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.

3.3 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.

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

3.7 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.

3.9 $\sum x = 5 + (-7) + 2 + 0 + (-9) + 16 + 10 + 7 = 24$
$\mu = (\sum x)/N = 24/8 = 3$
The ranked data are: −9 −7 0 2 5 7 10 16
Median = average of the 4th and 5th terms in ranked data set = (2 + 5)/2 = 3.50
This data set has no mode since all values occur exactly once.

3.11 $\bar{x} = (\sum x)/N = (205 + 265 + 176 + 314 + 243 + 192 + 297 + 357 + 238 + 281 + 342 + 259)/12 = 3{,}169/12 = 264.08$

The ranked data are: 176 192 205 238 243 259 265 281 297 314 342 357

Median = average of the 6th and 7th terms in ranked data set = 262

There is no mode since all values occur exactly once.

3.13 **a.** $\bar{x} = (\sum x)/n = (127 + 82 + 45 + 99 + 153 + 3261 + 77 + 108 + 68 + 278)/10 = 4298/10 = 429.80$ thousands of dollars.

The ranked data are: 45 68 77 82 99 108 127 153 278 3261

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 (68 + 77 + 82 + 99 + 108 + 127 + 153 + 278)/8 = 992/8 = 124 thousand dollars.

d. Median and trimmed mean are good measures to use here because of one outlier.

3.15 **a.** $\bar{x} = (\sum x)/n = (205 + 214 + 265 + 195 + 283 + 188 + 251 + 325 + 219 + 295)/10 = 2{,}440/10 = 244$ thousand dollars.

The ranked data are: 188 195 205 214 219 251 265 283 295 325

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 (195 + 205 + 214 + 219 + 251 + 265 + 283 + 295)/8 = 1,927/8 = 240.875 thousand dollars.

3.17 **a.** $\bar{x} = (\sum x)/n = (23 + 37 + 26 + 19 + 33 + 22 + 30 + 42 + 24 + 26 + 28 + 32 + 37 + 29 + 38 + 24 + 35 + 20 + 34 + 38)/20 = 597/20 = 29.85$ patients

The ranked data are: 19 20 22 23 24 24 26 26 28 29 30 32 33 34 35 37 37 38 38 42

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 (23 + 24 + 24 + 26 + 26 + 28 + 29 + 30 + 32 + 33 + 34 + 35 + 37 + 37)/14 = 418/14 = 29.86 patients.

3.19 The opinion that they will not allow their children to play football occurs the most often (360 times) and hence, is the mode.

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.

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

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.

3.27 Geometric mean =

$$\sqrt[n]{x_1 \cdot x_2 \cdot x_3 \cdot \ldots \cdot x_n} = \sqrt[5]{1.04\times1.03\times1.05\times1.06\times1.08} = \sqrt[5]{1.287625248} \approx 1.052$$

Then: Geometric mean – 1 = 1.052 – 1 = 0.052, so the mean inflation rate is 5.2%.

Section 3.2

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

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.

3.33 **a.** $\bar{x} = (\sum x)/n$ = (230 + 170 + 204 + 113 + 56 + 361 + 147)/7 = 1281/7 = \$183

Prices	Deviations from the Mean
\$230	230 – 183 = 47
\$170	170 – 183 = –13
\$204	204 – 183 = 21
\$113	113 – 183 = –70
\$56	56 – 183 = –127
\$361	361 – 183 = 178
\$147	147 – 183 = –36
	Sum = 0

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

b. $\sum x = 1281$, $\sum x^2 = 291{,}251$, and $n = 7$
Range = 361 – 56 = \$305

$$s^2 = \frac{\sum x^2 - \frac{(\sum x)^2}{n}}{n-1} = \frac{291{,}251 - \frac{(1281)^2}{7}}{7-1} \qquad s = \sqrt{9471.33} = \$97.32$$
$$= 9471.33$$

Coefficient of variation = (97.32/183) × 100% = 53.2%

3.35 a.

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

$\sum x = 1{,}113$ and $\sum x^2 = 64{,}871$ $n = 20$
Range = Largest value – Smallest value = 77 – 35 = 42 thousand dollars

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

$s = \sqrt{154.34} \approx 12.42$ 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.

