Chapter 3
Numerical Descriptive Measures

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 five measures of central tendency (mean, median, mode, trimmed mean, and weighted mean) 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 roughly 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 the skew or outliers.

3.9 $\Sigma x = 5 + (-7) + 2 + 0 + (-9) + 16 + 10 + 7$
$= 24$
$\mu = (\Sigma x)/N = 24/8 = 3$
Median = value of the 4.5^{th} term in ranked data = (2 + 5)/2 = 3.50
This data set has no mode.

3.10 $\bar{x} = (\Sigma x)/n$ = \$158,542/10 = \$15,854.20
Median = average of the 5th and 6th terms in ranked data set = \$14,539.50. There is no mode since all values occur exactly once.

3.11 $\bar{x} = (\Sigma x)/N$ = 3,169/12 = 264.08
Median = average of the 6th and 7th terms in ranked data set = 262. There is no mode since all values occur exactly once.

3.12 **a.** $\bar{x} = (\Sigma x)/n$ = 463/20 = 23.15 years
Median = average of the 10th and 11th terms in ranked data set = 21 years.
The mode is 5 and 27 since both values occur twice and all others just once.

b. For the 10% trimmed mean, we must remove 0.10(20) = 2 values from each end of the ranked data set. So, we discard 3, 5, 51, and 59 and average the remaining 16 values to get 345/16 = 21.5625 years.

3.13 **a.** $\bar{x} = (\Sigma x)/n$ = 4298/10 = 429.80 thousands of dollars.
Median = average of the 5th and 6th terms in ranked data set = 103.5 thousand dollars.

b. There is no mode because all values occur exactly once.

c. For the 10% trimmed mean, we must remove 0.10(10) = 1 value from each end of the ranked data set and then average the remaining 8 values to get 992/8 = 124 thousand dollars.

d. Median and trimmed mean are good measures to use where there is an outlier.

3.14 **a.** $\bar{x} = (\Sigma x)/n$ = 19,167/20 = 958.35 dollars
Median = average of the 10th and 11th terms in ranked data set = 990 dollars.

b. For the 20% trimmed mean, we must remove 0.20(20) = 4 values from each end of the ranked data set and average the remaining 16 values to get 15,466/16 = 966.63 dollars.

3.15 **a.** $\bar{x} = (\Sigma x)/n$ = 2,440/10 = 244 thousand dollars.
Median = average of the 5th and 6th terms in ranked data set = 235 thousand dollars.

b. For the 10% trimmed mean, we must remove 0.10(10) = 1 value from each end of the ranked data set and average the remaining 8 values to get 1,927/8 = 240.875 thousand dollars.

3.16 **a.** $\bar{x} = (\Sigma x)/n$ = 1,113/20 = 55.65 thousand dollars
Median = average of the 10th and 11th terms in ranked data set = 57.5 thousand dollars. The mode is 64 because it occurs twice and all other values occur just once.

b. For the 15% trimmed mean, we must remove 0.15(20) = 3 values from each end of the ranked data set and average the remaining 14 values to get 779/14 = 55.64 thousand dollars

3.17 **a.** $\bar{x} = (\Sigma x)/n$ = 597/20 = 29.85 patients
Median = average of the 10th and 11th terms in ranked data set = 29.5 patients
Each of 24, 26, 37, 38 are modes because they each occur twice and all other values exactly once.

b. For the 15% trimmed mean, we must remove 0.15(20) = 3 values from each end of the ranked data set and average the remaining 14 values to get 418/14 = 29.86 patients.

3.18 **a.** $\bar{x} = (\Sigma x)/n = 2{,}651/20 = 132.55$ mmHg
Median = average of the 10th and 11th terms in ranked data set = 135 mmHg. The mode is 144 because it occurs more often than all other values.

b. For the 10% trimmed mean, we must remove 0.10(20) = 2 values from each end of the ranked data set and average the remaining 16 values to get 132.6875 mmHg

3.19 The opinion that they will not allow their children to play football occurs the most often and hence, is the mode.

3.20 John's overall score = 0.30(75) + 0.05(52) + 0.10(85) + 0.15(74) + 0.40(81) = 77.1 out of 100.

3.21 The weighted mean =

$$\frac{1200(30)+1900(45)+1400(40)+2200(35)+1300(50)}{8000} = \frac{319,500}{8000} = 39.9375$$

So, the average price paid is about \$39.94.

3.22 $n_1 = 10$, $n_2 = 8$, $\bar{x}_1 = \$140$, $\bar{x}_2 = \$160$

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

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

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

3.25 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.

3.26 Sum of the prices paid by the seven passengers = (7)(361) = \$2527
Total price paid by the couple = 2527 – (420 + 210 + 333 + 695 + 485) = \$384
Price paid by each of the couple = 384/2 = \$192

3.27 Geometric mean = $\sqrt[n]{x_1 \cdot x_2 \cdot x_3 \cdot \ldots \cdot x_n} = \sqrt[5]{1.04 \cdot 1.03 \cdot 1.05 \cdot 1.06 \cdot 1.08} = \sqrt[5]{1.287625248} \approx 1.052$
Then, 1– Geometric mean = 1.052 –1= 0.052, so the mean inflation rate is 4.8%.

Section 3.2

3.28 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.

3.29 No, the value of the standard deviation cannot be negative, because the deviations from the mean are squared and, therefore, either positive or zero. The square root of the sum of these values must also be either positive or zero.

3.30 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:

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

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

3.31 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 and this class is considered to be the population of interest, 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.

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

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

3.33 **a.** $\bar{x} = (\Sigma x)/n = 840/7 = \$120$

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

Range = 170 – 56 = \$114

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

$$= 1712$$

Coefficient of variation = (41.38/120) $\times$ 100% = 34.5%

3.34 **a.**

x	x^2	x	x^2
23	529	12	144
9	81	31	961
12	144	5	25
21	441	10	100
24	576	27	729
6	36	9	81
33	1089	15	225
34	1156	16	256
17	289	30	900
3	9	38	1444

$\Sigma x = 375$ and $\Sigma x^2 = 9215$

$n = 20$

Range = Largest value – Smallest value = 38 – 3 = 35 years

$$s^2 = \frac{\sum x^2 - \frac{(\sum x)^2}{n}}{N} = \frac{9215 - \frac{(375)^2}{20}}{20} \approx 109.1875$$

$$s = \sqrt{109.1875} \approx 10.45 \text{ years}$$

b. The coefficient of variation = $\frac{s}{\bar{x}} \times 100\% = \frac{10.45}{375/20} \times 100\% \approx 55.73\%$.

c. These are population parameters because ALL employees of the company are used.

