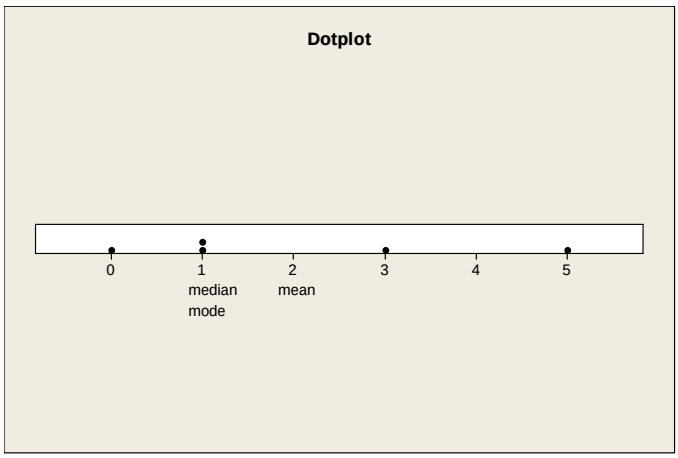
2: Describing Data with Numerical Measures

2.1 **a** The dotplot shown below plots the five measurements along the horizontal axis. Since there are two "1"s, the corresponding dots are placed one above the other. The approximate center of the data appears to be around 1.



b The mean is the sum of the measurements divided by the number of measurements, or

$$\bar{x} = \frac{\sum x_i}{n} = \frac{0+5+1+1+3}{5} = \frac{10}{5} = 2$$

To calculate the median, the observations are first ranked from smallest to largest: 0, 1, 1, 3, 5. Then since $n = 5$, the position of the median is $0.5(n+1) = 3$, and the median is the 3rd ranked measurement, or $m = 1$.

The mode is the measurement occurring most frequently, or mode = 1.

c The three measures in part **b** are located on the dotplot. Since the median and mode are to the left of the mean, we conclude that the measurements are skewed to the right.

2.2 **a** The mean is

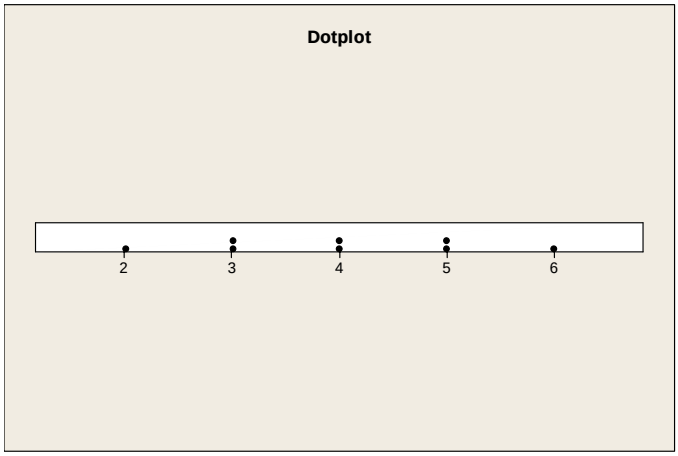
$$\bar{x} = \frac{\sum x_i}{n} = \frac{3+2+\cdots+5}{8} = \frac{32}{8} = 4$$

b To calculate the median, the observations are first ranked from smallest to largest:

2, 3, 3, 4, 4, 5, 5, 6

Since $n = 8$ is even, the position of the median is $0.5(n+1) = 4.5$, and the median is the average of the 4th and 5th measurements, or $m = (4+4)/2 = 4$.

c Since the mean and the median are equal, we conclude that the measurements are symmetric. The dotplot shown below confirms this conclusion.



2.3 **a** $\bar{x} = \dfrac{\sum x_i}{n} = \dfrac{58}{10} = 5.8$

b The ranked observations are: 2, 3, 4, 5, 5, 6, 6, 8, 9, 10. Since $n = 10$, the median is halfway between the 5th and 6th ordered observations, or $m = (5+6)/2 = 5.5$.

c There are two measurements, 5 and 6, which both occur twice. Since this is the highest frequency of occurrence for the data set, we say that the set is *bimodal* with modes at 5 and 6.

2.4 **a** $\bar{x} = \dfrac{\sum x_i}{n} = \dfrac{9455}{4} = 2363.75$ **b** $\bar{x} = \dfrac{\sum x_i}{n} = \dfrac{8280}{4} = 2070$

c The average premium cost in several different cities is not as important to the consumer as the average cost for a variety of consumers in his or her geographical area.

2.5 **a** Although there may be a few households who own more than one DVD player, the majority should own either 0 or 1. The distribution should be slightly skewed to the right.

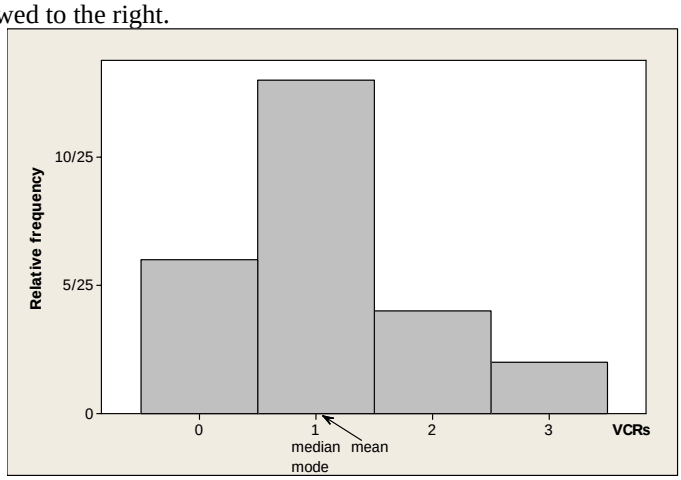
b Since most households will have only one DVD player, we guess that the mode is 1.

c The mean is

$$\bar{x} = \frac{\sum x_i}{n} = \frac{1+0+\cdots+1}{25} = \frac{27}{25} = 1.08$$

To calculate the median, the observations are first ranked from smallest to largest: There are six 0s, thirteen 1s, four 2s, and two 3s. Then since $n = 25$, the position of the median is $0.5(n+1) = 13$, which is the 13th ranked measurement, or $m = 1$. The mode is the measurement occurring most frequently, or mode = 1.

d The relative frequency histogram is shown below, with the three measures superimposed. Notice that the mean falls slightly to the right of the median and mode, indicating that the measurements are slightly skewed to the right.



2.6 **a** The stem and leaf plot below was generated by ***Minitab***. It is skewed to the right.

Stem-and-Leaf Display: Revenues

```
Stem-and-leaf of Revenues  N  = 10
Leaf Unit = 10000

 2   0  33
(4)  0  5555
 4   0
 4   0  889
 1   1
 1   1
 1   1
 1   1
 1   1  9
```

b The mean is

$$\bar{x} = \frac{\sum x_i}{n} = \frac{192604 + 91134 + \cdots + 38416}{10} = \frac{736951}{10} = 73{,}695.10$$

To calculate the median, notice that the observations are already ranked from smallest to largest. Then since $n = 10$, the position of the median is $0.5(n+1) = 5.5$, the average of the 5th and 6th ranked measurements or $m = (54848 + 52620)/2 = 53{,}734$.

c Since the mean is strongly affected by outliers, the median would be a better measure of center for this data set.

2.7 It is obvious that any one family cannot have 2.5 children, since the number of children per family is a quantitative discrete variable. The researcher is referring to the average number of children per family calculated for all families in the United States during the 1930s. The average does not necessarily have to be integer-valued.

2.8 **a** Similar to previous exercises. The mean is

$$\bar{x} = \frac{\sum x_i}{n} = \frac{0.99 + 1.92 + \cdots + 0.66}{14} = \frac{12.55}{14} = 0.896$$

b To calculate the median, rank the observations from smallest to largest. The position of the median is $0.5(n+1) = 7.5$, and the median is the average of the 7th and 8th ranked measurement or $m = (0.67 + 0.69)/2 = 0.68$.

c Since the mean is slightly larger than the median, the distribution is slightly skewed to the right.

2.9 The distribution of sports salaries will be skewed to the right, because of the very high salaries of some sports figures. Hence, the median salary would be a better measure of center than the mean.

2.10 **a** Similar to previous exercises.

$$\bar{x} = \frac{\sum x_i}{n} = \frac{2150}{10} = 215$$

b The ranked observations are shown below:

175	225
185	230
190	240
190	250
200	265

The position of the median is $0.5(n+1) = 5.5$ and the median is the average of the 5th and 6th observation or

$$\frac{200 + 225}{2} = 212.5$$

c Since there are no unusually large or small observations to affect the value of the mean, we would probably report the mean or average time on task.

2.11 **a** Similar to previous exercises.

$$\bar{x} = \frac{\sum x_i}{n} = \frac{85}{18} = 4.72$$

The ranked observations are:

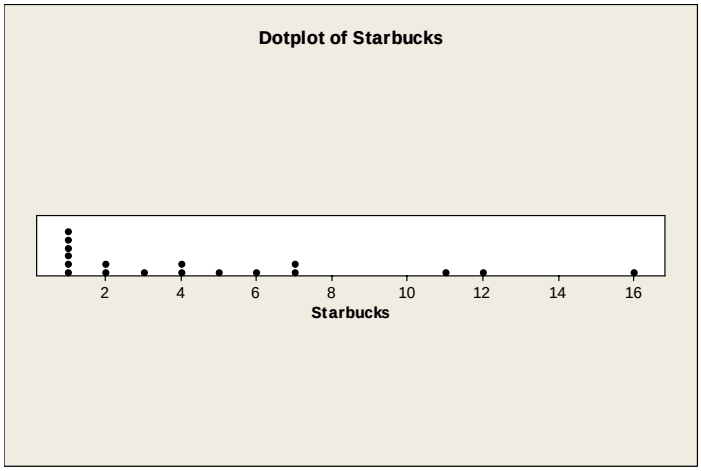
1	1	3	6	12
1	1	4	7	16
1	2	4	7	
1	2	5	11	

The median is the average of the 9th and 10th observations or

$m = (3 + 4)/2 = 3.5$ and the mode is the most frequently occurring observation—mode = 1.

b Since the mean is larger than the median, the data are skewed to the right.

c The dotplot is shown below. The distribution is skewed to the right.



2.12 **a** $\bar{x} = \dfrac{\sum x_i}{n} = \dfrac{19850}{10} = 1985$

b The ranked data are: 1200, 1300, 1350, 1500, 1800, 2000, 2200, 2600, 2900, 3000 and the median is the average of the 5th and 6th observations or

$$m = \frac{1800 + 2000}{2} = 1900$$

c Average cost would not be as important as many other variables, such as picture quality, sound quality, size, lowest cost for the best quality, and many other considerations.

2.13 **a** $\bar{x} = \dfrac{\sum x_i}{n} = \dfrac{12}{5} = 2.4$

b Create a table of differences, $(x_i - \bar{x})$ and their squares, $(x_i - \bar{x})^2$.

x_i	$x_i - \bar{x}$	$(x_i - \bar{x})^2$
2	–0.4	0.16
1	–1.4	1.96
1	–1.4	1.96
3	0.6	0.36
5	2.6	6.76
Total	0	11.20

Then

$$s^2 = \frac{\sum (x_i - \bar{x})^2}{n-1} = \frac{(2-2.4)^2 + \cdots + (5-2.4)^2}{4} = \frac{11.20}{4} = 2.8$$

c The sample standard deviation is the positive square root of the variance or

$$s = \sqrt{s^2} = \sqrt{2.8} = 1.673$$

d Calculate $\sum x_i^2 = 2^2 + 1^2 + \cdots + 5^2 = 40$. Then

$$s^2 = \frac{\sum x_i^2 - \dfrac{(\sum x_i)^2}{n}}{n-1} = \frac{40 - \dfrac{(12)^2}{5}}{4} = \frac{11.2}{4} = 2.8 \text{ and } s = \sqrt{s^2} = \sqrt{2.8} = 1.673.$$

The results of parts **a** and **b** are identical.

2.14 The results will vary from student to student, depending on their particular type of calculator. The results should agree with Exercise 2.13.

2.15 **a** The range is $R = 4 - 1 = 3$. **b** $\bar{x} = \dfrac{\sum x_i}{n} = \dfrac{17}{8} = 2.125$

c Calculate $\sum x_i^2 = 4^2 + 1^2 + \cdots + 2^2 = 45$. Then

$$s^2 = \frac{\sum x_i^2 - \dfrac{\left(\sum x_i\right)^2}{n}}{n-1} = \frac{45 - \dfrac{(17)^2}{8}}{7} = \frac{8.875}{7} = 1.2679 \text{ and } s = \sqrt{s^2} = \sqrt{1.2679} = 1.126.$$

2.16 **a** The range is $R = 6 - 1 = 5$. **b** $\bar{x} = \dfrac{\sum x_i}{n} = \dfrac{31}{8} = 3.875$

c Calculate $\sum x_i^2 = 3^2 + 1^2 + \cdots + 5^2 = 137$. Then

$$s^2 = \frac{\sum x_i^2 - \dfrac{\left(\sum x_i\right)^2}{n}}{n-1} = \frac{137 - \dfrac{(31)^2}{8}}{7} = \frac{16.875}{7} = 2.4107$$

and $s = \sqrt{s^2} = \sqrt{2.4107} = 1.55$.

d The range, $R = 5$, is $5/1.55 = 3.23$ standard deviations.

