CHAPTER 2: DESCRIPTIVE STATISTICS

Exercise 1. *The miles per gallon rating for 30 cars are shown below (lowest to highest).*
19, 19, 19, 20, 21, 21, 25, 25, 25, 26, 26, 28, 29, 31, 31, 32, 32, 33, 34, 35, 36, 37, 37, 38, 38, 38, 38, 41, 43, 43

Solution

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

Table 2.84

Exercise 2. *The height in feet of 25 trees is shown below (lowest to highest).*
25, 27, 33, 34, 34, 34, 35, 37, 37, 38, 39, 39, 39, 40, 41, 45, 46, 47, 49, 50, 50, 53, 53, 54, 54

Solution

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

Exercise 3. *The data are the prices of different laptops at an electronics store. Round each value to the nearest ten. Create a stem plot using the hundreds digits as the stems and the tens digits as the leaves.*
249, 249, 260, 265, 265, 280, 299, 299, 309, 319, 325, 326, 350, 350, 350, 365, 369, 389, 409, 459, 489, 559, 569, 570, 610

Solution

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

Table 2.85

Exercise 4. *The data are daily high temperatures in a town for one month.*
61, 61, 62, 64, 66, 67, 67, 67, 68, 69, 70, 70, 70, 71, 71, 72, 74, 74, 74, 75, 75, 75, 76, 76, 77, 78, 78, 79, 79, 95
For the next three exercises, use the data to construct a line graph.

Solution

Stem	Stem Leaf
6	1 1 2 4 6 7 7 7 8 9
7	0 0 0 1 1 2 4 4 4 5 5 5 6 6 7 8 8 9 9

8	
9	5

95 appears to be an outlier.

Exercise 5. *Use the data to construct a line graph. In a survey, 40 people were asked how many times they visited a store before making a major purchase. The results are shown in Table 2.37.*

Number of times in store Frequency	***Frequency***
1	*4*
2	*10*
3	*16*
4	*6*
5	*4*

Table 2.37

Solution

Figure 2.51

Exercise 6. *Use the data to construct a line graph. In a survey, several people were asked how many years it has been since they purchased a mattress. The results are shown in Table 2.38.*

Years since last purchase	***Frequency***
0	*2*
1	*8*
2	*13*
3	*22*
4	*16*
5	*9*

Table 2.38

Solution

Exercise 7. *Several children were asked how many TV shows they watch each day. The results of the survey are shown in Table 2.39.*

Number of TV Shows	***Frequency***
0	*12*
1	*18*
2	*36*
3	*7*
4	*2*

Table 2.39

Solution

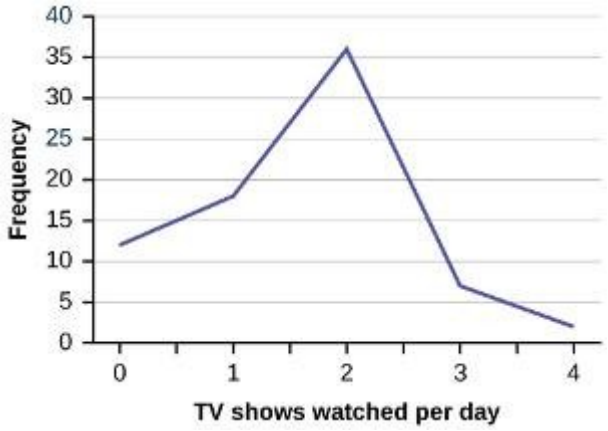


Figure 2.52

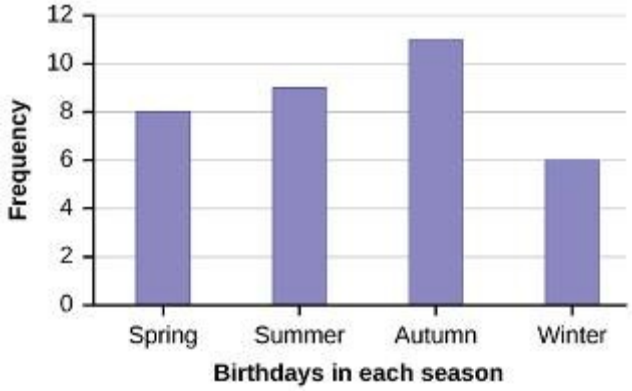
Exercise 8. *The students in Ms. Ramirez's math class have birthdays in each of the four seasons. Table 2.40 shows the four seasons, the number of students who have birthdays in each season, and the percentage (%) of students in each group. Construct a bar graph showing the number of students.*

Seasons	***Number of students***	***Proportion of population***
Spring	*8*	*24%*
Summer	*9*	*26%*

Autumn	*11*	*32%*
Winter	*6*	*18%*

Table 2.40

Solution



Exercise 9. *Using the data from Mrs. Ramirez's math class supplied in **Exercise 2.8**, construct a bar graph showing the percentages.*

