20.3 \text{ mg.}$$

Americans consume some brands much more often than others, but the 20 brands are all weighted equally in the calculations, so the statistics are not necessarily representative of the population of all cans of the same 20 brands consumed by Americans.

13. The range is  $34 - 8 = 26.0$  firefighters.

$$\text{The variance is } s^2 = \frac{14(5770) - (264)^2}{14(14-1)} = 60.9 \text{ firefighters}^2.$$

The standard deviation is  $s = \sqrt{60.9} = 7.8$  firefighters.

The data are time-series data, but the measures of variation do not reveal anything about a trend consisting of a pattern of change over time.

14. The range is  $10.4 - 8.6 = 1.80$  in.

$$\text{The variance is } s^2 = \frac{11(985.94) - (107)^2}{11(11-1)} = 0.27 \text{ in}^2$$

The standard deviation is  $s = \sqrt{0.27} = 0.52$  in.

Because the measurements were made in 1988, they are not necessarily representative of the current population of all Army women.

15. The range is  $57,261 - 53,323 = \$3938.00$ .

$$\text{The variance is } s^2 = \frac{10(30,113,141,176) - (548,620)^2}{10(10-1)} = 1,638,970.9 \text{ dollars}^2.$$

The standard deviation is  $s = \sqrt{1,638,970.9} = \$1280.2$ .

Because the data include only the 10 highest costs, the measures of variation don't tell us anything about the variation among costs for the population of all U.S. college tuitions.

16. The range is  $50 - 0 = 50.0$  cigarettes.

$$\text{The variance is } s^2 = \frac{50(4781) - (139)^2}{50(50-1)} = 89.7 \text{ (cigarettes)}^2.$$

The standard deviation is  $s = \sqrt{89.7} = 9.5$  cigarettes.

Because the selected subjects report the number of cigarettes smoked, it is very possible that the data are not at all accurate, so the results might not reflect the actual smoking behavior of California adults.

17. Systolic:  $\bar{x} = 127.6$ ,  $s = 18.6$ ; The coefficient of variation is  $\frac{18.6}{127.6} \cdot 100\% = 14.6\%$ .

$$\text{Diastolic: } \bar{x} = 73.6, s = 12.5; \text{ The coefficient of variation is } \frac{12.5}{73.6} \cdot 100\% = 16.9\%.$$

The variation is roughly about the same.

18. All units are (1000 cells/ $\mu$ L).

$$\text{White Blood Cell Counts: } \bar{x} = 6.73, s = 1.46; \text{ The coefficient of variation is } \frac{1.46}{6.73} \cdot 100\% = 21.7\%.$$

$$\text{Red Blood Cell Counts: } \bar{x} = 4.57, s = 0.42; \text{ The coefficient of variation is } \frac{0.42}{4.57} \cdot 100\% = 9.2\%.$$

White blood cell counts appear to vary more than red blood cell counts.

19. All units are (1000 cells/ $\mu$ L).

$$\text{Male: } \bar{x} = 6.17, s = 1.53; \text{ The coefficient of variation is } \frac{1.53}{6.17} \cdot 100\% = 24.8\%.$$

$$\text{Female: } \bar{x} = 7.35, s = 1.57; \text{ The coefficient of variation is } \frac{1.57}{7.35} \cdot 100\% = 21.3\%.$$

The variation is roughly about the same.

20. Single Line:  $\bar{x} = 429.0$ ,  $s = 28.6$ ; The coefficient of variation is  $\frac{28.6}{429.0} \cdot 100\% = 6.7\%$ .

Individual Lines:  $\bar{x} = 429.0$ ,  $s = 109.3$ ; The coefficient of variation is  $\frac{109.3}{429.0} \cdot 100\% = 25.5\%$ .

The single line has much less variation than with individual lines.

21. Using all data: Range = 112.0 mg/dL,  $s^2 = 238.3 \text{ (mg/dL)}^2$ ,  $s = 15.4 \text{ mg/dL}$

Excluding possible outlier of 138: Range = 87.0 mg/dL,  $s^2 = 215.2 \text{ (mg/dL)}^2$ ,  $s = 14.7 \text{ mg/dL}$

The measures of variation change, but not by substantial amounts.