3.35 **a.**

x	x^2	x	x^2
50	2500	74	5476
71	5041	40	1600
57	3249	67	4489
39	1521	44	1936
45	2025	77	5929
64	4096	61	3721
38	1444	58	3364
53	2809	55	3025
35	1225	64	4096
62	3844	59	3481

$\Sigma x = 1{,}113$ and $\Sigma x^2 = 64{,}871$

$n = 20$

Range = Largest value – Smallest value = 77 – 35 = 42 thousand dollars

$$s^2 = \frac{\sum x^2 - \frac{(\sum x)^2}{n}}{n-1} = \frac{64{,}871 - \frac{(1{,}113)^2}{20}}{20-1} \approx 154.34$$

$$s = \sqrt{164.34} \approx 12.42 \text{ thousand dollars}$$

b. The coefficient of variation = $\frac{s}{\bar{x}} \times 100\% = \frac{12.42}{1{,}113/20} \times 100\% \approx 22.3\%$

c. These values are statistics because they only reflect the annual salaries of 20 randomly selected health care workers, not all of them.

3.36 **a.**

x	x^2	x	x^2
23	529	28	784
37	1369	32	1024
26	676	37	1369
19	361	29	841
33	1089	38	1444
22	484	24	576
30	900	35	1225
42	1764	20	400
24	576	34	1156
26	676	38	1444

$\Sigma x = 597$ and $\Sigma x^2 = 18{,}687$

$n = 20$

Range = Largest value – Smallest value = 42 – 19 = 23 patients

$$s^2 = \frac{\sum x^2 - \frac{(\sum x)^2}{n}}{n-1} = \frac{18,687 - \frac{(597)^2}{20}}{20-1} \approx 45.608$$

$s = \sqrt{45.608} \approx 6.75$ patients

b. The coefficient of variation = $\frac{s}{\bar{x}} \times 100\% = \frac{6.75}{597/20} \times 100\% \approx 22.6\%$

3.37 **a.**

x	x^2	x	x^2
139	19,321	111	12,321
151	22,801	150	22,500
138	19,044	107	11,449
153	23,409	132	17,424
134	17,956	144	20,736
136	18,496	116	13,456
141	19,881	159	25,281
126	15,876	121	14,641
109	11,881	127	16,129
144	20,736	113	12,769

$\Sigma x = 2,651$ and $\Sigma x^2 = 356,107$

$n = 20$

Range = Largest value – Smallest value = 159 – 107 = 52 mmHg

$$s^2 = \frac{\sum x^2 - \frac{(\sum x)^2}{n}}{n-1} = \frac{356,107 - \frac{(2,651)^2}{20}}{20-1} \approx 248.261$$

$s = \sqrt{248.261} \approx 15.76$ mmHg

b. The coefficient of variation = $\frac{s}{\bar{x}} \times 100\% = \frac{15.76}{2,651/20} \times 100\% \approx 11.89\%$

3.38 **a.**

x	x^2
35	1225
10	100
22	484
38	1444
31	961
27	729
53	2809
44	1936
16	256
44	1936
25	625
12	144
$\Sigma x = 357$	$\Sigma x^2 = 12,649$

$n = 12$

Range = Largest value – Smallest value = 53 – 10 = 43 minutes

$$s^2 = \frac{\sum x^2 - \frac{(\sum x)^2}{n}}{N} = \frac{12,649 - \frac{(357)^2}{12}}{12} = 169.021 \quad \text{(parameter, not statistic)}$$

$s = \sqrt{169.021}$ = 13.0 minutes

b. The coefficient of variation = $\frac{s}{\bar{x}} \times 100\% = \frac{13.0}{357/12} \times 100\% \approx 43.7\%$

c. The large standard deviation suggests that the data is widely spread from the mean.

3.39 **a.**

x	x^2
15	225
26	676
16	256
36	1296
31	961
13	169
29	841
18	324
21	441
39	1521
Σx = 244	Σx^2 = 6,710

n = 10

Range = Largest value – Smallest value = 39 – 13 = 26 minutes

$$s^2 = \frac{\sum x^2 - \frac{(\sum x)^2}{n}}{N} = \frac{6,710 - \frac{(244)^2}{10}}{10} = 75.64 \quad \text{parameter, not statistic}$$

$s = \sqrt{75.64}$ = 8.70 minutes

b. The coefficient of variation = $\frac{s}{\bar{x}} \times 100\% = \frac{8.70}{244/10} \times 100\% \approx 35.64\%$

c. The large standard deviation suggests that the data is widely spread from the mean.

3.40 **a.**

x	x^2
205	42,025
265	70,225
176	30,976
314	98,596
243	59,049
192	36,864
297	88,209
357	127,449
238	56,644
281	78,961
342	116,964
259	67,081
Σx = 3,169	Σx^2 = 873,043

n = 12

Range = Largest value – Smallest value = 357 – 176 = 181 dollars

$$s^2 = \frac{\sum x^2 - \frac{(\sum x)^2}{n}}{n-1} = \frac{873,043 - \frac{(3,169)^2}{12}}{12-1} = 3,287.538$$

$s = \sqrt{3287.538}$ = 57.34 dollars

b. The coefficient of variation = $\frac{s}{\bar{x}} \times 100\% = \frac{57.34}{3,169/12} \times 100\% \approx 21.71\%$

3.41 **a.**

x	x^2
127	16,129
82	6724
45	2025
99	9801
153	23,409
3261	10, 634,121
77	5929
108	11,664
68	4624
278	77,284
Σx = 4,298	Σx^2 = 10,791,710

n = 10

Range = Largest value – Smallest value = 3,261 – 45 = 3,216 thousand dollars

$$s^2 = \frac{\sum x^2 - \frac{(\sum x)^2}{n}}{n-1} = \frac{10,791,710 - \frac{(4,298)^2}{10}}{10-1} = 993,825.51$$

$s = \sqrt{993,825.51}$ = 996.91 thousand dollars

b. The coefficient of variation = $\frac{s}{\bar{x}} \times 100\% = \frac{996.91}{4,298/10} \times 100\% \approx 231.95\%$

3.42

x	x^2
22	484
22	484
22	484
22	484
22	484
22	484
22	484
22	484
Σx = 176	Σx^2 = 3872

n = 8

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

$= 0$

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

3.43 For the yearly salaries of all employees,

$$\text{CV} = \frac{\sigma}{\mu} \times 100\% = \frac{6820}{62,350} \times 100\% = 10.94\%$$

For the years of experience of these employees, $\text{CV} = \frac{\sigma}{\mu} \times 100\% = \frac{2}{15} \times 100\% = 13.33\%$