Solution

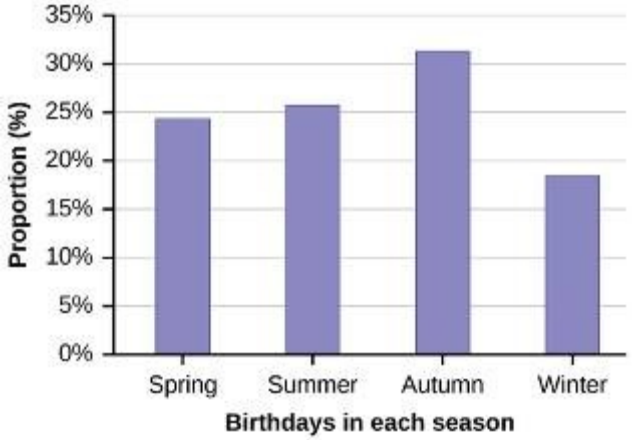


Figure 2.53

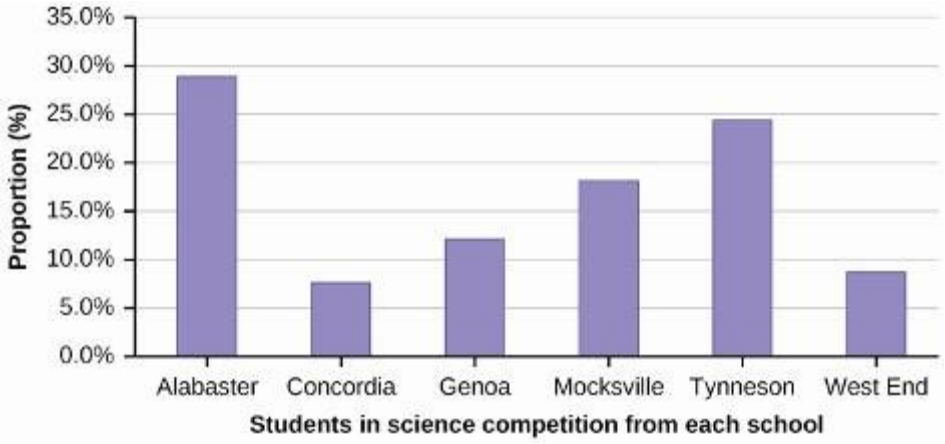
Exercise 10. *David County has six high schools. Each school sent students to participate in a county-wide science competition. Table 2.41 shows the percentage breakdown of competitors from each school, and the percentage of the entire student population of the county that goes to each school. Construct a bar graph that shows the population percentage of competitors from each school.*

High School	***Science competition population***	***Overall student population***
Alabaster	*28.9%*	*8.6%*
Concordia	*7.6%*	*23.2%*
Genoa	*12.1%*	*15.0%*
Mocksville	*18.5%*	*14.3%*
Tynneson	*24.2%*	*10.1%*

West End	*8.7%*	*28.8%*

Table 2.41

Solution



Exercise 11. *Use the data from the David County science competition supplied in Exercise 2.10. Construct a bar graph that shows the county-wide population percentage of students at each school.*

Solution

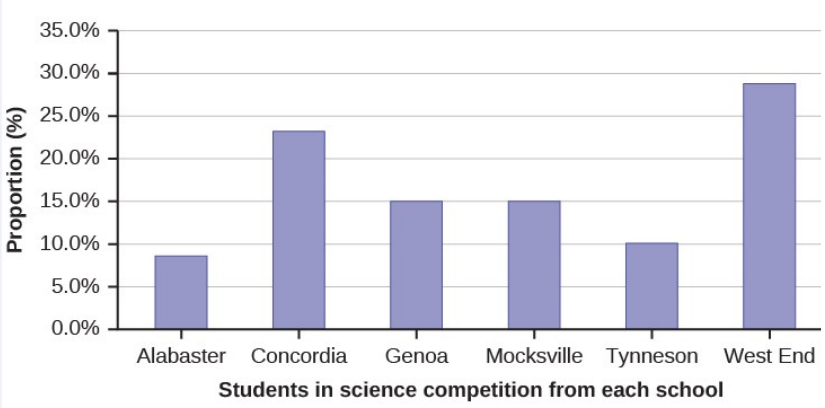


Figure 2.54FiFigure 2.54

Exercise 12. *Sixty-five randomly selected car salespersons were asked the number of cars they generally sell in one week. Fourteen people answered that they generally sell three cars; nineteen generally sell four cars; twelve generally sell five cars; nine generally sell six cars; eleven generally sell seven cars. Complete the table.*

Data Value (# cars)	***Frequency***	***Relative Frequency***	***Cumulative Relative Frequency***

Table 2.42

Solution

Data Value (# cars)	**Frequency**	**Relative Frequency**	**Cumulative Relative Frequency**
3	14	0.22	0.22
4	19	0.29	0.51
5	12	0.18	0.69
6	9	0.14	0.83
7	11	0.17	1.00

Exercise 13. *What does the frequency column in Table 2.42 sum to? Why?*

Solution 65

Exercise 14. *What does the relative frequency column in Table 2.42 sum to? Why?*

Solution 1

Exercise 15. *What is the difference between relative frequency and frequency for each data value in Table 2.42?*

Solution The relative frequency shows the *proportion* of data points that have each value. The frequency tells the *number* of data points that have each value.

Exercise 16. *What is the difference between cumulative relative frequency and relative frequency for each data value in Table 2.42?*

Solution The relative frequency shows the proportion of data points that have each value. The cumulative relative frequency tells the proportion of data points that are equal to or less than each value.

Exercise 17. *To construct the histogram for the data in Table 2.42, determine appropriate minimum and maximum x and y values and the scaling. Sketch the histogram. Label the horizontal and vertical axes with words. Include numerical scaling.*

Solution Answers will vary. One possible histogram is shown:

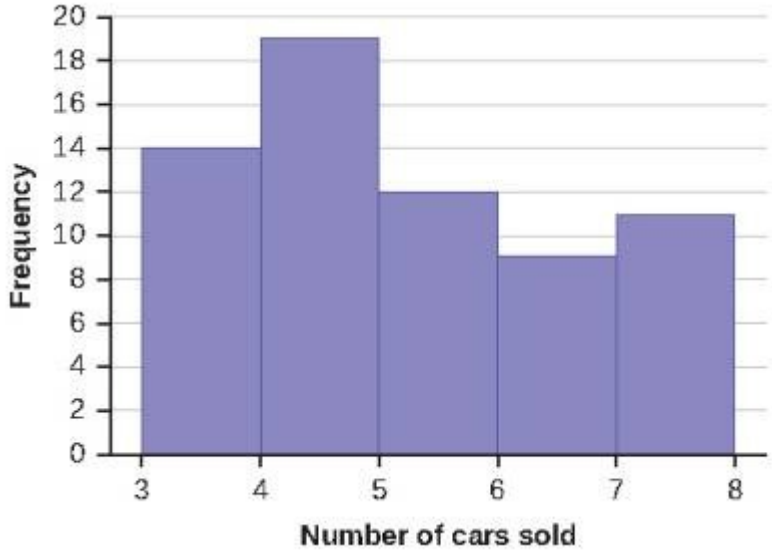


Figure 2.55

Exercise 18. *Construct a frequency polygon for the following:*
a.