3.37 **a.**

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

$\sum x = 2{,}651$ and $\sum x^2 = 356{,}107$ $n = 20$
Range = Largest value – Smallest value = 159 – 107 = 52 mmHg

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

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

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

3.39 **a.** $\sum x = 15 + 26 + 16 + 36 + 31 + 13 + 29 + 18 + 21 + 39 = 244$
$\sum x^2 = (15)^2 + (26)^2 + (16)^2 + (36)^2 + (31)^2 + (13)^2 + (29)^2 + (18)^2 + (21)^2 + (39)^2 =$ 6,710
$N = 10$
Range = Largest value – Smallest value = 39 – 13 = 26 minutes

$$\sigma^2 = \frac{\sum x^2 - \frac{(\sum x)^2}{N}}{N} = \frac{6{,}710 - \frac{(244)^2}{10}}{10} = 75.64$$

$$\sigma = \sqrt{75.64} = 8.70 \text{ minutes}$$

b. The coefficient of variation = $\frac{\sigma}{\bar{x}} \times 100\% = \frac{8.70}{244/10} \times 100\% \approx 35.66\%$

c. The large standard deviation suggests that the data are widely spread around the mean.

3.41 **a.** $\sum x = 127 + 82 + 45 + 99 + 153 + 3261 + 77 + 108 + 68 + 278 = 4{,}298$

$\sum x^2 = (127)^2 + (82)^2 + (45)^2 + (99)^2 + (153)^2 + (3261)^2 + (77)^2 + (108)^2 + (68)^2 + (278)^2$
$= 10{,}791{,}710$

$n = 10$

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 $= \frac{s}{\bar{x}} \times 100\% = \frac{996.91}{4{,}298/10} \times 100\% \approx 231.95\%$

3.43 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.

3.45 Data Set I: $\sum x = 12 + 25 + 37 + 8 + 41 = 123$

$\sum x^2 = (12)^2 + (25)^2 + (37)^2 + (8)^2 + (41)^2 = 3883$

Data Set II: $\sum x = 19 + 32 + 44 + 15 + 48 = 158$

$\sum x^2 = (19)^2 + (32)^2 + (44)^2 + (15)^2 + (48)^2 = 5850$

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

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

The standard deviations of the two data sets are equal. So, adding 7 to each corresponding value of the first data set, does not change the standard deviation.

Section 3.3

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

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

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

$$\sigma = \sqrt{12.5775} = 3.55$$

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
	$N = \sum f = 100$		$\sum mf = 1124$	$\sum m^2f = 16{,}272$

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

$$\sigma^2 = \frac{\sum m^2 f - \frac{\left(\sum mf\right)^2}{N}}{N} = \frac{16{,}272 - \frac{(1124)^2}{100}}{100} = 36.3824$$

$$\sigma = \sqrt{36.3824} = 6.03 \text{ hours}$$

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
	$n = \sum f = 50$		$\sum mf = 7500$	$\sum m^2f = 1{,}405{,}800$

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

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

$s = \sqrt{5730.6122} = \$75.70$

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 2019. The value $\sum mf = \$7500$ is the approximate total amount of the electric bills for all 50 families included in the sample.

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
	$n = \sum f = 400$		$\sum mf = 5211.5$	$\sum m^2 f = 99,127$

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

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

$s = \sqrt{78.2648} = 8.85$ hours

Section 3.4

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.

3.57 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 $\mu \pm 2\sigma = 230 \pm 82 = (148, 312)$.

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 $\mu \pm 2.5\sigma = 230 \pm 102.5 = (127.5, 332.5)$.

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 $\mu \pm 3\sigma = 230 \pm 123 = (107, 353)$.

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

3.61 **a.** **i.** Each of the two values is 20 minutes from μ = 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 μ = 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 $k \approx 3$.

$\mu - 3\sigma = 34 - 3(8) = 10$ and

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

Thus, the required interval is 10 to 58.

3.63 $\mu = 34$ and $\sigma = 8$

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

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 median of the (ranked) observations that are less than Q_2.
4. The third quartile, Q_3, is the value of the median of the (ranked) observations that are greater than 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.

