Chapter Three

Section 3.1

3.1 For a data set with an odd number of observations, first we rank the data set in increasing (or decreasing) order and then find the value of the middle term. This value is the median. For a data set with an even number of observations, first we rank the data set in increasing (or decreasing) order and then find the average of the two middle terms. The average gives the median.

3.2 A few values that are either very small or very large relative to the majority of the values in a data set are called **outliers** or **extreme values**. Suppose the exam scores for seven students are 73, 82, 95, 79, 22, 86, and 91. Then, 22 is an outlier because this value is very small compared to the other values. The median is a better measure of central tendency as compared to the mean for a data set that contains an outlier because the mean is affected much more by outliers than is the median.

3.3 Suppose the exam scores for seven students are 73, 82, 95, 79, 22, 86, and 91 points. Then,
Mean = (73 + 82 + 95 + 79 + 22 + 86 + 91)/7 = 75.43 points. If we drop the outlier (22),
Mean = (73 + 82 + 95 + 79 + 86 + 91)/6 = 84.33 points. This shows how an outlier can affect the value of the mean.

3.4 All three measures of central tendency (the mean, the median, and the mode) can be calculated for quantitative data. Note that the mode may or may not exist for a data set. However, only the mode (if it exists) can be found for a qualitative data set. Examples given in Sections 3.1.1, 3.1.2, and 3.1.3 of the text show these cases.

3.5 The mode can assume more than one value for a data set. Examples 3–8 and 3–9 of the text present such cases.

3.6 A quantitative data set will definitely have a mean and a median but it may or may not have a mode. Example 3–7 of the text presents a data set that has no mode.

3.7 For a symmetric histogram (with one peak), the values of the mean, median, and mode are all equal. Figure 3.2 of the text shows this case. For a histogram that is skewed to the right, the value of the mode is the smallest and the value of the mean is the largest. The median lies between the mode and the mean. Such a case is presented in Figure 3.3 of the text. For a histogram that is skewed to the left, the value of the mean is the smallest, the value of the mode is the largest, and the value of the median lies between the mean and the mode. Figure 3.4 of the text exhibits this case.

3.8 The median is the best measure to summarize this data set since it is not influenced by outliers.

3.9 $\sum x = 5 + (-7) + 2 + 0 + (-9) + 16 + 10 + 7 = 24$

$\mu = (\sum x)/N = 24/8 = 3$

Median = value of the 4.5th term in ranked data = (2 + 5)/2 = 3.50

This data set has no mode.

3.10 $\sum x = 14 + 18 - 10 + 8 + 8 - 16 = 22$

$\bar{x} = (\sum x)/n = 22/6 = 3.67$

Median = value of the 3.5th term in ranked data = (8+8)/2 = 8

Mode = 8

3.11 $\bar{x} = (\sum x)/n$ = 34,015/9 = \$3779.44

Median = value of the 5th term in ranked data set = \$3,250

3.12 $\bar{x} = (\sum x)/n$ = 16,269/9 = \$1807.67

Median = value of the 5th term in ranked data set = \$1,040

3.13 a. $\mu = (\sum x)/N$ = 13,649/50 = \$272.98 billion

Median = value of the 25.5th term in ranked data set = (158 + 166)/2 = \$162 billion

These values are population parameters because the data set includes all 50 states.

b. The mode is 34 and 216 since these values occur twice and no other value occurs more than once.

3.14 $\bar{x} = (\sum x)/n$ = 4407.2/6 = \$734.53 million

Median = value of the 3.5th term in ranked data set = (273.6 + 393.1)/2 = \$333.35 million

This data set has no mode because no value occurs more than once.

3.15 $\bar{x} = (\sum x)/n$ = 970.3/20 = 48.52 million pounds

Median = value of the 10.5th term in ranked data set = (36.9 + 38.3)/2 = 37.6 million pounds

3.16 $\bar{x} = (\sum x)/n$ = 81/12 = 6.75 cars

Median = value of the 6.5th term in ranked data set = (6 + 7)/2 = 6.5 cars

Mode = 3, 6, and 7 cars

3.17 $\bar{x} = (\sum x)/n$ = 3178/6 = \$529.67 million

Median = value of the 3.5th term in ranked data set = (231 + 668)/2 = \$449.5 million

This data set has no mode because no value appears more than once.

3.18 $\mu = (\sum x)/N = 610/15 = 40.67$ major penalties

Median = value of the 8th term in ranked data set = 38 major penalties

The mode is 35 since that value appears twice and no other value appears more than once.

3.19 $\bar{x} = (\sum x)/n = 35/12 = 2.92$ outages

Median = value of the 6.5th term in ranked data set = (2 + 3)/2 = 2.5 outages

Mode = 2 outages

3.20 $\bar{x} = (\sum x)/n = 91/15 = 6.07$ items

Median = value of the 8th term in ranked data set = 7 items

Mode = 7 items

These are sample statistics since they are calculated from the sample of 15 motorists.

3.21 $\bar{x} = (\sum x)/n = 294/10 = 29.4$ computer monitors

Median = value of the 5.5th term in ranked data set = (28 + 29)/2 = 28.5 computer monitors

Mode = 23 computer monitors

3.22 $\bar{x} = (\sum x)/n = 107/12 = 8.92$ students

Median = value of the 6.5th term in ranked data set = (9 + 9)/2 = 9 students

Mode = 6 and 9 students

3.23
a. $\bar{x} = (\sum x)/n = 21{,}636/12 = 1803$ tornadoes

Median = value of the 6.5th term in ranked data set = (1166 + 1374)/2 = 1270 tornadoes

b. The outlier is 5490. When we drop this value,

Mean = 16,146/11 = 1467.82 tornadoes

Median = value of the 6th term in ranked data set = 1166 tornadoes

As we observe, the mean is affected more by the outlier.

c. The median is a better measure because it is not as sensitive to outliers as the mean.

3.24
a. $\bar{x} = (\sum x)/n = 25/10 = 2.5$ women

Median = value of the 5.5th term in ranked data set = (2 + 2)/2 = 2 women

b. The outlier is 9. When we drop this value,

Mean = 16/9 = 1.78 women

Median = value of the 5th term in ranked data set = 2 women

As we observe, the mean is affected more by the outlier.

c. The median is a better measure because it is not as sensitive to outliers as the mean.

3.25 $n_1 = 10$, $n_2 = 8$, $\bar{x}_1 = \$140$, $\bar{x}_2 = \$160$

$$\bar{x} = \frac{n_1\bar{x}_1 + n_2\bar{x}_2}{n_1 + n_2} = \frac{(10)(140)+(8)(160)}{10+8} = \frac{2680}{18} = \$148.89$$

3.26 $n_1 = 18$, $n_2 = 20$, $\bar{x}_1 = \$144$, $\bar{x} = \$150$

$$\bar{x} = \frac{n_1\bar{x}_1 + n_2\bar{x}_2}{n_1 + n_2} = \frac{(18)(144)+(20)(\bar{x}_2)}{18+20} = 150\text{. Then,}$$

$$\bar{x}_2 = \frac{(n_1 + n_2)\bar{x} - n_1\bar{x}_1}{n_2} = \frac{(18+20)(150)-(18)(144)}{20} = 155.4$$

3.27 Total money spent by 10 persons = $\sum x = n\bar{x} = 10(105.50) = \$1055$

3.28 Total 2009 incomes of five families = $\sum x = n\bar{x} = 5(99{,}520) = \$497{,}600$

3.29 Sum of the ages of six persons = (6)(46) = 276 years, so the age of sixth person = 276 – (57 + 39 + 44 + 51 + 37) = 48 years.

3.30 Sum of the prices paid by the seven passengers = (7)(361) = $2527
Total price paid by the couple = 2527 – (402 + 210 + 333 + 695 + 485) = $384
Price paid by each of the couple = 384/2 = $192

3.31 For Data Set I: Mean = 123/5 = 24.6 For Data Set II: Mean = 158/5 = 31.6
The mean of the second data set is greater than the mean of the first data set by 7.

3.32 For Data Set I: Mean = 47/5 = 9.4 For Data Set II: Mean = 94/5 = 18.8
The mean of the second data set is twice the mean of the first data set.

3.33 The ranked data are: 19 23 26 31 38 39 47 49 53 67
By dropping 19 and 67, we obtain $\sum x$ = 23 + 26 + 31 + 38 + 39 + 47 + 49 + 53 = 306
10% Trimmed Mean = $(\sum x)/n$ = 306/8 = 38.25 years

3.34 The ranked data are: 184 195 245 259 271 297 307 309 323 365
369 387 390 410 438 457 578 590 679 795
To calculate the 20% trimmed mean, drop 20% of the smallest values and 20% of the largest values. This data set contains 20 values, and 20% of 20 is 4. Hence, drop the 4 smallest values and the 4 largest values. By dropping 184, 195, 245, 259, 578, 590, 679, and 795, we obtain
$\sum x$ = 271 + 297 + 307 + 309 + 323 + 365 + 369 + 387 + 390 + 410 + 438 + 457 = 4323
20% Trimmed Mean = $(\sum x)/n$ = 4323/12 = $306.25 thousand = $306,250

3.35 From the given information: $x_1 = 73$, $x_2 = 67$, $x_3 = 85$, $w_1 = w_2 = 1$, $w_3 = 2$

$$\text{Weighted mean} = \frac{\sum xw}{\sum w} = \frac{(73)(1)+(67)(1)+(85)(2)}{4} = \frac{310}{4} = 77.5$$

3.36 Geometric mean = $\sqrt[n]{x_1 \cdot x_2 \cdot x_3 \cdot \ldots \cdot x_n} = \sqrt[5]{1.04 \cdot 1.03 \cdot 1.05 \cdot 1.06 \cdot 1.08} = \sqrt[5]{1.287625248} = 1.05$

Then, 1– Geometric mean = 1.05 –1= .05, so the mean inflation rate is 5%.

Section 3.2

3.37 Suppose the exam scores for seven students are 73, 82, 95, 79, 22, 86, and 91.

Then, Range = Largest value – Smallest value = 95 – 22 = 73 points.

If we drop the outlier (22) and calculate the range,

Range = Largest value – Smallest value = 95 – 73 = 22 points.

Thus, when we drop the outlier, the range decreases from 73 to 22 points.

3.38 No, the value of the standard deviation cannot be negative, because the deviations from the mean are squared and, therefore, either positive or zero.

3.39 The value of the standard deviation is zero when all values in a data are the same. For example, suppose the exam scores of a sample of seven students are 82, 82, 82, 82, 82, 82, and 82. As this data set has no variation, the value of the standard deviation is zero for these observations. This is shown below:

$\sum x = 574$ and $\sum x^2 = 47{,}068$

$$s = \sqrt{\frac{\sum x^2 - \frac{(\sum x)^2}{n}}{n-1}} = \sqrt{\frac{47{,}068 - \frac{(574)^2}{7}}{7-1}} = \sqrt{\frac{47{,}068 - 47{,}068}{6}} = 0$$

3.40 A summary measure calculated for a population data set is called a **population parameter**. If the average exam score for all students enrolled in a statistics class is 75.3, then 75.3 is a population parameter. A summary measure calculated for a sample data set is called a **sample statistic**. If we took a random sample of 10 students in the statistics class and found the average exam score to be 77.1, this would be an example of a sample statistic.