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

3.44 For the SAT scores of the 100 students,

$$\text{CV} = \frac{s}{\bar{x}} \times 100\% = \frac{105}{3975} \times 100\% = 10.77\%$$

For the GPAs of these students,

$$\text{CV} = \frac{s}{\bar{x}} \times 100\% = \frac{0.22}{3.16} \times 100\% = 6.96\%$$

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

3.45

Data Set I		Data Set II	
x	x^2	x	x^2
12	144	19	361
25	625	32	1024
37	1369	44	1936
8	64	15	225
41	1681	48	2304
$\Sigma x = 123$	$\Sigma x^2 = 3883$	$\Sigma x = 158$	$\Sigma x^2 = 5850$

For Data Set I: $s = \sqrt{\dfrac{\sum x^2 - \dfrac{(\sum x)^2}{n}}{n-1}} = \sqrt{\dfrac{3883 - \dfrac{(123)^2}{5}}{5-1}} = \sqrt{214.300} = 14.64$

For Data Set II: $s = \sqrt{\dfrac{\sum x^2 - \dfrac{(\sum x)^2}{n}}{n-1}} = \sqrt{\dfrac{5850 - \dfrac{(158)^2}{5}}{5-1}} = \sqrt{214.300} = 14.64$

The standard deviations of the two data sets are equal.

Section 3.3

3.46 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.

3.47

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
	$N = \Sigma f = 40$		$\Sigma mf = 354$	$\Sigma m^2 f = 3636$

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

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

3.48

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
	$n = \Sigma f = 80$		$\Sigma mf = 752$	$\Sigma m^2 f = 10{,}048$

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

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

3.49

Hours Per Week	Number of Students	m	mf	m^2f
0 to less than 4	14	2	28	56
4 to less than 8	18	6	108	648
8 to less than 12	25	10	250	2500
12 to less than 16	18	14	252	3528
16 to less than 20	16	18	288	5184
20 to less than 24	9	22	198	4356
	$N = \Sigma f = 100$		$\Sigma mf = 1124$	$\Sigma m^2 f = 16{,}272$

$\mu = (\Sigma mf)/N = 1124/100 = 11.24$ hours

$$\sigma^2 = \frac{\sum m^2 f - \frac{\left(\sum mf\right)^2}{N}}{N-1} = \frac{16{,}272 - \frac{(1124)^2}{100}}{100} = 36.3824 \qquad s = \sqrt{36.3824} = 6.03 \text{ hours}$$

3.50

Miles Driven in 2012 (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
	$n = \Sigma f = 300$		$\Sigma mf = 5900$	$\Sigma m^2 f = 136{,}275$

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

$$s^2 = \frac{\sum m^2 f - \frac{\left(\sum mf\right)^2}{n}}{N} = \frac{136{,}275 - \frac{(5900)^2}{300}}{300} = 67.4722 \qquad s = \sqrt{67.4722} = 8.21 \text{ or } 8210 \text{ miles}$$

Parameter, not statistic

Each value in the column labeled *mf* gives the approximate total mileage (in thousands) 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 $\Sigma mf = 5900$ indicates that the total mileage for all 300 car owners was approximately 5,900,000 miles.

3.51

Amount of Electric Bill (dollars)	Number of Families	m	mf	m^2f
0 to less than 60	5	30	150	4500
60 to less than 120	16	90	1440	129,600
120 to less than 180	11	150	1650	247,500
180 to less than 240	10	210	2100	441,000
240 to less than 300	8	270	2160	583,200
	$n = \Sigma f = 50$		$\Sigma mf = 7500$	$\Sigma m^2f = 1{,}405{,}800$

$\bar{x} = (\Sigma mf)/n = 7500/50 = \$150$

$$s^2 = \frac{\sum m^2 f - \frac{\left(\sum mf\right)^2}{n}}{n-1} = \frac{1{,}405{,}800 - \frac{(7500)^2}{50}}{50-1} = 5730.6122 \qquad s = \sqrt{5730.6122} = \$75.70$$

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 \$150 for electricity in August 2012. The value $\Sigma mf = \$7500$ is the approximate total amount of the electric bills for all 50 families included in the sample.

3.52

x	f	m	mf	m^2f
200 to less than 500	38	350	13,300	4,655,000
500 to less than 800	105	650	68,250	44,362,500
800 to less than 1100	130	950	123,500	117,325,000
1100 to less than 1400	60	1250	75,000	93,750,000
1400 to less than 1700	42	1550	65,100	100,905,000
1700 to less than 2000	18	1850	33,300	61,605,000
2000 to less than 2300	7	2150	15,050	32,357,500
	$n = \Sigma f = 400$		$\Sigma mf = 393{,}500$	$\Sigma m^2f = 454{,}960{,}000$

$\bar{x} = (\Sigma mf)/n = 393{,}500/400 = 983.75$ dollars

$$s^2 = \frac{\sum m^2 f - \frac{\left(\sum mf\right)^2}{n}}{n-1} = \frac{454{,}960{,}000 - \frac{(393{,}500)^2}{400}}{400-1} = 170{,}061.0902$$

$s = \sqrt{170{,}061.0902} = 412.38$ dollars

3.53

Hours per Week	f	m	mf	m^2f
0 to less than 3.5	34	1.75	59.5	104.125
3.5 to less than 7.0	92	5.25	483	2535.75
7.0 to less than 10.5	55	8.75	481.25	4210.9375
10.5 to less than 14.0	83	12.25	1016.75	12,455.1875
14.0 to less than 28.0	121	21	2541	53,361
28.0 to less than 56.0	15	42	630	26,460
	$n = \Sigma f = 400$		$\Sigma mf = 5211.5$	$\Sigma m^2f = 99{,}127$

$\bar{x} = (\Sigma mf)/n = 5211.5/400 = 13.02875$ hours

$$s^2 = \frac{\sum m^2 f - \frac{\left(\sum mf\right)^2}{n}}{n-1} = \frac{99{,}127 - \frac{(5211.5)^2}{400}}{400-1} = 78.2648 \qquad s = \sqrt{78.2648} = 8.85 \text{ hours}$$

Section 3.4

3.54 Chebyshev's theorem is applied to find a lower bound for the area under a distribution's 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.

3.55 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.

3.56 For the interval $\bar{x} \pm 2s$: k = 2, and

$1 - \frac{1}{k^2} = 1 - \frac{1}{2^2} = 1 - 0.25 = 0.75$ or 75%. Thus, at least 75% of the observations fall in the interval = 74 ± 24 = (50, 98).