2.17 **a** The range is $R = 2.39 - 1.28 = 1.11$.

b Calculate $\sum x_i^2 = 1.28^2 + 2.39^2 + \cdots + 1.51^2 = 15.415$. Then

$$s^2 = \frac{\sum x_i^2 - \dfrac{\left(\sum x_i\right)^2}{n}}{n-1} = \frac{15.415 - \dfrac{(8.56)^2}{5}}{4} = \frac{.76028}{4} = .19007$$

and $s = \sqrt{s^2} = \sqrt{.19007} = .436$

c The range, $R = 1.11$, is $1.11/.436 = 2.5$ standard deviations.

2.18 **a** The range is $R = 343.50 - 162.64 = 180.86$. **b** $\bar{x} = \dfrac{\sum x_i}{n} = \dfrac{2940.2}{12} = 245.02$

c Calculate $\sum x_i^2 = 266.63^2 + 163.41^2 + \cdots + 230.46^2 = 763773.912$. Then

$$s^2 = \frac{\sum x_i^2 - \dfrac{\left(\sum x_i\right)^2}{n}}{n-1} = \frac{763773.912 - \dfrac{(2940.2)^2}{12}}{11} = 3943.264424$$

and $s = \sqrt{s^2} = \sqrt{3943.264424} = 62.795$.

2.19 **a** The range of the data is $R = 6 - 1 = 5$ and the range approximation with $n = 10$ is

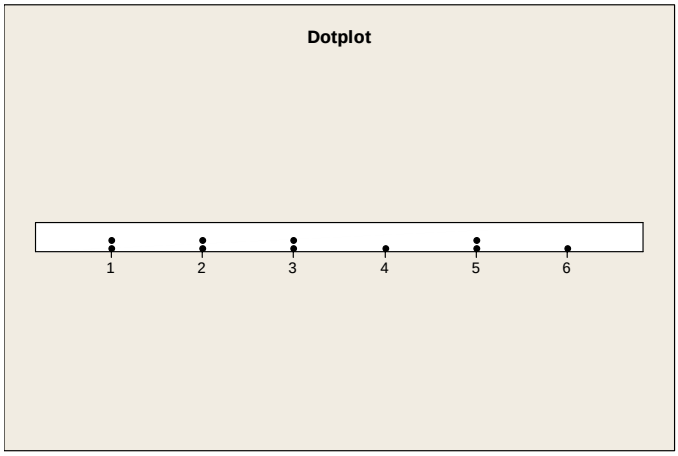
$$s \approx \frac{R}{3} = 1.67$$

b The standard deviation of the sample is

$$s = \sqrt{s^2} = \sqrt{\frac{\sum x_i^2 - \dfrac{\left(\sum x_i\right)^2}{n}}{n-1}} = \sqrt{\frac{130 - \dfrac{(32)^2}{10}}{9}} = \sqrt{3.0667} = 1.751$$

which is very close to the estimate for part **a**.

c-e From the dotplot on the next page, you can see that the data set is not mound-shaped. Hence you can use Tchebysheff's Theorem, but not the Empirical Rule to describe the data.



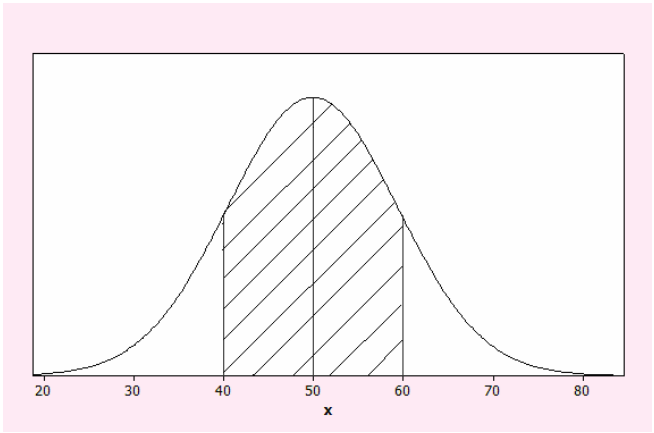
2.20 **a** First calculate the intervals:

$$\bar{x} \pm s = 36 \pm 3 \text{ or } 33 \text{ to } 39$$
$$\bar{x} \pm 2s = 36 \pm 6 \text{ or } 30 \text{ to } 42$$
$$\bar{x} \pm 3s = 36 \pm 9 \text{ or } 27 \text{ to } 45$$

According to the Empirical Rule, approximately 68% of the measurements will fall in the interval 33 to 39; approximately 95% of the measurements will fall between 30 and 42; approximately 99.7% of the measurements will fall between 27 and 45.

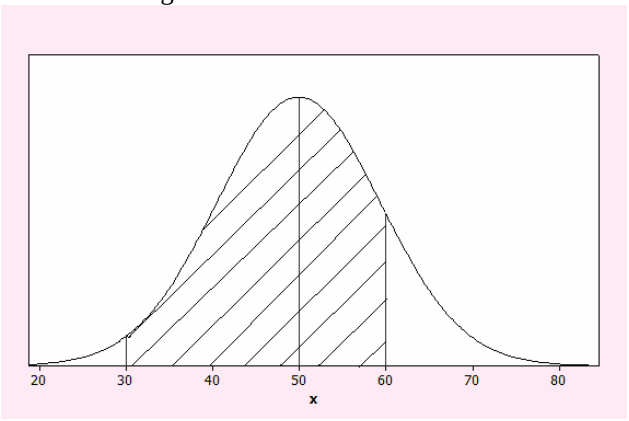
b If no prior information as to the shape of the distribution is available, we use Tchebysheff's Theorem. We would expect at least $\left(1-1/1^2\right) = 0$ of the measurements to fall in the interval 33 to 39; at least $\left(1-1/2^2\right) = 3/4$ of the measurements to fall in the interval 30 to 42; at least $\left(1-1/3^2\right) = 8/9$ of the measurements to fall in the interval 27 to 45.

2.21 **a** The interval from 40 to 60 represents $\mu \pm \sigma = 50 \pm 10$. Since the distribution is relatively mound-shaped, the proportion of measurements between 40 and 60 is 68% according to the Empirical Rule and is shown below.



b Again, using the Empirical Rule, the interval $\mu \pm 2\sigma = 50 \pm 2(10)$ or between 30 and 70 contains approximately 95% of the measurements.

c Refer to the figure below.



Since approximately 68% of the measurements are between 40 and 60, the symmetry of the distribution implies that 34% of the measurements are between 50 and 60. Similarly, since 95% of the measurements are between 30 and 70, approximately 47.5% are between 30 and 50. Thus, the proportion of measurements between 30 and 60 is

$$0.34 + 0.475 = 0.815$$

d From the figure in part **a**, the proportion of the measurements between 50 and 60 is 0.34 and the proportion of the measurements which are greater than 50 is 0.50. Therefore, the proportion that are greater than 60 must be

$$0.5 - 0.34 = 0.16$$

2.22 Since nothing is known about the shape of the data distribution, you must use Tchebysheff's Theorem to describe the data.

a The interval from 60 to 90 represents $\mu \pm 3\sigma$ which will contain at least 8/9 of the measurements.

b The interval from 65 to 85 represents $\mu \pm 2\sigma$ which will contain at least 3/4 of the measurements.

c The value $x = 65$ lies two standard deviations below the mean. Since at least 3/4 of the measurements are within two standard deviation range, *at most* 1/4 can lie outside this range, which means that at most 1/4 can be less than 65.

2.23 **a** The range of the data is $R = 1.1 - 0.5 = 0.6$ and the approximate value of s is

$$s \approx \frac{R}{3} = 0.2$$

b Calculate $\sum x_i = 7.6$ and $\sum x_i^2 = 6.02$, the sample mean is

$$\overline{x} = \frac{\sum x_i}{n} = \frac{7.6}{10} = .76$$

and the standard deviation of the sample is

$$s = \sqrt{s^2} = \sqrt{\frac{\sum x_i^2 - \frac{\left(\sum x_i\right)^2}{n}}{n-1}} = \sqrt{\frac{6.02 - \frac{(7.6)^2}{10}}{9}} = \sqrt{\frac{0.244}{9}} = 0.165$$

which is very close to the estimate from part **a**.

2.24 **a** The stem and leaf plot generated by ***Minitab*** shows that the data is roughly mound-shaped. Note however the gap in the center of the distribution and the two measurements in the upper tail.

Stem-and-Leaf Display: Weight

```
Stem-and-leaf of Weight  N  = 27
Leaf Unit = 0.010

 1   7   5
 2   8   3
 6   8   7999
 8   9   23
 13  9   66789
 13  10
(3)  10  688
 11  11  2244
 7   11  788
 4   12  4
 3   12  8
 2   13
 2   13  8
 1   14  1
```

b Calculate $\sum x_i = 28.41$ and $\sum x_i^2 = 30.6071$, the sample mean is

$$\overline{x} = \frac{\sum x_i}{n} = \frac{28.41}{27} = 1.052$$

and the standard deviation of the sample is

$$s = \sqrt{s^2} = \sqrt{\frac{\sum x_i^2 - \frac{\left(\sum x_i\right)^2}{n}}{n-1}} = \sqrt{\frac{30.6071 - \frac{(28.41)^2}{27}}{26}} = 0.166$$

c The following table gives the actual percentage of measurements falling in the intervals $\overline{x} \pm ks$ for $k = 1, 2, 3$.

k	$\overline{x} \pm ks$	Interval	Number in Interval	Percentage
1	1.052 ± 0.166	0.866 to 1.218	21	78%
2	1.052 ± 0.332	0.720 to 1.384	26	96%
3	1.052 ± 0.498	0.554 to 1.550	27	100%

d The percentages in part **c** do not agree too closely with those given by the Empirical Rule, especially in the one standard deviation range. This is caused by the lack of mounding (indicated by the gap) in the center of the distribution.

e The lack of any one-pound packages is probably a marketing technique intentionally used by the supermarket. People who buy slightly less than one-pound would be drawn by the slightly lower price, while those who need exactly one-pound of meat for their recipe might tend to opt for the larger package, increasing the store's profit.

2.25 According to the Empirical Rule, if a distribution of measurements is approximately mound-shaped,

a approximately 68% or 0.68 of the measurements fall in the interval $\mu \pm \sigma = 12 \pm 2.3$ or 9.7 to 14.3

b approximately 95% or 0.95 of the measurements fall in the interval $\mu \pm 2\sigma = 12 \pm 4.6$ or 7.4 to 16.6

c approximately 99.7% or 0.997 of the measurements fall in the interval $\mu \pm 3\sigma = 12 \pm 6.9$ or 5.1 to 18.9

Therefore, approximately 0.3% or 0.003 will fall outside this interval.

2.26 **a** The stem and leaf plots are shown below. The second set has a slightly higher location and spread.

Stem-and-Leaf Display: Method 1, Method 2

```
Stem-and-leaf of Method 1  N  = 10    Stem-and-leaf of Method 2  N  = 10
Leaf Unit = 0.00010                   Leaf Unit = 0.00010

 1   10  0
 3   11  00                           1  11  0
 4   12  0                            3  12  00
(4)  13  0000                         5  13  00
 2   14  0                            5  14  0
 1   15  0                            4  15  00
                                      2  16  0
                                      1  17  0
```

b *Method 1*: Calculate $\sum x_i = 0.125$ and $\sum x_i^2 = 0.001583$. Then $\overline{x} = \dfrac{\sum x_i}{n} = 0.0125$ and

$$s = \sqrt{s^2} = \sqrt{\frac{\sum x_i^2 - \dfrac{\left(\sum x_i\right)^2}{n}}{n-1}} = \sqrt{\frac{0.001583 - \dfrac{(0.125)^2}{10}}{9}} = 0.00151$$

Method 2: Calculate $\sum x_i = 0.138$ and $\sum x_i^2 = 0.001938$. Then $\overline{x} = \dfrac{\sum x_i}{n} = 0.0138$ and

$$s = \sqrt{s^2} = \sqrt{\frac{\sum x_i^2 - \dfrac{\left(\sum x_i\right)^2}{n}}{n-1}} = \sqrt{\frac{0.001938 - \dfrac{(0.138)^2}{10}}{9}} = 0.00193$$

The results confirm the conclusions of part **a**.

2.27 **a** The center of the distribution should be approximately halfway between 0 and 9 or $(0+9)/2 = 4.5$.

b The range of the data is $R = 9 - 0 = 9$. Using the range approximation, $s \approx R/4 = 9/4 = 2.25$.

c Using the data entry method the students should find $\overline{x} = 4.586$ and $s = 2.892$, which are fairly close to our approximations.

2.28 **a** Similar to previous exercises. The intervals, counts and percentages are shown in the table.

k	$\overline{x} \pm ks$	Interval	Number in Interval	Percentage
1	4.586 ± 2.892	1.694 to 7.478	43	61%
2	4.586 ± 5.784	–1.198 to 10.370	70	100%
3	4.586 ± 8.676	–4.090 to 13.262	70	100%

b The percentages in part **a** do not agree with those given by the Empirical Rule. This is because the shape of the distribution is not mound-shaped, but flat.

2.29 **a** Although most of the animals will die at around 32 days, there may be a few animals that survive a very long time, even with the infection. The distribution will probably be skewed right.

b Using Tchebysheff's Theorem, at least 3/4 of the measurements should be in the interval $\mu \pm \sigma \Rightarrow 32 \pm 72$ or 0 to 104 days.