3.41 Range = Largest value – Smallest value = 16 – (–9) = 25, $\sum x = 24$, $\sum x^2 = 564$ and $N = 8$

$$\sigma^2 = \frac{\sum x^2 - \frac{(\sum x)^2}{N}}{N} = \frac{564 - \frac{(24)^2}{8}}{8} = \frac{564 - 72}{8} = 61.5 \qquad \sigma = \sqrt{61.5} = 7.84$$

3.42 Range = Largest value – Smallest value = 18 – (–16) = 34, $\sum x = 22$, $\sum x^2 = 1004$, and $n = 6$

$$s^2 = \frac{\sum x^2 - \frac{(\sum x)^2}{n}}{n-1} = \frac{1004 - \frac{(22)^2}{6}}{6-1} = \frac{1004 - 80.6667}{5} = 184.6667 \qquad s = \sqrt{184.6667} = 13.59$$

3.43 a. $\bar{x} = (\sum x)/n = 72/8 = 9$ shoplifters caught

Shoplifters caught	Deviations from the Mean
7	7 – 9 = –2
10	10 – 9 = 1
8	8 – 9 = –1
3	3 – 9 = –6
15	15 – 9 = 6
12	12 – 9 = 3
6	6 – 9 = –3
11	11 – 9 = 2
	Sum = 0

Yes, the sum of the deviations from the mean is zero.

b. Range = Largest value – Smallest value = 15 – 3 = 12, $\sum x = 72$, $\sum x^2 = 748$, and $n = 8$

$$s^2 = \frac{\sum x^2 - \frac{(\sum x)^2}{n}}{n-1} = \frac{748 - \frac{(72)^2}{8}}{8-1} = 14.2857 \qquad s = \sqrt{14.2857} = 3.78$$

3.44 a. $\bar{x} = (\sum x)/n = 840/7 = \$120$

Prices	Deviations from the Mean
\$89	89 – 120 = –31
\$170	170 – 120 = 50
\$104	104 – 120 = –16
\$113	113 – 120 = –7
\$56	56 – 120 = –64
\$161	161 – 120 = 41
\$147	147 – 120 = 27
	Sum = 0

Yes, the sum of the deviations from the mean is zero.

b. $\sum x = 840$, $\sum x^2 = 111{,}072$, and $n = 7$

Range = 170 – 56 = \$114

$$s^2 = \frac{\sum x^2 - \frac{(\sum x)^2}{n}}{n-1} = \frac{111{,}072 - \frac{(840)^2}{7}}{7-1} = 1712 \qquad s = \sqrt{1712} = \$41.38$$

3.45 $\sum x = 81$, $\sum x^2 = 699$, and $n = 12$

Range = Largest value – Smallest value = 15 – 2 = 13 thefts

$$s^2 = \frac{\sum x^2 - \frac{(\sum x)^2}{n}}{n-1} = \frac{699 - \frac{(81)^2}{12}}{12-1} = 13.8409 \qquad s = \sqrt{13.8409} = 3.72 \text{ thefts}$$

3.46 $\sum x = 21{,}636$, $\sum x^2 = 56{,}052{,}418$, and $n = 12$

Range = Largest value – Smallest value = 5490 –1039 = 4451 tornadoes

$$s^2 = \frac{\sum x^2 - \frac{(\sum x)^2}{n}}{n-1} = \frac{56{,}052{,}418 - \frac{(21{,}636)^2}{12}}{12-1} = 1{,}549{,}337.2727$$

$s = \sqrt{1{,}549{,}337.2727} = 1244.72$ tornadoes

3.47 $\sum x = 291$, $\sum x^2 = 9171$, and $n = 10$

Range = Largest value – Smallest value = 41 – 14 = 27 pieces

$$s^2 = \frac{\sum x^2 - \frac{(\sum x)^2}{n}}{n-1} = \frac{9171 - \frac{(291)^2}{10}}{10-1} = 78.1 \qquad s = \sqrt{78.1} = 8.84 \text{ pieces}$$

3.48 $\sum x = 55$, $\sum x^2 = 397$, and $n = 9$

Range = Largest value – Smallest value = 10 – 2 = 8 collisions

$$s^2 = \frac{\sum x^2 - \frac{(\sum x)^2}{n}}{n-1} = \frac{397 - \frac{(55)^2}{9}}{9-1} = 7.6111 \qquad s = \sqrt{7.6111} = 2.76 \text{ collisions}$$

3.49 $\sum x = 27$, $\sum x^2 = 111$, and $n = 13$

Range = Largest value – Smallest value = 7 – 0 = 7 patients

$$s^2 = \frac{\sum x^2 - \frac{(\sum x)^2}{n}}{n-1} = \frac{111 - \frac{(27)^2}{13}}{13-1} = 4.5769 \qquad s = \sqrt{4.5769} = 2.14 \text{ patients}$$

3.50 $\sum x = 180$, $\sum x^2 = 3666$, and $n = 10$

Range = Largest value – Smallest value = 32 – 8 = 24 hotdogs

$$s^2 = \frac{\sum x^2 - \frac{(\sum x)^2}{n}}{n-1} = \frac{3666 - \frac{(180)^2}{10}}{10-1} = 47.3333 \qquad s = \sqrt{47.3333} = 6.88 \text{ hotdogs}$$

3.51 $\sum x = 80$, $\sum x^2 = 1552$, and $n = 8$

Range = Largest value – Smallest value = 23 – (–7) = 30º Fahrenheit

$$s^2 = \frac{\sum x^2 - \frac{(\sum x)^2}{n}}{n-1} = \frac{1552 - \frac{(80)^2}{8}}{8-1} = 107.4286 \qquad s = \sqrt{107.4286} = 10.36^\circ \text{ Fahrenheit}$$

3.52 $\sum x = 64$, $\sum x^2 = 580$, and $n = 10$

Range = Largest value – Smallest value = 14 – 0 = 14 hours

$$s^2 = \frac{\sum x^2 - \frac{(\sum x)^2}{n}}{n-1} = \frac{580 - \frac{(64)^2}{10}}{10-1} = 18.9333 \qquad s = \sqrt{18.9333} = 4.35 \text{ hours}$$

3.53 $\sum x = 450$, $\sum x^2 = 21{,}616$, and $n = 10$

Range = Largest value – Smallest value = 69 – 31 = 38 points

$$s^2 = \frac{\sum x^2 - \frac{(\sum x)^2}{n}}{n-1} = \frac{21{,}616 - \frac{(450)^2}{10}}{10-1} = 151.7778 \qquad s = \sqrt{151.7778} = 12.32 \text{ points}$$

3.54 $\sum x = 9127$, $\sum x^2 = 11{,}861{,}223$, and $n = 8$

Range = Largest value – Smallest value = 1800 – 489 = \$1311 thousand

$$s^2 = \frac{\sum x^2 - \frac{(\sum x)^2}{n}}{n-1} = \frac{11{,}861{,}223 - \frac{(9127)^2}{8}}{8-1} = 206{,}922.4107 \quad s = \sqrt{260{,}922.4107} = \$454.89 \text{ thousand}$$

3.55 $\sum x = 176$, $\sum x^2 = 3872$, and $n = 8$

$$s = \sqrt{\frac{\sum x^2 - \frac{(\sum x)^2}{n}}{n-1}} = \sqrt{\frac{3872 - \frac{(176)^2}{8}}{8-1}} = \sqrt{\frac{3872 - 3872}{7}} = 0$$

The standard deviation is zero because all these data values are the same and there is no variation among them.

3.56 $\sum x = 114$, $\sum x^2 = 2166$, and $n = 6$

$$s = \sqrt{\frac{\sum x^2 - \frac{(\sum x)^2}{n}}{n-1}} = \sqrt{\frac{2166 - \frac{(114)^2}{6}}{6-1}} = \sqrt{\frac{2166 - 2166}{5}} = 0$$

The standard deviation is zero because all these data values are the same and there is no variation among them.

3.57 For the yearly salaries of all employees, CV = $(\sigma/\mu) \times 100\%$ = (6,820/62,350) × 100 = 10.94%

For the years of experience of these employees, CV = $(\sigma/\mu) \times 100\%$ = (2/15) × 100 = 13.33%

The relative variation in salaries is lower than that in years of experience.

3.58 For the SAT scores of the 100 students, CV = $(s/\bar{x}) \times 100\%$ = (105/975) × 100 = 10.77%

For the GPAs of these students, CV = $(s/\bar{x}) \times 100\%$ = (.22/3.16) × 100 = 6.96%

The relative variation in SAT scores is higher than that in GPAs.

3.59 For Data Set I: $\sum x = 123$, $\sum x^2 = 3883$, and $n = 5$

$$s = \sqrt{\frac{\sum x^2 - \frac{(\sum x)^2}{n}}{n-1}} = \sqrt{\frac{3883 - \frac{(123)^2}{5}}{5-1}} = \sqrt{214.300} = 14.64$$

For Data Set II: $\sum x = 158$, $\sum x^2 = 5850$, and $n = 5$

$$s = \sqrt{\frac{\sum x^2 - \frac{(\sum x)^2}{n}}{n-1}} = \sqrt{\frac{5850 - \frac{(158)^2}{5}}{5-1}} = \sqrt{214.300} = 14.64$$

The standard deviations of the two data sets are equal.

3.60 For Data Set I: $\sum x = 47$, $\sum x^2 = 507$, and $N = 5$

$$\sigma = \sqrt{\frac{\sum x^2 - \frac{(\sum x)^2}{N}}{N}} = \sqrt{\frac{507 - \frac{(47)^2}{5}}{5}} = \sqrt{13.04} = 3.61$$

For Data Set II: $\sum x = 94$, $\sum x^2 = 2028$, and $N = 5$

$$\sigma = \sqrt{\frac{\sum x^2 - \frac{(\sum x)^2}{N}}{N}} = \sqrt{\frac{2028 - \frac{(94)^2}{5}}{5}} = \sqrt{52.16} = 7.22$$

The standard deviation of the second data set is twice the standard deviation of the first data set.

Section 3.3

3.61 The values of the mean and standard deviation for a grouped data set are the approximate values of the mean and standard deviation. The exact values of the mean and standard deviation are obtained only when ungrouped data are used.

3.62

x	f	m	mf	m^2f
2 – 4	5	3	15	45
5 – 7	9	6	54	324
8 – 10	14	9	126	1134
11 – 13	7	12	84	1008
14 – 16	5	15	75	1125
	$N = \sum f = 40$		$\sum mf = 354$	$\sum m^2f = 3636$

$\mu = (\sum mf)/N = 354/40 = 8.85$

$$\sigma^2 = \frac{\sum m^2 f - \frac{(\sum mf)^2}{N}}{N} = \frac{3636 - \frac{(354)^2}{40}}{40} = 12.5775 \qquad \sigma = \sqrt{12.5775} = 3.55$$

3.63

x	f	m	mf	m^2f
0 to less than 4	17	2	34	68
4 to less than 8	23	6	138	828
8 to less than 12	15	10	150	1500
12 to less than 16	11	14	154	2156
16 to less than 20	8	18	144	2592
20 to less than 24	6	22	132	2904
	$n = \sum f = 80$		$\sum mf = 752$	$\sum m^2 f = 10{,}048$

$\bar{x} = (\sum mf)/n = 752/80 = 9.40$

$$s^2 = \frac{\sum m^2 f - \frac{(\sum mf)^2}{n}}{n-1} = \frac{10{,}048 - \frac{(752)^2}{80}}{80-1} = 37.7114 \qquad s = \sqrt{37.7114} = 6.14$$

3.64

Amount of Telephone Bill (dollars)	Number of Families	m	mf	m^2f
40 to less than 70	9	55	495	27,225
70 to less than 100	11	85	935	79,475
100 to less than 130	16	115	1840	211,600
130 to less than 160	10	145	1450	210,250
160 to less than 190	4	175	700	122,500
	$n = \sum f = 50$		$\sum mf = 5420$	$\sum m^2 f = 651{,}050$

$\bar{x} = (\sum mf)/n = 5420/50 = \$108.40$

$$s^2 = \frac{\sum m^2 f - \frac{(\sum mf)^2}{n}}{n-1} = \frac{651{,}050 - \frac{(5420)^2}{50}}{50-1} = 1296.3673 \qquad s = \sqrt{1296.3673} = \$36.01$$

3.65

Hours Per Week	Number of Students	m	mf	m^2f
0 to less than 5	7	2.5	17.5	43.75
5 to less than 10	12	7.5	90.0	675.00
10 to less than 15	15	12.5	187.5	2343.75
15 to less than 20	13	17.5	227.5	3981.25
20 to less than 25	8	22.5	180.0	4050.00
25 to less than 30	5	27.5	137.5	3781.25
	$N = \sum f = 60$		$\sum mf = 840$	$\sum m^2 f = 14{,}875$

$\mu = (\sum mf)/N = 840/60 = 14$ hours

$$\sigma^2 = \frac{\sum m^2 f - \frac{(\sum mf)^2}{N}}{N-1} = \frac{14{,}875 - \frac{(840)^2}{60}}{60} = 51.9167 \qquad \sigma = \sqrt{51.9167} = 7.21 \text{ hours}$$

3.66

Weight (pounds)	Number of Babies	m	mf	m^2f
3 to less than 5	5	4	20	80
5 to less than 7	30	6	180	1080
7 to less than 9	40	8	320	2560
9 to less than 11	20	10	200	2000
11 to less than 13	5	12	60	720
	$N = \sum f = 100$		$\sum mf = 780$	$\sum m^2f = 6440$

$\mu = (\sum mf)/N = 780/100 = 7.80$ pounds

$$\sigma^2 = \frac{\sum m^2 f - \frac{(\sum mf)^2}{N}}{N} = \frac{6440 - \frac{(780)^2}{100}}{100} = 3.5600 \qquad \sigma = \sqrt{3.5600} = 1.89 \text{ pounds}$$

3.67

Miles Driven in 2009 (in thousands)	Number of Car Owners	m	mf	m^2f
0 to less than 5	7	2.5	17.5	43.75
5 to less than 10	26	7.5	195.0	1462.50
10 to less than 15	59	12.5	737.5	9218.75
15 to less than 20	71	17.5	1242.5	21,743.75
20 to less than 25	62	22.5	1395.0	31,387.50
25 to less than 30	39	27.5	1072.5	29,493.75
30 to less than 35	22	32.5	715.0	23,237.50
35 to less than 40	14	37.5	525.0	19,687.50
	$n = \sum f = 300$		$\sum mf = 5900$	$\sum m^2f = 136{,}275$

$\bar{x} = (\sum mf)/n = 5900/300 = 19.67$ or 19,670 miles

$$s^2 = \frac{\sum m^2 f - \frac{(\sum mf)^2}{n}}{n-1} = \frac{136{,}275 - \frac{(5900)^2}{300}}{300-1} = 67.6979 \qquad s = \sqrt{67.6979} = 8.23 \text{ or } 8230 \text{ miles}$$

Each value in the column labeled *mf* gives the approximate total mileage for the car owners in the corresponding class. For example, the value of $mf = 17.5$ for the first class indicates that the seven car owners in this class drove a total of approximately 17,500 miles. The value $\sum mf = 5900$ indicates that the total mileage for all 300 car owners was approximately 5,900,000 miles.

