This is “Relative Position of Data”, section 2.4 from the book Beginning Statistics (v. 1.0). For details on it (including licensing), click here.

For more information on the source of this book, or why it is available for free, please see the project's home page. You can browse or download additional books there.

Has this book helped you? Consider passing it on:

Help Creative Commons

Creative Commons supports free culture from music to education. Their licenses helped make this book available to you.

Help a Public School

DonorsChoose.org helps people like you help teachers fund their classroom projects, from art supplies to books to calculators.

Previous Section | Table of Contents | Next Section

2.4 Relative Position of Data

LEARNING OBJECTIVES

1. To learn the concept of the relative position of an element of a data set.
2. To learn the meaning of each of two measures, the percentile rank and the *z*-score, of the relative position of a measurement and how to compute each one.
3. To learn the meaning of the three quartiles associated to a data set and how to compute them.
4. To learn the meaning of the five-number summary of a data set, how to construct the box plot associated to it, and how to interpret the box plot.

When you take an exam, what is often as important as your actual score on the exam is the way your score compares to other students’ performance. If you made a 70 but the average score (whether the mean, median, or mode) was 85,

you did relatively poorly. If you made a 70 but the average score was only 55 then you did relatively well. In general, the significance of one observed value in a data set strongly depends on how that value compares to the other observed values in a data set. Therefore we wish to attach to each observed value a number that measures its relative position.

Percentiles and Quartiles

Anyone who has taken a national standardized test is familiar with the idea of being given both a score on the exam and a “percentile ranking” of that score. You may be told that your score was 625 and that it is the 85th percentile. The first number tells how you actually did on the exam; the second says that 85% of the scores on the exam were less than or equal to your score, 625.

Definition

Given an observed value x in a data set, x is the ***P*th percentile** *of the data if the percentage of the data that are less than or equal to x is P. The number P is the* **percentile rank** *of x.*

EXAMPLE 13

What percentile is the value 1.39 in the data set of ten GPAs considered in Note 2.12 "Example 3" in Section 2.2 "Measures of Central Location"? What percentile is the value 3.33?

Solution:

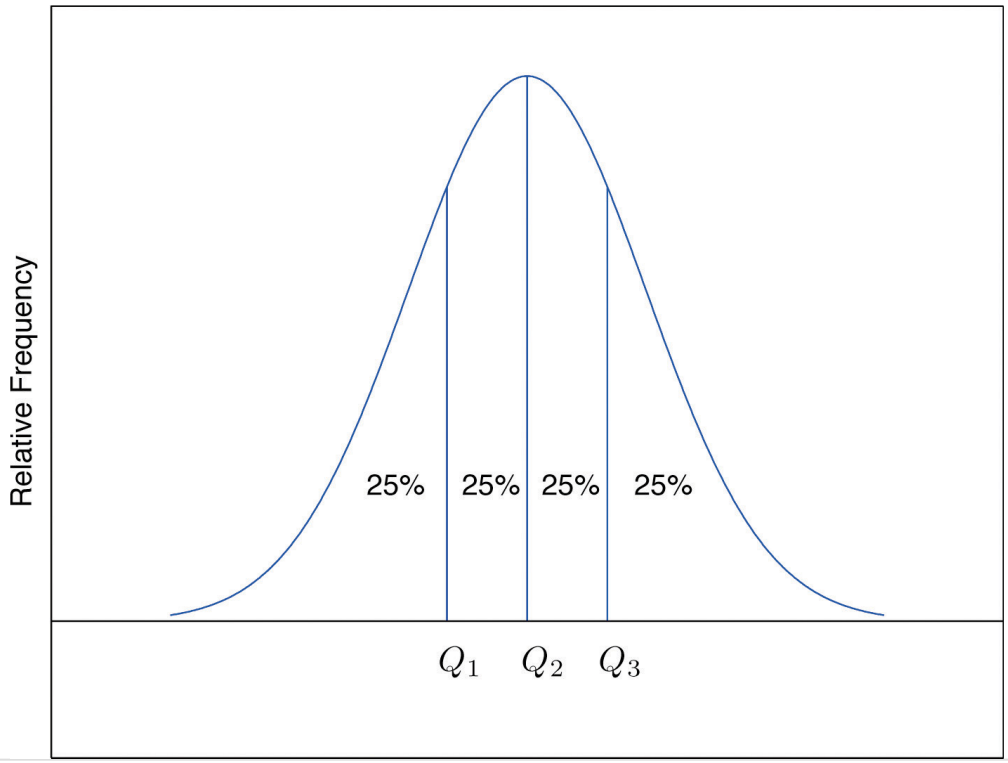
The data written in increasing order are

$$1.39 \;\; 1.76 \;\; 1.90 \;\; 2.12 \;\; 2.53 \;\; 2.71 \;\; 3.00 \;\; 3.33 \;\; 3.71 \;\; 4.00$$