22. Using all data: Range = 212.0 mg/dL,  $s^2 = 1238.3 \text{ (mg/dL)}^2$ ,  $s = 35.2 \text{ mg/dL}$

Excluding possible outlier of 251: Range = 184.0 mg/dL,  $s^2 = 1179.0 \text{ (mg/dL)}^2$ ,  $s = 34.3 \text{ mg/dL}$

The measures of variation change, but not by substantial amounts.

23. Range =  $3.10^\circ\text{F}$ ,  $s^2 = 0.39 \text{ (}^\circ\text{F)}^2$ ,  $s = 0.62^\circ\text{F}$

24. Range = 4600.0 g,  $s^2 = 480,848.1 \text{ g}^2$ ,  $s = 693.4 \text{ g}$ ; All of the weights end in 00, so they are all rounded to the nearest 100 grams. This suggests that the results should be rounded as follows:

Range = 4600.0 g,  $s^2 = 480,850 \text{ g}^2$ ,  $s = 690 \text{ g}$ .

25. The rule of thumb standard deviation is  $s \approx \frac{138 - 26}{4} = 28.0 \text{ mg/dL}$ , which is much larger than  $s = 15.4 \text{ mg/dL}$  found by using all of the data.

26. The rule of thumb standard deviation is  $s \approx \frac{251 - 39}{4} = 53.0 \text{ mg/dL}$ , which is much larger than  $s = 35.2 \text{ mg/dL}$  found by using all of the data.

27. The rule of thumb standard deviation is  $s \approx \frac{99.6 - 96.5}{4} = 0.78^\circ\text{F}$ , which is not substantially different from  $s = 0.62^\circ\text{F}$  found by using all of the data.

28. The rule of thumb standard deviation is  $s \approx \frac{4900 - 300}{4} = 1150.0 \text{ g}$ , which differs from  $s = 693.4 \text{ g}$  found by using all of the data by a considerable amount. Several of the lowest weights correspond to premature births, and they cause the range to be larger, with the resulting estimate being larger.

29. Significantly low values are less than or equal to  $74.0 - 2(12.5) = 49.0$  beats per minute, and significantly high values are greater than or equal to  $74.0 + 2(12.5) = 99.0$  beats per minute. A pulse rate of 44 beats per minute is significantly low.

30. Significantly low values are less than or equal to  $69.6 - 2(11.3) = 47.0$  beats per minute, and significantly high values are greater than or equal to  $69.6 + 2(11.3) = 92.2$  beats per minute. A pulse rate of 50 beats per minute is neither significantly low or high.

31. Significantly low values are less than or equal to  $77.32 - 2(1.29) = 24.74 \text{ cm}$ , and significantly high values are greater than or equal to  $77.32 + 2(1.29) = 29.90 \text{ cm}$ . A foot length of 30 cm is significantly high.

32. Significantly low values are less than or equal to  $98.20 - 2(0.62) = 96.96^\circ\text{F}$ , and significantly high values are greater than or equal to  $98.20 + 2(0.62) = 99.44^\circ\text{F}$ . A body temperature of  $100^\circ\text{F}$  is significantly high.



$$33. s = \sqrt{\frac{103(1 \cdot 49.5^2 + \cdots + 1 \cdot 649.5^2) - (1 \cdot 49.5 + \cdots + 1 \cdot 649.5)^2}{103(103-1)}} = 68.4, \text{ which is somewhat far from the exact value of } 59.5.$$

$$34. s = \sqrt{\frac{147(25 \cdot 149.5^2 + \cdots + 2 \cdot 549.5^2) - (25 \cdot 149.5 + \cdots + 2 \cdot 549.5)^2}{147(147-1)}} = 69.5, \text{ which is not substantially far from the exact value of } 65.5.$$

$$35. a. \mu = \frac{9+10+20}{3} = 13 \text{ cigarettes and } \sigma^2 = \frac{(9-13)^2 + (10-13)^2 + (20-13)^2}{3} = 24.7 \text{ cigarettes}^2$$

b. The nine possible samples of two values are the following:  $\{(9, 9), (9, 10), (9, 20), (10, 9), (10, 10), (10, 20), (20, 9), (20, 10), (20, 20)\}$  which have the following corresponding sample variances:  $\{0, 0.5, 60.5, 0.5, 0, 50, 60.5, 50, 0\}$ , that have a mean of  $s^2 = 2.47 \text{ cigarettes}^2$ .