3.68

Amount of Electric Bill (dollars)	Number of Families	m	mf	m^2f
0 to less than 40	5	20	100	2,000
40 to less than 80	16	60	960	57,600
80 to less than 120	11	100	1100	110,000
120 to less than 160	10	140	1400	196,000
160 to less than 200	8	180	1440	259,200
	$n = \sum f = 50$		$\sum mf = 5000$	$\sum m^2f = 624{,}800$

$\bar{x} = (\sum mf)/n = 5000/50 = \$100$

$$s^2 = \frac{\sum m^2 f - \frac{(\sum mf)^2}{n}}{n-1} = \frac{624{,}800 - \frac{(5000)^2}{50}}{50-1} = 2546.9388 \qquad s = \sqrt{2546.9388} = \$50.47$$

The values in the column labeled *mf* give the approximate total amounts of electric bills for the families belonging to corresponding classes. For example, the five families belonging to the first class paid a total of $100 for electricity in August 2009. The value $\sum mf = \$5000$ is the approximate total amount of the electric bills for all 50 families included in the sample.

3.69

x	f	m	mf	m^2f
0 to less than 20	14	10	140	1400
20 to less than 40	18	30	540	16,200
40 to less than 60	9	50	450	22,500
60 to less than 80	5	70	350	24,500
80 to less than 100	4	90	360	32,400
	$n = \sum f = 50$		$\sum mf = 1840$	$\sum m^2 f = 97{,}000$

$\bar{x} = (\sum mf)/n = 1840/50 = 36.80$ minutes

$$s^2 = \frac{\sum m^2 f - \frac{(\sum mf)^2}{n}}{n-1} = \frac{97{,}000 - \frac{(1840)^2}{50}}{50-1} = 597.7143 \qquad s = \sqrt{597.7143} = 24.45 \text{ minutes}$$

3.70

Number of Errors	f	m	mf	m^2f
0	11	0	0	0
1	14	1	14	14
2	9	2	18	36
3	7	3	21	63
4	3	4	12	48
5	1	5	5	25
	$N = \sum f = 45$		$\sum mf = 70$	$\sum m^2 f = 186$

$\mu = (\sum mf)/N = 70/45 = 1.56$ errors

$$\sigma^2 = \frac{\sum m^2 f - \frac{(\sum mf)^2}{N}}{N} = \frac{186 - \frac{(70)^2}{45}}{45} = 1.7136 \qquad \sigma = \sqrt{1.7136} = 1.31 \text{ errors}$$

3.71 a. $\bar{x} = \sum x/n = 1946.66/14 = \$139.05$ per barrel

b.

Price of Oil per Barrel	f	m	mf
128.00 to less than 131.00	2	129.5	259.0
131.00 to less than 134.00	0	132.5	0.0
134.00 to less than 137.00	3	135.5	406.5
137.00 to less than 140.00	2	138.5	277.0
140.00 to less than 143.00	3	141.5	424.5
143.00 to less than 146.00	4	144.5	578.0
	$n = \sum f = 14$		$\sum mf = 1945.0$

c. $\bar{x} = (\sum mf)/n = 1945/14 = \$138.93$ per barrel

d. The two means are not equal because the second method uses approximations (mid points of the range) and the first one does not. This leads to slightly different results.

Section 3.4

3.72 Chebyshev's theorem is applied to find a lower bound for the area under a distribution curve between two points that are on opposite sides of the mean and at the same distance from the mean. According to this theorem, for any number k greater than 1, at least $(1 - (1/k^2))$ of the data values lie within k standard deviations of the mean.

3.73 The empirical rule is applied to a bell–shaped distribution. According to this rule, approximately
(1) 68% of the observations lie within one standard deviation of the mean.
(2) 95% of the observations lie within two standard deviations of the mean.
(3) 99.7% of the observations lie within three standard deviations of the mean.

3.74 For the interval $\bar{x} \pm 2s$: $k = 2$, and $1 - \frac{1}{k^2} = 1 - \frac{1}{(2)^2} = 1 - .25 = .75$ or 75%. Thus, at least 75% of the observations fall in the interval $\bar{x} \pm 2s$.

For the interval $\bar{x} \pm 2.5s$: $k = 2.5$, and $1 - \frac{1}{k^2} = 1 - \frac{1}{(2.5)^2} = 1 - .16 = .84$ or 84%. Thus, at least 84% of the observations fall in the interval $\bar{x} \pm 2.5s$.

For the interval $\bar{x} \pm 3s$: $k = 3$, and $1 - \frac{1}{k^2} = 1 - \frac{1}{(3)^2} = 1 - .11 = .89$ or 89%. Thus, at least 89% of the observations fall in the interval $\bar{x} \pm 3s$.

3.75 For the interval $\mu \pm 2\sigma$: $k = 2$, and $1 - \frac{1}{k^2} = 1 - \frac{1}{(2)^2} = 1 - .25 = .75$ or 75%. Thus, at least 75% of the observations fall in the interval $\mu \pm 2\sigma$.

For the interval $\mu \pm 2.5\sigma$: $k = 2.5$, and $1 - \frac{1}{k^2} = 1 - \frac{1}{(2.5)^2} = 1 - .16 = .84$ or 84%. Thus, at least 84% of the observations fall in the interval $\mu \pm 2.5\sigma$.

For the interval $\mu \pm 3\sigma$: $k = 3$, and $1 - \frac{1}{k^2} = 1 - \frac{1}{(3)^2} = 1 - .11 = .89$ or 89%. Thus, at least 89% of the observations fall in the interval $\mu \pm 3\sigma$.

3.76 Approximately 68% of the observations fall in the interval $\mu \pm \sigma$, approximately 95% fall in the interval $\mu \pm 2\sigma$, and about 99.7% fall in the interval $\mu \pm 3\sigma$.

3.77 Approximately 68% of the observations fall in the interval $\bar{x} \pm s$, approximately 95% fall in the interval $\bar{x} \pm 2s$, and about 99.7% fall in the interval $\bar{x} \pm 3s$.

3.78 a. Each of the two values is 40 minutes from μ = 220. Hence,

$$k = 40/20 = 2 \text{ and } 1 - \frac{1}{k^2} = 1 - \frac{1}{(2)^2} = 1 - .25 = .75 \text{ or } 75\%.$$

Thus, at least 75% of the runners ran the race in 180 to 260 minutes.

b. Each of the two values is 60 minutes from μ = 220. Hence,

$$k = 60/20 = 3 \text{ and } 1 - \frac{1}{k^2} = 1 - \frac{1}{(3)^2} = 1 - .11 = .89 \text{ or } 89\%.$$

Thus, at least 89% of the runners ran the race in 160 to 280 minutes.

c. Each of the two values is 50 minutes from μ = 220. Hence,

$$k = 50/20 = 2.5 \text{ and } 1 - \frac{1}{k^2} = 1 - \frac{1}{(2.5)^2} = 1 - .16 = .84 \text{ or } 84\%.$$

Thus, at least 84% of the runners ran this race in 170 to 270 minutes.

3.79 a. Each of the two values is \$1.2 million from μ = \$2.3 million. Hence,

$$k = 1.2/.6 = 2 \text{ and } 1 - \frac{1}{k^2} = 1 - \frac{1}{(2)^2} = 1 - .25 = .75 \text{ or } 75\%.$$

Thus, at least 75% of all companies had 2009 gross sales of \$1.1 to \$3.5 million.

b. Each of the two values is \$1.5 million from μ = \$2.3 million. Hence,

$$k = 1.5/.6 = 2.5 \text{ and } 1 - \frac{1}{k^2} = 1 - \frac{1}{(2.5)^2} = 1 - .16 = .84 \text{ or } 84\%.$$

Thus, at least 84% of all companies had 2009 gross sales of \$.8 to \$3.8 million.

c. Each of the two values is \$1.8 million from μ = \$2.3 million. Hence,

$$k = 1.8/.6 = 3 \text{ and } 1 - \frac{1}{k^2} = 1 - \frac{1}{(3)^2} = 1 - .11 = .89 \text{ or } 89\%.$$

Thus, at least 89% of all companies had 2009 gross sales of \$.5 to \$ 4.1 million.

3.80 μ = \$9500 and σ = \$2600

a. i. Each of the two values is \$5200 from μ = \$9500. Hence,

$$k = 5200/2600 = 2 \text{ and } 1 - \frac{1}{k^2} = 1 - \frac{1}{(2)^2} = 1 - .25 = .75 \text{ or } 75\%.$$

Thus, at least 75% of all households have credit card debt between \$4300 and \$14,700.

ii. Each of the two values is \$6500 from μ = \$9500. Hence,

k = 6500/2600 = 2.5 and $1 - \frac{1}{k^2} = 1 - \frac{1}{(2.5)^2}$ = 1 – .16 = .84 or 84%.

Thus, at least 84% of all households have credit card debt between \$3000 and \$16,000.

b. $1 - \frac{1}{k^2}$ = .89 gives $\frac{1}{k^2}$ = 1 – .89 = .11 or $k^2 = \frac{1}{.11}$, so $k \approx 3$.

$\mu - 3\sigma$ = 9500 – 3(2600) = \$1700 and $\mu + 3\sigma$ = 9500 + 3(2600) = \$17,300

Thus, the required interval is \$1700 to \$17,300.

3.81 a. i. Each of the two values is \$680 from μ = \$2365. Hence,

k = 680/340 = 2 and $1 - \frac{1}{k^2} = 1 - \frac{1}{(2)^2}$ = 1 – .25 = .75 or 75%.

Thus, at least 75% of all homeowners pay a monthly mortgage of \$1685 to \$3045.

ii. Each of the two values is \$1020 from μ = \$2365. Hence,

k = 1020/340 = 3 and $1 - \frac{1}{k^2} = 1 - \frac{1}{(3)^2}$ = 1 – .11 = .89 or 89%.

Thus, at least 89% of all homeowners pay a monthly mortgage of \$1345 to \$3385.

b. $1 - \frac{1}{k^2}$ = .84 gives $\frac{1}{k^2}$ = 1 – .84 = .16 or $k^2 = \frac{1}{.16}$ so k = 2.5.

$\mu - 2.5\sigma$ = 2365 – 2.5(340) = \$1515 and $\mu + 2.5\sigma$ = 2365 +2.5(340) = \$3215

Thus, the required interval is \$1515 to \$3215.

3.82 μ = 44 months and σ = 3 months.

a. The interval 41 to 47 months is $\mu - \sigma$ to $\mu + \sigma$. Hence, approximately 68% of the batteries have a life of 41 to 47 months.

b. The interval 38 to 50 months is $\mu - 2\sigma$ to $\mu + 2\sigma$. Hence, approximately 95% of the batteries have a life of 38 to 50 months.

c. The interval 35 to 53 months is $\mu - 3\sigma$ to $\mu + 3\sigma$. Hence, approximately 99.7% of the batteries have a life of 35 to 53 months.

3.83 μ = \$3423 and σ = \$520

a. The interval \$1863 to \$4983 is $\mu - 3\sigma$ to $\mu + 3\sigma$. Hence, approximately 99.7% of employees have annual premium payments between \$1863 and \$4983.

b. The interval \$2903 to \$3943 is $\mu - \sigma$ to $\mu + \sigma$. Hence, approximately 68% of employees have annual premium payments between \$2903 and \$3943.

c. The interval \$2383 to \$4463 is $\mu - 2\sigma$ to $\mu + 2\sigma$. Hence, approximately 95% of employees have annual premium payments between \$2383 and \$4463.

3.84 μ = \$105 and σ = \$20

a. i. The interval \$85 to \$125 is $\mu-\sigma$ to $\mu+\sigma$. Hence, approximately 68% of all college textbooks are priced between \$85 and \$125.

ii. The interval \$65 to \$145 is $\mu-2\sigma$ to $\mu+2\sigma$. Hence, approximately 95% of all college textbooks are priced between \$65 and \$145.

b. $\mu-3\sigma$ = 105 – 3(20) = \$45 and $\mu+3\sigma$ = 105 + 3(20) = \$165. The interval that contains the prices of 99.7% of college textbooks is \$45 to \$165.

3.85 μ = 72 mph and σ = 3 mph

a. i. The interval 63 to 81 mph is $\mu-3\sigma$ to $\mu+3\sigma$. Hence, about 99.7% of speeds of all vehicles are between 63 and 81 mph.

ii. The interval 69 to 75 mph is $\mu-\sigma$ to $\mu+\sigma$. Hence, about 68% of the speeds of all vehicles are between 69 and 75 mph.

b. $\mu-2\sigma$ = 72 – 2(3) = 66 mph and $\mu+2\sigma$ = 72 + 2(3) = 78 mph. The interval that contains the speeds of 95% of the vehicles is 66 to 78 mph.

Section 3.5

3.86 To find the three quartiles:

1. Rank the given data set in increasing order.
2. Find the median using the procedure in Section 3.1.2. The median is the second quartile, Q_2.
3. The first quartile, Q_1, is the value of the middle term among the (ranked) observations that are less than Q_2.
4. The third quartile, Q_3, is the value of the middle term among the (ranked) observations that are greater that Q_2.