The only data value that is less than or equal to 1.39 is 1.39 itself. Since 1 is 1⁄10 = .10 or 10% of 10, the value 1.39 is the 10th percentile. Eight data values are less than or equal to 3.33. Since 8 is 8⁄10 = .80 or 80% of 10, the value 3.33 is the 80th percentile.

The Pth percentile cuts the data set in two so that approximately P% of the data lie below it and $(100 - P)$% of the data lie above it. In particular, the three percentiles that cut the data into fourths, as shown in Figure 2.12 "Data Division by Quartiles", are called the **quartiles**. The following simple computational definition of the three quartiles works well in practice.

Figure 2.12 *Data Division by Quartiles*



Definition

For any data set:

1. *The* **second quartile** Q_2 *of the data set is its median.*
2. *Define two subsets:*
    1. *the lower set: all observations that are strictly less than* Q_2*;*
    2. *the upper set: all observations that are strictly greater than* Q_2*.*
3. *The* **first quartile** Q_1 *of the data set is the median of the lower set.*
4. *The* **third quartile** Q_3 *of the data set is the median of the upper set.*

EXAMPLE 14

Find the quartiles of the data set of GPAs of Note 2.12 "Example 3" in Section 2.2 "Measures of Central Location".

Solution:

As in the previous example we first list the data in numerical order:

$$1.39 \;\; 1.76 \;\; 1.90 \;\; 2.12 \;\; 2.53 \;\; 2.71 \;\; 3.00 \;\; 3.33 \;\; 3.71 \;\; 4.00$$

This data set has *n* = 10 observations. Since 10 is an even number, the median is the mean of the two middle observations: $\tilde{x} = (2.53 + 2.71)/2 = 2.62$. Thus the second quartile is $Q_2 = 2.62$. The lower and upper subsets are

$$\begin{aligned} \text{Lower:} \quad L &= \{1.39, 1.76, 1.90, 2.12, 2.53\} \\ \text{Upper:} \quad U &= \{2.71, 3.00, 3.33, 3.71, 4.00\} \end{aligned}$$

Each has an odd number of elements, so the median of each is its middle observation. Thus the first quartile is $Q_1 = 1.90$, the median of *L*, and the third quartile is $Q_3 = 3.33$, the median of *U*.

EXAMPLE 15

Adjoin the observation 3.88 to the data set of the previous example and find the quartiles of

the new set of data.

Solution:

As in the previous example we first list the data in numerical order:

$$1.39 \;\; 1.76 \;\; 1.90 \;\; 2.12 \;\; 2.53 \;\; 2.71 \;\; 3.00 \;\; 3.33 \;\; 3.71 \;\; 3.88 \;\; 4.00$$

This data set has 11 observations. The second quartile is its median, the middle value 2.71. Thus $Q_2 = 2.71$. The lower and upper subsets are now

$$\begin{aligned} \text{Lower:} \quad L &= \{1.39, 1.76, 1.90, 2.12, 2.53\} \\ \text{Upper:} \quad U &= \{3.00, 3.33, 3.71, 3.88, 4.00\} \end{aligned}$$

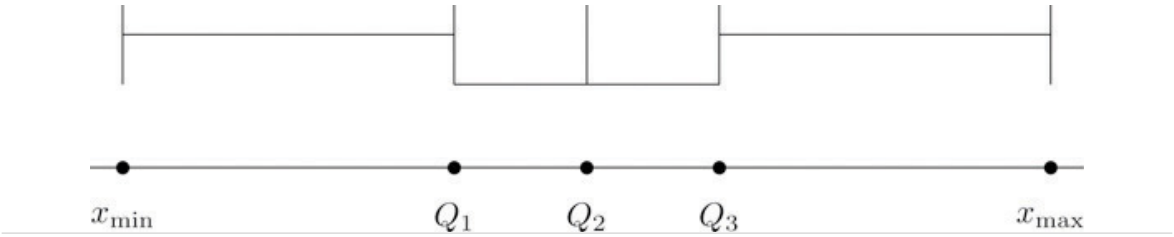
The lower set *L* has median the middle value 1.90, so $Q_1 = 1.90$. The upper set has median the middle value 3.71, so $Q_3 = 3.71$.