Pulse Rates for Women	***Frequency***
60–69	*12*
70–79	*14*
80–89	*11*
90–99	*1*

100–109	*1*
110–119	*0*
120–129	*1*

Table 2.43

b.

Actual Speed in a 30 MPH Zone	***Frequency***
42–45	*25*
46–49	*14*
50–53	*7*
54–57	*3*
58–61	*1*

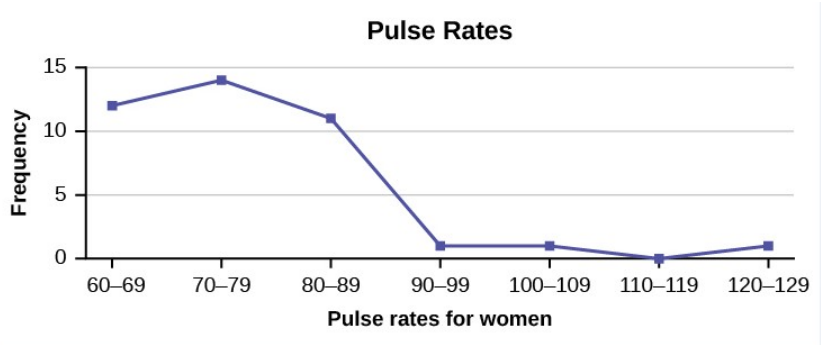
Table 2.44

c.

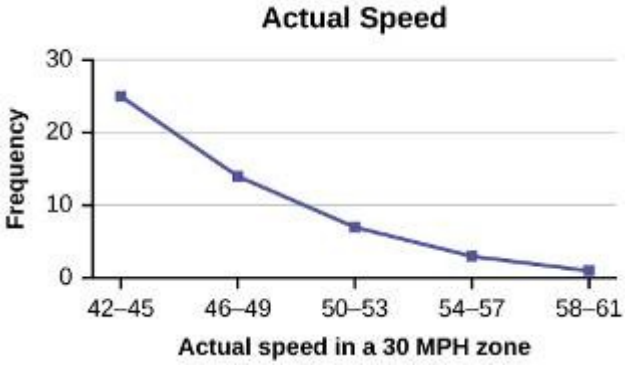
Tar (mg) in Nonfiltered Cigarettes	***Frequency***
10–13	*1*
14–17	*0*
18–21	*15*
22–25	*7*
26–29	*2*

Table 2.45

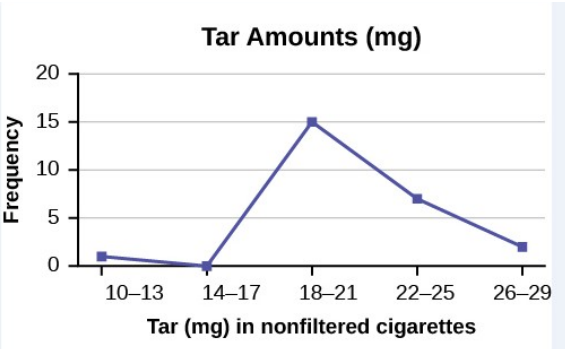
Solution a. Find the midpoint for each class. These will be graphed on the *x*-axis. The frequency values will be graphed on the *y*-axis values.



b. Find the midpoint for each class. These will be graped on the *x*-axis. The frequency values will be graphed on the *y*-axis values



c. Find the midpoint for each class. These will be graped on the *x*-axis. The frequency values will be graphed on the *y*-axis values.



Exercise 19. *Construct a frequency polygon from the frequency distribution for the 50 highest ranked countries for depth of hunger.*

Depth of Hunger	***Frequency***
230–259	*21*

260–289	*13*
290–319	*5*
320–349	*7*
350–379	*1*
380–409	*1*
410–439	*1*

Table 2.46

Solution Find the midpoint for each class. These will be graphed on the *x*-axis. The frequency values will be graphed on the *y*-axis values.

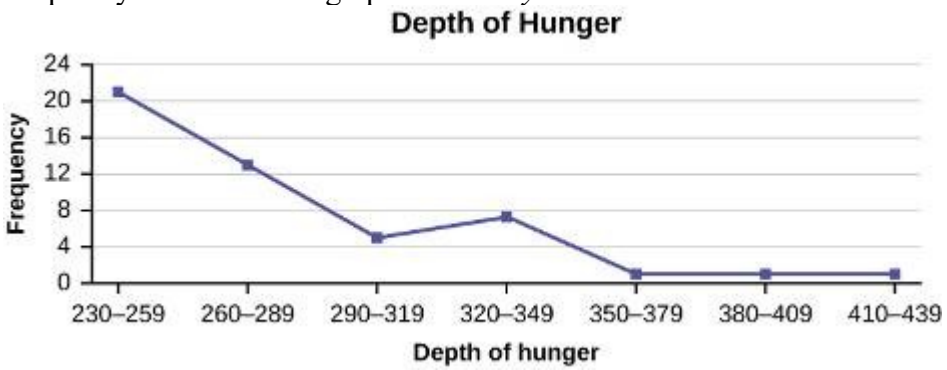


Figure 2.56

Exercise 20. *Use the two frequency tables to compare the life expectancy of men and women from 20 randomly selected countries. Include an overlayed frequency polygon and discuss the shapes of the distributions, the center, the spread, and any outliers. What can we conclude about the life expectancy of women compared to men?*