Examples 3–20 and 3–21 of the text exhibit how to calculate the three quartiles for data sets with an even and odd number of observations, respectively.

3.87 The **interquartile range** (IQR) is given by $Q_3 - Q_1$, where Q_1 and Q_3 are the first and third quartiles, respectively. Examples 3–20 and 3–21 of the text show how to find the IQR for a data set.

3.88 Given a data set of *n* values, to find the k^{th} percentile (P_k):

1. Rank the given data in increasing order.
2. Calculate *kn*/ 100. Then, P_k is the term that is approximately (*kn*/100) in the ranking. If *kn*/ 100 falls between two consecutive integers *a* and *b*, it may be necessary to average the a^{th} and b^{th} values in the ranking to obtain P_k.

3.89 If x_i is a particular observation in the data set, the **percentile rank of x_i** is the percentage of the values in the data set that are less than x_i. Thus,

$$\text{Percentile rank of } x_i = \frac{\text{Number of values less than } x_i}{\text{Total number of values in the data set}} \times 100$$

3.90 The ranked data are: 5 5 7 8 8 9 10 10 11 11 12 14 18 21 25

a. The three quartiles are $Q_1 = 8$, $Q_2 = 10$, and $Q_3 = 14$

IQR = $Q_3 - Q_1 = 14 - 8 = 6$

b. $kn/100 = 82(15)/100 = 12.30 \approx 12$

Thus, the 82nd percentile can be approximated by the value of the 12th term in the ranked data, which is 14. Therefore, $P_{82} = 14$.

c. Six values in the given data are smaller than 10. Hence, the percentile rank of 10 = (6/15) × 100 = 40%.

3.91 The ranked data are: 68 68 69 69 71 72 73 74 75 76 77 78 79

a. The three quartiles are $Q_1 = (69 + 69)/2 = 69$, $Q_2 = 73$, and $Q_3 = (76 + 77)/2 = 76.5$

IQR = $Q_3 - Q_1 = 76.5 - 69 = 7.5$

b. $kn/100 = 35(13)/100 = 4.55 \approx 4.5$

Thus, the 35th percentile can be approximated by the average of the fourth and fifth terms in the ranked data. Therefore, $P_{35} = (69 + 71)/2 = 70$.

c. Four values in the given data set are smaller than 70. Hence, the percentile rank of 70 = (4/13) × 100 = 30.77%.

3.92 The ranked data are: 41 42 43 44 44 45 46 46 47 47 48 48
48 49 50 50 51 51 52 52 52 53 53 54 56

a. The three quartiles are $Q_1 = (45 + 46)/2 = 45.5$, $Q_2 = 48$, and $Q_3 = (52 + 52)/2 = 52$

IQR = $Q_3 - Q_1 = 52 - 45.5 = 6.5$

b. $kn/100 = 53(25)/100 = 13.25 \approx 13$

Thus, the 53rd percentile can be approximated by the value of the thirteenth term in the ranked data, which is 48. Therefore, $P_{53} = 48$.

c. Fourteen values in the given data are less than 50. Therefore, the percentile rank of 50 = (14/25) × 100 = 56%.

3.93 The ranked data are:

318	336	337	339	362	363	366	369	372	375
378	381	384	385	386	387	390	393	395	403
405	409	417	431	433	434	438	444	461	480

a. The quartiles are $Q_1 = 369$, $Q_2 = (386 + 387)/2 = 386.5$, and $Q_3 = 417$

IQR = $Q_3 - Q_1 = 417 - 369 = 48$

b. $kn/100 = 57(30)/100 = 17.1 \approx 17$

Thus, the 57^{th} percentile can be approximated by the value of the 17^{th} term in the ranked data, which is 390. Therefore, $P_{57} = 390$.

c. Twenty-two values in the given data are smaller than 417. Hence, the percentile rank of 417 = (22/30) × 100 = 73.33%.

3.94 The ranked data are: 3 5 6 6 7 9 9 10 11 12 14 15

a. The three quartiles are $Q_1 = (6+6)/2 = 6$, $Q_2 = (9+9)/2 = 9$, and $Q_3 = (11+12)/2 = 11.5$

$IQR = Q_3 - Q_1 = 11.5 - 6 = 5.5$

The value 10 lies between Q_2 and Q_3, which means it lies in the third 25% group from the bottom in the ranked data set.

b. $kn/100 = 55(12)/100 = 6.6 \approx 6.5$

Thus, the 55^{th} percentile can be approximated by the average of the sixth and seventh terms in the ranked data. Therefore, $P_{55} = (9 + 9)/2 = 9$.

c. Four values in the given data set are less than 7. Hence, the percentile rank of 7 = (4/12) × 100 = 33.33%.

3.95 The ranked data are: 20 22 23 23 23 23 24 25 26 26 27 27 27 28 28
29 29 31 31 31 32 33 33 33 34 35 35 36 37 43

a. The three quartiles are $Q_1 = 25$, $Q_2 = (28+29)/2 = 28.5$, and $Q_3 = 33$

$IQR = Q_3 - Q_1 = 33 - 25 = 8$

The value 31 lies between Q_2 and Q_3, which means that it is in the third 25% group from the bottom in the (ranked) data set.

b. $kn/100 = 65(30)/100 \approx 19.5$

Thus, the 65^{th} percentile may be approximated by the average of the nineteenth and twentieth terms in the ranked data. Therefore, $P_{65} = (31 + 31)/2 = 31$. Thus, we can state that the number of computer monitors produced by Nixon Corporation is less than or equal to 31 for approximately 65% of the days in this sample.

c. Twenty values in the given data are less than 32. Hence, the percentile rank of 32 = (20/30) × 100 = 66.67%. Thus, on 66.67% of the days in this sample, fewer than 32 monitors were produced. Hence, for 100 – 66.67 = 33.33% of the days, the company produced 32 or more monitors.

3.96 The ranked data are: 3 3 4 5 5 6 7 7 8 8 8 9 9 10 10 11 11 12 12 16

a. The three quartiles are $Q_1 = (5 + 6)/2 = 5.5$, $Q_2 = (8 + 8)/2 = 8$, and $Q_3 = (10 + 11)/2 = 10.5$

$IQR = Q_3 - Q_1 = 10.5 - 5.5 = 5$

The value 4 lies below Q_1, which indicates that it is in the bottom 25% group in the (ranked) data set.

b. $kn/100 = 25(20)/100 = 5$

Thus, the 25th percentile may be approximated by the value of the fifth term in the ranked data, which is 5. Therefore, $P_{25} = 5$. Thus, the number of new cars sold at this dealership is less than or equal to 5 for approximately 25% of the days in this sample.

c. Thirteen values in the given data are less than 10. Hence, the percentile rank of 10 = (13/20) × 100 = 65%. Thus, on 65% of the days in the sample, this dealership sold fewer than 10 cars.

3.97 The ranked data are:

356	422	430	468	494	533	572	600	604	617	625
628	639	647	690	702	728	747	749	772	797	805

a. The three quartiles are

$Q_1 = 533$, $Q_2 = (625 + 628)/2 = 626.5$, and $Q_3 = 728$

$IQR = Q_3 - Q_1 = 728 - 533 = 195$

The value 617 lies between Q_1 and Q_2, which indicates that it is in the second 25% group from the bottom in the (ranked) data set.

b. $kn/100 = 30(22)/100 = 6.6 \approx 6.5$

Thus, the 30th percentile may be approximated by the average of the 6th and 7th terms in the ranked data. Therefore, $P_{30} = (533 + 572)/2 = 552.5$. Thus, approximately 30% of the credit scores of applicants are at or below 552.5.

c. Five values in the given data are smaller than 533. Hence, the percentile rank of 533 = (5/22) × 100 = 22.73%. Thus, 22.73% of the credit scores of applicants are below 533.

Section 3.6

3.98 A **box–and–whisker plot** is based on five summary measures: the median, the first quartile, the third quartile, and the smallest and largest value in the data set between the lower and upper inner fences.

3.99 The ranked data are:

22	24	25	28	31	32	34	35	36	41	42	43
47	49	52	55	58	59	61	61	63	65	73	98

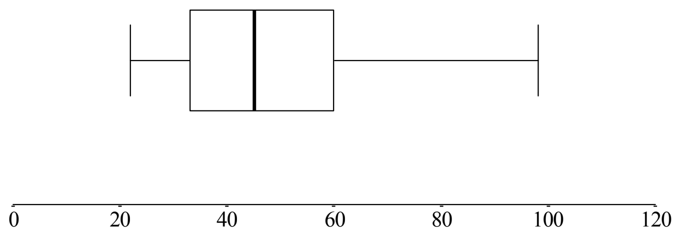
Median = (43 + 47)/2 = 45, $Q_1 = (32 + 34)/2 = 33$, and $Q_3 = (59 + 61)/2 = 60$,

$IQR = Q_3 - Q_1 = 60 - 33 = 27$, $1.5 \times IQR = 1.5 \times 27 = 40.5$,

Lower inner fence = $Q_1 - 40.5 = 33 - 40.5 = -7.5$,

Upper inner fence = $Q_3 + 40.5 = 60 + 40.5 = 100.5$

The smallest and largest values within the two inner fences are 22 and 98, respectively. The data set has no outliers. The box–and–whisker plot is shown below.



3.100 The ranked data are: 3 6 7 8 11 13 14 15 16 18 19 23 26 29 30 31 33 42 62 75

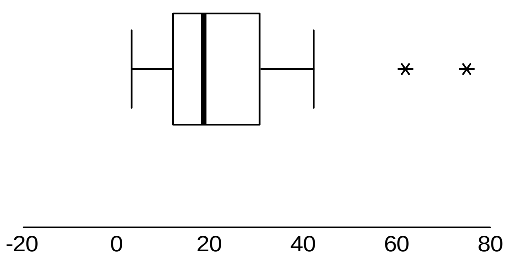
Median = (18 + 19)/2 = 18.5, Q_1 = (11 + 13)/2 = 12, and Q_3 = (30 + 31)/2 = 30.5,

IQR = $Q_3 - Q_1$ = 30.5 – 12 = 18.5, 1.5 × IQR = 1.5 × 18.5 = 27.75,

Lower inner fence = Q_1 – 27.75 = 12 – 27.75= – 15.75,

Upper inner fence = Q_3 + 27.75= 30.5 + 27.75= 58.25

The smallest and the largest values within the two inner fences are 3 and 42, respectively.



There are two outliers, 62 and 75. To classify them, we compute:

3.0 × IQR = 3.0 × 18.5 = 55.5. Hence, the upper outer fence is: Q_3 + 55.5 = 86

Since 62 and 75 are both less than 86, they are within the upper outer fence and are called mild outliers.

3.101 The ranked data are: 3 5 5 6 8 10 14 15 16 17 17 19 21 22 23 25 30 31 31 34

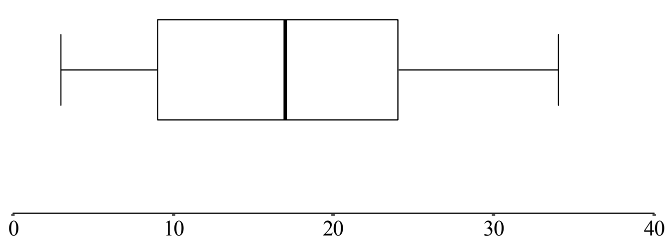
Median = (17 + 17)/2 = 17, Q_1 = (8 + 10)/2 = 9, and Q_3 = (23 + 25)/2 = 24,

IQR = $Q_3 - Q_1$ = 24 – 9 = 15, 1.5 × IQR = 1.5 × 15 = 22.5,

Lower inner fence = Q_1 – 22.5 = 9 – 22.5 = –13.5,

Upper inner fence = Q_3 + 22.5 = 24 + 22.5 = 46.5

The smallest and the largest values within the two inner fences are 3 and 34, respectively. The data set contains no outliers.



The data are nearly symmetric.

3.102 The ranked data are: 356 422 430 468 494 533 572 600 604 617 625
628 639 647 690 702 728 747 749 772 797 805

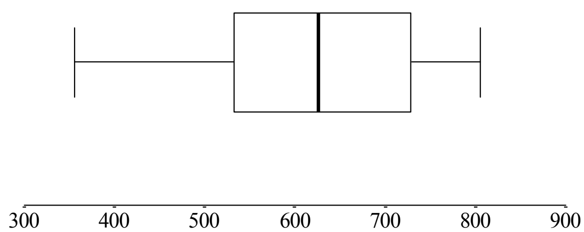
Median = (625 + 628)/2 = 626.5, Q_1 = 533, and Q_3 = 728

IQR = $Q_3 - Q_1$ = 728 – 533 = 195, 1.5 × IQR = 1.5 × 195 = 292.5,

Lower inner fence = Q_1 – 292.5 = 533 – 292.5 = 240.5,

Upper inner fence = Q_3 + 292.5 = 728 + 292.5 = 1020.5

The smallest and the largest values within the two inner fences are 356 and 805, respectively. The data set contains no outliers.



The data are skewed slightly to the left.