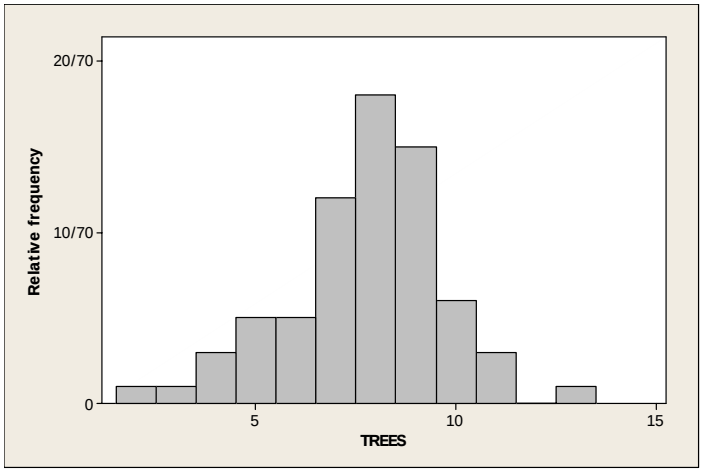
2.30 **a** The value of x is $\mu - \sigma = 32 - 36 = -4$.

b The interval $\mu \pm \sigma$ is 32 ± 36 should contain approximately $(100 - 68) = 34\%$ of the survival times, of which 17% will be longer than 68 days and 17% less than –4 days.

c The latter is clearly impossible. Therefore, the approximate values given by the Empirical Rule are not accurate, indicating that the distribution cannot be mound-shaped.

2.31 **a** We choose to use 12 classes of length 1.0. The tally and the relative frequency histogram follow.

Class i	Class Boundaries	Tally	f_i	Relative frequency, f_i/n
1	2 to < 3	1	1	1/70
2	3 to < 4	1	1	1/70
3	4 to < 5	111	3	3/70
4	5 to < 6	11111	5	5/70
5	6 to < 7	11111	5	5/70
6	7 to < 8	11111 11111 11	12	12/70
7	8 to < 9	11111 11111 11111 111	18	18/70
8	9 to < 10	11111 11111 11111	15	15/70
9	10 to < 11	11111 1	6	6/70
10	11 to < 12	111	3	3/70
11	12 to < 13		0	0
12	13 to < 14	1	1	1/70



b Calculate $n = 70,\ \sum x_i = 541$ and $\sum x_i^2 = 4453$. Then $\overline{x} = \dfrac{\sum x_i}{n} = \dfrac{541}{70} = 7.729$ is an estimate of μ.

c The sample standard deviation is

$$s = \sqrt{\frac{\sum x_i^2 - \dfrac{\left(\sum x_i\right)^2}{n}}{n-1}} = \sqrt{\frac{4453 - \dfrac{(541)^2}{70}}{69}} = \sqrt{3.9398} = 1.985$$

The three intervals, $\overline{x} \pm ks$ for $k = 1, 2, 3$ are calculated below. The table shows the actual percentage of measurements falling in a particular interval as well as the percentage predicted by Tchebysheff's Theorem and the Empirical Rule. Note that the Empirical Rule should be fairly accurate, as indicated by the mound-shape of the histogram in part **a**.

k	$\overline{x} \pm ks$	Interval	Fraction in Interval	Tchebysheff	Empirical Rule
1	7.729 ± 1.985	5.744 to 9.714	50/70 = 0.71	at least 0	≈ 0.68
2	7.729 ± 3.970	3.759 to 11.699	67/70 = 0.96	at least 0.75	≈ 0.95
3	7.729 ± 5.955	1.774 to 13.684	70/70 = 1.00	at least 0.89	≈ 0.997

2.32 **a** Calculate $R = 1.92 - 0.53 = 1.39$ so that $s \approx R/4 = 1.39/4 = 0.3475$.

b Calculate $n = 14,\ \sum x_i = 12.55$ and $\sum x_i^2 = 13.3253$. Then

$$s^2 = \frac{\sum x_i^2 - \dfrac{\left(\sum x_i\right)^2}{n}}{n-1} = \frac{13.3253 - \dfrac{(12.55)^2}{14}}{13} = 0.1596 \text{ and } s = \sqrt{0.15962} = 0.3995$$

which is fairly close to the approximate value of s from part **a**.

2.33 **a-b** Calculate $R = 93 - 51 = 42$ so that $s \approx R/4 = 42/4 = 10.5$.

c Calculate $n = 30$, $\sum x_i = 2145$ and $\sum x_i^2 = 158,345$. Then

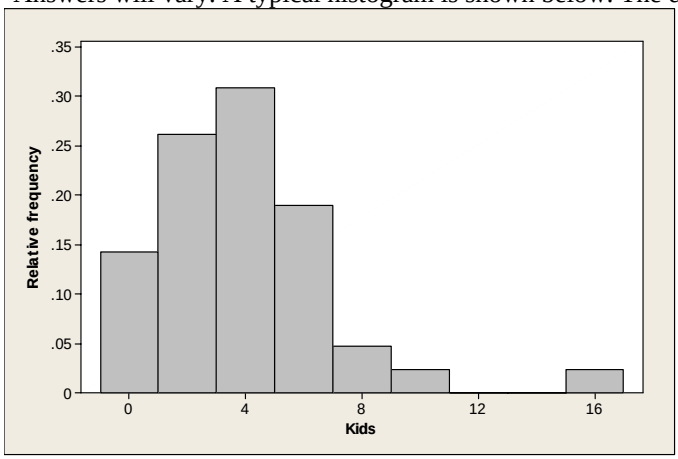
$$s^2 = \frac{\sum x_i^2 - \frac{\left(\sum x_i\right)^2}{n}}{n-1} = \frac{158,345 - \frac{(2145)^2}{30}}{29} = 171.6379 \text{ and } s = \sqrt{171.6379} = 13.101$$

which is fairly close to the approximate value of *s* from part **b**.

d The two intervals are calculated below. The proportions agree with Tchebysheff's Theorem, but are not to close to the percentages given by the Empirical Rule. (This is because the distribution is not quite mound-shaped.)

k	$\bar{x} \pm ks$	Interval	Fraction in Interval	Tchebysheff	Empirical Rule
2	71.5 ± 26.20	45.3 to 97.7	30/30 = 1.00	at least 0.75	≈ 0.95
3	71.5 ± 39.30	32.2 to 110.80	30/30 = 1.00	at least 0.89	≈ 0.997

2.34 **a** Answers will vary. A typical histogram is shown below. The distribution is skewed to the right.



b Calculate $n = 42$, $\sum x_i = 151$ and $\sum x_i^2 = 897$. Then

$$\bar{x} = \frac{\sum x_i}{n} = \frac{151}{42} = 3.60,$$

$$s^2 = \frac{\sum x_i^2 - \frac{\left(\sum x_i\right)^2}{n}}{n-1} = \frac{897 - \frac{(151)^2}{42}}{41} = 8.63705$$

and $s = \sqrt{8.63705} = 2.94$

c The three intervals, $\bar{x} \pm ks$ for $k = 1, 2, 3$ are calculated below. The table shows the actual percentage of measurements falling in a particular interval as well as the percentage predicted by Tchebysheff's Theorem and the Empirical Rule. Note that the Empirical Rule is not very accurate for the first interval, since the histogram in part **a** is skewed.

k	$\bar{x} \pm ks$	Interval	Fraction in Interval	Tchebysheff	Empirical Rule
1	3.60 ± 2.94	.66 to 6.54	32/42 = .76	at least 0	≈ 0.68
2	3.60 ± 5.88	–2.28 to 9.48	40/42 = .95	at least 0.75	≈ 0.95
3	3.60 ± 8.82	–5.22 to 12.42	41/42 = .976	at least 0.89	≈ 0.997

2.35 **a** Calculate $R = 2.39 - 1.28 = 1.11$ so that $s \approx R/2.5 = 1.11/2.5 = .444$.

b In Exercise 2.17, we calculated $\sum x_i = 8.56$ and $\sum x_i^2 = 1.28^2 + 2.39^2 + \cdots + 1.51^2 = 15.415$. Then

$$s^2 = \frac{\sum x_i^2 - \frac{\left(\sum x_i\right)^2}{n}}{n-1} = \frac{15.451 - \frac{(8.56)^2}{5}}{4} = \frac{.76028}{4} = .19007$$

and $s = \sqrt{s^2} = \sqrt{.19007} = .436$, which is very close to our estimate in part **a**.

2.36 **a** Answers will vary. A typical stem and leaf plot is generated by ***Minitab.***

Stem-and-Leaf Display: Favre

```
Stem-and-leaf of Favre  N  = 16
Leaf Unit = 1.0

 1   0  5
 1   1
 4   1  579
(8)  2  01222244
 4   2  568
 1   3  1
```

b Calculate $n = 16,\ \sum x_i = 343$ and $\sum x_i^2 = 7875$. Then $\overline{x} = \dfrac{\sum x_i}{n} = \dfrac{343}{16} = 21.44$,

$$s^2 = \frac{\sum x_i^2 - \dfrac{\left(\sum x_i\right)^2}{n}}{n-1} = \frac{7875 - \dfrac{(343)^2}{16}}{15} = 34.79583$$

and $s = \sqrt{s^2} = \sqrt{34.795833} = 5.90$.

c Calculate $\overline{x} \pm 2s \Rightarrow 21.44 \pm 11.80$ or 9.64 to 33.24. From the original data set, 15 of the 16 measurements, or about 94% fall in this interval.

2.37 **a** Calculate $n = 15,\ \sum x_i = 21$ and $\sum x_i^2 = 49$. Then $\overline{x} = \dfrac{\sum x_i}{n} = \dfrac{21}{15} = 1.4$ and

$$s^2 = \frac{\sum x_i^2 - \dfrac{\left(\sum x_i\right)^2}{n}}{n-1} = \frac{49 - \dfrac{(21)^2}{15}}{14} = 1.4$$

b Using the frequency table and the grouped formulas, calculate

$$\sum x_i f_i = 0(4) + 1(5) + 2(2) + 3(4) = 21$$

$$\sum x_i^2 f_i = 0^2(4) + 1^2(5) + 2^2(2) + 3^2(4) = 49$$

Then, as in part **a**,

$$\overline{x} = \frac{\sum x_i f_i}{n} = \frac{21}{15} = 1.4$$

$$s^2 = \frac{\sum x_i^2 f_i - \dfrac{\left(\sum x_i f_i\right)^2}{n}}{n-1} = \frac{49 - \dfrac{(21)^2}{15}}{14} = 1.4$$

2.38 Use the formulas for grouped data given in Exercise 2.37. Calculate $n = 17$, $\sum x_i f_i = 79$, and $\sum x_i^2 f_i = 393$. Then,

$$\overline{x} = \frac{\sum x_i f_i}{n} = \frac{79}{17} = 4.65$$

$$s^2 = \frac{\sum x_i^2 f_i - \dfrac{\left(\sum x_i f_i\right)^2}{n}}{n-1} = \frac{393 - \dfrac{(79)^2}{17}}{16} = 1.6176 \text{ and } s = \sqrt{1.6176} = 1.27$$

2.39 **a** The data in this exercise have been arranged in a frequency table.

x_i	0	1	2	3	4	5	6	7	8	9	10
f_i	10	5	3	2	1	1	1	0	0	1	1

Using the frequency table and the grouped formulas, calculate

$$\sum x_i f_i = 0(10) + 1(5) + \cdots + 10(1) = 51$$

$$\sum x_i^2 f_i = 0^2(10) + 1^2(5) + \cdots + 10^2(1) = 293$$

Then

$$\bar{x} = \frac{\sum x_i f_i}{n} = \frac{51}{25} = 2.04$$

$$s^2 = \frac{\sum x_i^2 f_i - \frac{\left(\sum x_i f_i\right)^2}{n}}{n-1} = \frac{293 - \frac{(51)^2}{25}}{24} = 7.873 \text{ and } s = \sqrt{7.873} = 2.806.$$

b-c The three intervals $\bar{x} \pm ks$ for $k = 1, 2, 3$ are calculated in the table along with the actual proportion of measurements falling in the intervals. Tchebysheff's Theorem is satisfied and the approximation given by the Empirical Rule are fairly close for $k = 2$ and $k = 3$.

k	$\bar{x} \pm ks$	Interval	Fraction in Interval	Tchebysheff	Empirical Rule
1	2.04 ± 2.806	–0.766 to 4.846	21/25 = 0.84	at least 0	≈ 0.68
2	2.04 ± 5.612	–3.572 to 7.652	23/25 = 0.92	at least 0.75	≈ 0.95
3	2.04 ± 8.418	–6.378 to 10.458	25/25 = 1.00	at least 0.89	≈ 0.997

2.40 The sorted data set, along with the positions of the quartiles and the quartiles themselves are shown in the table.

Sorted Data Set	***n***	**Position of Q_1**	**Position of Q_3**	**Lower quartile, Q_1**	**Upper quartile, Q_3**
.13, 16, .21, .28, .34, .76, .88	7	.25(8) = 2	.75(8) = 6	.16	.76
1.0, 1.7, 2.0, 2.1, 2.3, 2.8, 2.9, 4.4, 5.1, 6.5, 8.8	11	.25(12) = 3	.75(12) = 9	2.0	5.1

2.41 The data have already been sorted. Find the positions of the quartiles, and the measurements that are just above and below those positions. Then find the quartiles by interpolation.