In addition to the three quartiles, the two extreme values, the minimum $x_{\min}$ and the maximum $x_{\max}$ are also useful in describing the entire data set. Together these five numbers are called the **five-number summary** of the data set:

$$\{x_{\min}, Q_1, Q_2, Q_3, x_{\max}\}$$

The five-number summary is used to construct a **box plot** as in Figure 2.13 "The Box Plot". Each of the five numbers is represented by a vertical line segment, a box is formed using the line segments at Q_1 and Q_3 as its two vertical sides, and two horizontal line segments are extended from the vertical segments marking Q_1 and Q_3 to the adjacent extreme values. (The two horizontal line segments are referred to as “whiskers,” and the diagram is sometimes called a “box and whisker plot.”) We caution the reader that there are other types of box plots that differ somewhat from the ones we are constructing, although all are based on the three quartiles.

Figure 2.13 The Box Plot



Note that the distance from Q_1 to Q_3 is the length of the interval over which the middle half of the data range. Thus it has the following special name.

Definition

The **interquartile range (IQR)** *is the quantity*

$$IQR = Q_3 - Q_1$$

EXAMPLE 16

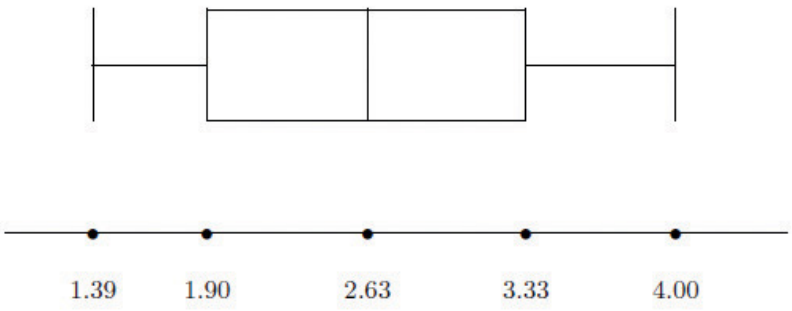
Construct a box plot and find the IQR for the data in Note 2.44 "Example 14".

Solution:

From our work in Note 2.44 "Example 14" we know that the five-number summary is

$$x_{\min} = 1.39 \quad Q_1 = 1.90 \quad Q_2 = 2.62 \quad Q_3 = 3.33 \quad x_{\max} = 4.00$$

The box plot is



The interquartile range is $\text{IQR} = 3.33 - 1.90 = 1.43$.

z-scores

Another way to locate a particular observation x in a data set is to compute its distance from the mean in units of standard deviation.

Definition

The ***z*-score** *of an observation x is the number z given by the computational formula*

$$z = \frac{x - \overline{x}}{s} \quad \textit{or} \quad z = \frac{x - \mu}{\sigma}$$

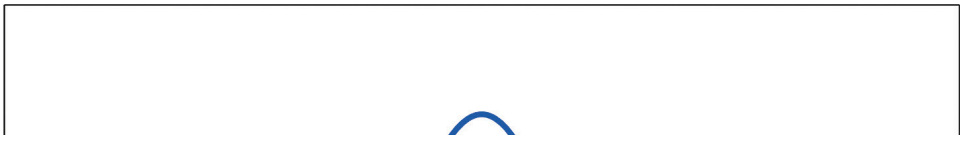
according to whether the data set is a sample or is the entire population.