3.103 The ranked data are: 22.5 25.0 26.8 27.0 29.8 51.2 64.0 80.3 94.4 97.7 112.0 261.7

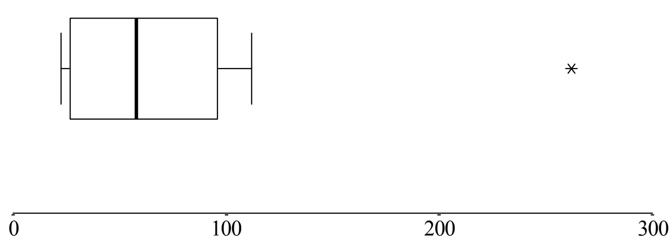
Median = (51.2 + 64.0)/2 = 57.6, Q_1 = (26.8 + 27.0)/2 = 26.9, and Q_3 = (94.4 + 97.7)/2 = 96.05,

IQR = $Q_3 - Q_1$ = 96.05 – 26.9 = 69.15, 1.5 × IQR = 1.5 × 69.15 = 103.725,

Lower inner fence = Q_1 – 103.725 = 26.9 – 103.725 = –76.825,

Upper inner fence = Q_3 + 103.725 = 96.05 + 103.725 = 199.775

The smallest and largest values within the two inner fences are 22.5 and 112, respectively. The value 261.7 is an outlier.



The data are skewed to the right.

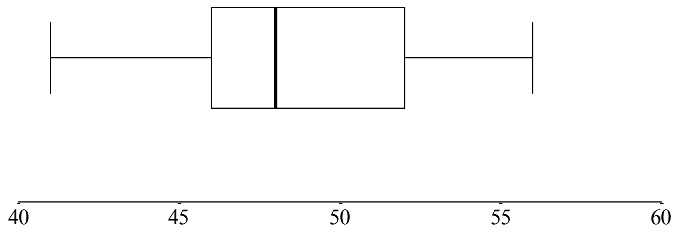
3.104 The ranked data are: 41 42 43 44 44 45 46 46 47 47 48 48
48 49 50 50 51 51 52 52 52 53 53 54 56

Median = 48, Q_1 = (45 + 46)/2 = 45.5, and Q_3 = (52 + 52) = 52,

IQR = $Q_3 - Q_1$ = 52 – 45.5 = 6.5, 1.5 × IQR = 1.5 × 6.5 = 9.75,

Lower inner fence = Q_1 – 9.75 = 45.5 – 9.75 = 35.75, Upper inner fence = Q_3 + 9.75 = 52 + 9.75 = 61.75

The smallest and largest values within the two inner fences are 41 and 56, respectively. There are no outliers.



The data are skewed slightly to the right.

3.105 The ranked data are:

318	336	337	339	362	363	366	369	372	375
378	381	384	385	386	387	390	393	395	403
405	409	417	431	433	434	438	444	461	480

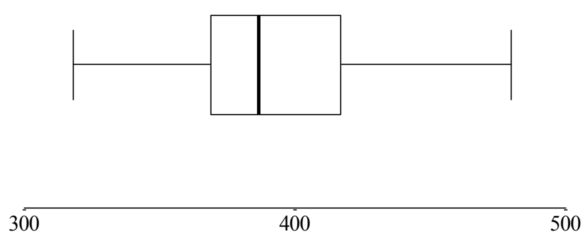
Median = (386 + 387)/2 = 386.5, Q_1 = 369, and Q_3 = 417,

IQR = $Q_3 - Q_1$ = 417 – 369 = 48, 1.5 × IQR = 1.5 × 48 = 72,

Lower inner fence = Q_1 – 72 = 369 – 72 = 297,

Upper inner fence = Q_3 + 72 = 417 + 72 = 489

The smallest and largest values within the two inner fences are 318 and 480, respectively. There are no outliers.



The data are skewed slightly to the right.

3.106 The ranked data are: 3 5 6 6 7 9 9 10 11 12 14 15

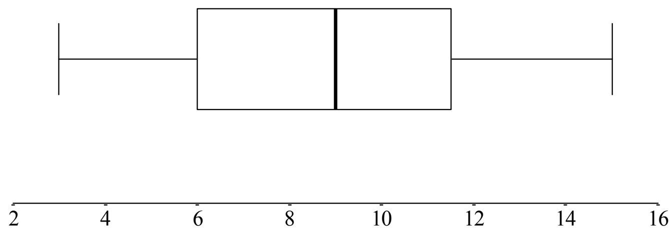
Median = (9 + 9)/2 = 9, Q_1 = (6 + 6)/2 = 6, and Q_3 = (11 + 12)/2 = 11.5,

IQR = $Q_3 - Q_1$ = 11.5 – 6 = 5.5, 1.5 × IQR = 1.5 × 5.5 = 8.25,

Lower inner fence = Q_1 – 8.25 = 6 – 8.25 = –2.25,

Upper inner fence = Q_3 + 8.25 = 11.5 + 8.25 = 19.75

The smallest and largest values within the two inner fences are 3 and 15, respectively. There are no outliers.



The data are nearly symmetric.

3.107 The ranked data are: 20 21 22 23 23 23 23 24 25 26 26 27 27 27 27
28 28 28 29 29 31 31 31 32 33 33 33 34 35 35

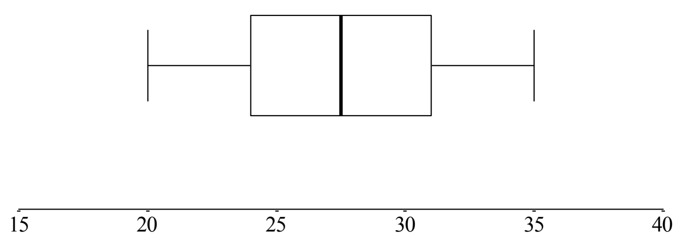
Median = (27 + 28)/2 = 27.5, Q_1 = 24, and Q_3 = 31,

IQR = $Q_3 - Q_1$ = 31 – 24 = 7, 1.5 × IQR = 1.5 × 7 = 10.5,

Lower inner fence = Q_1 – 10.5 = 24 – 10.5 = 13.5,

Upper inner fence = Q_3 + 10.5 = 31 + 10.5 = 41.5

The smallest and largest values within the two inner fences are 20 and 35, respectively. There are no outliers.



The data are nearly symmetric.

3.108 The ranked data are: 3 3 4 5 5 6 7 7 8 8 8 9 9 10 10 11 11 12 12 16

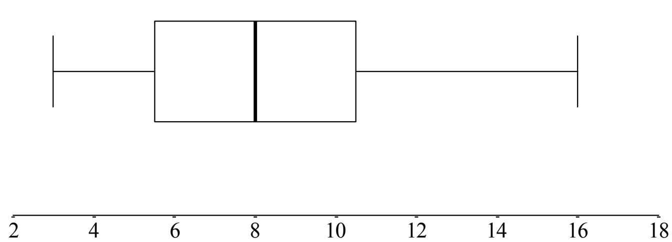
Median = (8 + 8)/2 = 8, Q_1 = (5 + 6)/2 = 5.5, and Q_3 = (10 + 11)/2 = 10.5,

IQR = $Q_3 - Q_1$ = 10.5 – 5.5 = 5, 1.5 × IQR = 1.5 × 5 = 7.5,

Lower inner fence = Q_1 – 7.5 = 5.5 – 7.5 = –2,

Upper inner fence = Q_3 + 7.5 = 10.5 + 7.5 = 18

The smallest and largest values within the two inner fences are 3 and 16, respectively. There are no outliers.



The data are skewed to the right.

Supplementary Exercises

3.109 a. $\bar{x} = (\sum x)/n$ = 1065/10 = \$106.5 thousand

Median = value of the 5.5th term in ranked data = (74 + 78)/2= \$76 thousand

b. Yes, 382 is an outlier. After dropping this value,

$\bar{x} = (\sum x)/n$ = 683/9 = \$75.89 thousand

Median = value of the 5th term in ranked data = \$74 thousand

The value of the mean changes by a larger amount.

c. The median is a better summary measure for these data since it is influenced less by outliers.

3.110 a. $\bar{x} = (\sum x)/n$ = 46.22/10 = 4.622 seconds

Median = value of the 5.5th term in ranked data = (4.26 + 4.74)/2 = 4.5 seconds

This data set has no mode as no value occurs more than once.

b. Range = Largest value – Smallest value = 7.20 – 2.80 = 4.40 seconds

$$s^2 = \frac{\sum x^2 - \frac{(\sum x)^2}{n}}{n-1} = \frac{228.5904 - \frac{(46.22)^2}{10}}{10-1} = 1.6624 \qquad s = \sqrt{1.6624} = 1.29 \text{ seconds}$$

3.111 a. $\bar{x} = (\sum x)/n$ = 19,736/10 = 1973.6 points

Median = value of the 5.5th term in ranked data = (1857 + 1978)/2 = 1917.5 points

This data set has no mode as no value occurs more than once.

b. Range = Largest value – Smallest value = 2323 – 1779 = 544 points

$$s^2 = \frac{\sum x^2 - \frac{(\sum x)^2}{n}}{n-1} = \frac{39{,}329{,}734 - \frac{(19{,}736)^2}{10}}{10-1} = 42{,}084.9333$$

$$s = \sqrt{42{,}084.9333} = 205.15 \text{ points}$$

3.112 a. $\bar{x} = (\sum x)/n$ = 88/12 = 7.33 citations

Median = value of the 6.5th term in ranked data = (7 + 8)/2 = 7.5 citations

Mode = 4, 7, and 8 citations

b. Range = Largest value – Smallest value = 14 – 0 = 14 citations

$$s^2 = \frac{\sum x^2 - \frac{(\sum x)^2}{n}}{n-1} = \frac{834 - \frac{(88)^2}{12}}{12-1} = 17.1515 \qquad s = \sqrt{17.1515} = 4.14 \text{ citations}$$

c. The values of the summary measures in parts a and b are sample statistics because the data are based on a sample of 12 drivers.

3.113

Rainfall	Number of Cities	m	mf	m^2f
0 to less than 2	6	1	6	6
2 to less than 4	10	3	30	90
4 to less than 6	20	5	100	500
6 to less than 8	7	7	49	343
8 to less than 10	4	9	36	324
10 to less than 12	3	11	33	363
	$N = \sum f = 50$		$\sum mf = 254$	$\sum m^2f = 1626$

$\bar{x} = (\sum mf)/n = 254/50 = 5.08$ inches

$$s^2 = \frac{\sum m^2 f - \frac{(\sum mf)^2}{n}}{n-1} = \frac{1626 - \frac{(254)^2}{50}}{50-1} = 6.8506 \qquad s = \sqrt{6.8506} = 2.62 \text{ inches}$$

The values of these summary measures are sample statistics since they are based on a sample of 50 cities.

3.114

Time	Number of Students	m	mf	m^2f
0 to less than 4	1	2	2	4
4 to less than 8	7	6	42	252
8 to less than 12	15	10	150	1500
12 to less than 16	18	14	252	3528
16 to less than 20	6	18	108	1944
20 to less than 24	3	22	66	1452
	$N = \sum f = 50$		$\sum mf = 620$	$\sum m^2f = 8680$

$\bar{x} = (\sum mf)/n = 620/50 = 12.4$ minutes

$$s^2 = \frac{\sum m^2 f - \frac{(\sum mf)^2}{n}}{n-1} = \frac{8680 - \frac{(620)^2}{50}}{50-1} = 20.2449 \qquad s = \sqrt{20.2449} = 4.50 \text{ minutes}$$

The values of these summary measures are sample statistics since they are based on a sample of 50 students.

3.115 a i. Each of the two values is 40 minutes from $\mu = 200$. Hence,

$k = 40/20 = 2$ and $1 - \frac{1}{k^2} = 1 - \frac{1}{(2)^2} = 1 - .25 = .75$ or 75%.

Thus, at least 75% of the students will learn the basics in 160 to 240 minutes.

ii. Each of the two values is 60 minutes from $\mu = 200$. Hence,

$k = 60/20 = 3$ and $1 - \frac{1}{k^2} = 1 - \frac{1}{(3)^2} = 1 - .11 = .89$ or 89%.

Thus, at least 89% of the students will learn the basics in 140 to 260 minutes.

b. $1 - \frac{1}{k^2} = .75$ gives $\frac{1}{k^2} = 1 - .75 = .25$ or $k^2 = \frac{1}{.25}$, so $k = 2$

$\mu - 2\sigma$ = 200 – 2(20) = 160 minutes and $\mu + 2\sigma$ = 200 + 2(20) = 240 minutes

Thus, the required interval is 160 to 240 minutes.

3.116 a. i. Each of the two values is 900 hours from μ = \$1750 hours. Hence,

k = 900/450 = 2 and $1 - \frac{1}{k^2} = 1 - \frac{1}{(2)^2}$ = 1 – .25 = .75 or 75%.

Thus, at least 75% of Americans watched between 850 and 2650 hours of television.

ii. Each of the two values is 1350 hours from μ = 1750 hours. Hence,

k = 1350/450 = 3 and $1 - \frac{1}{k^2} = 1 - \frac{1}{(3)^2}$ = 1 – .11 = .89 or 89%.

Thus, at least 89% of households watched between 400 and 3100 hours of television.

b. $1 - \frac{1}{k^2}$ = .84 gives $\frac{1}{k^2}$ = 1 – .84 = .16 or $k^2 = \frac{1}{.16}$, so k = 2.5.