3.67 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 value of the term that is approximately the ($kn/100$)th term in the ranking. If $kn/100$ falls between two consecutive integers a and b, use the b^{th} value in the ranking to obtain P_k.

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 71. Hence, the percentile rank of $71 = (4/13) \times 100 = 30.77\%$.

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 above Q_3, which indicates that it is in the top 25% group in the (ranked) data set.

b. $kn/100 = 91(40)/100 = 36.4$
Thus, the 91st percentile can be approximated by the 37^{th} 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 $40 = (12/40) \times 100 = 30\%$. Therefore, the number of text message was 40 or higher on 70% of the days.

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

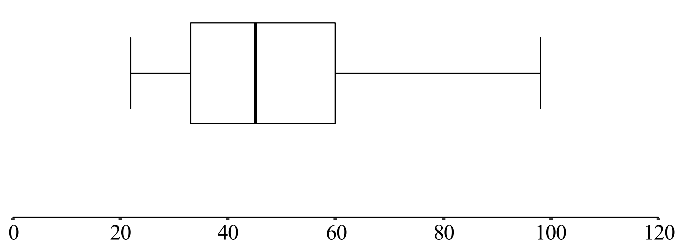
a. The quartiles are:
Q_1 = average of 5th and 6th term in ranked data set = (44+ 45)/2 = 44.5
Q_2 = average of 10th and 11th term in ranked data set = (57 + 58)/2 = 57.5
Q_3 = average of 15th and 16th term in ranked data set = (64 + 64)/2 = 64
IQR = $Q_3 - Q_1$ = 64 – 44.5 = 19.5
The value 57 lies between 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 30th percentile is the value of the 6th 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%.

<u>Section 3.6</u>

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.

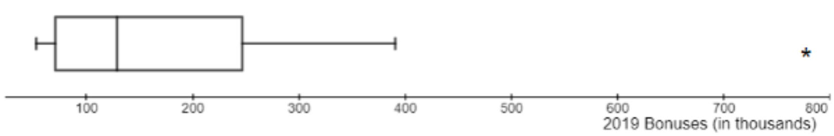


3.77 The ranked data are:

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 = $Q_3 - Q_1 = 187$,
$1.5 \times$ IQR = $1.5 \times 187 = 280.5$,
Lower inner fence = $Q_1 - 280.25 = -209.5$
Upper inner fence = $Q_3 + \ 280.5 = 538.5$
The data are skewed to the right (that is, toward larger values).

The smallest and the largest values within the two inner fences are 53 and 391, respectively. The largest value exceeds the upper fence and so, is an outlier.

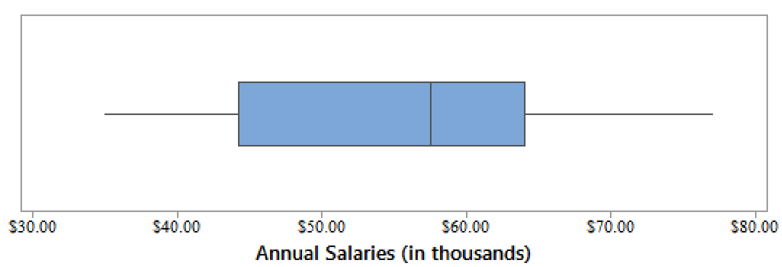


3.79 The ranked data are:

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

Median = $(57 + 58)/2 = 57.5$, $Q_1 = (44 + 45)/2 = 44.5$ and $Q_3 = (64 + 64)/2 = 64$
IQR = $Q_3 - Q_1 = 19.5$,
$1.5 \times$ IQR = $1.5 \times 19.5 = 29.25$,
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 slightly skewed to the left because the lower 50% of the values are spread over a larger range than the upper 50% of the values.