Sorted Data Set	**Position of Q_1**	**Above and below**	**Q_1**	**Position of Q_3**	**Above and below**	**Q_3**
1, 1.5, 2, 2, 2.2	.25(6) = 1.5	1 and 1.5	1.25	.75(6) = 4.5	2 and 2.2	2.1
0, 1.7, 1.8, 3.1, 3.2, 7, 8, 8.8, 8.9, 9, 10	.25(12) = 3	None	1.8	.75(12) = 9	None	8.9
.23, .30, .35, .41, .56, .58, .76, .80	.25(9) = 2.25	.30 and .35	.30 + .25(.05) = .3125	.75(9) = 6.75	.58 and .76	.58 + .75(.18) = .7150

2.42 The ordered data are:

0, 1, 3, 4, 4, 5, 6, 6, 7, 7, 8

a With $n = 12$, the median is in position $0.5(n+1) = 6.5$, or halfway between the 6th and 7th observations. The lower quartile is in position $0.25(n+1) = 3.25$ (one-fourth of the way between the 3rd and 4th observations) and the upper quartile is in position $0.75(n+1) = 9.75$ (three-fourths of the way between the 9th and 10th observations). Hence, $m = (5+6)/2 = 5.5$, $Q_1 = 3 + 0.25(4-3) = 3.25$ and $Q_3 = 6 + 0.75(7-6) = 6.75$. Then the five-number summary is

Min	Q_1	Median	Q_3	Max
0	3.25	5.5	6.75	8

and

$$IQR = Q_3 - Q_1 = 6.75 - 3.25 = 3.50$$

b Calculate $n = 12$, $\sum x_i = 57$ and $\sum x_i^2 = 337$. Then $\bar{x} = \dfrac{\sum x_i}{n} = \dfrac{57}{12} = 4.75$ and the sample standard deviation is

$$s = \sqrt{\frac{\sum x_i^2 - \dfrac{\left(\sum x_i\right)^2}{n}}{n-1}} = \sqrt{\frac{337 - \dfrac{(57)^2}{12}}{11}} = \sqrt{6.022727} = 2.454$$

c For the smaller observation, $x = 0$,

$$z\text{-score} = \frac{x - \bar{x}}{s} = \frac{0 - 4.75}{2.454} = -1.94$$

and for the largest observation, $x = 8$,

$$z\text{-score} = \frac{x - \bar{x}}{s} = \frac{8 - 4.75}{2.454} = 1.32$$

Since neither *z*-score exceeds 2 in absolute value, none of the observations are unusually small or large.

2.43 The ordered data are:

0, 1, 5, 6, 7, 8, 9, 10, 12, 12, 13, 14, 16, 19, 19

With $n = 15$, the median is in position $0.5(n+1) = 8$, so that $m = 10$. The lower quartile is in position $0.25(n+1) = 4$ so that $Q_1 = 6$ and the upper quartile is in position $0.75(n+1) = 12$ so that $Q_3 = 14$. Then the five-number summary is

Min	Q_1	Median	Q_3	Max
0	6	10	14	19

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

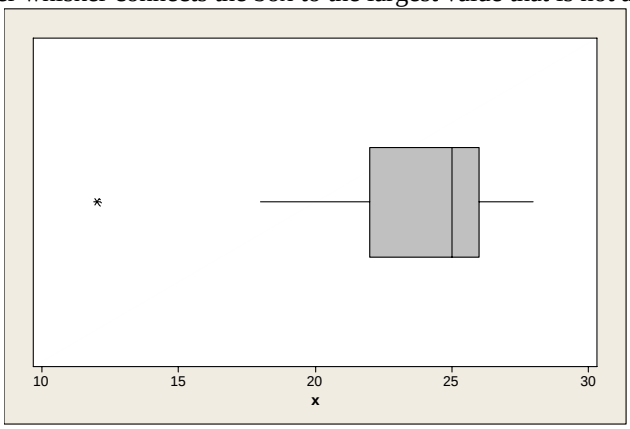
2.44 The ordered data are:

12, 18, 22, 23, 24, 25, 25, 26, 26, 27, 28

For $n = 11$, the position of the median is $0.5(n+1) = 0.5(11+1) = 6$ and $m = 25$. The positions of the quartiles are $0.25(n+1) = 3$ and $0.75(n+1) = 9$, so that $Q_1 = 22$, $Q_3 = 26$, and $IQR = 26 - 22 = 4$.
The *lower and upper fences* are:

$$Q_1 - 1.5IQR = 22 - 6 = 16$$
$$Q_3 + 1.5IQR = 26 + 6 = 32$$

The only observation falling outside the fences is $x = 12$ which is identified as an outlier. The box plot is shown below. The lower whisker connects the box to the smallest value that is not an outlier, $x = 18$. The upper whisker connects the box to the largest value that is not an outlier or $x = 28$.



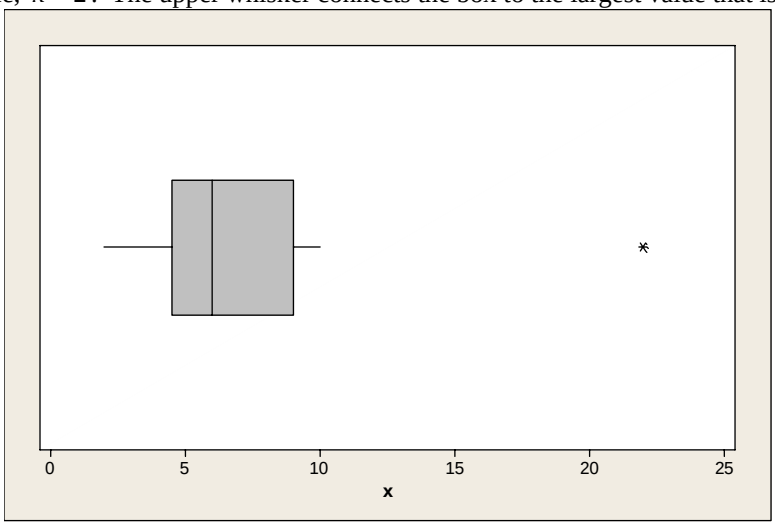
2.45 The ordered data are:

2, 3, 4, 5, 6, 6, 6, 7, 8, 9, 9, 10, 22

For $n = 13$, the position of the median is $0.5(n+1) = 0.5(13+1) = 7$ and $m = 6$. The positions of the quartiles are $0.25(n+1) = 3.5$ and $0.75(n+1) = 10.5$, so that $Q_1 = 4.5$, $Q_3 = 9$, and $IQR = 9 - 4.5 = 4.5$. The *lower and upper fences* are:

$$Q_1 - 1.5IQR = 4.5 - 6.75 = -2.25$$
$$Q_3 + 1.5IQR = 9 + 6.75 = 15.75$$

The value $x = 22$ lies outside the upper fence and is an outlier. The box plot is shown below. The lower whisker connects the box to the smallest value that is not an outlier, which happens to be the minimum value, $x = 2$. The upper whisker connects the box to the largest value that is not an outlier or $x = 10$.



2.46 From Section 2.6, the 69th percentile implies that 69% of all students scored below your score, and only 31% scored higher.

2.47 **a** The ordered data are shown below:

1.70	101.00	209.00	264.00	316.00	445.00
1.72	118.00	218.00	278.00	318.00	481.00
5.90	168.00	221.00	286.00	329.00	485.00
8.80	180.00	241.00	314.00	397.00	
85.40	183.00	252.00	315.00	406.00	

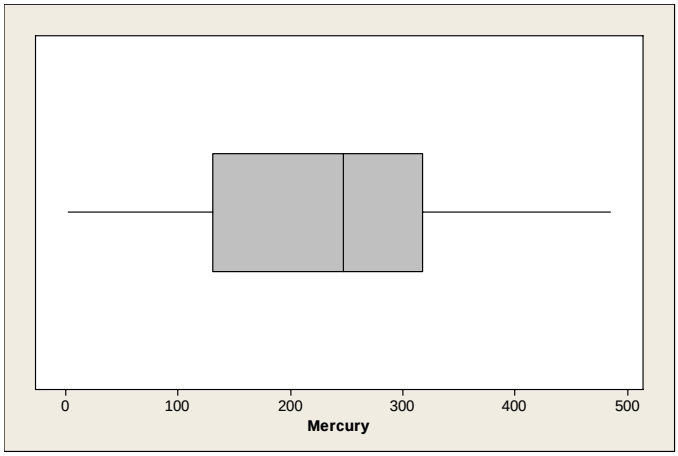
For $n = 28$, the position of the median is $0.5(n+1) = 14.5$ and the positions of the quartiles are $0.25(n+1) = 7.25$ and $0.75(n+1) = 21.75$. The lower quartile is ¼ the way between the 7th and 8th measurements or $Q_1 = 118 + 0.25(168 - 118) = 130.5$ and the upper quartile is ¾ the way between the 21st and 22nd measurements or $Q_3 = 316 + 0.75(318 - 316) = 317.5$. Then the five-number summary is

Min	Q_1	Median	Q_3	Max
1.70	130.5	246.5	317.5	485

b Calculate $IQR = Q_3 - Q_1 = 317.5 - 130.5 = 187$. Then the *lower and upper fences* are:

$$Q_1 - 1.5IQR = 130.5 - 280.5 = -150$$
$$Q_3 + 1.5IQR = 317.5 + 280.5 = 598$$

The box plot is shown below. Since there are no outliers, the whiskers connect the box to the minimum and maximum values in the ordered set.



c-d The boxplot does not identify any of the measurements as outliers, mainly because the large variation in the measurements cause the IQR to be large. However, the student should notice the extreme difference in the magnitude of the first four observations taken on young dolphins. These animals have not been alive long enough to accumulate a large amount of mercury in their bodies.

2.48 **a** See Exercise 2.24b.

b For $x = 1.38$,

$$z\text{-score} = \frac{x - \bar{x}}{s} = \frac{1.38 - 1.05}{0.17} = 1.94$$

while for $x = 1.41$,

$$z\text{-score} = \frac{x - \bar{x}}{s} = \frac{1.41 - 1.05}{0.17} = 2.12$$

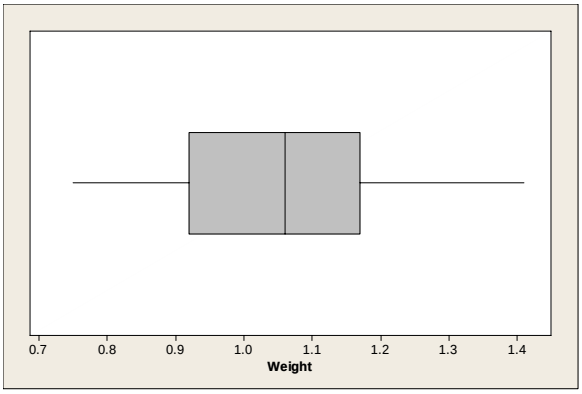
The value $x = 1.41$ would be considered somewhat unusual, since its *z*-score exceeds 2 in absolute value.

c For $n = 27$, the position of the median is $0.5(n+1) = 0.5(27+1) = 14$ and $m = 1.06$. The positions of the quartiles are $0.25(n+1) = 7$ and $0.75(n+1) = 21$, so that $Q_1 = 0.92$, $Q_3 = 1.17$, and $IQR = 1.17 - 0.92 = 0.25$.

The *lower and upper fences* are:

$$Q_1 - 1.5IQR = 0.92 - 0.375 = 0.545$$
$$Q_3 + 1.5IQR = 1.17 + 0.375 = 1.545$$

The box plot is shown below. Since there are no outliers, the whiskers connect the box to the minimum and maximum values in the ordered set.



Since the median line is almost in the center of the box, the whiskers are nearly the same lengths, the data set is relatively symmetric.

2.49 **a** For $n = 16$, the position of the median is $0.5(n+1) = 8.5$ and the positions of the quartiles are $0.25(n+1) = 4.25$ and $0.75(n+1) = 12.75$. The lower quartile is ¼ the way between the 4th and 5th measurements and the upper quartile is ¾ the way between the 12th and 13th measurements. The sorted measurements are shown below.

Favre: 5, 15, 17, 19, 20, 21, 22, 22, 22, 22, 24, 24, 25, 26, 28, 31

Peyton Manning: 14, 14, 20, 20, 21, 21, 21, 22, 25, 25, 25, 26, 27, 29, 30, 32

For Brett Favre,

$m = (22+22)/2 = 22$, $Q_1 = 19 + 0.25(20-19) = 19.25$ and $Q_3 = 24 + 0.75(25-24) = 24.75$.

For Peyton Manning,

$m = (22+25)/2 = 23.50$, $Q_1 = 20 + 0.25(21-20) = 20.25$ and $Q_3 = 26 + 0.75(27-26) = 26.75$.

Then the five-number summaries are

	Min	Q_1	Median	Q_3	Max
Favre	5	19.25	22	24.75	31
Manning	14	20.25	23.5	26.75	32

b For Brett Favre, calculate $IQR = Q_3 - Q_1 = 24.75 - 19.25 = 5.5$. Then the *lower and upper fences* are:

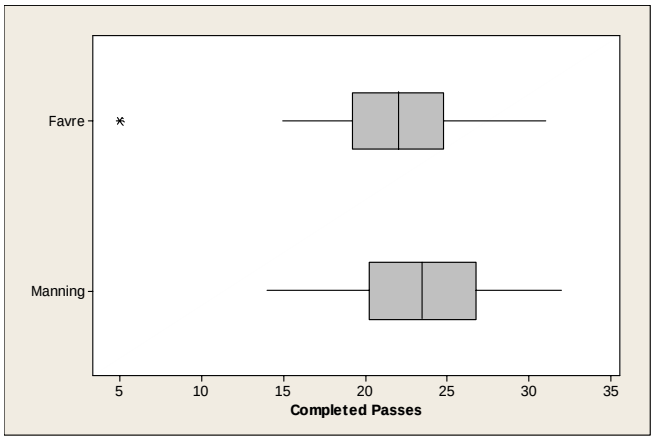
$$Q_1 - 1.5IQR = 19.25 - 8.25 = 11$$
$$Q_3 + 1.5IQR = 24.75 + 8.25 = 33$$

and $x = 5$ is an outlier.