$\mu - 2.5\sigma$ = 1750 – 2.5(450) = 625 hours and $\mu + 2.5\sigma$ = 1750 + 2.5(450) = 2875 hours

Thus, the required interval is 625 to 2875 hours.

3.117 μ = 200 minutes and σ = 20 minutes

a. i. The interval 180 to 220 minutes is $\mu - \sigma$ to $\mu + \sigma$. Thus, approximately 68% of the students will learn the basics in 180 to 220 minutes.

ii. The interval 160 to 240 minutes is $\mu - 2\sigma$ to $\mu + 2\sigma$. Hence, approximately 95% of the students will learn the basics in 160 to 240 minutes.

b. $\mu - 3\sigma$ = 200 – 3(20) = 140 minutes and $\mu + 3\sigma$ = 200 + 3(20) = 260 minutes. The interval that contains the learning time of 99.7% of the students is 140 to 260 minutes.

3.118 μ = \$134,000 and σ = \$12,000

a. i. The interval \$98,000 to \$170,000 is $\mu - 3\sigma$ to $\mu + 3\sigma$. Thus, approximately 99.7% of all such employees have salaries between \$98,000 and \$170,000.

ii. The interval \$110,000 to \$158,000 is $\mu - 2\sigma$ to $\mu + 2\sigma$. Thus, approximately 95% of all such employees have salaries between \$110,000 and \$158,000.

b. $\mu - \sigma$ = 134,000 – 12,000 = \$122,000 and $\mu + \sigma$ = 134,000 + 12,000 = \$146,000. The interval that contains the salaries of 68% of all such CPAs is \$122,000 to \$146,000.

3.119 The ranked data are: 56 59 60 68 74 78 84 97 107 382

a. The three quartiles are Q_1 = 60, Q_2 = (74 + 78)/2 = 76, and Q_3 = 97

IQR = $Q_3 - Q_1$ = 97 – 60 = 37

The value 60 lies at Q_1, which indicates that it is on the line that separates the bottom 25% group in the (ranked) data set from the second 25% group.

b. $kn/100 = 70(10)/100 = 7$
Thus, the 70th percentile occurs at the seventh term in the ranked data, which is 84. Therefore, $P_{70} = 84$. This means that about 70% of the values in the data set are smaller than or equal to 84.

c. Seven values in the given data are smaller than 97. Hence, the percentile rank of 97 = (7/10) × 100 = 70%. This means approximately 70% of the values in the data set are less than 97.

3.120 The ranked data are: 1779 1788 1791 1817 1857 1978 1989 2164 2250 2323

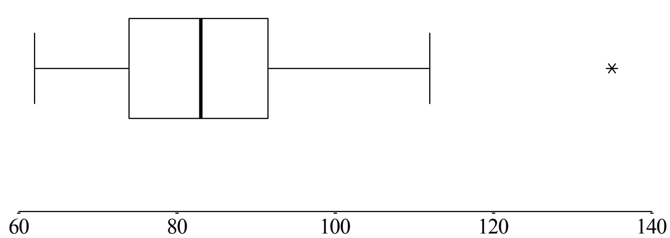
a. The three quartiles are $Q_1 = 1791$, $Q_2 = (1857 + 1978)/2 = 1917.5$, and $Q_3 = 2164$
$IQR = Q_3 - Q_1 = 2164 - 1791 = 373$
The value 1978 is between Q_2 and Q_3, which indicates it is in the third 25% group from the bottom in this (ranked) data set.

b. $kn/100 = 70(10) / 100 = 7$. Thus, the 70th percentile occurs at the seventh term in the ranked data, which is 1989. Therefore, $P_{70} = 1989$. Thus, approximately 70% of these players scored less than or equal to 1989 total points during the 2007-08 regular season.

c. One value in the given data is less than 1788. Hence, the percentile rank of 1788 = (1/10) × 100 = 10%. Thus, 10% of the players scored fewer than 1788 total points during the 2007-08 regular season.

3.121 The ranked data are: 62 67 72 73 75 77 81 83 84 85 90 93 107 112 135

Median = 83, $Q_1 = 73$, and $Q_3 = 93$,
$IQR = Q_3 - Q_1 = 93 - 73 = 20$, $1.5 \times IQR = 1.5 \times 20 = 30$,
Lower inner fence = $Q_1 - 30 = 73 - 30 = 43$,
Upper inner fence = $Q_3 + 30 = 93 + 30 = 123$
The smallest and largest values within the two inner fences are 62 and 112, respectively. The value 135 is an outlier.

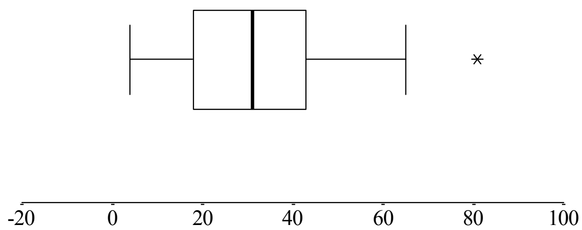


The data are skewed to the right.

3.122 The ranked data are: 4 8 9 16 18 21 23 24 30 32 33 38 42 43 44 55 65 81

Median = (30 + 32)/2 = 31, $Q_1 = 18$, and $Q_3 = 43$,
$IQR = Q_3 - Q_1 = 43 - 18 = 25$, $1.5 \times IQR = 1.5 \times 25 = 37.5$,
Lower inner fence = $Q_1 - 12 = 18 - 37.5 = -19.5$
Upper inner fence = $Q_3 + 12 = 43 + 37.5 = 80.5$

The smallest and largest values in the data set within the two inner fences are 4 and 65, respectively. The value 81 is an outlier.



The data are skewed to the right.

3.123 Let y = Melissa's score on the final exam. Then, her grade is $\frac{75+69+87+y}{5}$. To get a B, she needs this to be at least 80. So we solve,

$$80 = \frac{75+69+87+y}{5}$$

$5(80) = 75 + 69 + 87 + y$

$400 = 231 + y$

$y = 169$

Thus, the minimum score that Melissa needs on the final exam in order to get a B grade is 169 out of 200 points.

3.124 a. Let y = amount that Jeffery suggests. Then, to insure the outcome Jeffery wants, we need

$$\frac{y+12{,}000(5)}{6} = 20{,}000$$

$y + 12{,}000(5) = 6(20{,}000)$

$y + 60{,}000 = 120{,}000$

$y = 60{,}000$

So, Jeffery would have to suggest \$60,000 be awarded to the plaintiff.

b. To prevent a juror like Jeffery from having an undue influence on the amount of damage to be awarded to the plaintiff, the jury could revise its procedure by throwing out any amounts that are outliers and then recalculate the mean, or by using the median, or by using a trimmed mean.

3.125 a. Since $\bar{x} = (\sum x)/n$, we have $76 = (\sum x)/5$, so $\sum x = 5(76) = 380$ inches. If we replace the tallest player by a substitute who is two inches taller, the sum of the new heights is 380 + 2 = 382 inches. Thus, the new mean is $\bar{x}$ = 382/5 = 76.4 inches. Since Range = Largest value – Smallest value, and the largest value has increased by two while the smallest value is unchanged, the range has increased by two. Thus, the new range is 11 + 2 = 13 inches. The median is the height of the third player (if their heights are ranked) and this does not change. So, the median remains 78 inches.

b. If we replace the tallest player by a substitute who is four inches shorter, then by reasoning similar to that in part a, we have a new mean of $\bar{x}$ = 376/5 = 75.2 inches. You cannot determine the new median or range with only the information given. We do not know how the new player's height compares to the rest of the players on the team; we have no knowledge of whether the substitute is now the tallest player or not.

3.126 a. To calculate how much time the trip requires, divide miles driven by miles per hour for each 100 mile segment. Then, time = 100/52 + 100/65 + 100/58 = 1.92 + 1.54 + 1.72 = 5.18 hours.

b. Linda's average speed for the 300 mile trip is not equal to (52 + 65 + 58)/3 = 58.33 mph. This would assume that she spent an equal amount of time on each 100 mile segment, which is not true, because her average speed is different on each segment. Linda's average speed for the entire 300 mile trip is given by (miles driven)/(elapsed time) = 300/5.18 = 57.92 mph.

3.127 The mean price per barrel of oil purchased in that week is

Mean = [(1000)(51) + (200)(64) + (100)(70)]/1300 = 70,800/1300 = $54.46 per barrel

3.128 The method of calculating the mean is wrong in this case because it does not take into account the fact that the homeowner bought different amounts of heating oil in the four deliveries. The correct method of calculating the mean is

$\mu = (\sum mf)/n$ = [(2.22)(209) + (2.34)(182) + (2.41)(157) + (2.43)(149)]/(209 + 182 + 157 + 149)

= $2.34 per gallon.

3.129 a. Mean = (9.4 + 9.5 + 9.5 + 9.5 + 9.6)/5 = 9.5

b. The percentage of trimmed mean is (2/7 × 100)/2 ≈ 28.6/2 = 14.3 % since we dropped two of the seven values.

c. Suppose gymnast B has the following scores: 9.4, 9.4, 9.5, 9.5, 9.5, 9.5, and 9.9. Then, the mean for the gymnast B is μ = (9.4 + 9.4 + 9.5 + 9.5 + 9.5 + 9.5 + 9.9)/7 = 9.5286, and the mean for gymnast A is μ = (9.4 + 9.7 + 9.5 + 9.5 + 9.4 + 9.6 + 9.5)/7 = 9.5143. So, Gymnast B would win if all seven scores were counted. The trimmed mean of B is (9.4 + 9.5 + 9.5 + 9.5 + 9.5)/5 = 9.4800. This is less than the trimmed mean for A (9.500), so gymnast A would win using the trimmed mean.

3.130 a. Total amount spent per month by the 2000 shoppers = (14)(8)(1100) + (18)(11)(900) = $301,400

b. Total number of trips per month by the 2000 shoppers = (8)(1100) + (11)(900) = 18,700

Mean number of trips per month per shopper = 18,700/2000 = 9.35 trips

c. Mean amount spent per person per month by shoppers aged 12-17 = 301,400/2000 = $150.70

3.131 a. For people age 30 and under, we have the following death rates from heart attack:

Country A: $\dfrac{\text{number of deaths}}{\text{population}} \times 1000 = \dfrac{1}{40} \times 1000 = 25$

Country B: $\dfrac{\text{number of deaths}}{\text{population}} \times 1000 = \dfrac{.5}{25} \times 1000 = 20$

So the death rate for people 30 and under is lower in Country B.

b. For people age 31 and older, the death rates from heart attack are as follows:

Country A: $\dfrac{\text{number of deaths}}{\text{population}} \times 1000 = \dfrac{2}{20} \times 1000 = 100$

Country B: $\dfrac{\text{number of deaths}}{\text{population}} \times 1000 = \dfrac{3}{35} \times 1000 = 85.7$

Thus, the death rate for Country A is greater than that for Country B for people age 31 and older.

c. The overall death rates are as follows:

Country A: $\dfrac{\text{number of deaths}}{\text{population}} \times 1000 = \dfrac{3}{60} \times 1000 = 50$

Country B: $\dfrac{\text{number of deaths}}{\text{population}} \times 1000 = \dfrac{3.5}{60} \times 1000 = 58.3$

Thus, overall the death rate for country A is *lower* than the death rate for Country B.

d. In both countries people age 30 and under have a lower percentage of death due to heart attack than people age 31 and over. Country A has 2/3 of its population age 30 and under while more than 1/2 of the people in Country B are age 31 and over. Thus, more people in Country B than in County A fall into the higher risk group which drives up Country B's overall death rate from heart attacks.

3.132 Total distance for the first 100 students = (100)(8.73) = 873 miles

Total distance for all 103 students = 873 + 11.5 + 7.6 + 10.0 = 902.1 miles

Mean distance for all 103 students = 902.1/103 = 8.76 miles

3.133 $\mu = 70$ minutes and $\sigma = 10$

a. Using Chebyshev's theorem, we need to find k so that

$1 - \dfrac{1}{k^2} = .50$ gives $\dfrac{1}{k^2} = 1 - .50 = .50$ or $k^2 = \dfrac{1}{.50} = 2$, so $k = \sqrt{2} \approx 1.4$.

Thus, at least 50% of the scores are within 1.4 standard deviations of the mean.

b. Using Chebyshev's theorem, we first find k so that at least 1 –.20 = .80 of the scores are within k standard deviations of the mean.

$1 - \dfrac{1}{k^2} = .80$ gives $\dfrac{1}{k^2} = 1 - .80 = .20$ or $k^2 = \dfrac{1}{.20} = 5$, so $k = \sqrt{5} \approx 2.2$.

Thus, at least 80% of the scores are within 2.2 standard deviations of the mean, but this means that at most 10% of the scores are greater than 2.2 standard deviations above the mean.

3.134 a. Since we are dealing with a bell-shaped distribution and we know that 16% of all students scored above 85, which is $\mu + 15$, we must also have that 16% of all students scored below $\mu - 15 = 55$. Therefore, the remaining 68% of students scored between 55 and 85. By the empirical rule, we know that approximately 68% of the scores fall in the interval $\mu - \sigma$ to $\mu + \sigma$, so we have $\mu - \sigma = 70 - \sigma = 55$ and $\mu + \sigma = 70 + \sigma = 85$. Thus, $\sigma = 15$.