Life Expectancy at Birth – Women	***Frequency***
49–55	*3*
56–62	*3*
63–69	*1*

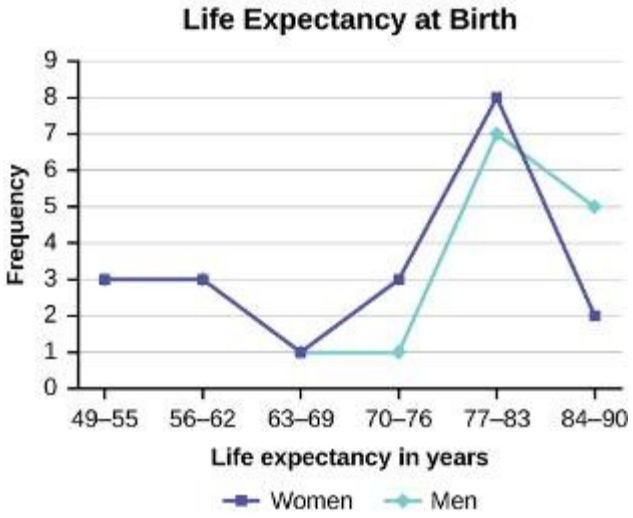
70–76	*3*
77–83	*8*
84–90	*2*

Table 2.47

Life Expectancy at Birth - Men	***Frequency***
49–55	*3*
56–62	*3*
63–69	*1*
70–76	*1*
77–83	*7*
84–90	*5*

Table 2.48

Solution



From the shape of the distributions, we see that the graphs for men and women are both skewed to the left. This means that a larger proportion of the adults in this sample die before the age of 80 than after the age of 80. Calculating the mean life expectancy for women, we have an average life expectancy of 71.6 years with a standard deviation of 11.93 years. For the men, we have an average

life expectancy of 73.4 years with a standard deviation of 13.14 years. We can conclude that the women tend to live slightly longer than men in this sample.

Exercise 21. *Construct a times series graph for (a) the number of male births, (b) the number of female births, and (c) the total number of births.*

Sex/ Year	*1855*	*1856*	*1857*	*1858*	*1859*	*1860*	*1861*
Female	*45,545*	*49,582*	*50,257*	*50,324*	*51,915*	*51,220*	*52,403*
Male	*47,804*	*52,239*	*53,158*	*53,694*	*54,628*	*54,409*	*54,606*
Total	*93,349*	*101,821*	*103,415*	*104,018*	*106,543*	*105,629*	*107,009*

Table 2.49

Sex/ Year	*1862*	*1863*	*1864*	*1865*	*1866*	*1867*	*1868*	*1869*
Female	*51,812*	*53,115*	*54,959*	*54,850*	*55,307*	*55,527*	*56,292*	*55,033*
Male	*55,257*	*56,226*	*57,374*	*58,220*	*58,360*	*58,517*	*59,222*	*58,321*
Total	*107,069*	*109,341*	*57,374*	*113,070*	*113,667*	*58,517*	*115,514*	*113,354*

Table 2.50

Sex/ Year	*1870*	*1871*	*1872*	*1873*	*1874*	*1875*
Female	*56,431*	*56,099*	*57,472*	*58,233*	*60,109*	*60,146*
Male	*58,959*	*60,029*	*61,293*	*61,467*	*63,602*	*63,432*
Total	*115,390*	*116,128*	*118,765*	*119,700*	*123,711*	*123,578*

Table 2.51

Solution

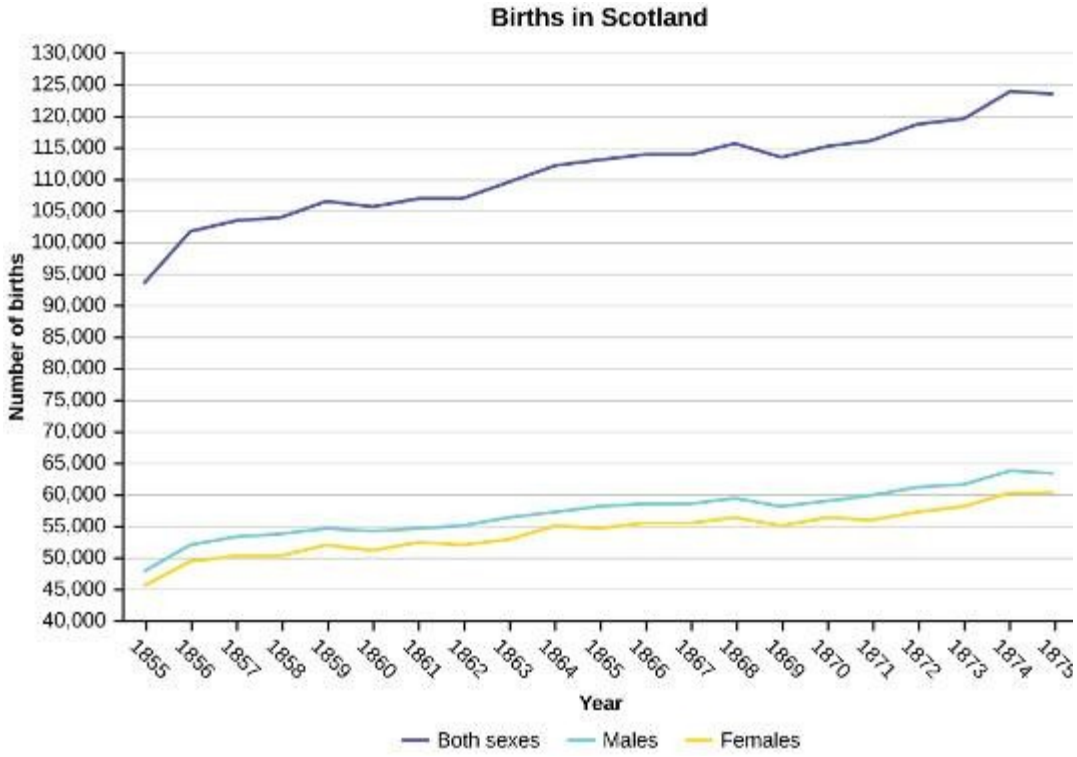


Figure 2.57

Exercise 22. *The following data sets list full time police per 100,000 citizens along with homicides per 100,000 citizens for the city of Detroit, Michigan during the period from 1961 to 1973.*