For the interval $\bar{x} \pm 2.5s$: k = 2.5, and

$1 - \frac{1}{k^2} = 1 - \frac{1}{2.5^2} = 1 - 0.16 = 0.84$ or 84%. Thus, at least 84% of the observations fall in the interval $\bar{x} \pm 2.5s$ = 74 ± 30 = (44, 104).

For the interval $\bar{x} \pm 3s$: k = 3, and

$1 - \frac{1}{k^2} = 1 - \frac{1}{3^2} = 1 - 0.11 = 0.89$ or 89%. Thus, at least 89% of the observations fall in the interval $\bar{x} \pm 3s$ = 74 ± 36 = (38, 100)

3.57 For the interval $\mu \pm 2\sigma$: k = 2, and

$1 - \frac{1}{k^2} = 1 - \frac{1}{2^2} = 1 - 0.25 = 0.75$ or 75%. Thus, at least 75% of the observations fall in the interval $\mu \pm 2\sigma$ = 230 ± 82 = (148, 312).

For the interval $\mu \pm 2.5\sigma$: k = 2.5, and

$1 - \frac{1}{k^2} = 1 - \frac{1}{2.5^2} = 1 - 0.16 = 0.84$ or 84%. Thus, at least 84% of the observations fall in the interval $\mu \pm 2.5\sigma$ = 230 ± 102.5 = (127.5, 332.5).

For the interval $\mu \pm 3\sigma$: k = 3, and

$1 - \frac{1}{k^2} = 1 - \frac{1}{3^2} = 1 - 0.11 = 0.89$ or 89%. Thus, at least 89% of the observations fall in the interval $\mu \pm 3\sigma$ = 230 ± 123 = (107, 353).

3.58 Approximately 68% of the observations fall in the interval $\mu \pm \sigma$ = 310 ± 27 = (273, 347), approximately 95% fall in the interval $\mu \pm 2\sigma$ = 310 ± 74 = (236, 384), and about 99.7% fall in the interval $\mu \pm 3\sigma$ = 310 ± 111 = (199, 421).

3.59 Approximately 68% of the observations fall in the interval $\bar{x} \pm s$ = 82 ± 16 = (66, 98), approximately 95% fall in the interval $\bar{x} \pm 2s$ = 82 ± 32 = (50, 114), and about 99.7% fall in the interval $\bar{x} \pm 3s$ = 82 ± 48 = (34, 130).

3.60 **a.** Each of the two values is 40 minutes from μ = 220. Hence, k = 40/20 = 2 and $1 - \frac{1}{k^2} = 1 - \frac{1}{2^2} = 1 - 0.25 = 0.75$ 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 m = 220. Hence, k = 60/20 = 3 and $1 - \frac{1}{k^2} = 1 - \frac{1}{3^2} = 1 - 0.11 = 0.89$ 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 m = 220. Hence, k = 50/20 = 2.5 and $1 - \frac{1}{k^2} = 1 - \frac{1}{2.5^2} = 1 - 0.16 = 0.84$ or 84%. Thus, at least 84% of the runners ran this race in 170 to 270 minutes.

3.61 **a.** **i.** Each of the two values is 20 minutes from m = 34 minutes. Hence,

k = 20/8 = 2.5 and $1 - \frac{1}{k^2} = 1 - \frac{1}{2.5^2} = 1 - 0.16 = 0.84$ or 84%. Thus, at least 84% of all workers have commuting times between 14 and 54 minutes.

ii. Each of the two values is 16 minutes from m = 34 minutes. Hence,

k = 16/8 = 2 and

$1 - \frac{1}{k^2} = 1 - \frac{1}{2^2} = 1 - 0.25 = 0.75$ or 75%. Thus, at least 75% of all workers have commuting times between 18 and 50 minutes.

b. $1 - \frac{1}{k^2} = 0.89$ gives $\frac{1}{k^2} = 1 - 0.89 = 0.11$ or $k^2 = \frac{1}{0.11}$, so k » 3.

$\mu - 3\sigma$ = 34 – 3(8) = 10 and

$\mu + 3\sigma$ = 34 + 3(8) = 58.

Thus, the required interval is 10 to 58.

3.62 **a.** **i.** Each of the two values is \$680 from m = \$2365. Hence, k = 680/340 = 2 and

$1 - \frac{1}{k^2} = 1 - \frac{1}{2^2} = 1 - 0.25 = 0.75$ 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 m = \$2365. Hence,

k = 1020/340 = 3 and

$1 - \frac{1}{k^2} = 1 - \frac{1}{3^2} = 1 - 0.11 = 0.89$ or 89%. Thus, at least 89% of all homeowners pay a monthly mortgage of \$1345 to \$3385.

b. $1 - \frac{1}{k^2} = 0.84$ gives $\frac{1}{k^2} = 1 - 0.84 = 0.16$ or $k^2 = \frac{1}{0.16}$ 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.

3.63 μ = 34 and σ = 8

a. The interval 10 to 58 is $\mu - 3\sigma$ to $\mu + 3\sigma$. Hence, approximately 99.7% of all workers have commuting times between 10 and 58 minutes.

b. The interval 26 to 42 is $\mu - \sigma$ to $\mu + \sigma$. Hence, approximately 68% of all workers have commuting times between 26 and 42 minutes.

c. The interval 18 to 50 is $\mu - 2\sigma$ to $\mu + 2\sigma$. Hence, approximately 95% of all workers have commuting times between 18 and 50 minutes

3.64 μ = \$180 and σ = \$30

a. **i.** The interval \$150 to \$210 is $\mu - \sigma$ to $\mu + \sigma$. Hence, approximately 68% of all college textbooks are priced between \$150 and \$210.

ii. The interval \$120 to \$240 is $\mu - 2\sigma$ to $\mu + 2\sigma$. Hence, approximately 95% of all college textbooks are priced between \$120 and \$240.

b. $\mu - 3\sigma = 180 - 3(30) = \$90$ and $\mu + 3\sigma = 180 + 3(30) = \$270$

The interval that contains the prices of 99.7% of college textbooks is \$90 to \$270.

Section 3.5

3.65 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.

3.66 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.

3.67 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.

3.68 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,

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

3.69 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. $kn/100 = 35(13)/100 = 4.55$

Thus, the 35^{th} percentile can be approximated by the 5^{th} term in the ranked data. Therefore, $P_{35} = 71$.