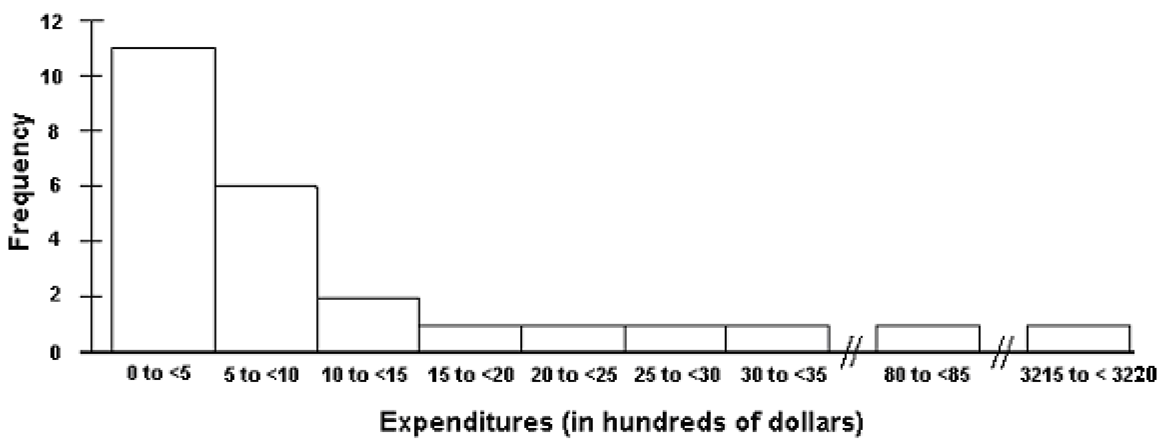
b. We know that 95% of the scores are between 60 and 80 and that $\mu = 70$. By the empirical rule, 95% of the scores fall in the interval $\mu - 2\sigma$ to $\mu + 2\sigma$. Then

$60 = \mu - 2\sigma = 70 - 2\sigma$ and $80 = \mu + 2\sigma = 70 + 2\sigma$. Therefore, $10 = 2\sigma$ and $\sigma = 5$.

3.135 a.

$\bar{x}$ = \$13,872	Q_1 = \$50	s^2 = 4,110,408,767
Median = \$500	Q_3 = \$1400	s = \$64,112.47
Mode = \$0	IQR = \$1350	Lowest = \$0
	Range = \$321,500	Highest = \$321,500

Below is the histogram for the given data.



The vacation expenditures are strongly skewed to the right. Most of the expenditures are relatively small (\$3400 or less) but there are two extreme outliers (\$8200 and \$321,500).

b. Neither the mode (\$0) nor the mean (\$13,872) are typical of these expenditures. Thus, the median (\$500) is the best indicator of the average family's vacation expenditures.

3.136 a. Mean = \$600.35, Median = \$90, and Mode = \$0

b. The mean is the largest.

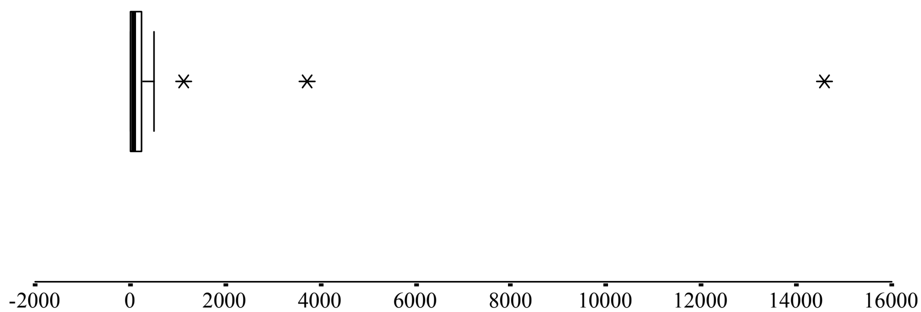
c. Q_1 = \$0, Q_3 = \$272.50, IQR = \$272.50, 1.5 × IQR = \$408.75

Lower Inner fence is $Q_1 - 408.75 = 0 - 408.75 = -408.75$

Upper Inner Fence is $Q_3 + 408.75 = 272.50 + 408.75 = 681.25$

The largest and smallest values within the two inner fences are 0 and 501, respectively. There are three outliers at 1127, 3709 and 14,589.

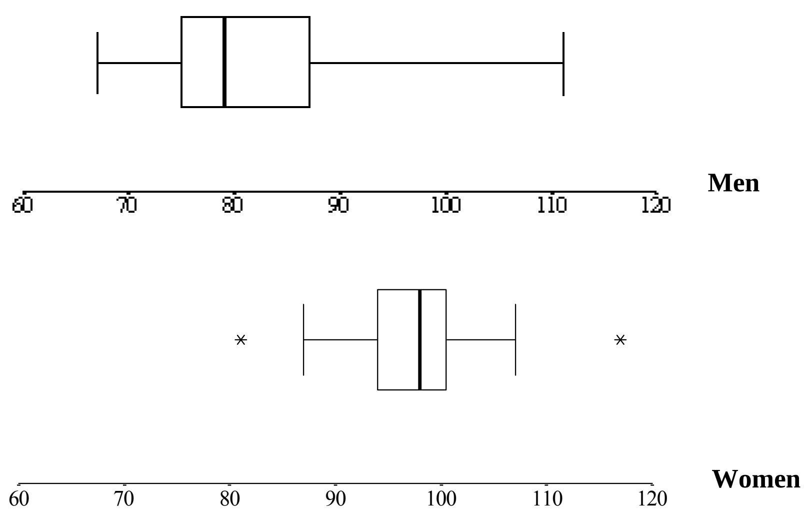
Below is the box–and–whisker plot for the given data.



The data are strongly skewed to the right.

b. Because the data are skewed to the right, the insurance company should use the mean when considering the center of the data as it is more affected by the extreme values. The insurance company would want to use a measure that takes into consideration the possibility of extremely large losses.

3.137 a.



The box–and–whisker plots show that the men's scores tend to be lower and more varied than the women's scores. The men's scores are skewed to the right, while the women's are more nearly symmetric.

b. **Men**

$\bar{x}$ = 82

Median = 79

Modes = 75, 79, and 92

Range = 45

Women

$\bar{x}$ = 97.53

Median = 98

Modes = 94 and 100

Range = 36

$s^2 = 145.8750$	$s^2 = 71.2667$
$s = 12.08$	$s = 8.44$
$Q_1 = 73.5$	$Q_1 = 94$
$Q_3 = 89.5$	$Q_3 = 101$
IQR = 16	IQR = 7

These numerical measures confirm the observations based on the box–and–whisker plots.

3.138 a. Since $\bar{x} = (\sum x)/n$, we have $n = (\sum x)/\bar{x} = 12{,}372/51.55 = 240$ pieces of luggage.

b. Since $\bar{x} = (\sum x)/n$, we have $(\sum x) = n\bar{x} = (7)(81) = 567$ points. Let x = seventh student's score. Then, $x + 81 + 75 + 93 + 88 + 82 + 85 = 567$. Hence, $x + 504 = 567$, so $x = 567 - 504 = 63$.

3.139 a. The total enrollment in the 25 freshman engineering classes is (24)(25) + 150 = 750. Then, the mean size of these 25 classes is 750/25 = 30.

b. Each student attends five classes with total enrollment of 25 + 25 + 25 + 25 + 150 = 250. Then, the mean size of the class is 250/5 = 50.

The means in parts a and b are not equal because:

1) From the college's point of view, the large class of 150 is just one of 25 classes, so its influence on the mean is strongly offset by the 24 small classes. This leads to a relatively small mean of 30 students per class.
2) From the point of view of each student, the larger class is one of just five, so it has a stronger influence on the mean. This results in a larger mean of 50.

3.140 For all students:

$n = 44$, $\sum x = 6597$, $\sum x^2 = 1{,}030{,}639$, and Median = 147.50 pounds

$\bar{x} = (\sum x)/n = 6597/44 = 149.93$ pounds

$$s = \sqrt{\frac{\sum x^2 - \frac{(\sum x)^2}{n}}{n-1}} = \sqrt{\frac{1{,}030{,}639 - \frac{(6597)^2}{44}}{44-1}} = 31.0808 \text{ pounds}$$

For men only:

$n = 22$, $\sum x = 3848$, $\sum x^2 = 680{,}724$ and Median = 179 pounds

$\bar{x} = (\sum x)/n = 3848/22 = 174.91$ pounds

$$s = \sqrt{\frac{\sum x^2 - \frac{(\sum x)^2}{n}}{n-1}} = \sqrt{\frac{680{,}724 - \frac{(3848)^2}{22}}{22-1}} = 19.1160 \text{ pounds}$$

For women:

$n = 22$, $\sum x = 2749$, $\sum x^2 = 349{,}915$ and Median = 123 pounds

$\bar{x} = (\sum x)/n = 2749/22 = 124.95$ pounds

$$s = \sqrt{\frac{\sum x^2 - \frac{(\sum x)^2}{n}}{n-1}} = \sqrt{\frac{349{,}915 - \frac{(2749)^2}{22}}{22-1}} = 17.4778 \text{ pounds}$$

In this case, the median may be more informative than the mean, since it is less influenced by extremely high or low weights. As one might expect, the mean and median weights for men are higher than those of women. For the entire group, the mean and median weights are about midway between the corresponding values for men and women. The standard deviations are roughly the same for men and women. The standard deviation for the whole group is much larger than for men or women only, due to the fact that it includes the lower weights of women and the heavier weights of men.

3.141 $\mu = 6$ inches and $\sigma = 2$ inches

a. Each of the two values is 3 inches from $\mu = 6$ inches. Hence,

$k = 3/2 = 1.5$ and $1 - \frac{1}{k^2} = 1 - \frac{1}{(1.5)^2} = 1 - .444 = .556$ or 55.6%.

Thus, at least 55.6% of the fish are between 3 and 9 inches in length.

b. $1 - \frac{1}{k^2} = .84$ gives $\frac{1}{k^2} = 1 - .84 = .16$ or $k^2 = \frac{1}{.16}$, so $k = 2.5$

$\mu - 2.5\sigma = 6 - 2.5(2) = 1$ inch and $\mu + 2.5\sigma = 6 + 2.5(2) = 11$ inches

Thus, the required interval is 1 to 11 inches.

c. 100 – 36 = 64% of the fish have lengths *inside* the required interval. Then

$1 - \frac{1}{k^2} = .64$ gives $\frac{1}{k^2} = 1 - .64 = .36$ or $k^2 = \frac{1}{.36}$, so $k = 1.67$

$\mu - 1.67\sigma = 6 - 1.67(2) = 2.66$ inches and $\mu + 1.67\sigma = 6 + 1.67(2) = 9.34$ inches

Thus, the required interval is 2.66 to 9.34 inches.

3.142 The given data are: 3 6 9 12 18 15 11 10 15 25 21 26 38 41 62

Ranked data are: 3 6 9 10 11 12 15 15 18 21 25 26 38 41 62

a. $\bar{x} = 20.80$ thousand miles, Median = 15 thousand miles, and Mode = 15 thousand miles

b. Range = 59 thousand miles, $s^2 = 249.03$, $s = 15.78$ thousand miles

c. $Q_1 = 10$ thousand miles and $Q_3 = 26$ thousand miles

d. $IQR = Q_3 - Q_1 = 26 - 10 = 16$ thousand miles

Since the interquartile range is based on the middle 50% of the observations it is not affected by outliers. The standard deviation, however, is strongly affected by outliers. Thus, the interquartile range is preferable in applications in which a measure of variation is required that is unaffected by extreme values.

3.143 a. The mean, median, and standard deviations of the weights of males and females in grams are:

Men	**Women**
$\bar{x}$ = 76,188.65	$\bar{x}$ = 54,428.95
Median = 77,970.61	Median = 53,577.57
s = 8326.7209	s = 7613.1618

The mean, median, and standard deviations of the weights of males and females in stones are:

Men	**Women**
$\bar{x}$ = 12.49	$\bar{x}$ = 8.93
Median = 12.79	Median = 8.79
s = 1.3654	s = 1.2484

b. Converting the answers from Problem 3.140 to grams yields:

Men	**Women**
$\bar{x}$ = 76,189.05	$\bar{x}$ = 54,426.97
Median = 77,970.61	Median = 53,577.57
s = 8326.7384	s = 7613.1549

Converting the answers from Problem 3.140 to stones yields:

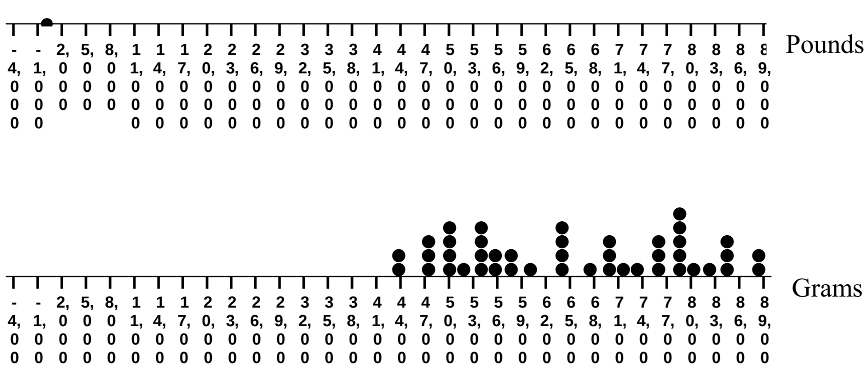
Men	**Women**
$\bar{x}$ = 12.49	$\bar{x}$ = 8.93
Median = 12.79	Median = 8.79
s = 1.3686	s = 1.2484

The answers are the same as those from part a with the exception of rounding error.

c. When converting from a larger unit to a smaller unit, the summary measures get larger by the conversion factor. When converting from a smaller unit to a larger unit, the summary measures get smaller by the conversion factor.

d. The distribution in units of pounds has more variability than that of stones. This is so because we converted from a smaller unit to a larger unit; hence, the standard deviation was reduced by the amount of the conversion factor (s for stones = s for pounds/14).

e.