For Peyton Manning, calculate $IQR = Q_3 - Q_1 = 26.75 - 20.25 = 6.5$. Then the *lower and upper fences* are:

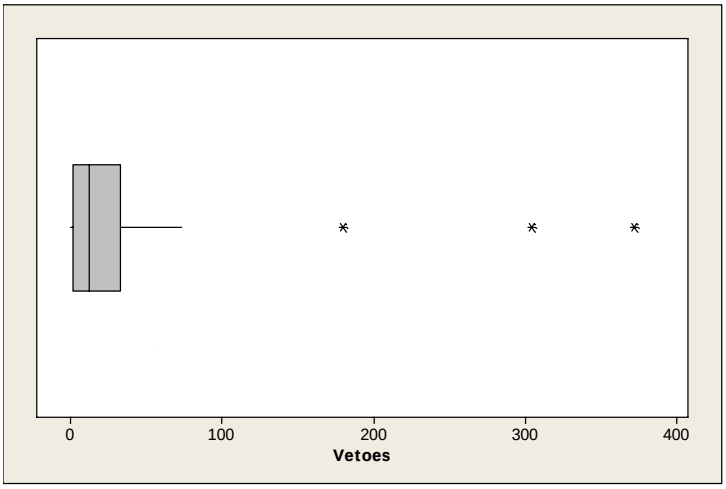
$$Q_1 - 1.5IQR = 20.25 - 9.75 = 10.5$$
$$Q_3 + 1.5IQR = 26.75 + 9.75 = 36.5$$

and there are no outliers. The box plots are shown below.



c Answers will vary. The Favre distribution is relatively symmetric except for the one outlier; the Manning distribution is roughly symmetric, probably mound-shaped. The Manning distribution is slightly more variable; Manning has a higher median number of completed passes.

2.50 Answers will vary from student to student. The distribution is skewed to the right with three outliers (Truman, Cleveland and F. Roosevelt). The box plot is shown on the next page.



2.51 **a** Just by scanning through the 20 measurements, it seems that there are a few unusually small measurements, which would indicate a distribution that is skewed to the left.

b The position of the median is $0.5(n+1) = 0.5(25+1) = 10.5$ and $m = (120+127)/2 = 123.5$. The mean is

$$\overline{x} = \frac{\sum x_i}{n} = \frac{2163}{20} = 108.15$$

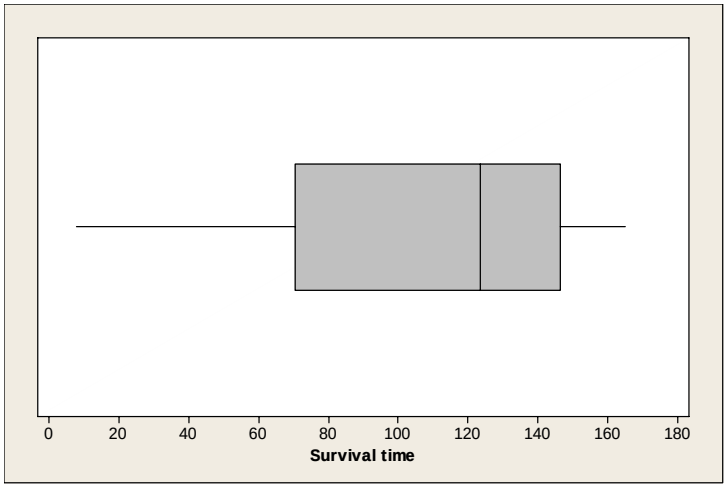
which is smaller than the median, indicate a distribution skewed to the left.

c The positions of the quartiles are $0.25(n+1) = 5.25$ and $0.75(n+1) = 15.75$, so that $Q_1 = 65 - .25(87-65) = 70.5$, $Q_3 = 144 + .75(147-144) = 146.25$, and $IQR = 146.25 - 70.5 = 75.75$.

The *lower and upper fences* are:

$$Q_1 - 1.5IQR = 70.5 - 113.625 = -43.125$$
$$Q_3 + 1.5IQR = 146.25 + 113.625 = 259.875$$

The box plot is shown below. There are no outliers. The long left whisker and the median line located to the right of the center of the box indicates that the distribution that is skewed to the left.



2.52 **a** The sorted data is:

162.64, 163.41, 187.16, 208.99, 219.41, 226.80
230.46, 266.63, 289.17, 306.55, 335.48, 343.50

The positions of the median and the quartiles are $0.5(n+1) = 6.5$, $0.25(n+1) = 3.25$ and $0.75(n+1) = 9.75$,

$m = (226.80 + 230.46)/2 = 228.63$

so that $Q_1 = 187.16 + .25(208.99 - 187.16) = 192.6175$

$Q_3 = 289.17 + .75(306.55 - 289.17) = 302.205$

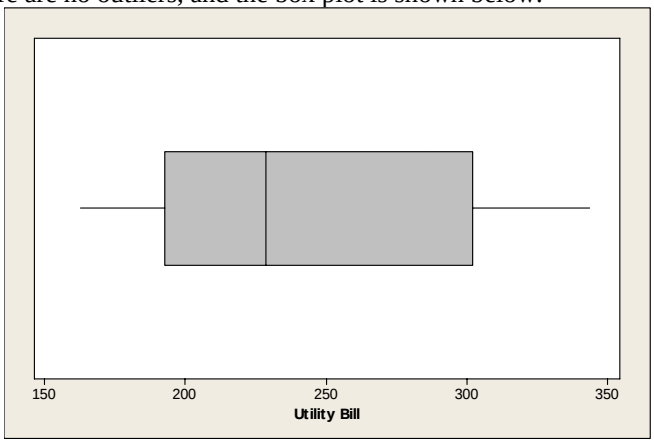
and $IQR = 302.205 - 192.6175 = 109.5875$.

The *lower and upper fences* are:

$$Q_1 - 1.5IQR = 192.6175 - 164.38125 = 28.236$$

$$Q_3 + 1.5IQR = 302.205 + 164.38125 = 466.586$$

There are no outliers, and the box plot is shown below.



b Because of the slightly longer right whisker and the median line to the left of center, the distribution is slightly skewed to the right.

2.53 Answers will vary. The student should notice the outliers in the female group, and that the median female temperature is higher than the median male temperature.

2.54 **a** Calculate $n = 14$, $\sum x_i = 367$ and $\sum x_i^2 = 9641$. Then $\bar{x} = \dfrac{\sum x_i}{n} = \dfrac{367}{14} = 26.214$ and

$$s = \sqrt{\frac{\sum x_i^2 - \dfrac{\left(\sum x_i\right)^2}{n}}{n-1}} = \sqrt{\frac{9641 - \dfrac{(367)^2}{14}}{13}} = 1.251$$

b Calculate $n = 14$, $\sum x_i = 366$ and $\sum x_i^2 = 9644$. Then $\bar{x} = \dfrac{\sum x_i}{n} = \dfrac{366}{14} = 26.143$ and

$$s = \sqrt{\frac{\sum x_i^2 - \dfrac{\left(\sum x_i\right)^2}{n}}{n-1}} = \sqrt{\frac{9644 - \dfrac{(366)^2}{14}}{13}} = 2.413$$

c The centers are roughly the same; the Sunmaid raisins appear slightly more variable.

2.55 **a** The ordered sets are shown below:

Generic					Sunmaid				
24	25	25	25	26	22	24	24	24	24
26	26	26	26	27	25	25	27	28	28
27	28	28	28		28	28	29	30	

For $n = 14$, the position of the median is $0.5(n+1) = 0.5(14+1) = 7.5$ and the positions of the quartiles are $0.25(n+1) = 3.75$ and $0.75(n+1) = 11.25$, so that

Generic: $m = 26$, $Q_1 = 25$, $Q_3 = 27.25$, and $IQR = 27.25 - 25 = 2.25$

Sunmaid: $m = 26$, $Q_1 = 24$, $Q_3 = 28$, and $IQR = 28 - 24 = 4$

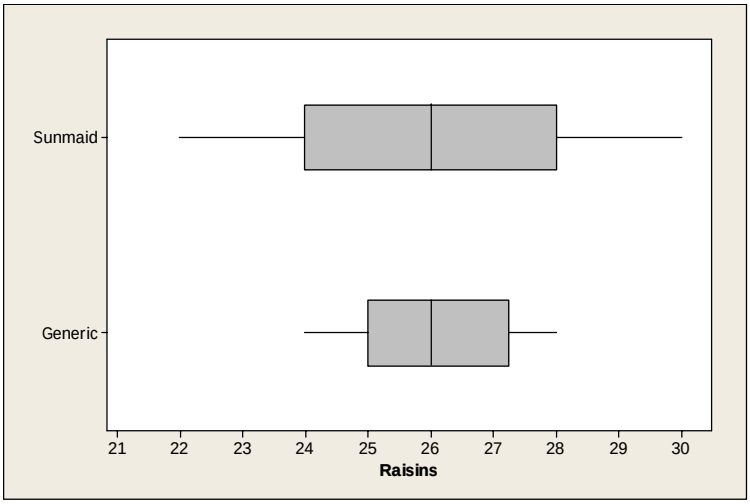
b **Generic:** *Lower and upper fences* are:

$$Q_1 - 1.5IQR = 25 - 3.375 = 21.625$$
$$Q_3 + 1.5IQR = 27.25 + 3.375 = 30.625$$

Sunmaid: *Lower and upper fences* are:

$$Q_1 - 1.5IQR = 24 - 6 = 18$$
$$Q_3 + 1.5IQR = 28 + 6 = 34$$

The box plots are shown below. There are no outliers.



d If the boxes are not being underfilled, the average size of the raisins is roughly the same for the two brands. However, since the number of raisins is more variable for the Sunmaid brand, it would appear that some of the Sunmaid raisins are large while others are small. The individual sizes of the generic raisins are not as variable.

2.56 **a** Calculate the range as $R = 15 - 1 = 14$. Using the range approximation, $s \approx R/4 = 14/4 = 3.5$.

b Calculate $n = 25$, $\sum x_i = 155.5$ and $\sum x_i^2 = 1260.75$. Then

$$\bar{x} = \frac{\sum x_i}{n} = \frac{155.5}{25} = 6.22 \text{ and}$$
$$s = \sqrt{\frac{\sum x_i^2 - \frac{\left(\sum x_i\right)^2}{n}}{n-1}} = \sqrt{\frac{1260.75 - \frac{(155.5)^2}{25}}{24}} = 3.497$$

which is very close to the approximation found in part **a**.

c Calculate $\bar{x} \pm 2s = 6.22 \pm 6.994$ or –0.774 to 13.214. From the original data, 24 measurements or $(24/25)100 = 96\%$ of the measurements fall in this interval. This is close to the percentage given by the Empirical Rule.

2.57 **a** The largest observation found in the data from Exercise 1.26 is 32.3, while the smallest is 0.2. Therefore the range is $R = 32.3 - 0.2 = 32.1$.

b Using the range, the approximate value for *s* is: $s \approx R/4 = 32.1/4 = 8.025$.

c Calculate $n = 50$, $\sum x_i = 418.4$ and $\sum x_i^2 = 6384.34$. Then

$$s = \sqrt{\frac{\sum x_i^2 - \frac{\left(\sum x_i\right)^2}{n}}{n-1}} = \sqrt{\frac{6384.34 - \frac{(418.4)^2}{50}}{49}} = 7.671$$

2.58 **a** Refer to Exercise 2.57. Since $\sum x_i = 418.4$, the sample mean is

$$\bar{x} = \frac{\sum x_i}{n} = \frac{418.4}{50} = 8.368$$

The three intervals of interest is shown in the following table, along with the number of observations which fall in each interval.

k	$\bar{x} \pm ks$	Interval	Number in Interval	Percentage
1	8.368 ± 7.671	0.697 to 16.039	37	74%
2	8.368 ± 15.342	–6.974 to 23.710	47	94%
3	8.368 ± 23.013	–14.645 to 31.381	49	98%

b The percentages falling in the intervals do agree with Tchebysheff's Theorem. At least 0 fall in the first interval, at least $3/4 = 0.75$ fall in the second interval, and at least $8/9 = 0.89$ fall in the third. The percentages are not too close to the percentages described by the Empirical Rule (68%, 95%, and 99.7%).
c The Empirical Rule may be unsuitable for describing these data. The data distribution does not have a strong mound-shape (see the relative frequency histogram in the solution to Exercise 1.26), but is skewed to the right.