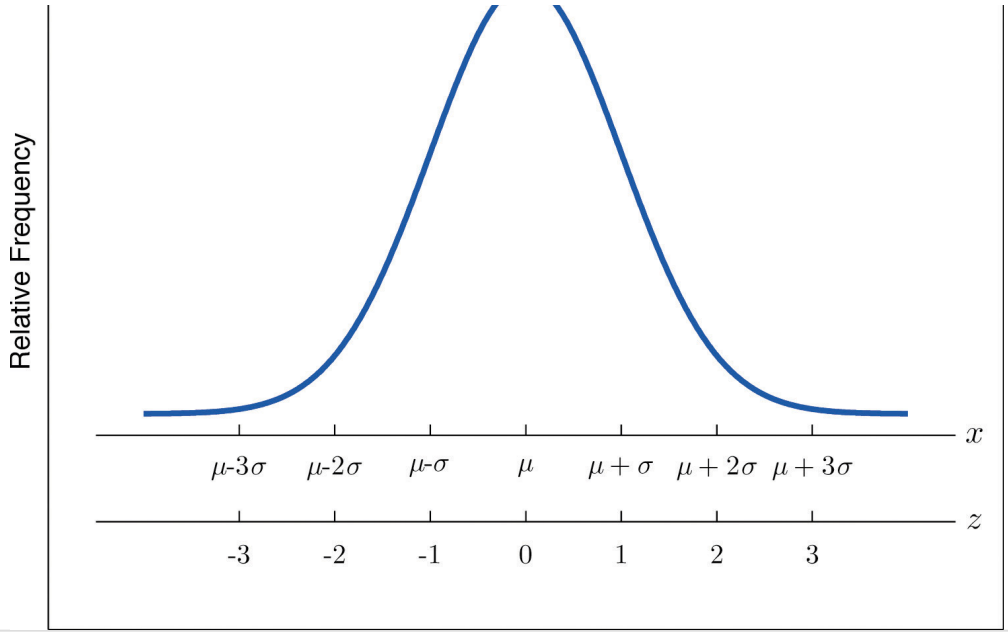
The formulas in the definition allow us to compute the z-score when x is known. If the z-score is known then x can be recovered using the corresponding inverse formulas

$$x = \overline{x} + sz \quad \text{or} \quad x = \mu + \sigma z$$

The z-score indicates how many standard deviations an individual observation x is from the center of the data set, its mean. If z is negative then x is below average. If z is 0 then x is equal to the average. If z is positive then x is above average. See Figure 2.14.

Figure 2.14 *x-Scale versus z-Score*





EXAMPLE 17

Find the *z*-scores for all ten observations in the GPA sample data in Note 2.12 "Example 3" in Section 2.2 "Measures of Central Location".

$$1.90 \;\; 3.00 \;\; 2.53 \;\; 3.71 \;\; 2.12 \;\; 1.76 \;\; 2.71 \;\; 1.39 \;\; 4.00 \;\; 3.33$$

Solution:

For these data $\bar{x} = 2.645$ and $s = 0.8674$. The first observation $x = 1.9$ in the data set has *z*-score

$$z = \frac{x - \bar{x}}{s} = \frac{1.9 - 2.645}{0.8674} = -0.8589$$

which means that $x = 1.90$ is 0.8589 standard deviations *below* the sample mean. The second observation $x = 3.00$ has *z*-score

$$z = \frac{x - \bar{x}}{s} = \frac{3.00 - 2.645}{0.8674} = 0.4093$$

which means that $x = 3.00$ is 0.4093 standard deviations *above* the sample mean. Repeating

the process for the remaining observations gives the full set of *z*-scores

$$-0.86 \quad 0.41 \quad -0.13 \quad 1.23 \quad -0.61 \quad -1.02 \quad 0.07 \quad -1.45 \quad 1.56 \quad 0.79$$

EXAMPLE 18

Suppose the mean and standard deviation of the GPAs of all currently registered students at a college are $\mu = 2.70$ and $\sigma = 0.50$. The *z*-scores of the GPAs of two students, Antonio and Beatrice, are $z = -0.62$ and $z = 1.28$, respectively. What are their GPAs?

Solution:

Using the second formula right after the definition of *z*-scores we compute the GPAs as

$$\text{Antonio: } x = \mu + z \quad \sigma = 2.70 + (-0.62)(0.50) = 2.39$$
$$\text{Beatrice: } x = \mu + z \quad \sigma = 2.70 + (1.28)(0.50) = 3.34$$

KEY TAKEAWAYS

- The percentile rank and *z*-score of a measurement indicate its relative position with regard to the other measurements in a data set.
- The three quartiles divide a data set into fourths.
- The five-number summary and its associated box plot summarize the location and distribution of the data.

EXERCISES

BASIC

1. Consider the data set

69	92	68	77	80
93	75	76	82	100
70	85	88	85	96
53	70	70	82	85

a. Find the percentile rank of 82.

b. Find the percentile rank of 68.

2. Consider the data set

8.5	8.2	7.0	7.0	4.9
9.6	8.5	8.8	8.5	8.7
6.5	8.2	7.6	1.5	9.3
8.0	7.7	2.9	9.2	6.9

a. Find the percentile rank of 6.5.

b. Find the percentile rank of 7.7.