c. The population variances of the nine samples above are  $\{0, 0.25, 30.25, 0.25, 0, 25, 30.25, 25, 0\}$  that have a mean of  $s^2 = 12.3 \text{ cigarettes}^2$ .

d. Part (b), because repeated samples result in variances that target the same value ( $24.7 \text{ cigarettes}^2$ ) as the population variance. Use division by  $n-1$ .

e. No. The mean of the sample variances ( $24.7 \text{ cigarettes}^2$ ) equals the population variance ( $24.7 \text{ cigarettes}^2$ ), but the mean of the sample standard deviations ( $3.5 \text{ cigarettes}$ ) does not equal the population standard deviation ( $5.0 \text{ cigarettes}$ ).

36. The mean absolute deviation of the population is 4.7 cigarettes. With repeated samplings of size 2, the nine different possible samples have mean absolute deviations of 0, 0, 0, 0.5, 0.5, 5, 5, 5.5, 5.5. With many such samples, the mean of those nine results is 2.4 cigarettes, showing that the sample mean absolute deviations tend to center about the value of 2.4 cigarettes instead of the mean absolute deviation of the population, which is 4.7 cigarettes. The sample mean deviations do not target the mean deviation of the population. This is not good. This indicates that a sample mean absolute deviation is not a good estimator of the mean absolute deviation of a population.

### Section 3-3: Measures of Relative Standing and Boxplots

1. James' height is 4.07 standard deviations above the mean.
2. The minimum height is 155 cm, the first quartile,  $Q_1$ , is 169.1 cm, the second quartile,  $Q_2$ , (or the median) is 173.8 cm, the third quartile,  $Q_3$ , is 179.4 cm, and the maximum height is 193.3 cm.
3. The bottom boxplot represents weights of women, because it depicts weights that are generally lower.
4. 2.00 should be preferred, because it is 2.00 standard deviations above the mean and would correspond to the highest of the five different possible scores.
5. a. The difference is  $104 - 74 = 30$  bpm.  
 b.  $\frac{30}{12.5} = 2.74$  standard deviations  
 c.  $z = 2.74$   
 d. The pulse rate of 104 beats per minute is significantly high.

6. a. The difference is  $36 - 74 = -38$  BPM.  
 b.  $\frac{38}{12.5} = 3.04$  standard deviations  
 c.  $z = -3.04$   
 d. The pulse rate of 36 beats per minute is significantly low.
7. a. The difference is  $96.50 - 98.20 = -1.7^\circ\text{F}$ .  
 b.  $\frac{1.7}{0.62} = 2.74$  standard deviations  
 c.  $z = -2.74$   
 d. The temperature of  $96.5^\circ\text{F}$  is significantly low.
8. a. The difference is  $98.60 - 98.20 = 0.4^\circ\text{F}$ .  
 b.  $\frac{0.4}{0.62} = 0.65$  standard deviations  
 c.  $z = 0.65$   
 d.  $Q_3$  is not significantly high.
9. Significantly low scores are less than or equal to  $21.1 - 2(5.1) = 10.9$ , and significantly high scores are greater than or equal to  $21.1 + 2(5.1) = 31.3$ .
10. Significantly low scores are less than or equal to  $25.2 - 2(6.4) = 12.4$ , and significantly high scores are greater than or equal to  $25.2 + 2(6.4) = 38.0$ .
11. Significantly low weights are less than or equal to  $3152.0 - 2(693.4) = 1765.2$  g, and significantly high weights are greater than or equal to  $3152.0 + 2(693.4) = 4538.8$  g.
12. Significantly low hip breadths are less than or equal to  $36.6 - 2(2.5) = 31.6$  cm, and significantly high hip breadths are greater than or equal to  $36.6 + 2(2.5) = 41.6$  cm.
13. The tallest man's  $z$  score is  $z = \frac{251 - 174.12}{7.10} = 10.83$  and the shortest man's  $z$  score is  $z = \frac{54.6 - 174.12}{7.10} = -16.83$ . Chandra Bahadur Dangi has the more extreme height because his  $z$  score of  $-16.83$  is farther from the mean than the  $z$  score of  $10.83$  for Sultan Kosen.
14. The female has a higher red blood cell count because her  $z$  score is  $z = \frac{5.23 - 4.349}{0.402} = 2.19$ , which is a higher number than the  $z$  score of  $z = \frac{5.58 - 4.719}{0.490} = 1.76$  for the male.
15. The male has a more extreme birth weight because his  $z$  score is  $z = \frac{1500 - 3272.8}{660.2} = -2.19$ , which is a lower number than the  $z$  score of  $z = \frac{1500 - 3037.1}{706.3} = -2.18$  for the female.
16. Julianne Moore had the more extreme age since her  $z$  score is  $z = \frac{54 - 36.2}{11.5} = 1.55$ , which is farther from the mean than the  $z$  score of  $z = \frac{33 - 44.1}{8.9} = -1.25$  for Eddie Redmayne.