2.59 The ordered data are shown below.

0.2	2.0	4.3	8.2	14.7
0.2	2.1	4.4	8.3	16.7
0.3	2.4	5.6	8.7	18.0
0.4	2.4	5.8	9.0	18.0
1.0	2.7	6.1	9.6	18.4
1.2	3.3	6.6	9.9	19.2
1.3	3.5	6.9	11.4	23.1
1.4	3.7	7.4	12.6	24.0
1.6	3.9	7.4	13.5	26.7
1.6	4.1	8.2	14.1	32.3

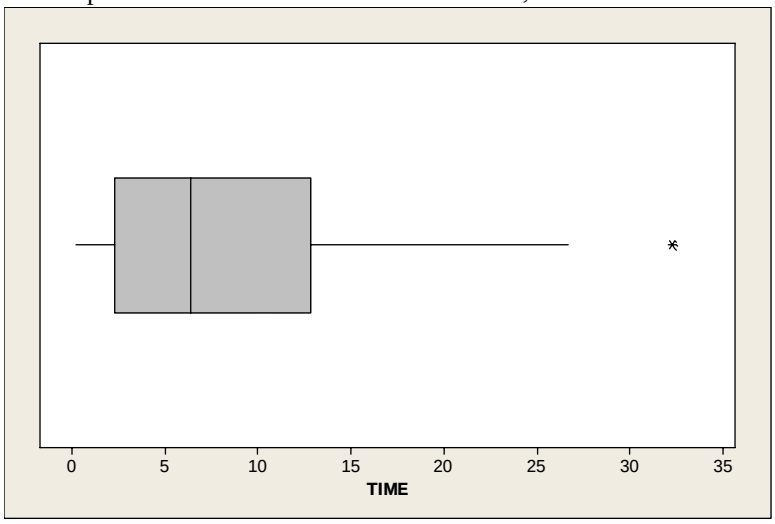
Since $n = 50$, the position of the median is $0.5(n+1) = 25.5$ and the positions of the lower and upper quartiles are $0.25(n+1) = 12.75$ and $0.75(n+1) = 38.25$.

Then $m = (6.1 + 6.6)/2 = 6.35$, $Q_1 = 2.1 + 0.75(2.4 - 2.1) = 2.325$ and $Q_3 = 12.6 + 0.25(13.5 - 12.6) = 12.825$. Then $IQR = 12.825 - 2.325 = 10.5$.

The *lower and upper fences* are:

$$Q_1 - 1.5IQR = 2.325 - 15.75 = -13.425$$
$$Q_3 + 1.5IQR = 12.825 + 15.75 = 28.575$$

and the box plot is shown below. There is one outlier, $x = 32.3$. The distribution is skewed to the right.



2.60 **a** For $n = 14$, the position of the median is $0.5(n+1) = 7.5$ and the positions of the quartiles are $0.25(n+1) = 3.75$ and $0.75(n+1) = 11.25$. The lower quartile is ¾ the way between the 3rd and 4th measurements or $Q_1 = 0.60 + 0.75(0.63 - 0.60) = 0.6225$ and the upper quartile is ¼ the way between the 11th and 12th measurements or $Q_3 = 1.12 + 0.25(1.23 - 1.12) = 1.1475$.

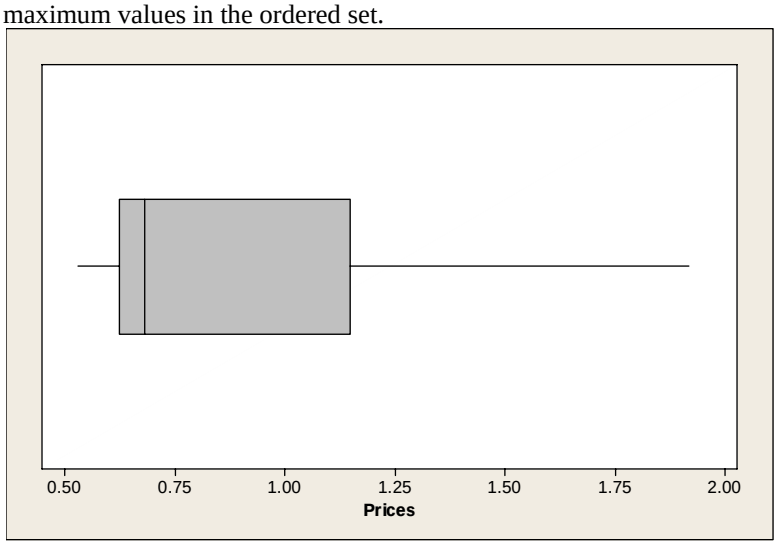
Then the five-number summary is

Min	Q_1	Median	Q_3	Max
0.53	0.6225	0.68	1.1475	1.92

b Calculate $IQR = Q_3 - Q_1 = 1.1475 - 0.6225 = 0.5250$. Then the *lower and upper fences* are:

$$Q_1 - 1.5IQR = 0.6225 - 0.7875 = -0.165$$
$$Q_3 + 1.5IQR = 1.1475 + 0.7875 = 1.935$$

The box plot is shown below. Since there are no outliers, the whiskers connect the box to the minimum and maximum values in the ordered set.



c Calculate $n = 14$, $\sum x_i = 12.55$, $\sum x_i^2 = 13.3253$. Then

$$\bar{x} = \frac{\sum x_i}{n} = \frac{12.55}{14} = 0.896 \text{ and}$$

$$s = \sqrt{\frac{\sum x_i^2 - \frac{\left(\sum x_i\right)^2}{n}}{n-1}} = \sqrt{\frac{13.3253 - \frac{(12.55)^2}{14}}{13}} = 0.3995$$

The *z*-score for $x = 1.92$ is

$$z = \frac{x - \bar{x}}{s} = \frac{1.92 - 0.896}{0.3995} = 2.56$$

which is somewhat unlikely. This observation does not appear as an outlier in the box plot.

2.61 First calculate the intervals:

$\bar{x} \pm s = 0.17 \pm 0.01$ or 0.16 to 0.18
$\bar{x} \pm 2s = 0.17 \pm 0.02$ or 0.15 to 0.19
$\bar{x} \pm 3s = 0.17 \pm 0.03$ or 0.14 to 0.20

a If no prior information as to the shape of the distribution is available, we use Tchebysheff's Theorem. We would expect at least $\left(1 - 1/1^2\right) = 0$ of the measurements to fall in the interval 0.16 to 0.18; at least

$\left(1-1/2^2\right)=3/4$ of the measurements to fall in the interval 0.15 to 0.19; at least $\left(1-1/3^2\right)=8/9$ of the measurements to fall in the interval 0.14 to 0.20.

b According to the Empirical Rule, approximately 68% of the measurements will fall in the interval 0.16 to 0.18; approximately 95% of the measurements will fall between 0.15 to 0.19; approximately 99.7% of the measurements will fall between 0.14 and 0.20. Since mound-shaped distributions are so frequent, if we do have a sample size of 30 or greater, we expect the sample distribution to be mound-shaped. Therefore, in this exercise, we would expect the Empirical Rule to be suitable for describing the set of data.

c If the chemist had used a sample size of four for this experiment, the distribution would not be mound-shaped. Any possible histogram we could construct would be non-mound-shaped. We can use at most 4 classes, each with frequency 1, and we will not obtain a histogram that is even close to mound-shaped. Therefore, the Empirical Rule would not be suitable for describing $n=4$ measurements.

2.62 Since it is not obvious that the distribution of amount of chloroform per liter of water in various water sources is mound-shaped, we cannot make this assumption. Tchebysheff's Theorem can be used, however, and the necessary intervals and fractions falling in these intervals are given in the table.

k	$\bar{x}\pm ks$	Interval	Tchebysheff
1	34 ± 53	–19 to 87	at least 0
2	34 ± 106	–72 to 40	at least 0.75
3	34 ± 159	–125 to 193	at least 0.89

2.63 The following information is available:

$$n=400,\ \bar{x}=600,\ s^2=4900$$

The standard deviation of these scores is then 70, and the results of Tchebysheff's Theorem follow:

k	$\bar{x}\pm ks$	Interval	Tchebysheff
1	600 ± 70	530 to 670	at least 0
2	600 ± 140	460 to 740	at least 0.75
3	600 ± 210	390 to 810	at least 0.89

If the distribution of scores is mound-shaped, we use the Empirical Rule, and conclude that approximately 68% of the scores would lie in the interval 530 to 670 (which is $\bar{x}\pm s$). Approximately 95% of the scores would lie in the interval 460 to 740.

2.64 **a** Calculate $n=10,\ \sum x_i=68.5, \sum x_i^2=478.375$. Then

$$\bar{x}=\frac{\sum x_i}{n}=\frac{68.5}{10}=6.85 \text{ and}$$

$$s=\sqrt{\frac{\sum x_i^2-\frac{\left(\sum x_i\right)^2}{n}}{n-1}}=\sqrt{\frac{478.375-\frac{(68.5)^2}{10}}{9}}=1.008$$

b The z-score for $x=8.5$ is

$$z=\frac{x-\bar{x}}{s}=\frac{8.5-6.85}{1.008}=1.64$$

This is not an unusually large measurement.

c The most frequently recorded measurement is the mode or x = 7 hours of sleep.

d For $n=10$, the position of the median is $0.5(n+1)=5.5$ and the positions of the quartiles are $0.25(n+1)=2.75$ and $0.75(n+1)=8.25$. The sorted data are:

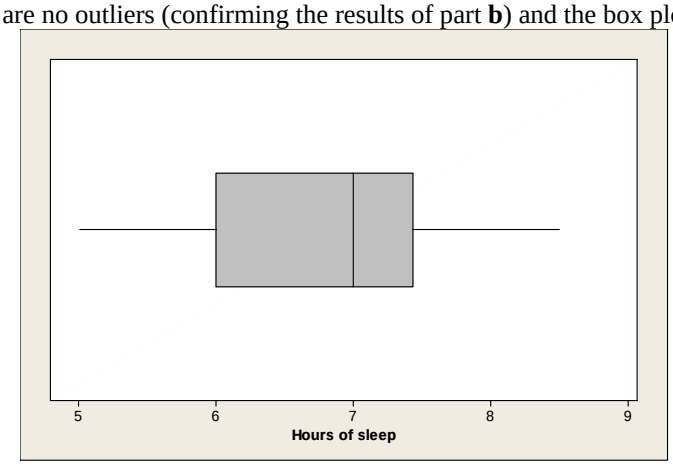
5, 6, 6, 6.75, 7, 7, 7, 7.25, 8, 8.5

Then $m=(7+7)/2=7,\ Q_1=6+0.75(6-6)=6$ and $Q_3=7.25+0.25(8-7.25)=7.4375$.

Then $IQR = 7.4375 - 6 = 1.4375$ and the *lower and upper fences* are:

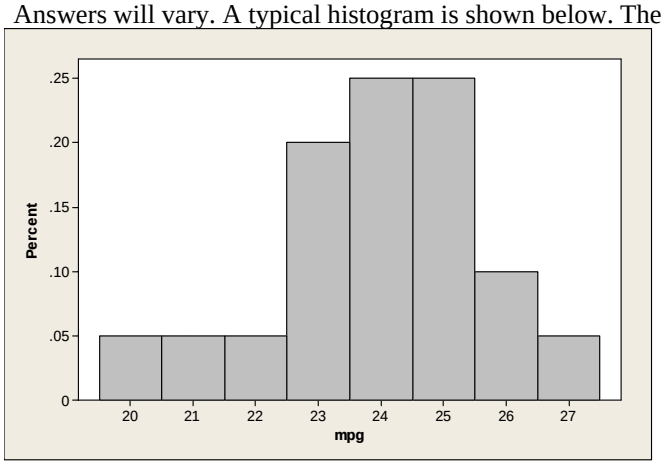
$$Q_1 - 1.5IQR = 6 - 2.15625 = 3.84$$
$$Q_3 + 1.5IQR = 7.4375 + 2.15625 = 9.59$$

There are no outliers (confirming the results of part **b**) and the box plot is shown below.



2.65 **a** Max = 27, Min = 20.2 and the range is $R = 27 - 20.2 = 6.8$.

b Answers will vary. A typical histogram is shown below. The distribution is slightly skewed to the left.



c Calculate $n = 20,\ \sum x_i = 479.2,\ \sum x_i^2 = 11532.82$. Then

$$\overline{x} = \frac{\sum x_i}{n} = \frac{479.2}{20} = 23.96$$

$$s = \sqrt{\frac{\sum x_i^2 - \frac{\left(\sum x_i\right)^2}{n}}{n-1}} = \sqrt{\frac{11532.82 - \frac{(479.2)^2}{20}}{19}} = \sqrt{2.694} = 1.641$$

d The sorted data is shown below:

```
20.2   21.3   22.2   22.7   22.9
23.1   23.2   23.6   23.7   24.2
24.4   24.4   24.6   24.7   24.7
24.9   25.3   25.9   26.2   27.0
```

The *z*-scores for $x = 20.2$ and $x = 27$ are

$$z = \frac{x - \overline{x}}{s} = \frac{20.2 - 23.96}{1.641} = -2.29 \text{ and } z = \frac{x - \overline{x}}{s} = \frac{27 - 23.96}{1.641} = 1.85$$

Since neither of the *z*-scores are greater than 3 in absolute value, the measurements are not judged to be outliers.

e The position of the median is $0.5(n+1) = 10.5$ and the median is m = (24.2 + 24.4)/2 = 24.3

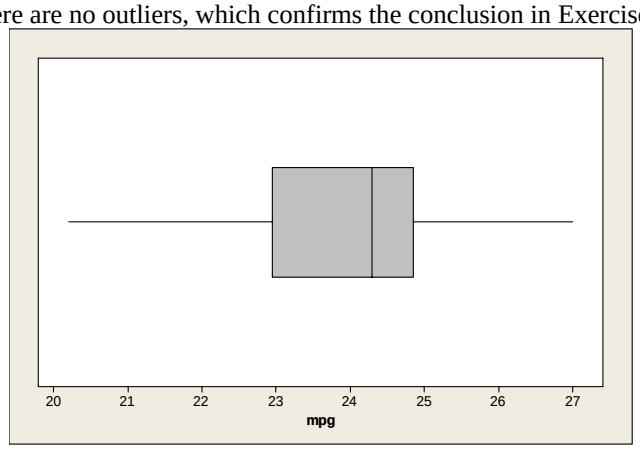
f The positions of the quartiles are $0.25(n+1) = 5.25$ and $0.75(n+1) = 15.75$. Then

$Q_1 = 22.9 + 0.25(23.1 - 22.9) = 22.95$ and $Q_3 = 24.7 + 0.75(24.9 - 24.7) = 24.85.$

2.66 Refer to Exercise 2.65. Calculate $IQR = 24.85 - 22.95 = 1.9$. The *lower and upper fences* are:

$$Q_1 - 1.5IQR = 22.95 - 2.85 = 20.10$$
$$Q_3 + 1.5IQR = 24.85 + 2.85 = 27.70$$

There are no outliers, which confirms the conclusion in Exercise 2.65. The box plot is shown below.