3. Consider the data set represented by the ordered stem and leaf diagram

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

a. Find the percentile rank of the grade 75.

b. Find the percentile rank of the grade 57.

4. Is the 90th percentile of a data set always equal to 90%? Why or why not?

5. The 29th percentile in a large data set is 5.

a. Approximately what percentage of the observations are less than 5?

b. Approximately what percentage of the observations are greater than 5?

6. The 54th percentile in a large data set is 98.6.

a. Approximately what percentage of the observations are less than 98.6?

b. Approximately what percentage of the observations are greater than 98.6?

7. In a large data set the 29th percentile is 5 and the 79th percentile is 10. Approximately what percentage of observations lie between 5 and 10?

8. In a large data set the 40th percentile is 125 and the 82nd percentile is 158. Approximately what percentage of observations lie between 125 and 158?

9. Find the five-number summary and the IQR and sketch the box plot for the sample represented by the stem and leaf diagram in Figure 2.2 "Ordered Stem and Leaf Diagram".

10. Find the five-number summary and the IQR and sketch the box plot for the sample explicitly displayed in Note 2.20 "Example 7" in Section 2.2 "Measures of Central Location".

11. Find the five-number summary and the IQR and sketch the box plot for the sample represented by the data frequency table

x	1	2	5	8	9
f	5	2	3	6	4

12. Find the five-number summary and the IQR and sketch the box plot for the sample represented by the data frequency table

x	−5	−3	−2	−1	0	1	3	4	5
f	2	1	3	2	4	1	1	2	1

13. Find the *z*-score of each measurement in the following sample data set.

$$-5 \quad 6 \quad 2 \quad -1 \quad 0$$

14. Find the *z*-score of each measurement in the following sample data set.

$$1.6 \quad 5.2 \quad 2.8 \quad 3.7 \quad 4.0$$

15. The sample with data frequency table

x	1	2	7
f	1	2	1

has mean $\bar{x} = 3$ and standard deviation $s \approx 2.71$. Find the *z*-score for every value in the sample.

16. The sample with data frequency table

x	−1	0	1	4
f	1	1	3	1

has mean $\bar{x} = 1$ and standard deviation $s \approx 1.67$. Find the *z*-score for every value in the sample.

17. For the population

$$0 \quad 0 \quad 2 \quad 2$$

compute each of the following.

a. The population mean μ.

b. The population variance σ^2.

c. The population standard deviation σ.

d. The z-score for every value in the population data set.

18. For the population

$$0.5 \;\; 2.1 \;\; 4.4 \;\; 1.0$$

compute each of the following.

a. The population mean μ.

b. The population variance σ^2.

c. The population standard deviation σ.

d. The z-score for every value in the population data set.

19. A measurement x in a sample with mean $\bar{x} = 10$ and standard deviation $s = 3$ has z-score $z = 2$. Find x.

20. A measurement x in a sample with mean $\bar{x} = 10$ and standard deviation $s = 3$ has z-score $z = -1$. Find x.

21. A measurement x in a population with mean $\mu = 2.3$ and standard deviation $\sigma = 1.3$ has z-score $z = 2$. Find x.

22. A measurement x in a sample with mean $\mu = 2.3$ and standard deviation $\sigma = 1.3$ has z-score $z = -1.2$. Find x.

APPLICATIONS

23. The weekly sales for the last 20 weeks in a kitchen appliance store for an electric automatic rice cooker are

$$\begin{array}{ccccc} 20 & 15 & 14 & 14 & 18 \\ 15 & 19 & 12 & 13 & 9 \\ 15 & 17 & 16 & 16 & 18 \\ 19 & 15 & 15 & 16 & 15 \end{array}$$

a. Find the percentile rank of 15.

b. If the sample accurately reflects the population, then what percentage of weeks would an inventory of 15 rice cookers be adequate?

24. The table shows the number of vehicles owned in a survey of 52 households.

x	0	1	2	3	4	5	6	7
f	2	12	15	11	6	3	1	2

a. Find the percentile rank of 2.

b. If the sample accurately reflects the population, then what percentage of households have at most two vehicles?