Year	*1961*	*1962*	*1963*	*1964*	*1965*	*1966*	*1967*
Police	*260.35*	*269.8*	*272.04*	*272.96*	*272.51*	*261.34*	*268.89*
Homicides	*8.6*	*8.9*	*8.52*	*8.89*	*13.07*	*14.57*	*21.36*

Table 2.52

Year	*1968*	*1969*	*1970*	*1971*	*1972*	*1973*
Police	*295.99*	*319.87*	*341.43*	*356.59*	*376.69*	*390.19*
Homicides	*28.03*	*31.49*	*37.39*	*46.26*	*47.24*	*52.33*

Table 2.53

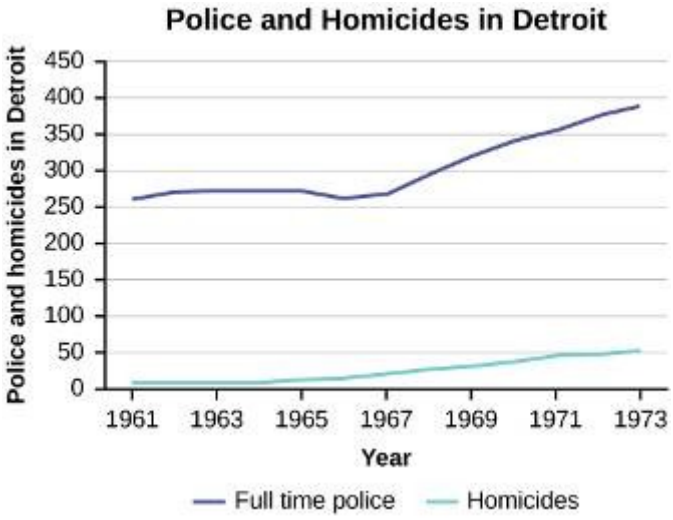
a. Construct a double time series graph using a common x-axis for both sets of data.

b. Which variable increased the fastest? Explain.

c. Did Detroit's increase in police officers have an impact on the murder rate? Explain.

Solution

a.



b. From the double time series plot, we can see that both police staff levels and murder rates increased between 1961 and 1973. If we look at the slopes for both lines, we can see the slope associated with police staffing is steeper than the slope for homicide rates after the year 1967. Before 1967, the two lines have very similar slopes.

c. Even though we see a greater rate of change in the number of police officers as compared to the homicide rate, it is difficult to say that the increased levels of police officers impacted the rate of homicides. Perhaps the increased rate of homicides impacted the level of police staffing.

Exercise 23.

Listed are 29 ages for Academy Award winning best actors in order from smallest to largest.
18; 21; 22; 25; 26; 27; 29; 30; 31; 33; 36; 37; 41; 42; 47; 52; 55; 57; 58; 62; 64; 67; 69; 71; 72; 73; 74; 76; 77
a. Find the 40^{th} percentile.
b. Find the 78^{th} percentile.

Solution

a. The 40^{th} percentile is 37 years.
b. The 78^{th} percentile is 70 years.

Exercise 24.

Listed are 32 ages for Academy Award winning best actors in order from smallest to largest. 18; 18; 21; 22; 25; 26; 27; 29; 30; 31; 31; 33; 36; 37; 37; 41; 42; 47; 52; 55; 57; 58; 62; 64; 67; 69; 71; 72; 73; 74; 76; 77
a. Find the percentile of 37.
b. Find the percentile of 72.

Solution

a. 37 is the 44^{th} percentile.
b. 72 is the 86^{th} percentile.

Exercise 25. *Jesse was ranked 37th in his graduating class of 180 students. At what percentile is Jesse's ranking?*

Solution Jesse graduated 37th out of a class of 180 students. There are 180 – 37 = 143 students ranked below Jesse. There is one rank of 37. $x = 143$ and $y = 1$. $\frac{x+0.5y}{n}(100) = \frac{143+0.5(1)}{180}(100) = 79.72$. Jesse's rank of 37 puts him at the 80th percentile.

Exercise 26. *a. For runners in a race, a low time means a faster run. The winners in a race have the shortest running times. Is it more desirable to have a finish time with a high or a low percentile when running a race?*
b. The 20th percentile of run times in a particular race is 5.2 minutes. Write a sentence interpreting the 20th percentile in the context of the situation.
c. A bicyclist in the 90th percentile of a bicycle race completed the race in 1 hour and 12 minutes. Is he among the fastest or slowest cyclists in the race? Write a sentence interpreting the 90th percentile in the context of the situation.

Solution a. For runners in a race it is more desirable to have a low percentile for a finish time. A low percentile means a short time which is faster.
b. 20% of runners finished the race in 5.2 minutes or less. 80% of runners finished the race in 5.2 minutes or longer.
c. He is among the slowest cyclists because 90% of cyclists were faster than him. 90% of cyclists had a finish time of one hour, 12 minutes or less. Only 10% of cyclists had a finish time of one hour, 12 minutes or longer

Exercise 27. *a. For runners in a race, a higher speed means a faster run. Is it more desirable to have a speed with a high or a low percentile when running a race?*
b. The 40th percentile of speeds in a particular race is 7.5 miles per hour. Write a sentence interpreting the 40th percentile in the context of the situation.