2.67 **a** The range is $R = 71 - 40 = 31$ and the range approximation is

$$s \approx R/4 = 31/4 = 7.75$$

b Calculate $n = 10,\ \sum x_i = 592$, $\sum x_i^2 = 36014$. Then

$$\bar{x} = \frac{\sum x_i}{n} = \frac{592}{10} = 59.2$$

$$s = \sqrt{\frac{\sum x_i^2 - \frac{\left(\sum x_i\right)^2}{n}}{n-1}} = \sqrt{\frac{36014 - \frac{(592)^2}{10}}{9}} = \sqrt{107.5111} = 10.369$$

The sample standard deviation calculated above is of the same order as the approximated value found in part **a**.

c The ordered set is:

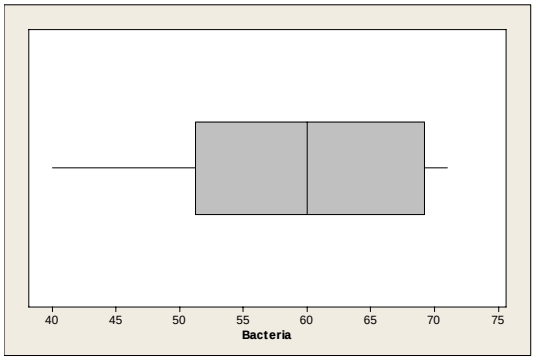
40, 49, 52, 54, 59, 61, 67, 69, 70, 71

Since $n = 10$, the positions of m, Q_1, and Q_3 are 5.5, 2.75 and 8.25 respectively, and $m = (59 + 61)/2 = 60$, $Q_1 = 49 + 0.75(52 - 49) = 51.25$, $Q_3 = 69.75$ and $IQR = 69.75 - 51.25 = 18.5$.

The *lower and upper fences* are:

$$Q_1 - 1.5IQR = 51.25 - 27.75 = 23.5$$
$$Q_3 + 1.5IQR = 69.75 + 27.75 = 97.50$$

and the box plot is shown on the next page. There are no outliers and the data set is slightly skewed left.



2.68 The results of the Empirical Rule follow:

k	$\bar{x} \pm ks$	Interval	Empirical Rule
1	420 ± 5	415 to 425	approximately 0.68
2	420 ± 10	410 to 430	approximately 0.95
3	420 ± 15	405 to 435	approximately 0.997

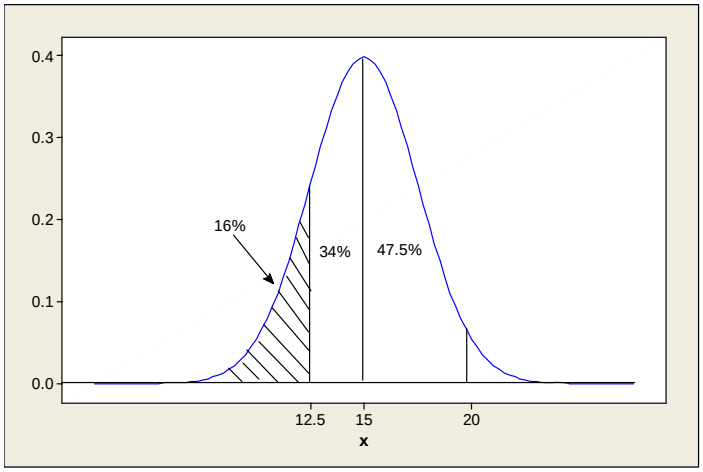
Notice that we are assuming that attendance follows a mound-shaped distribution and hence that the Empirical Rule is appropriate.

2.69 If the distribution is mound-shaped, then almost all of the measurements will fall in the interval $\mu \pm 3\sigma$, which is an interval 6σ in length. That is, the range of the measurements should be approximately 6σ. In this case, the range is $800 - 200 = 600$, so that $\sigma \approx 600/6 = 100$.

2.70 They are probably referring to the average number of times that men and women go camping per year.

2.71 The stem lengths are approximately normal with mean 15 and standard deviation 2.5.

a In order to determine the percentage of roses with length less than 12.5, we must determine the proportion of the curve which lies within the shaded area in the figure below. Using the Empirical Rule, the proportion of the area between 12.5 and 15 is half of 0.68 or 0.34. Hence, the fraction below 12.5 would be $0.5 - 0.34 = 0.16$ or 16%.



b Refer to the figure shown above. Again we use the Empirical Rule. The proportion of the area between 12.5 and 15 is half of 0.68 or 0.34, while the proportion of the area between 15 and 20 is half of 0.95 or 0.475. The total area between 12.5 and 20 is then 0.34 + 0.475 = .815 or 81.5%.

2.72 **a** The range is $R = 172 - 108 = 64$ and the range approximation is

$$s \approx R/4 = 64/4 = 16$$

b Calculate $n = 15,\ \sum x_i = 2041, \sum x_i^2 = 281{,}807$. Then

$$\bar{x} = \frac{\sum x_i}{n} = \frac{2041}{15} = 136.07$$

$$s = \sqrt{\frac{\sum x_i^2 - \frac{\left(\sum x_i\right)^2}{n}}{n-1}} = \sqrt{\frac{281{,}807 - \frac{(2041)^2}{15}}{14}} = \sqrt{292.495238} = 17.102$$

c According to Tchebysheff's Theorem, with $k = 2$, at least 3/4 or 75% of the measurements will lie within $k = 2$ standard deviations of the mean. For this data, the two values, a and b, are calculated as

$$\bar{x} \pm 2s \Rightarrow 136.07 \pm 2(17.10) \Rightarrow 137.07 \pm 34.20 \text{ or } a = 101.87 \text{ and } b = 170.27.$$

2.73 The diameters of the trees are approximately mound-shaped with mean 14 and standard deviation 2.8.

a The value $x = 8.4$ lies two standard deviations below the mean, while the value $x = 22.4$ is three standard deviations above the mean. Use the Empirical Rule. The fraction of trees with diameters between 8.4 and 14 is half of 0.95 or 0.475, while the fraction of trees with diameters between 14 and 22.4 is half of 0.997 or 0.4985. The total fraction of trees with diameters between 8.4 and 22.4 is

$$0.475 + 0.4985 = .9735$$

b The value $x = 16.8$ lies one standard deviation above the mean. Using the Empirical Rule, the fraction of trees with diameters between 14 and 16.8 is half of 0.68 or 0.34, and the fraction of trees with diameters greater than 16.8 is

$$0.5 - 0.34 = .16$$

2.74 **a** The range is $R = 19 - 4 = 15$ and the range approximation is

$$s \approx R/4 = 15/4 = 3.75$$

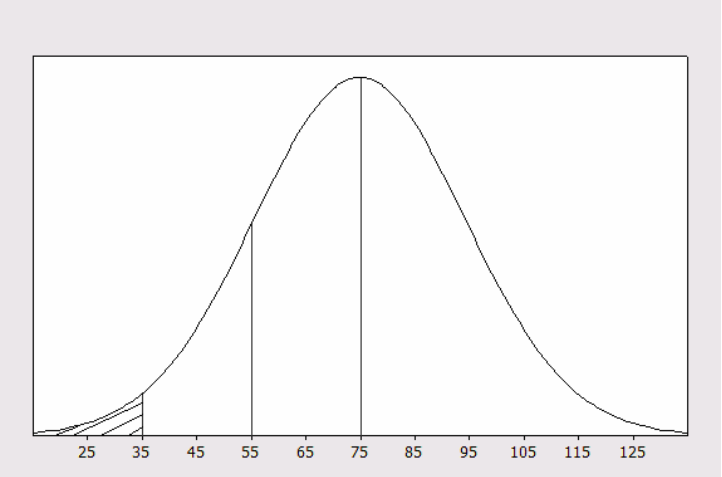
b Calculate $n = 15,\ \sum x_i = 175, \sum x_i^2 = 2237$. Then

$$\bar{x} = \frac{\sum x_i}{n} = \frac{175}{15} = 11.67$$

$$s = \sqrt{\frac{\sum x_i^2 - \frac{\left(\sum x_i\right)^2}{n}}{n-1}} = \sqrt{\frac{2237 - \frac{(175)^2}{15}}{14}} = \sqrt{13.95238} = 3.735$$

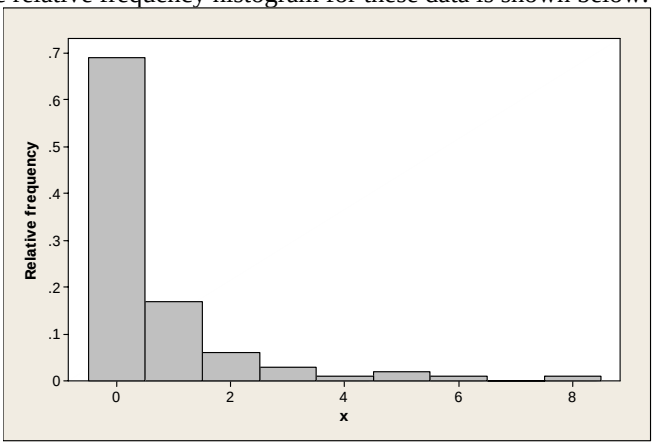
c Calculate the interval $\bar{x} \pm 2s \Rightarrow 11.67 \pm 2(3.735) \Rightarrow 11.67 \pm 7.47$ or 4.20 to 19.14. Referring to the original data set, the fraction of measurements in this interval is 14/15 = .93.

2.75 **a** It is known that duration times are approximately normal, with mean 75 and standard deviation 20. In order to determine the probability that a commercial lasts less than 35 seconds, we must determine the fraction of the curve which lies within the shaded area in the figure on the next page. Using the Empirical Rule, the fraction of the area between 35 and 75 is half of 0.95 or 0.475. Hence, the fraction below 35 would be $0.5 - 0.475 = 0.025$.



b The fraction of the curve area that lies above the 55 second mark may again be determined by using the Empirical Rule. Refer to the figure in part **a**. The fraction between 55 and 75 is 0.34 and the fraction above 75 is 0.5. Hence, the probability that a commercial lasts longer than 55 seconds is $0.5 + 0.34 = 0.84$.

2.76 **a** The relative frequency histogram for these data is shown below.



b Refer to the formulas given in Exercise 2.37. Using the frequency table and the grouped formulas, calculate $n = 100$, $\sum x_i f_i = 66$, $\sum x_i^2 f_i = 234$. Then

$$\bar{x} = \frac{\sum x_i f_i}{n} = \frac{66}{100} = 0.66$$

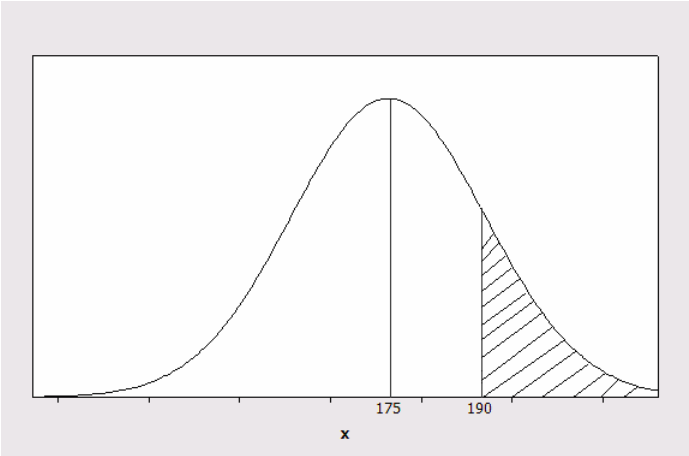
$$s^2 = \frac{\sum x_i^2 f_i - \frac{\left(\sum x_i f_i\right)^2}{n}}{n-1} = \frac{234 - \frac{(66)^2}{100}}{99} = 1.9236 \text{ and } s = \sqrt{1.9236} = 1.39.$$

c The three intervals, $\bar{x} \pm ks$ for $k = 2, 3$ are calculated in the table along with the actual proportion of measurements falling in the intervals. Tchebysheff's Theorem is satisfied and the approximation given by the Empirical Rule are fairly close for $k = 2$ and $k = 3$.

k	$\bar{x} \pm ks$	Interval	Fraction in Interval	Tchebysheff	Empirical Rule
2	0.66 ± 2.78	–2.12 to 3.44	95/100 = 0.95	at least 0.75	≈ 0.95
3	0.66 ± 4.17	–3.51 to 4.83	96/100 = 0.96	at least 0.89	≈ 0.997

2.77 **a** The percentage of colleges that have between 145 and 205 teachers corresponds to the fraction of measurements expected to lie within two standard deviations of the mean. Tchebysheff's Theorem states that this fraction will be at least ¾ or 75%.

b If the population is normally distributed, the Empirical Rule is appropriate and the desired fraction is calculated. Referring to the normal distribution shown below, the fraction of area lying between 175 and 190 is 0.34, so that the fraction of colleges having more than 190 teachers is $0.5 - 0.34 = 0.16$.