25. For two months Cordelia records her daily commute time to work each day to the nearest minute and obtains the following data:

x	26	27	28	29	30	31	32
f	3	4	16	12	6	2	1

Cordelia is supposed to be at work at 8:00 a.m. but refuses to leave her house before 7:30 a.m.

a. Find the percentile rank of 30, the time she has to get to work.

b. Assuming that the sample accurately reflects the population of *all* of Cordelia's commute times, use your answer to part (a) to predict the proportion of the work days she is late for work.

26. The mean score on a standardized grammar exam is 49.6; the standard deviation is 1.35. Dromio is told that the *z*-score of his exam score is −1.19.

a. Is Dromio's score above average or below average?

b. What was Dromio's actual score on the exam?

27. A random sample of 49 invoices for repairs at an automotive body shop is taken. The data are arrayed in the stem and leaf diagram shown. (Stems are thousands of dollars, leaves are hundreds, so that for example the largest observation is 3,800.)

3	5 6 8
3	0 0 1 1 2 4
2	5 6 6 7 7 8 8 9 9
2	0 0 0 0 1 2 2 4
1	5 5 5 6 6 7 7 7 8 8 9
1	0 0 1 3 4 4 4
0	5 6 8 8
0	4

For these data, $\Sigma x = 101{,}100$, $\Sigma x^2 = 244{,}830{,}000$.

a. Find the *z*-score of the repair that cost $1,100.

b. Find the *z*-score of the repairs that cost $2,700.

28. The stem and leaf diagram shows the time in seconds that callers to a telephone-order center were on hold before their call was taken.

0	0 0 0 0 0 0 1 1 1 1 1 1 1 2 2 2 2 2 3 3 3 3 3 3 3 4 4 4 4 4
0	5 5 5 5 5 5 5 5 5 6 6 6 6 6 6 6 6 6 6 7 7 7 7 7 7 8 8 8 9 9
1	0 0 1 1 1 1 2 2 2 2 4 4
1	5 6 6 8 9
2	2 4
2	5
3	0

a. Find the quartiles.

b. Give the five-number summary of the data.

c. Find the range and the IQR.

ADDITIONAL EXERCISES

29. Consider the data set represented by the ordered stem and leaf diagram

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

a. Find the three quartiles.

b. Give the five-number summary of the data.

c. Find the range and the IQR.

30. For the following stem and leaf diagram the units on the stems are thousands and the units on the leaves are hundreds, so that for example the largest observation is 3,800.

3	5 6 8
3	0 0 1 1 2 4
2	5 6 6 7 7 8 8 9 9
2	0 0 0 0 1 2 2 4
1	5 5 5 6 6 7 7 7 8 8 9
1	0 0 1 3 4 4 4
0	5 6 8 8
0	4

a. Find the percentile rank of 800.

b. Find the percentile rank of 3,200.

31. Find the five-number summary for the following sample data.

x	26	27	28	29	30	31	32
f	3	4	16	12	6	2	1

32. Find the five-number summary for the following sample data.

x	1	2	3	4	5	6	7	8	9	10
f	384	208	98	56	28	12	8	2	3	1

33. For the following stem and leaf diagram the units on the stems are thousands and the units on the leaves are hundreds, so that for example the largest observation is 3,800.

3	5 6 8
3	0 0 1 1 2 4
2	5 6 6 7 7 8 8 9 9
2	0 0 0 0 1 2 2 4
1	5 5 5 6 6 7 7 7 8 8 9
1	0 0 1 3 4 4 4
0	5 6 8 8
0	4

a. Find the three quartiles.

b. Find the IQR.

c. Give the five-number summary of the data.

34. Determine whether the following statement is true. “In any data set, if an observation x_1 is greater than another observation x_2, then the *z*-score of x_1 is greater than the *z*-score of x_2 . ”

35. Emilia and Ferdinand took the same freshman chemistry course, Emilia in the fall, Ferdinand in the spring. Emilia made an 83 on the common final exam that she took, on which the mean was 76 and the standard deviation 8. Ferdinand made a 79 on the common final exam that he took, which was more difficult, since the mean was 65 and the standard deviation 12. The one who has a higher *z*-score did relatively better. Was it Emilia or Ferdinand?

36. Refer to the previous exercise. On the final exam in the same course the following semester, the mean is 68 and the standard deviation is 9. What grade on the exam matches Emilia’s performance? Ferdinand’s?