Solution a. For runners in a race it is more desirable to have a high percentile for speed. A high percentile means a higher speed which is faster.
b. 40% of runners ran at speeds of 7.5 miles per hour or less (slower). 60% of runners ran at speeds of 7.5 miles per hour or more (faster).

Exercise 28. *On an exam, would it be more desirable to earn a grade with a high or low percentile? Explain.*

Solution On an exam, a high percentile is preferable; higher percentiles correspond to higher grades on the exam.

Exercise 29. *Mina is waiting in line at the Department of Motor Vehicles (DMV). Her wait time of 32 minutes is the 85th percentile of wait times. Is that good or bad?*

Write a sentence interpreting the 85th percentile in the context of this situation.

Solution When waiting in line at the DMV, the 85th percentile would be a long wait time compared to the other people waiting. 85% of people had shorter wait times than Mina. In this context, Mina would prefer a wait time corresponding to a lower percentile. 85% of people at the DMV waited 32 minutes or less. 15% of people at the DMV waited 32 minutes or longer.

Exercise 30. *In a survey collecting data about the salaries earned by recent college graduates, Li found that her salary was in the 78th percentile. Should Li be pleased or upset by this result? Explain.*

Solution Li should be pleased; her salary is relatively high compared to other recent college graduates. 78% of recent college graduates earn less than Li does. 22% of recent college graduates earn more than Li does.

Exercise 31. *In a study collecting data about the repair costs of damage to automobiles in a certain type of crash tests, a certain model of car had $1,700 in damage and was in the 90th percentile. Should the manufacturer and the consumer be pleased or upset by this result? Explain and write a sentence that interprets the 90th percentile in the context of this problem.*

Solution The manufacturer and the consumer would be upset. This is a large repair cost for the damages, compared to the other cars in the sample.
INTERPRETATION: 90% of the crash tested cars had damage repair costs of $1700 or less; only 10% had damage repair costs of $1700 or more.

Exercise 32. *The University of California has two criteria used to set admission standards for freshman to be admitted to a college in the UC system:*
a. Students' GPAs and scores on standardized tests (SATs and ACTs) are entered into a formula that calculates an "admissions index" score. The admissions index score is used to set eligibility standards intended to meet the goal of admitting the top 12% of high school students in the state. In this context, what percentile does the top 12% represent?
b. Students whose GPAs are at or above the 96th percentile of all students at their high school are eligible (called eligible in the local context), even if they are not in the top 12% of all students in the state. What percentage of students from each high school are "eligible in the local context"?

Solution a. The top 12% of students are those who are at or above the **88th percentile** of admissions index scores.
b. The **top 4%** of students' GPAs are at or above the 96th percentile, making the top 4% of students "eligible in the local context".

Exercise 33. *Suppose that you are buying a house. You and your realtor have determined that the most expensive house you can afford is the 34th percentile. The 34th*

percentile of housing prices is $240,000 in the town you want to move to. In this town, can you afford 34% of the houses or 66% of the houses?

Solution You can afford 34% of houses. 66% of the houses are too expensive for your budget. INTERPRETATION: 34% of houses cost $240,000 or less. 66% of houses cost $240,000 or more.

Exercise 34. *Sixty-five randomly selected car salespersons were asked the number of cars they generally sell in one week. Fourteen people answered that they generally sell three cars; nineteen generally sell four cars; twelve generally sell five cars; nine generally sell six cars; eleven generally sell seven cars.*
First quartile = _____

Solution 4

Exercise 35. *Sixty-five randomly selected car salespersons were asked the number of cars they generally sell in one week. Fourteen people answered that they generally sell three cars; nineteen generally sell four cars; twelve generally sell five cars; nine generally sell six cars; eleven generally sell seven cars.*
Second quartile = median = 50^{th} percentile = _____

Solution 4

Exercise 36. *Sixty-five randomly selected car salespersons were asked the number of cars they generally sell in one week. Fourteen people answered that they generally sell three cars; nineteen generally sell four cars; twelve generally sell five cars; nine generally sell six cars; eleven generally sell seven cars.*
Third quartile = _____

Solution 6

Exercise 37. *Sixty-five randomly selected car salespersons were asked the number of cars they generally sell in one week. Fourteen people answered that they generally sell three cars; nineteen generally sell four cars; twelve generally sell five cars; nine generally sell six cars; eleven generally sell seven cars.*
Interquartile range (IQR) = _____ – _____ = _____

Solution 6 – 4=2

Exercise 38. *Sixty-five randomly selected car salespersons were asked the number of cars they generally sell in one week. Fourteen people answered that they generally sell three cars; nineteen generally sell four cars; twelve generally sell five cars; nine generally sell six cars; eleven generally sell seven cars.*
10^{th} percentile = _____

Solution 3

Exercise 39. *Sixty-five randomly selected car salespersons were asked the number of cars they generally sell in one week. Fourteen people answered that they generally sell three cars; nineteen generally sell four cars; twelve generally sell five cars; nine generally sell six cars; eleven generally sell seven cars.*
70^{th} percentile = _____

Solution 6

Exercise 40. *Sixty-five randomly selected car salespersons were asked the number of cars they generally sell in one week. Fourteen people answered that they generally sell three cars; nineteen generally sell four cars; twelve generally sell five cars; nine generally sell six cars; eleven generally sell seven cars.*
Construct a box plot below. Use a ruler to measure and scale accurately.

Solution Min: 3
Q_1: 4
M: 4
Q_3: 6
Max: 7
For box plot, check student's solution.

Exercise 41. *Sixty-five randomly selected car salespersons were asked the number of cars they generally sell in one week. Fourteen people answered that they generally sell three cars; nineteen generally sell four cars; twelve generally sell five cars; nine generally sell six cars; eleven generally sell seven cars.*

Looking at your box plot, does it appear that the data are concentrated together, spread out evenly, or concentrated in some areas, but not in others? How can you tell?