3.81 **a.** $\bar{x} = (\sum x)/n = 1129/20 = \$56.45$ thousand

Median = average of the 10^{th} and 11^{th} terms of the ranked data set = (48+50)/2 = \$49 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(20) = 2 values from each end of the ranked data set and average the remaining 16 values to get 891/16 = \$55.69 thousand

c.

x	x^2
27	729
27	729
28	784
36	1296
38	1444
40	1600
40	1600
43	1849
43	1849
48	2304
50	2500
58	3364
62	3844
72	5184
77	5929
84	7056
86	7396
86	7396
90	8100
94	8836
$\sum x = 1129$	$\sum x^2 = 73,789$

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

$$\sigma^2 = \frac{\sum x^2 - \frac{\left(\sum x\right)^2}{N}}{N} = \frac{73,789 - \frac{(1129)^2}{20}}{20} = 502.8475$$

$\sigma = \sqrt{502.8475} = \$22.42$ thousand

Coefficient of variation = $\frac{\sigma}{\bar{x}} \times 100\% = \frac{22.42}{56.45} \times 100\% = 39.72\%$

3.83 Weighted mean =

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

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

3.85

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
	$n = \sum f = 50$		$\sum mf = 254$	$\sum m^2f = 1626$

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

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

$s = \sqrt{6.8506} = 2.62$ inches

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

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.

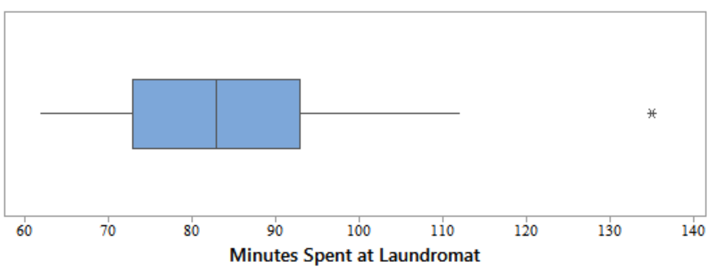
3.89 The ranked data are: 56 59 60 68 74 78 84 97 107 382

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.

3.91 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 $Q_3 = 93$,
$IQR = Q_3 - Q_1 = 93 - 73 = 20$
$1.5 \times IQR = 1.5 \times 20 = 30$,
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.

Advanced Exercises

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.

3.95 **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

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

a. Using Chebyshev's theorem, we need to find k so that at least 50% 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.

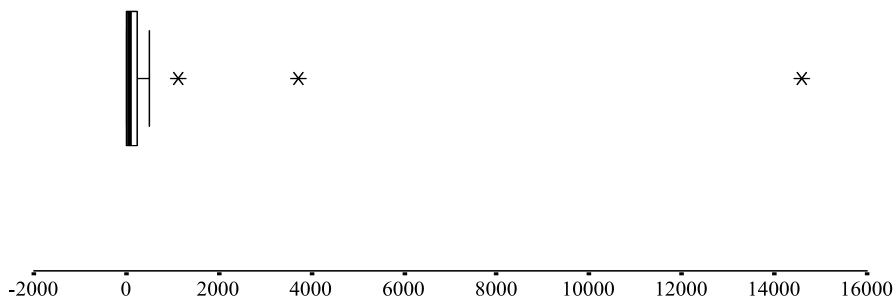
$1 - \frac{1}{k^2} = 0.80$ gives $\frac{1}{k^2} = 1 - 0.80 = 0.20$ or $k^2 = \frac{1}{0.20} = 5$, so $k = \sqrt{5}$ » 2.24

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.

3.99 **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 high 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.

3.101 **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 $x = 567 - 504 = 63$.

3.103 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.

<u>Self–Review Test</u>

1. b

3. c

5. b

7. a

9. b

11. b

13. a

15. **a.** $\bar{x} = (\sum x)/n = 420/20 = 21$ 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(20) = 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
$\sum x = 420$	$\sum x^2 = 18{,}978$

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

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

$s = \sqrt{534.63} = 23.12$ times

d. Coefficient of variation = $\frac{s}{\overline{x}} \times 100\% = \frac{23.12}{21} \times 100\% = 110.10\%$

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

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.

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: $n = 5$, $\sum x = 335$, and $\sum x^2 =$ 22,445.

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

21. **a.** i. Each of the two values is 32.4 minutes from $\mu = 91.8$ 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 $\mu = 91.8$ 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.

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%.

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

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

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 $n = 4$

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

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.