37. Rosencrantz and Guildenstern are on a weight-reducing diet. Rosencrantz, who weighs 178 lb, belongs to an age and body-type group for which the mean weight is 145 lb and the standard deviation is 15 lb. Guildenstern, who weighs 204 lb, belongs to an age and body-type group for which the mean weight is 165 lb and the standard deviation is 20 lb. Assuming *z*-scores are good measures for comparison in this context, who is more overweight for his age and body type?

LARGE DATA SET EXERCISES

38. Large Data Set 1 lists the SAT scores and GPAs of 1,000 students.

    http://www.flatworldknowledge.com/sites/all/files/data1.xls

a. Compute the three quartiles and the interquartile range of the 1,000 SAT scores.
b. Compute the three quartiles and the interquartile range of the 1,000 GPAs.

39. Large Data Set 10 records the scores of 72 students on a statistics exam.

http://www.flatworldknowledge.com/sites/all/files/data10.xls

a. Compute the five-number summary of the data.
b. Describe in words the performance of the class on the exam in the light of the result in part (a).

40. Large Data Sets 3 and 3A list the heights of 174 customers entering a shoe store.

http://www.flatworldknowledge.com/sites/all/files/data3.xls

http://www.flatworldknowledge.com/sites/all/files/data3A.xls

a. Compute the five-number summary of the heights, without regard to gender.
b. Compute the five-number summary of the heights of the men in the sample.
c. Compute the five-number summary of the heights of the women in the sample.

41. Large Data Sets 7, 7A, and 7B list the survival times in days of 140 laboratory mice with thymic leukemia from onset to death.

http://www.flatworldknowledge.com/sites/all/files/data7.xls

http://www.flatworldknowledge.com/sites/all/files/data7A.xls

http://www.flatworldknowledge.com/sites/all/files/data7B.xls

a. Compute the three quartiles and the interquartile range of the survival times for all mice, without regard to gender.
b. Compute the three quartiles and the interquartile range of the survival times for the 65 male mice (separately recorded in Large Data Set 7A).
c. Compute the three quartiles and the interquartile range of the survival times for the 75 female mice (separately recorded in Large Data Set 7B).

ANSWERS

1. a. 60.

   b. 10.

3. a. 59.

   b. 23.

5. a. 29.

   b. 71.

7. 50%.

9. $x_{\min} = 25, Q_1 = 70, Q_2 = 77.5, Q_3 = 90, x_{\max} = 100, \text{IQR} = 20$

11. $x_{\min} = 1, Q_1 = 1.5, Q_2 = 6.5, Q_3 = 8, x_{\max} = 9, \text{IQR} = 6.5$

13. −1.3, 1.39, 0.4, −0.35, −0.11.

15. $z = -0.74$ for $x = 1$, $z = -0.37$ for $x = 2$, $z = 1.48$ for $x = 7$.

17. a. 1.

    b. 1.

    c. 1.

    d. $z = -1$ for $x = 0$, $z = 1$ for $x = 2$.

19. 16.

21. 4.9.

23. a. 55.

    b. 55.

25. a. 93.

    b. 0.07.

27. a. −1.11.

    b. 0.73.

29. a. $Q_1 = 59, Q_2 = 70, Q_3 = 81$.

    b. $x_{\min} = 39, Q_1 = 59, Q_2 = 70, Q_3 = 81, x_{\max} = 100$.

    c. $R = 61, \text{IQR} = 22$.

31. $x_{\min} = 26, Q_1 = 28, Q_2 = 28, Q_3 = 29, x_{\max} = 32$.

33. a. $Q_1 = 1450, Q_2 = 2000, Q_3 = 2800$.

    b. $\text{IQR} = 1350$.

    c. $x_{\min} = 400, Q_1 = 1450, Q_2 = 2000, Q_3 = 2800, x_{\max} = 3800$.

35. Emilia: $z = .875$, Ferdinand: $z = 1.1\overline{6}$.

37. Rosencrantz: $z = 2.2$, Guildenstern: $z = 1.95$. Rosencrantz is more overweight for his age and body type.

39. a. $x_{min} = 15, Q_1 = 51, Q_2 = 67, Q_3 = 82$, and $x_{max} = 97$.

    b. The data set appears to be skewed to the left.

41. a. $Q_1 = 440, Q_2 = 552.5, Q_3 = 661$, and $IQR = 221$.

    b. $Q_1 = 641, Q_2 = 667, Q_3 = 700$, and $IQR = 59$.

    c. $Q_1 = 407, Q_2 = 448, Q_3 = 504$, and $IQR = 97$.