Solution More than 25% of salespersons sell four cars in a typical week. You can see this concentration in the box plot because the first quartile is equal to the median. The top 25% and the bottom 25% are spread out evenly; the whiskers have the same length.

Exercise 42. *Find the mean for the following frequency tables.*
a.

Grade	***Frequency***
49.5–59.5	*2*

59.5–69.5	*3*
69.5–79.5	*8*
79.5–89.5	*12*
89.5–99.5	*5*

Table 2.54

b.

Daily Low Temperature	***Frequency***
49.5–59.5	*53*
59.5–69.5	*32*
69.5–79.5	*15*
79.5–89.5	*1*
89.5–99.5	*0*

Table 2.55

c.

Points per Game	***Frequency***
49.5–59.5	*14*
59.5–69.5	*32*
69.5–79.5	*15*
79.5–89.5	*23*
89.5–99.5	*2*

Table 2.56

Solution

a. Find the mean for the following frequency tables

◦ Find the midpoints:

$$\frac{49.5+59.5}{2}, \frac{59.5+69.5}{2}, \frac{69.5+79.5}{2}, \frac{79.5+89.5}{2}, \frac{89.5+99.5}{2}$$

◦ Find the sum of the products of each midpoint and its associated frequency

54.5(2) + 64.5(3) + 74.5(8) + 84.5(12) + 94.5(5)

◦ Divide by the total number of frequencies

$$\frac{54.5(2)+64.5(3)+74.5(8)+84.5(12)+94.5(5)}{2+3+8+12+5}$$

$$\bar{x} = \frac{2385}{30} = 79.5$$

b. The mean of the sample is

◦ Find the midpoints:

$$\frac{49.5+59.5}{2}, \frac{59.5+69.5}{2}, \frac{69.5+79.5}{2}, \frac{79.5+89.5}{2}, \frac{89.5+99.5}{2}$$

◦ Find the sum of the products of each midpoint and its associated frequency

54.5(53) + 64.5(32) + 74.5(15) + 84.5(1) + 94.5(0)

◦ Divide by the total number of frequencies

$$\frac{54.5(53)+64.5(32)+74.5(15)+84.5(1)+94.5(0)}{53+32+15+1+0}$$

$$\bar{x} = \frac{6154.5}{101} = 60.94$$

c. The mean of the sample is

Points per Game	**Frequency**
49.5–59.5	14
59.5–69.5	32
69.5–79.5	15
79.5–89.5	23
89.5–99.5	2

◦ Find the midpoints:

$$\frac{49.5+59.5}{2}, \frac{59.5+69.5}{2}, \frac{69.5+79.5}{2}, \frac{79.5+89.5}{2}, \frac{89.5+99.5}{2}.$$

◦ Find the sum of the products of each midpoint and its associated frequency

54.5(14) + 64.5(32) + 74.5(15) + 84.5(23) + 94.5(2)

◦ Divide by the total number of frequencies

$$\frac{54.5(14)+64.5(32)+74.5(15)+84.5(23)+94.5(2)}{14+32+15+23+2}$$

The mean of the sample is $\bar{x} = \frac{6077}{86} = 70.66$.

Exercise 43.

The following data show the lengths of boats moored in a marina. The data are ordered from smallest to largest: 16; 17; 19; 20; 20; 21; 23; 24; 25; 25; 25; 26; 26; 27; 27; 27; 28; 29; 30; 32; 33; 33; 34; 35; 37; 39; 40

Calculate the mean.

Solution Mean: 16 + 17 + 19 + 20 + 20 + 21 + 23 + 24 + 25 + 25 + 25 + 26 + 26 + 27 + 27 + 27 + 28 + 29 + 30 + 32 + 33 + 33 + 34 + 35 + 37 + 39 + 40 = 738, $\frac{738}{27} = 27.33$

Exercise 44. *The following data show the lengths of boats moored in a marina. The data are ordered from smallest to largest: 16; 17; 19; 20; 20; 21; 23; 24; 25; 25; 25; 26; 26; 27; 27; 27; 28; 29; 30; 32; 33; 33; 34; 35; 37; 39; 40*
Identify the median.

Solution Starting at the smallest value, the median is the 14th term, which is 27.

Exercise 45. *The following data show the lengths of boats moored in a marina. The data are ordered from smallest to largest: 16; 17; 19; 20; 20; 21; 23; 24; 25; 25; 25; 26; 26; 27; 27; 27; 28; 29; 30; 32; 33; 33; 34; 35; 37; 39; 40*
Identify the mode.

Solution The most frequent lengths are 25 and 27, which occur three times. Mode = 25, 27

Exercise 46. *Sixty-five randomly selected car salespersons were asked the number of cars they generally sell in one week. Fourteen people answered that they generally sell three cars; nineteen generally sell four cars; twelve generally sell five cars; nine generally sell six cars; eleven generally sell seven cars. Calculate the following:*

sample mean = $\bar{x}$ = _____

Solution 4.75

Exercise 47. *Sixty-five randomly selected car salespersons were asked the number of cars they generally sell in one week. Fourteen people answered that they generally sell three cars; nineteen generally sell four cars; twelve generally sell five cars; nine generally sell six cars; eleven generally sell seven cars. Calculate the following:*

median = _____

Solution 4

Exercise *Sixty-five randomly selected car salespersons were asked the number of cars*

48. *they generally sell in one week. Fourteen people answered that they generally sell three cars; nineteen generally sell four cars; twelve generally sell five cars; nine generally sell six cars; eleven generally sell seven cars. Calculate the following:*

mode = _____

Solution 4

Exercise 49. *State whether the data are symmetrical, skewed to the left, or skewed to the right.*

1; 1; 1; 2; 2; 2; 2; 3; 3; 3; 3; 3; 3; 3; 3; 4; 4; 4; 5; 5

Solution The data are symmetrical. The median is 3 and the mean is 2.85. They are close, and the mode lies close to the middle of the data, so the data are symmetrical.