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

3.70 The ranked data are:

427	441	530	595	699	716	872	933	934	1046
1065	1125	1127	1187	1234	1274	1353	1480	1630	2199

a. The quartiles are:

Q_1 = average of 5th and 6th term in ranked data set = (699 + 716)/2 = 707.5

Q_2 = average of 10th and 11th term in ranked data set = (1046 + 1065)/2 = 1055.5

Q_3 = average of 15th and 16th term in ranked data set = (1234 + 1274)/2 = 1254

IQR = $Q_3 - Q_1$ = 1254 – 707.5 = 546.5

b. *kn*/100 = 57(20)/100 = 11.4
Thus, the 57th percentile can be approximated by the value of the 12th term in the ranked data, which is 1125. Therefore, P_{57} = 1125.

c. Nine values in the given data are smaller than 1046. Hence, the percentile rank of 1046 = (9/20) × 100 = 45%. This means that 45% of the data are less than 1046.

3.71 The ranked data are:

32	33	33	34	35	36	37	37	37	37
38	39	40	41	41	42	42	42	43	44
44	45	45	45	47	47	47	47	47	48
48	49	50	50	51	52	53	54	59	61

a. The three quartiles are Q_1 = (37 + 38)/2 = 37.5, Q_2 = (44 + 44)/2 = 44, and Q_3 = (48 + 48)/2 = 48
IQR = $Q_3 - Q_1$ = 48 – 37.5 = 10.5
The value 49 lies between Q_2 and Q_3, which means at least 50% of the data are smaller and at least 25% of the data are larger than 49.

b. *kn*/100 = 91(40)/100 = 36.4
Thus, the 91st percentile can be approximated by the 37th term in the ranked data. Therefore, P_{91} = 53. This means that 91% of the values in the data set are less than 53.

c. Twelve values in the given data set are less than 40. Hence, the percentile rank of 40 = (12/40) × 100 = 30%. Therefore, the number of text message was 40 or higher on 70% of the days.

3.72 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.

3.73 The ranked data are:

35	38	39	40	44	45	50	53	55	57
58	59	61	62	64	64	67	71	74	77

a. The quartiles are:

Q_1 = average of 5th and 6th term in ranked data set = (44+ 45)/2 = 44.5

Q_2 = average of 10^{th} and 11^{th} term in ranked data set = (57 + 58)/2 = 57.5

Q_3 = average of 15^{th} and 16^{th} term in ranked data set = (64 + 64)/2 = 64

IQR = $Q_3 - Q_1$ = 64 – 44.5 = 19.5

The value 57 lies between Q_1 and Q_2 which means at least 25% of the data are smaller and at least 50% of the data are larger than 57.

b. $kn/100$ = 30(20)/100 = 6
Thus, the 30^{th} percentile is the value of the 6^{th} term in the ranked data, which is 45. Therefore, P_{30} = 45.

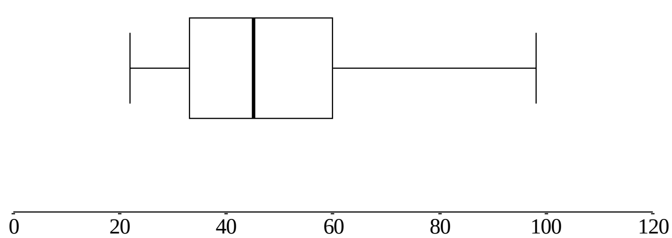
c. Twelve values in the given data are smaller than 61. Hence, the percentile rank of 61 = (12/20) × 100 = 60%.

Section 3.6

3.74 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.

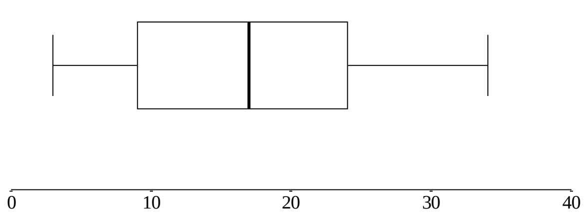
3.75 The ranked data are: 22 24 25 28 31 32 34 35 36 41 42 43
47 49 52 55 58 59 61 61 63 65 73 98

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 × IQR = 1.5 × 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.



3.76 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.

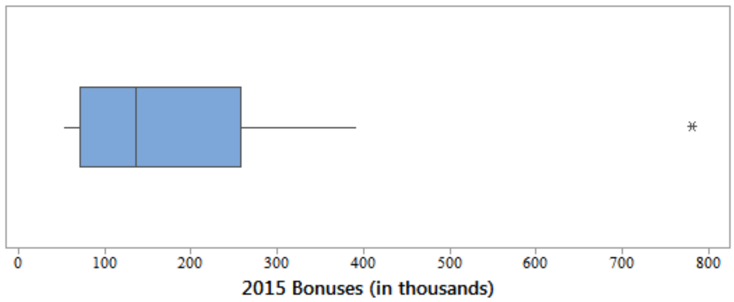


The data are nearly symmetric.

3.77 The ranked data are:

53	61	67	71	89	107	122	136	175	208
247	258	361	391	781					

Median = 136, Q_1 = 71, and Q_3 = 258
IQR = $Q_3 - Q_1$ = 187, 1.5 × IQR = 1.5 × 187 = 280.5,
Lower inner fence = Q_1 – 280.25 = –209.5
Upper inner fence = Q_3 + 280.5 = 538.5
The largest value exceeds the upper fence and so, is an outlier.

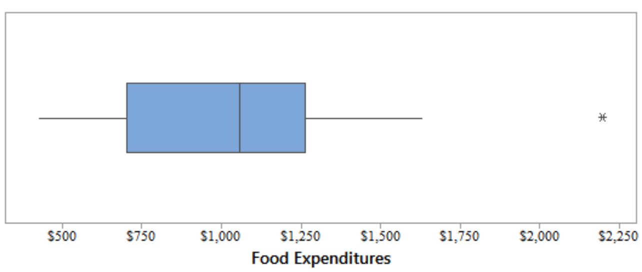


The data are skewed to the right (that is, toward smaller values).

3.78 The ranked data are:

427	441	530	595	699	716	872	933	934	1046
1065	1125	1127	1187	1234	1274	1353	1480	1630	2199

Median = (1046+1065)/2 = 1055.5, Q_1 = (600 + 716)/2 = 707.5, and Q_3 = (1274 + 1234)/2 = 1254
IQR = $Q_3 - Q_1$ = 546.5, 1.5 × IQR = 1.5 × 546.5 = 819.75,
Lower inner fence = Q_1 – 819.75 = –112.25
Upper inner fence = Q_3 + 819.75 = 2073.75
The largest value exceeds the upper fence and so, is an outlier.