17. For 61.0 in.,  $\frac{27}{54} \cdot 100 = 50$ , so it is the 50th percentile.

18. For 47.0 in.,  $\frac{9}{54} \cdot 100 = 17$ , so it is the 17th percentile.

19. For 70.0 in.,  $\frac{44}{54} \cdot 100 = 81$ , so it is the 81st percentile.

20. For 58.0 in.,  $\frac{21}{54} \cdot 100 = 39$ , so it is the 39th percentile.

21.  $L = \frac{60 \cdot 54}{100} = 32.4$ , so  $P_{60} = 63.0$  in.

22.  $L = \frac{25 \cdot 54}{100} = 13.5$ , so  $Q_1 = P_{25} = 50.0$  in. (Tech: Excel: 50.5 in., Minitab: 49.75)

23.  $L = \frac{75 \cdot 54}{100} = 40.5$ , so  $Q_3 = P_{75} = 66.5$  in. (Tech: Excel: 66.625 in., Minitab: 66.63)

24.  $L = \frac{40 \cdot 54}{100} = 21.6$ , so  $P_{40} = 58$  in. (Tech: Excel: 58.2 in.)

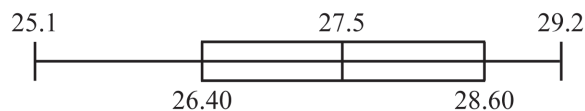
25.  $L = \frac{50 \cdot 54}{100} = 27$ , so  $P_{50} = \frac{60.5 + 61.0}{2} = 60.75$  in.

26.  $L = \frac{75 \cdot 54}{100} = 40.5$ , so  $P_{75} = 66.5$  in. (Tech: Excel: 66.125 in., Minitab: 66.625)

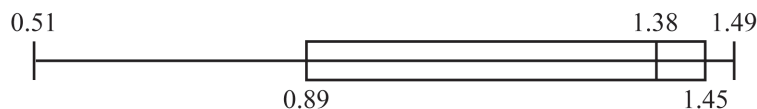
27.  $L = \frac{25 \cdot 54}{100} = 13.5$ , so  $P_{25} = 50.0$  in. (Tech: Excel: 50.5 in., Minitab: 49.75)

28.  $L = \frac{85 \cdot 54}{100} = 45.9$ , so  $P_{85} = 70.5$  in. (Tech: Excel: 70.575 in., Minitab: 71.625)

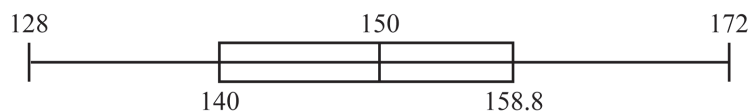
29. The five number summary is 25.1 cm, 26.40 cm, 27.50 cm, 28.60 cm, 29.2 cm.