2.78 We must estimate *s* and compare with the student's value of 0.263. In this case, $n = 20$ and the range is $R = 17.4 - 16.9 = 0.5$. The estimated value for *s* is then

$$s \approx R/4 = 0.5/4 = 0.125$$

which is less than 0.263. It is important to consider the magnitude of the difference between the "rule of thumb" and the calculated value. For example, if we were working with a standard deviation of 100, a difference of 0.142 would not be great. However, the student's calculation is twice as large as the estimated value. Moreover, two standard deviations, or $2(0.263) = 0.526$, already exceeds the range. Thus, the value $s = 0.263$ is probably incorrect. The correct value of *s* is

$$s = \sqrt{\frac{\sum x_i^2 - \frac{\left(\sum x_i\right)^2}{n}}{n-1}} = \sqrt{\frac{5851.95 - \frac{117032.41}{20}}{19}} = \sqrt{0.0173} = 0.132$$

2.79 Notice that two (Sosa and McGuire) of the four players have relatively symmetric distributions. The whiskers are the same length and the median line is close to the middle of the box. The variability of the distributions is similar for all four players, but Barry Bonds has a distribution with a long right whisker, meaning that there may be an unusually large number of homers during one of his seasons. The distribution for Babe Ruth is slightly different from the others. The median line to the right of middle indicates a distribution skewed to the left; that there were a few seasons in which his homerun total was unusually low. In fact, the median number of homeruns for the other three players are all about 34-35, while Babe Ruth's median number of homeruns is closer to 40.

2.80 **a** Use the information in the exercise. For 2001, $IQR = 16.5$, and the upper fence is

$$Q_3 + 1.5IQR = 41.50 + 24.75 = 66.25$$

For 2006, $IQR = 20$, and the upper fence is

$$Q_3 + 1.5IQR = 45.00 + 30.00 = 75.00$$

b The upper fence is different in 2006, so that the record number of homers, *x* = 73 is no longer an outlier, although it is still the most homers ever hit in a single season!

2.81 **a** Calculate $n = 50$, $\sum x_i = 418$, so that $\bar{x} = \frac{\sum x_i}{n} = \frac{418}{50} = 8.36$.

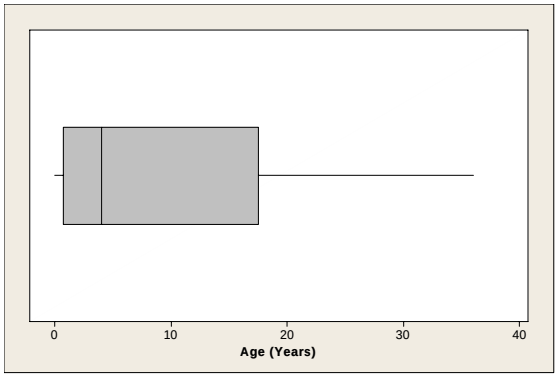
b The position of the median is $.5(n+1) = 25.5$ and $m = (4 + 4)/2 = 4$.

c Since the mean is larger than the median, the distribution is skewed to the right.

d Since $n = 50$, the positions of Q_1 and Q_3 are .25(51) = 12.75 and .75(51) = 38.25, respectively Then $Q_1 = 0 + 0.75(1-0) = 12.75$, $Q_3 = 17 + .25(19-17) = 17.5$ and $IQR = 17.5 - .75 = 16.75$.
The *lower and upper fences* are:

$$Q_1 - 1.5IQR = .75 - 25.125 = -24.375$$
$$Q_3 + 1.5IQR = 17.5 + 25.125 = 42.625$$

and the box plot is shown below. There are no outliers and the data is skewed to the right.



2.82 Each bulleted statement produces a percentile.

- x = hourly pay for salespeople. The value $x = 10.41$ is the 50th percentile or the median.
- x = hours worked per week by workers ages 16 and older. The value $x = 40$ is the (100-69) = 31st percentile.
- x = salary for Associate Professors of mathematics in the U.S.. The value $x = 91,823$ is the 75th percentile or the upper quartile.

2.83 Answers will vary. Students should notice that the distribution of baseline measurements is relatively mound-shaped. Therefore, the Empirical Rule will provide a very good description of the data. A measurement which is further than two or three standard deviations from the mean would be considered unusual.

2.84 **a** Calculate $n = 25$, $\sum x_i = 104.9$, $\sum x_i^2 = 454.810$. Then

$$\bar{x} = \frac{\sum x_i}{n} = \frac{104.9}{25} = 4.196$$

$$s = \sqrt{\frac{\sum x_i^2 - \frac{\left(\sum x_i\right)^2}{n}}{n-1}} = \sqrt{\frac{454.810 - \frac{(104.9)^2}{25}}{24}} = \sqrt{.610} = .781$$

b The ordered data set is shown below:

2.5	3.0	3.1	3.3	3.6
3.7	3.8	3.8	3.9	3.9
4.1	4.2	4.2	4.2	4.3
4.3	4.4	4.7	4.7	4.8
4.8	5.2	5.3	5.4	5.7

c The *z*-scores for $x = 2.5$ and $x = 5.7$ are

$$z = \frac{x - \bar{x}}{s} = \frac{2.5 - 4.196}{.781} = -2.17 \text{ and } z = \frac{x - \bar{x}}{s} = \frac{5.7 - 4.196}{.781} = 1.93$$

Since neither of the *z*-scores are greater than 3 in absolute value, the measurements are not judged to be unusually large or small.

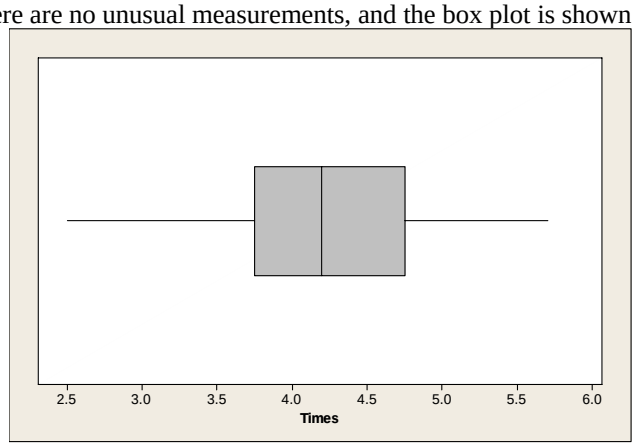
2.85 **a** For $n = 25$, the position of the median is $0.5(n+1) = 13$ and the positions of the quartiles are $0.25(n+1) = 6.5$ and $0.75(n+1) = 19.5$. Then $m = 4.2$, $Q_1 = (3.7 + 3.8)/2 = 3.75$ and $Q_3 = (4.7 + 4.8)/2 = 4.75$. Then the five-number summary is

Min	Q_1	Median	Q_3	Max
2.5	3.75	4.2	4.75	5.7

b-c Calculate $IQR = Q_3 - Q_1 = 4.75 - 3.75 = 1$. Then the *lower and upper fences* are:

$$Q_1 - 1.5IQR = 3.75 - 1.5 = 2.25$$
$$Q_3 + 1.5IQR = 4.75 + 1.5 = 6.25$$

There are no unusual measurements, and the box plot is shown below.



d Answers will vary. A stem and leaf plot, generated by ***Minitab,*** is shown below. The data is roughly mound-shaped.

Stem-and-Leaf Display: Times

```
Stem-and-leaf of Times  N  = 25
Leaf Unit = 0.10

 1   2  5
 4   3  013
 10  3  678899
(7)  4  1222334
 8   4  7788
 4   5  234
 1   5  7
```

2.86 **a** When the applet loads, the mean and median are shown in the upper left-hand corner:
$\bar{x} = 6.6$ and $m = 6.0$

b When the largest value is changed to $x = 13$, $\bar{x} = 7.0$ and $m = 6.0$.

c When the largest value is changed to $x = 33$, $\bar{x} = 11.0$ and $m = 6.0$. The mean is larger when there is one unusually large measurement.

d Extremely large values cause the mean to increase, but not the median.

2.87 **a-b** As the value of x gets smaller, so does the mean.

c The median does not change until the green dot is smaller than $x = 10$, at which point the green dot becomes the median.

d The largest and smallest possible values for the median are $5 \le m \le 10$.

2.88 **a** When the applet loads, the mean and median are shown in the upper left-hand corner:
$\bar{x} = 31.6$ and $m = 32.0$

b When the smallest value is changed to $x = 25$, $\bar{x} = 31.2$ and $m = 32.0$.

c When the smallest value is changed to $x = 5$, $\bar{x} = 27.2$ and $m = 32.0$. The mean is smaller when there is one unusually small measurement.

d $x = 29.0$

e The largest and smallest possible values for the median are $32 \le m \le 34$.

f Extremely small values cause the mean to decrease, but not the median.

2.89 Answers will vary from student to student. Students should notice that, when the estimators are compared in the long run, the standard deviation when dividing by $n-1$ is closer to $\sigma = 29.2$. When dividing by *n*, the estimate is closer to 23.8.

2.90 **a** Answers will vary from student to student. Students should notice that, when the estimators are compared in the long run, the standard deviation when dividing by $n-1$ is closer to $\sigma = 29.2$. When dividing by *n*, the estimate is closer to 27.5.

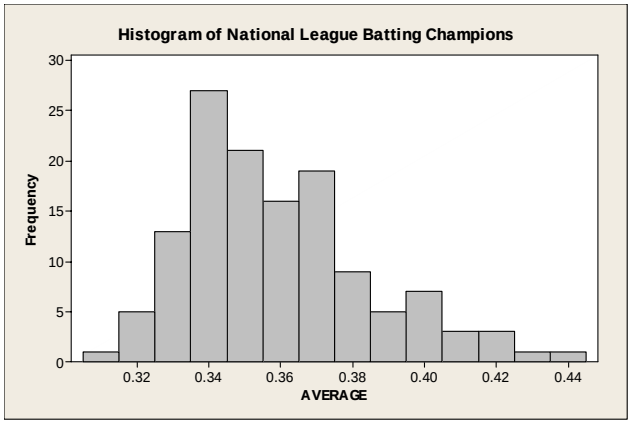
b When the sample size is larger, the estimate is not as far from the true value $\sigma = 29.2$. The difference between the two estimators is less noticeable.

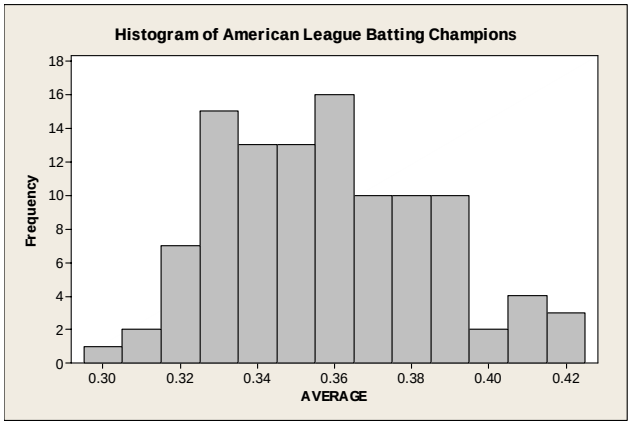
2.91 The box plot shows a distribution that is skewed to the left, but with one outlier to the **right** of the other observations ($x = 520$).

2.92 The box plot shows a distribution that is slightly skewed to the right, with no outliers. The student should estimate values for *m*, Q_1, and Q_3 that are close to the true values: $m = 12$, $Q_1 = 8.75$, and $Q_3 = 18.5$.

Case Study: The Boys of Summer

1 The ***Minitab*** computer package was used to analyze the data. In the printout below, various descriptive statistics as well as histograms and box plots are shown.



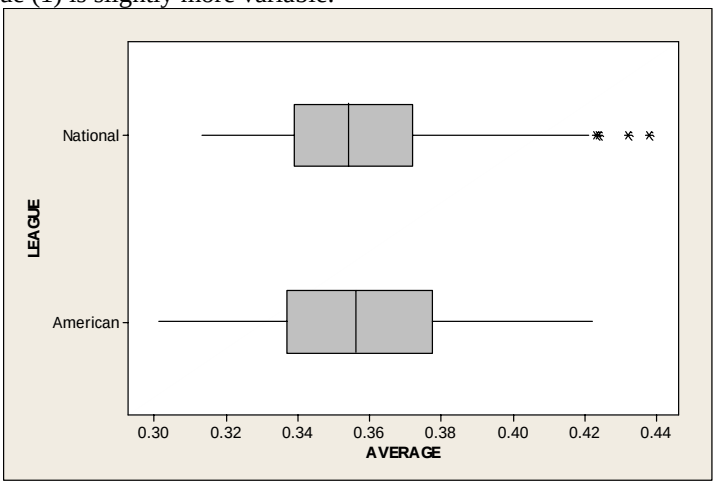


Descriptive Statistics: AVERAGE

Variable	LEAGUE	N	N*	Mean	SE Mean	StDev	Minimum	Q1	Median
AVERAGE	0	131	0	0.35878	0.00225	0.02575	0.31300	0.33900	0.35400
	1	106	0	0.35753	0.00263	0.02703	0.30100	0.33675	0.35600

Variable	LEAGUE	Q3	Maximum
AVERAGE	0	0.37200	0.43800
	1	0.37725	0.42200

2 Notice that the mean percentage of hits is almost the same for the two leagues, but that the American League (1) is slightly more variable.



3 The box plot shows that there are three outliers in the National League (0).

4 In summary, except for the two outliers, there is very little difference between the two leagues.