The distribution in units of grams has more variability than that of pounds. This is so because we converted from a larger unit to a smaller unit; hence, the standard deviation was increased by the amount of the conversion factor (s for grams = s for pounds × 435.59). Because of the large difference in units, retaining the same scale for the stacked plots displays all of the data values for pounds in one location with no variability visible.

3.144 $\bar{x}$ = 49.012 hours and s = 5.080 hours

a. For 75%: $1 - \frac{1}{k^2} = .75$ gives $\frac{1}{k^2} = 1 - .75 = .25$ or $k^2 = \frac{1}{.25}$, so $k = 2$

$\bar{x} - 2s = 49.012 - 2(5.080) = 38.85$ and $\bar{x} + 2s = 49.012 + 2(5.080) = 59.17$

Thus, the required interval is 38.85 to 59.17 hours.

For 88.89%: $1 - \frac{1}{k^2} = .8889$ gives $\frac{1}{k^2} = 1 - .8889 = .1111$ or $k^2 = \frac{1}{.1111}$, so $k \approx 3$

$\bar{x} - 3s = 49.012 - 3(5.080) = 33.77$ and $\bar{x} + 3s = 49.012 + 3(5.080) = 64.25$

Thus, the required interval is 33.77 to 64.25 hours.

For 93.75%: $1 - \frac{1}{k^2} = .9375$ gives $\frac{1}{k^2} = 1 - .9375 = .0625$ or $k^2 = \frac{1}{.0625}$, so $k = 4$

$\bar{x} - 4s = 49.012 - 4(5.080) = 28.69$ and $\bar{x} + 4s = 49.012 + 4(5.080) = 69.33$

Thus, the required interval is 28.69 to 69.33 hours.

b. 100% of the data falls into each of the intervals calculated in part a.

$\bar{x} - s = 49.012 - (5.080) = 43.93$ and $\bar{x} + s = 49.012 + (5.080) = 54.09$

56% of the observations fall within one standard deviation of the mean.

c. The endpoints provided by Chebyshev's Theorem are not useful since each of these intervals contain all of the data points.

d. With the change in the sample mean and standard deviations, the required intervals are 35.41 to 63.81 hours for 75%, 28.31 to 70.91 hours for 88.89%, and 21.21 to 78.01 hours for 93.75%. Each of these intervals contains 98% of the data which is a small change from 100%. The only value not included in these intervals is the outlier at 84.4 hours. Now, 80% of the observations fall within one standard deviation of the mean (between 42.51 and 56.71). This is a relatively large increase from the 56% found in part b.

e. Using the upper endpoint of 58.7, we have 58.7 = 49.012 + k(5.08). Then k = 1.907. We would have to go 1.907 standard deviations about the mean to capture all 50 data values. By Chebyshev's Theorem, the lower bound for the percentage of data that would fall in this interval is

$$1 - \frac{1}{k^2} = 1 - \frac{1}{(1.907)^2} = 1 - .2750 = .7250, \text{ or } 72.50\%.$$

3.145 This golfer's score was not an outlier; therefore, her score must be less than the value of the upper inner fence. From Exercise 3.137, Q_3 = 101 and IQR = 7. Then, 1.5 × IQR = 1.5 × 7 = 10.5 and Q_3 + 10.5 =

111.5. Since this golfer had the uniquely highest score, and the next highest score was 107, she shot between 108 and 111.

3.146 $\sum x/10 = -645.5/10 = -64.55$. This number represents the average returns for the stocks of these 10 companies for October 2008.

Self–Review Test

1. b	**2.** a and d	**3.** c	**4.** c	**5.** b	**6.** b	**7.** a
8. a	**9.** b	**10.** a	**11.**b	**12.** c	**13.** a	**14.** a

15. $n = 10$, $\sum x = 109$, and $\sum x^2 = 1775$

$\bar{x} = (\sum x)/\, n = 109/10 = 10.9$ times

Median = value of the 5.5th term in ranked data = (7 + 9)/2 = 8 times

Mode = 6 times

Range = Largest value – Smallest value = 28 – 2 =26 times

$$s^2 = \frac{\sum x^2 - \frac{(\sum x)^2}{n}}{n-1} = \frac{1775 - \frac{(109)^2}{10}}{10-1} = 65.2111 \qquad s = \sqrt{65.2111} = 8.08 \text{ times}$$

16. Suppose the exam scores for seven students are 73, 82, 95, 79, 22, 86, and 91 points. Then, Mean = (73 + 82 + 95 + 79 + 22 + 86 + 91)/7 = 75.43 points. If we drop the outlier (22), Mean = (73 + 82 + 95 + 79 + 86 + 91)/6 = 84.33 points. This shows how an outlier can affect the value of the mean.

17. Suppose the exam scores for seven students are 73, 82, 95, 79, 22, 86, and 91 points. Then,

Range = Largest value – Smallest value = 95 – 22 = 73 points.

If we drop the outlier (22) and calculate the range,

Range = Largest value – Smallest value = 95 – 73 = 22 points.

Thus, when we drop the outlier, the range decreases from 73 to 22 points.

18. The value of the standard deviation is zero when all the values in a data set are the same. For example, suppose the heights (in inches) of five women are: 67 67 67 67 67

This data set has no variation. As shown below the value of the standard deviation is zero for this data set. For these data: $n = 5$, $\sum x = 335$, and $\sum x^2 = 22{,}445$.

$$s = \sqrt{\frac{\sum x^2 - \frac{(\sum x)^2}{n}}{n-1}} = \sqrt{\frac{22{,}445 - \frac{(335)^2}{5}}{5-1}} = \sqrt{\frac{22{,}445 - 22{,}445}{4}} = 0$$

19. a. The frequency column gives the number of weeks for which the number of computers sold was in the corresponding class.

b. For the given data: $n = 25$, $\sum mf = 486.50$, and $\sum m^2 f = 10{,}524.25$

$\bar{x} = (\sum mf)/n = 486.50/25 = 19.46$ computers

$$s^2 = \frac{\sum m^2 f - \frac{(\sum mf)^2}{n}}{n-1} = \frac{10{,}524.25 - \frac{(486.50)^2}{25}}{25-1} = 44.0400,\ s = \sqrt{44.0400} = 6.64 \text{ computers}$$

20. a. i. Each of the two values is 5.5 years from $\mu = 7.3$ years. Hence,

$$k = 5.5/2.2 = 2.5 \text{ and } 1 - \frac{1}{k^2} = 1 - \frac{1}{(2.5)^2} = 1 - .16 = .84 \text{ or } 84\%$$

Thus, at least 84% of the cars are 1.8 to 12.8 years old.

ii. Each of the two values is 6.6 years from $\mu = 7.3$ years. Hence

$$k = 6.6/2.2 = 3 \text{ and } 1 - \frac{1}{k^2} = 1 - \frac{1}{(3)^2} = 1 - .11 = .89 \text{ or } 89\%$$

Thus, at least 89% of the cars are .7 to 13.9 years old.

b. $1 - \frac{1}{k^2} = .75$ gives $\frac{1}{k^2} = 1 - .75 = .25$ or $k^2 = \frac{1}{2.5}$, so $k = 2$

$\mu - 2\sigma = 7.3 - 2(2.2) = 2.9$ hours and $\mu + 2\sigma = 7.3 + 2(2.2) = 11.7$ hours

Thus, the required interval is 2.9 to 11.7 years.

21. $\mu = 7.3$ years and $\sigma = 2.2$ years

a. i. The intervals 5.1 to 9.5 years is $\mu - \sigma$ to $\mu + \sigma$. Hence, approximately 68% of the cars are 5.1 to 9.5 years old.

ii. The interval .7 to 13.9 years is $\mu - 3\sigma$ to $\mu + 3\sigma$. Hence, approximately 99.7% of the cars are .7 to 13.9 years.

b. $\mu - 2\sigma = 7.3 - 2(2.2) = 2.9$ hours and $\mu + 2\sigma = 7.3 + 2(2.2) = 11.7$ hours. The interval that contains ages of 95% of the cars will be 2.9 to 11.7 years old.

22. The ranked data are: 0 1 2 3 4 5 7 8 10 11 12 13 14 15 20

a. The three quartiles are $Q_1 = 3$, $Q_2 = 8$, and $Q_3 = 13$.

IQR = $Q_3 - Q_1 = 13 - 3 = 10$.

The value 4 lies between Q_1 and Q_2, which indicates that this value is in the second from the bottom 25% group in the ranked data.

b. $kn/100 = 60(15)/100 = 9$. Thus, the 60^{th} percentile may be represented by the value of the ninth term in the ranked data, which is 10. Therefore, $P_{60} = 10$. Thus, approximately 60% of the half hour time periods had fewer than or equal to 10 passengers set off the metal detectors during this day.

c. Ten values in the given data are less than 12. Hence, the percentile rank of 12 is (10/15) × 100 = 66.67%. Thus, 66.67% of the half hour time periods had fewer than 12 passengers set off the metal detectors during this day.

23. The ranked data are: 0 1 2 3 4 5 7 8 10 11 12 13 14 15 20

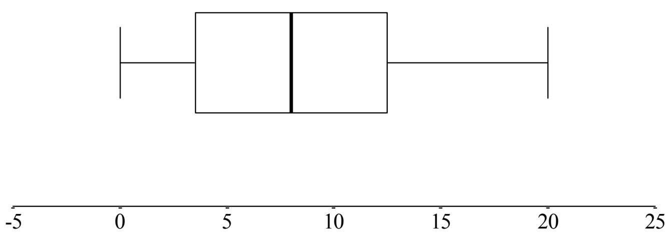
Median = 8, Q_1 = 3, and Q_3 = 13,

IQR = $Q_3 - Q_1$ = 13 – 3 = 10, 1.5 × IQR = 1.5 × 10 =15,

Lower inner fence = Q_1 – 15 = 3 – 15 = –12,

Upper inter fence = Q_3 + 15 = 13 + 15 = 28

The smallest and largest values in the data set within the two inner fences are 0 and 20, respectively. The data does not contain any outliers.



The data are skewed slightly to the right.

24. From the given information: $n_1 = 15$, $n_2 = 20$, $\bar{x}_1$ = \$1035, $\bar{x}_2$ = \$1090

$$\bar{x} = \frac{n_1\bar{x}_1 + n_1\bar{x}_2}{n_1 + n_2} = \frac{(15)(1035)+(20)(1090)}{15+20} = \frac{37{,}325}{35} = \$1066.43$$

25. Sum of the GPAs of five students = (5)(3.21) = 16.05

Sum of the GPAs of four students = 3.85 + 2.67 + 3.45 + 2.91 = 12.88

GPA of the fifth student = 16.05 – 12.88 = 3.17

26. The ranked data are: 207 238 287 293 349 366 463 479 538 2534

Thus, to find the 10% trimmed mean, we drop the smallest value and the largest value (10% of 10 is 1) and find the mean of the remaining 8 values. For these 8 values,

$\sum x$ = 238 + 287 + 293 + 349 + 366 + 463 + 479 + 538 = 3013

10% trimmed mean = $(\sum x)/8$ = 3013/8 = \$376.63 thousand = \$376,630. The 10% trimmed mean is a better summary measure for these data than the mean of all 10 values because it eliminates the effect of the outlier, 2534.

27. a. For Data Set I: $\bar{x} = (\sum x)/n$ = 79/4 = 19.75

For Data Set II: $\bar{x} = (\sum x)/n$ = 67/4 = 16.75

The mean of Data Set II is smaller than the mean of Data Set I by 3.

b. For Data Set I: $\sum x = 79$, $\sum x^2 = 1945$, and $n = 4$

$$s = \sqrt{\frac{\sum x^2 - \frac{(\sum x)^2}{n}}{n-1}} = \sqrt{\frac{1945 - \frac{(79)^2}{4}}{4-1}} = 11.32$$

c. For Data Set II: $\sum x = 67$, $\sum x^2 = 1507$, and $n = 4$

$$s = \sqrt{\frac{\sum x^2 - \frac{(\sum x)^2}{n}}{n-1}} = \sqrt{\frac{1507 - \frac{(67)^2}{4}}{4-1}} = 11.32$$

The standard deviations of the two data sets are equal.