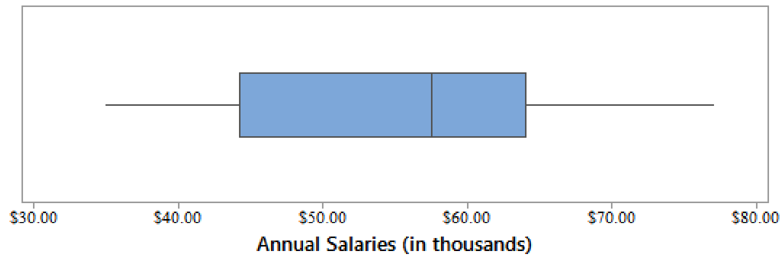
The data are skewed to the right (that is, toward larger values).

3.79 The ranked data are:

35	38	39	40	44	45	50	53	55	57
58	59	61	62	64	64	67	71	74	77

Median = (57 + 58)/2 = 57.5, Q_1 = 44.4, and Q_3 = 64
IQR = $Q_3 - Q_1$ = 19.5, 1.5 × IQR = 1.5 × 19.5 = 29.25,
Lower inner fence = Q_1 – 29.25 = 15.25
Upper inner fence = Q_3 + 29.25 = 93.25

The smallest and the largest values within the two inner fences are 35 and 77, respectively. The data set contains no outliers.

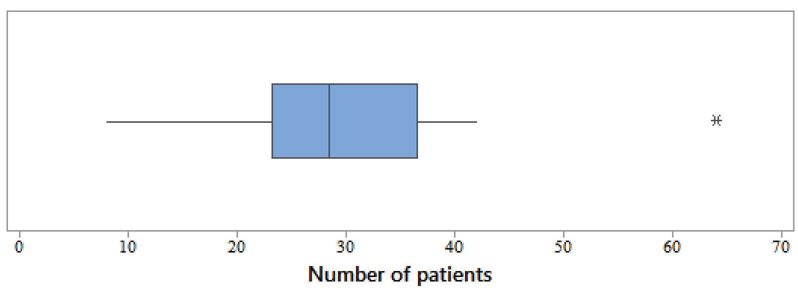


The data are skewed to the left.

3.80 The ranked data are:

8	16	19	20	22	23	24	24	26	26
28	29	30	32	33	34	35	37	37	38
38	42	64							

Median = (28 + 29)/2 = 28.5, Q_1 = 23.5, and Q_3 = 36
IQR = $Q_3 - Q_1$ = 12.5, 1.5 × IQR = 1.5 × 12.5 = 18.75,
Lower inner fence = Q_1 – 18.75 = 4.75
Upper inner fence = Q_3 + 18.75 = 54.75
The largest value exceeds the upper fence and so, is an outlier.



The data are skewed to the right.

3.81 **a.** $\bar{x} = (\Sigma x)/n$ = 1209/22 = $54.95 thousand
Median = average of the 11th and 12th terms of the ranked data set = (45+48)/2 = $46.5 thousand
The modes are 27, 40, 43, and 86 since they each occur twice and all other values once.

b. For the 10% trimmed mean, we must remove 0.10(22) = 2.2, or 2, values from each end of the ranked data set and average the remaining 18 values to get 971/18 = $53.94 thousand

c.

x	x^2
27	729
27	729
28	784
35	1225
36	1296
38	1444
40	1600
40	1600
43	1849
43	1849

45	2025
48	2304
50	2500
58	3364
62	3844
72	5184
77	5929
84	7056
86	7396
86	7396
90	8100
94	8836
$\Sigma x = 1209$	$\Sigma x^2 = 77{,}039$

Range = Largest value – Smallest value
= 94 – 27 = $67 thousand

$$s^2 = \frac{\sum x^2 - \frac{\left(\sum x\right)^2}{n}}{n-1} = \frac{77{,}039 - \frac{(1209)^2}{22}}{22-1} = 504.71 \qquad s = \sqrt{504.71} = \$22.47 \text{ thousand}$$

Coefficient of variation = $\frac{s}{\bar{x}} \times 100\% = \frac{22.47}{54.95} \times 100\% = 40.89\%$

3.82 **a.**

x	x^2
4	16
8	64
0	0
3	9
11	121
7	49
4	16
14	196
8	64
13	169
7	49
9	81
$\Sigma x = 88$	$\Sigma x^2 = 834$

$\bar{x} = (\Sigma x)/n = 88/12 = 7.33$ 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

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

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.

3.83 Weighted mean =

$$\frac{483(\$1630) + 1324(\$625) + 856(\$899) + 633(\$1178) + 394(\$1727) + 1138(\$923)}{4828}$$

$$= \frac{\$4,860,820}{4828} \approx \$1,006.80$$

3.84 Weighted mean =

$$\frac{3216(\$425) + 1828(\$1299) + 4036(\$369) + 3142(\$681) + 1662(\$1999)}{13,884}$$

$$= \frac{\$10,692,696}{13,884} \approx \$770.15$$

3.85

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
	$N = \Sigma f = 50$		$\Sigma mf = 254$	$\Sigma m^2f = 1626$

$\overline{x}$ = $(\Sigma mf)/n$ = 254/50 = 5.08 inches

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

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

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

$k = 40/20 = 2$ and $1 - \frac{1}{k^2} = 1 - \frac{1}{(2)^2} = 1 - 0.25 = 0.75$ 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 $1 - \frac{1}{k^2} = 1 - \frac{1}{(3)^2} = 1 - 0.11 = 0.89$ or 89%. Thus, at least 89% of the students will learn the basics in 140 to 260 minutes.

b. $1 - \frac{1}{k^2} = 0.84$ gives $\frac{1}{k^2} = 1 - 0.84 = 0.16$ or $k^2 = \frac{1}{0.16}$, so $k = 2.5$.
$\mu - 2.5\sigma = 200 - 2.5(20) = 150$ minutes and $\mu + 2.5\sigma = 200 + 2.5(20) = 250$ minutes. Thus, the required interval is 150 to 250 minutes.

3.87 **a.** **i.** Each of the two values is 15 minutes from $\mu = 30$ minutes. Hence,
$k = 15/6 = 2.5$ and $1 - \frac{1}{k^2} = 1 - \frac{1}{(2.5)^2} = 1 - 0.16 = 0.84$ or 84%. Thus, at least 84% of patients had waiting times between 15 and 45 minutes.

ii. Each of the two values is 18 minutes from $\mu = 30$ minutes. Hence, $k = 18/6 = 3$ and
$1 - \frac{1}{k^2} = 1 - \frac{1}{3^2} = 1 - 0.11 = 0.89$ or 89%. Thus, at least 89% of patients had waiting times between 12 and 48 minutes.

b. $1 - \frac{1}{k^2} = 0.75$ gives $\frac{1}{k^2} = 1 - 0.75 = 0.25$ or $k^2 = \frac{1}{0.25}$, so $k = 2$.
$\mu - 2\sigma = 30 - 2(6) = 18$ minutes and $\mu + 2\sigma = 30 + 2(6) = 42$ minutes.