30. The five number summary is 0.51 W/kg, 0.890 W/kg, 1.380 W/kg, 1.450 W/kg, 1.49 W/kg.  
(Excel:  $Q_1 = 0.965$  W/kg,  $Q_3 = 1.435$  W/kg.)

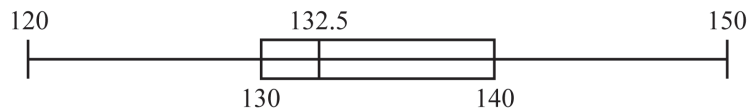


31. The five number summary is 128 mBq, 140.0 mBq, 150.0 mBq, 158.5 mBq, 172 mBq  
(Excel:  $Q_1 = 141.0$  mBq and  $Q_3 = 157.25$  mBq.)

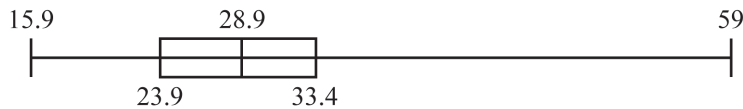
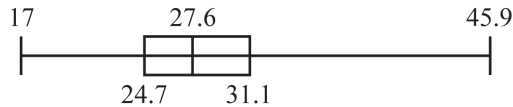


34 Chapter 3: Describing, Exploring, and Comparing Data

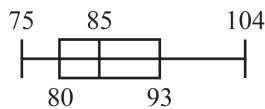
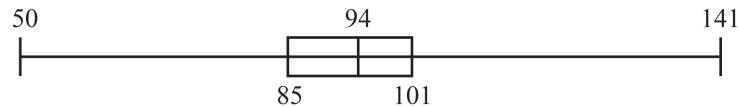
32. The five number summary is 120 mm Hg, 130.0 mm Hg, 132.5 mm Hg, 140.0 mm Hg, 150 mm Hg.



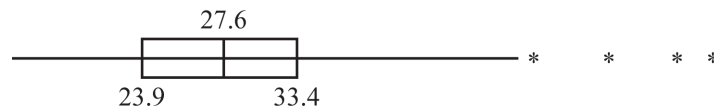
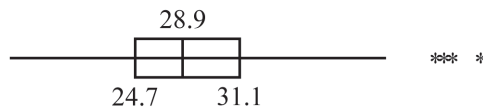
33. The top boxplot represents BMI values for males. The two boxplots do not appear to be very different, so BMI values of males and females appear to be about the same, except for a few very high BMI values for females that caused the boxplot to extend farther to the right.



34. The low lead level group represented in the top boxplot has much more variation and the IQ scores tend to be higher than the IQ scores from the high lead level group.



35. Bottom boxplot represents females. The two boxplots are not dramatically different. The outliers for females are 48.0 in., 52.6 in., 56.8 in., and 59.0 in. The outliers for males are 43.2 in., 43.7 in., 44.2 in., 45.9 in.



### Chapter Quick Quiz

- The sample mean is  $\bar{x} = \frac{4+5+6+6+7+7+7+7+8+8+8+8}{12} = 6.8$  hours.
- The median is  $\frac{7+7}{2} = 7.0$  hours.
- The modes are 7 hours and 8 hours.
- The variance is  $(1.3 \text{ hours})^2 = 1.7 \text{ hours}^2$
- Yes, because 0 hours is substantially less than all of the other data values.
- $z = \frac{5-6.3}{1.4} = -0.93$

7. 75% or 60 sleep times
8. minimum, first quartile  $Q_1$ , second quartile  $Q_2$  (or median), third quartile  $Q_3$ , maximum
9.  $s \approx \frac{10-4}{4} = 1.5$  hours
10.  $\bar{x}, \mu, s, \sigma, s^2, \sigma^2$