Thus, the required interval is 18 to 42 minutes.

3.88 $\mu = 200$ minutes and $\sigma = 20$ 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.

3.89 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 74 falls between Q_1 and Q_2, which indicates that it is at least as large as 25% of the data and no larger than 50% of the data.

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.

3.90 The ranked data are:

27	27	28	35	36	38	40	40	43	43
45	48	50	58	62	72	77	84	86	86
90	94								

a. The quartiles are:

Q_1 = 6th term in ranked data set 38

Q_2 = average of 11th and 12th term in ranked data set = (45 + 48)/2 = 46.5

Q_3 = 17th term in ranked data set = 77

IQR = $Q_3 - Q_1$ = 77 – 38 = 39

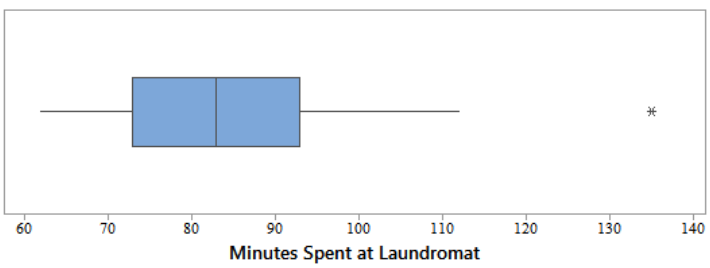
The value 77 is Q_3 which means it lies in the fourth 25% group from the bottom in the ranked data set.

So, it is at least as large as 75% of the data.

b. *kn*/100 = 18(22)/100 = 3.96
Thus, the 18th percentile can be approximated by the value of the 4th term in the ranked data, which is 35. Therefore, P_{18} = 35.

c. Fifteen values in the given data are smaller than 72. Hence, the percentile rank of 72 = (15/22) × 100 = 68%.

3.91 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 × IQR = 1.5 × 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.

3.92 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.

3.93 **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.

3.94 **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.

3.95 **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

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

Country A:

$$\frac{\text{number of deaths}}{\text{population}} \times 1000$$

$$= \frac{1}{40} \times 1000 = 25$$

Country B:

$$\frac{\text{number of deaths}}{\text{population}} \times 1000$$

$$= \frac{0.5}{25} \times 1000 = 20$$

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: $\frac{\text{number of deaths}}{\text{population}} \times 1000 = \frac{2}{20} \times 1000$

$$= 100$$

Country B:

$$\frac{\text{number of deaths}}{\text{population}} \times 1000 = \frac{3}{35} \times 1000$$

$$= 85.7$$

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: $\frac{\text{number of deaths}}{\text{population}} \times 1000 = \frac{3}{60} \times 1000$

$$= 50$$

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

$$= 58.3$$

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.

3.97 $\mu = 70$ and $\sigma = 10$

a. Using Chebyshev's theorem, we need to find k so that at least 1 – 0.50 = 0.5 of the scores are within k standard deviations of the mean.

$1 - \frac{1}{k^2} = 0.50$ gives $\frac{1}{k^2} = 1 - 0.50 = 0.50$ or $k^2 = \frac{1}{0.50} = 2$, so $k = \sqrt{2} \approx 1.41$.

Thus, at least 50% of the scores are within 1.41 standard deviations of the mean.

b. Using Chebyshev's theorem, we first find k so that at least 1 –0.20 =0 .80 of the scores are within k standard deviations of the mean.

$1 - \frac{1}{k^2} = 0.80$ gives $\frac{1}{k^2} = 1 - 0.80 = 0.20$ or $k^2 = \frac{1}{0.20} = 5$, so $k = \sqrt{5}$ » 2.24

Thus, at least 80% of the scores are within 2.24 standard deviations of the mean, but this means that at most 10% of the scores are greater than 2.24 standard deviations above the mean.

3.98 **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, $\sigma = 15$.

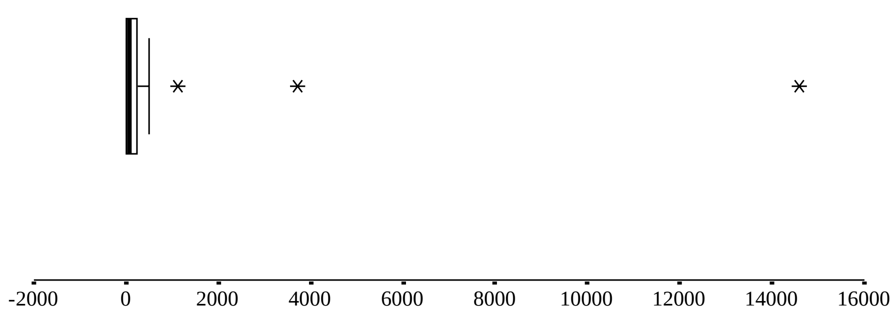
b. We know that 95% of the scores are between 60 and 80 and that $\mu = 70$. By the empirical rule, 95% of the scores fall in the interval $\mu - 2\sigma$ to $\mu + 2\sigma$. Then $60 = \mu - 2\sigma = 70 - 2\sigma$ and $80 = \mu + 2\sigma = 70 + 2\sigma$. Therefore, $10 = 2\sigma \Rightarrow \sigma = 5$.

3.99 **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.

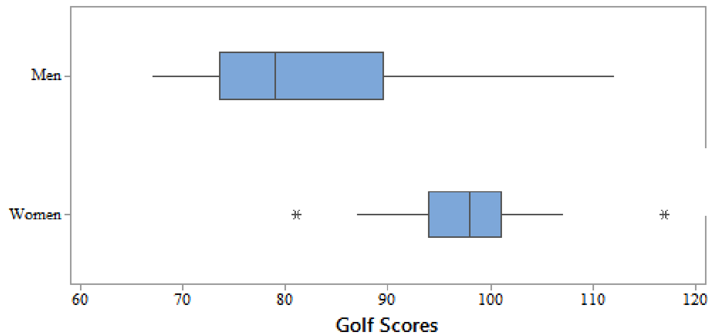
Below is the box–and–whisker plot for the given data.



The data are strongly skewed to the right.

d. 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.

3.100 **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	**Women**
$\bar{x} = 82$	$\bar{x} = 97.53$
Median = 79	Median = 98
Modes = 75, 79, and 92	Modes = 94 and 100
Range = 45	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.

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

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

3.102 For all students: $n = 44$, $\Sigma x = 6597$, $\Sigma x^2 = 1{,}030{,}639$, and median = 147.50 pounds

$\bar{x} = (\Sigma x)/n = 6597/44 = 149.93$ 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:

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

$\bar{x} = (\Sigma x)/n = 3848/22 = 174.91$ 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:

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

$\bar{x} = (\Sigma x)/n = 2749/22 = 124.95$ 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.

3.103 The 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.

3.104 $\bar{x}$ = 49.012 hours and s = 5.080 hours

a. For 75%: $1 - \frac{1}{k^2} = 0.75$ gives

$\frac{1}{k^2}$ = 1 – 0.75 = 0.25 or $k^2 = \frac{1}{0.25}$, 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%: $1 - \frac{1}{k^2} = 0.8889$ gives

$\frac{1}{k^2}$ = 1 – 0.8889 = .1111 or $k^2 = \frac{1}{0.1111}$, 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%: $1 - \frac{1}{k^2} = 0.9375$ gives

$\frac{1}{k^2}$ = 1 – 0.9375 = 0.0625 or $k^2 = \frac{1}{0.0625}$, so $k = 4$.
= 49.012 – 4(5.080) = 28.69 and $\bar{x} + 4s$ = 49.012 + 4(5.080) = 69.33
Thus, $\bar{x} - 4s$ 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
Twenty-eight or 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, 39 or 78% 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

$$1-\frac{1}{k^2}=1-\frac{1}{(1.907)^2}=1-0.2750$$

= 0.7250, or 72.50%.

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. **a.** $\bar{x}$ = $(\Sigma x)/n$ = 420/20 = 21 times
Median = average of the 10th and 11th terms of the ranked data set = (13 + 14)/2 = 13.5 times
The modes are 5, 8, and 14 since they each occur twice and all other values once.

b. For the 10% trimmed mean, we must remove 0.10(2) = 2 values from each end of the ranked data set and average the remaining 16 values to get 251/16 = 15.6875 times

c.

x	x^2
1	1
5	25
5	25
6	36
7	49
8	64
8	64
9	81
10	100
13	169
14	196
14	196
18	324
19	381
21	441
26	676
32	1024
41	1681
72	5184
91	8281
Σx = 420	Σx^2 = 18,978

Range = Largest value – Smallest value
= 91 – 1 = 90 times

$$s^2 = \frac{\sum x^2 - \frac{\left(\sum x\right)^2}{n}}{n-1} = \frac{18{,}978 - \frac{(420)^2}{20}}{20-1} \quad s = \sqrt{534.63} = 23.12 \text{ times}$$
$$= 534.63$$

d. Coefficient of variation = $\frac{s}{\bar{x}} \times 100\% = \frac{23.12}{21} \times 100\% = 110.11\%$

e. These are sample statistics because a subset of all people using debit cards was used, not ALL such people.

16. Weighted mean =
(2842($2055) + 4364($1165)
+ 3946($1459) + 1629($2734)
+ 3871($1672)) / 16,652 = $1,657.914

17. 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), then mean = (73 + 82 + 95 + 79 + 86 + 91)/6 = 84.33 points. This shows how an outlier can affect the value of the mean.

18. 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.

19. 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, Σx = 335, and $\Sigma x^2 = 22{,}445$.

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

20. **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, Σmf = 486.50, and $\Sigma m^2 f = 10{,}524.25$
$\bar{x}$ = $(\Sigma mf)/n$ = 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 computers

21. **a.** i. Each of the two values is 32.4 minutes from μ = 91.8 minutes. Hence, k = 32.4/16.2 = 2 and $1-\frac{1}{k^2}=1-\frac{1}{2^2}=1-0.25=0.75$ or 75%. Thus, 75% of the members spent between 59.4 and 124.2 minutes at the health club.

ii. Each of the two values is 40.5 minutes from μ = 91.8 minutes. Hence, k = 40.5/16.2 = 2.5 and $1-\frac{1}{k^2}=1-\frac{1}{2.5^2}=1-0.16=0.84$ or 84%. Thus, 84% of the members spent between 51.3 and 132.3 minutes at the health club.

b. $1-\frac{1}{k^2}=0.89 \Rightarrow 1-0.89=\frac{1}{k^2} \Rightarrow 0.11=\frac{1}{k^2} \Rightarrow k^2=\frac{1}{0.11} \Rightarrow k^2 \approx 9 \Rightarrow k=3$

$\mu-3\sigma=91.8-3(16.2)=43.2$ minutes and $\mu+3\sigma=91.8+3(16.2)=140.4$

Thus, the required interval is 43.2 minutes to 140.4 minutes.

22. μ = 7.3 years and σ = 2.2 years

a. i. The interval 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 0.7 to 13.9 years is $\mu-3\sigma$ to $\mu+3\sigma$. Hence, approximately 99.7% of the cars are 0.7 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.

23. The ranked data are:

45	48	49	50	52	54	55	56	56	58
61	63	64	66	70	74	77	79		

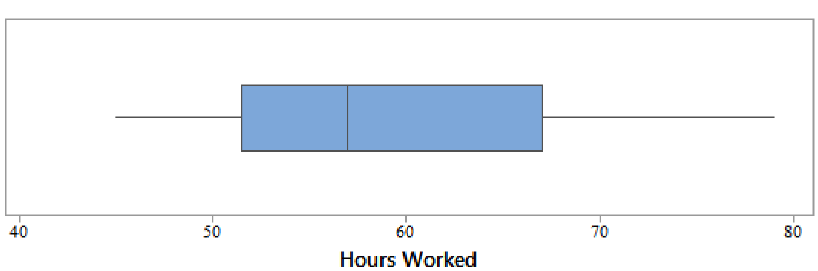
a. Median = (56 + 58)/2 = 57, Q_1 = 52, and Q_3 = 66
IQR = $Q_3 - Q_1$ = 66 – 52 = 14,

The value 54 lies between Q_1 and the Median, so it is in the second 25% group from the bottom of the ranked data set. This means that 25% of the data is less than 54 and that at least 50% of the data is larger than 54.

b. kn/100 = 60(18)/100 = 10.8
Thus, the 60th percentile can be approximated by the 11th term in the ranked data. Therefore, P_{60} = 61. This means that 60% of the values in the data set are less than 61.

c. Twelve values in the given data set are less than 64. Hence, the percentile rank of 64 = (12/18) × 100 = 66.7% or about 67%.

24.



The data is skewed to the right.

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

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

26. 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

27. **a.** For Data Set I: $\bar{x} = (\Sigma x)/n = 79/4 = 19.75$
For Data Set II: $\bar{x} = (\Sigma x)/n = 67/4 = 16.75$
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: $\Sigma x = 67$, $\Sigma x^2 = 1507$, 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.