**Review Exercises**

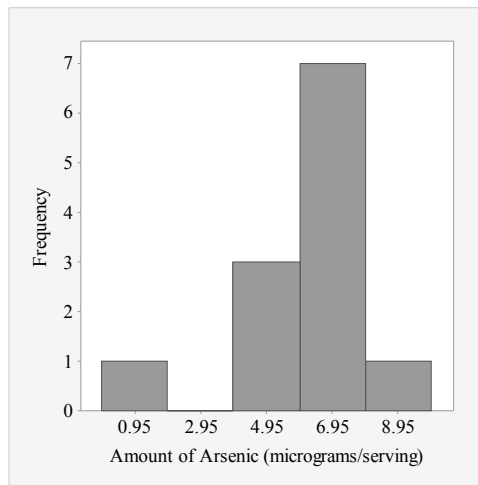
1. a. The mean is  $\bar{x} = \frac{1497+1538+1550+1571+1642}{12} = 1559.6$  mm.  
 b. The median is 1550.0 mm.  
 c. There is no mode.  
 d. The midrange is  $\frac{1497+1642}{2} = 1569.5$  mm.  
 e. The range is  $1642 - 1497 = 145.0$  mm.  
 f.  $s = \sqrt{\frac{(1497-1559.6)^2 + (1538-1559.6)^2 + (1550-1559.6)^2 + (1571-1559.6)^2 + (1642-1559.6)^2}{12-1}} = 53.4$  mm  
 g.  $s^2 = 53.4^2 = 2849.3$  mm<sup>2</sup>
2.  $z = \frac{1642-1559.6}{53.4} = 1.54$ ; The eye height is not significantly low or high because its  $z$  score is between 2 and -2, so it is within 2 standard deviations of the mean.
3. The mean is  $\bar{x} = \frac{12+14+22+27+40}{5} = 23.0$ . The numbers don't measure or count anything. They are used as replacements for the names of the categories, so the numbers are at the nominal level of measurement. In this case the mean is a meaningless statistic.
4. The male  $z$  score is  $z = \frac{3400-3273}{660} = 0.19$ . The female  $z$  score is  $z = \frac{3200-3037}{706} = 0.23$ . The female has the larger relative birth because the female has the larger  $z$  score.
5. The outlier is 646. The mean and standard deviation with the outlier included are  $\bar{x} = 267.8$  and  $s = 131.6$ . With the outlier excluded, the values are  $\bar{x} = 230.0$  and  $s = 42.0$ . Both statistics changed by a substantial amount, so here the outlier has a very strong effect on the mean and standard deviation.
6. The minimum value is 119 mm, the first quartile is 128 mm, the second quartile (or median) is 131 mm, the third quartile is 135 mm, and the maximum value is 141 mm.
7. Significantly low heights are  $97.5 - 2(6.9) = 83.7$  cm or less; significantly high heights are 111.3 cm or greater. The height of  $97.5 + 2(6.9) = 87.8$  cm is not significant, so the physician should not be concerned.
8. The median would be better because it is not affected much by the one very large income.

**Cumulative Review Exercises**

1.

Arsenic ( $\mu\text{g}$ )	Frequency
0.0 – 1.9	1
2.0 – 3.9	0
4.0 – 5.9	3
6.0 – 7.9	7
8.0 – 9.9	1

2.



3.

1.	5
2.	
3.	
4.	9
5.	4 4
6.	1 3 6 7 9
7.	3 8
8.	2

4. a. The mean is  $\bar{x} = \frac{1.5 + 4.9 + 5.4 + 5.4 + 6.1 + 6.3 + 6.6 + 6.7 + 6.9 + 7.3 + 7.8 + 8.2}{12} = 6.09 \mu g$ .

b. The median is  $\frac{6.3 + 6.6}{2} = 6.45 \mu g$ .

c.  $s = \sqrt{\frac{(1.5 - 6.09)^2 + (4.9 - 6.09)^2 + \dots + (7.8 - 6.09)^2 + (8.2 - 6.09)^2}{12 - 1}} = 1.75 \mu g$

d.  $s^2 = 1.75^2 = 3.06 (\mu g)^2$

e. The range is  $8.2 - 1.5 = 6.70 \mu g$ .

5. a. Mode, because the others are numerical measures that require data at the interval or ratio levels of measurement.  
 b. convenience sampling  
 c. More consistency can be achieved by lowering the standard deviation. (It is also important to keep the mean at an acceptable level.)