Exercise 50. *State whether the data are symmetrical, skewed to the left, or skewed to the right.*

16; 17; 19; 22; 22; 22; 22; 22; 23

Solution The data are skewed left. The median is 22 and the mean is 20.6. They are not close, and the mode lies to the right of the middle of the data, so the data are skewed left.

Exercise 51. *State whether the data are symmetrical, skewed to the left, or skewed to the right.*

87; 87; 87; 87; 87; 88; 89; 89; 90; 91

Solution The data are skewed right. The median is 87.5 and the mean is 88.2. Even though they are close, the mode lies to the left of the middle of the data, and there are many more instances of 87 than any other number, so the data are skewed right.

Exercise 52. *State whether the data are symmetrical, skewed to the left, or skewed to the right.*

When the data are skewed left, what is the typical relationship between the mean and median?

Solution When the data are skewed left, the mean is smaller than the median.

Exercise 53. *State whether the data are symmetrical, skewed to the left, or skewed to the right.*

When the data are symmetrical, what is the typical relationship between the mean and median?

Solution When the data are symmetrical, the mean and median are close or the same.

Exercise 54. *State whether the data are symmetrical, skewed to the left, or skewed to the right.*

What word describes a distribution that has two modes?

Solution bimodal

Exercise 55. *Describe the shape of this distribution.*

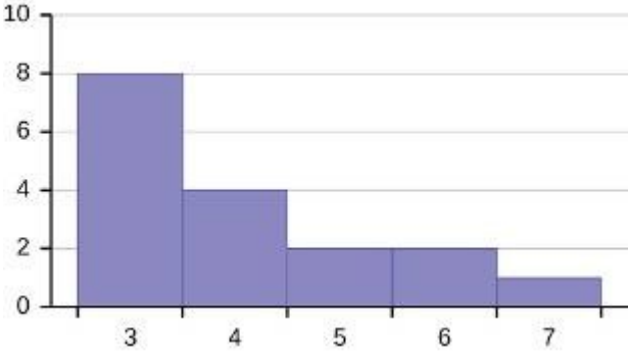


Figure 2.32

Solution The distribution is skewed right because it looks pulled out to the right.

Exercise 56. *Describe the relationship between the mode and the median of this distribution.*

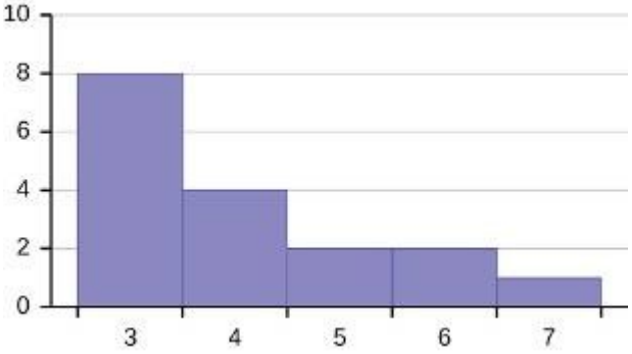


Figure 2.33

Solution The mode is three and is less than the median, which is four.

Exercise 57. *Describe the relationship between the mean and the median of this distribution.*

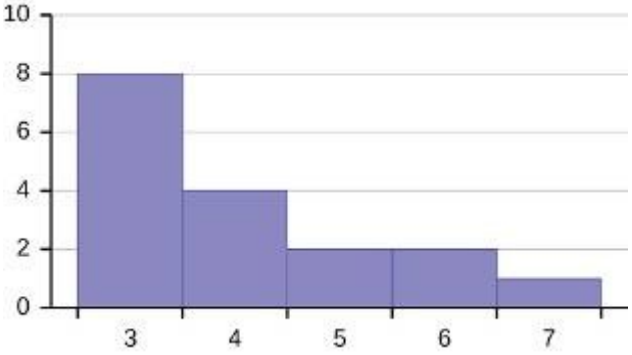


Figure 2.34

Solution The mean is 4.1 and is slightly greater than the median, which is four.

Exercise 58. *Describe the shape of this distribution.*

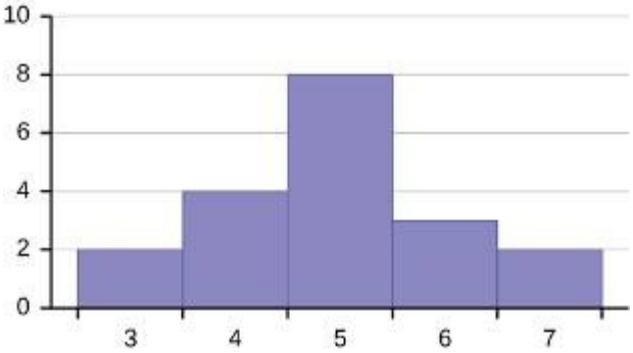


Figure 2.35

Solution The distribution is symmetrical because it peaks at the center.

Exercise 59. *Describe the relationship between the mode and the median of this distribution.*

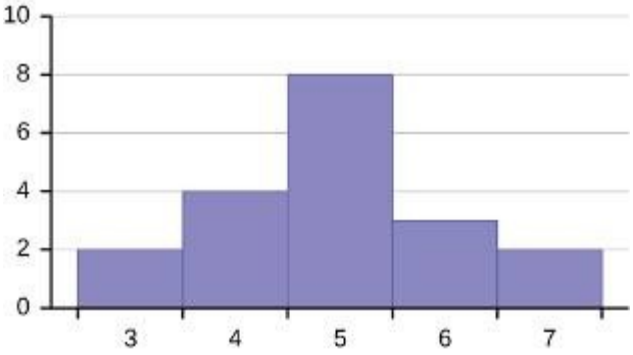


Figure 2.36

Solution The mode and the median are the same. In this case, they are both five.

Exercise 60. *Are the mean and the median the exact same in this distribution? Why or why not?*

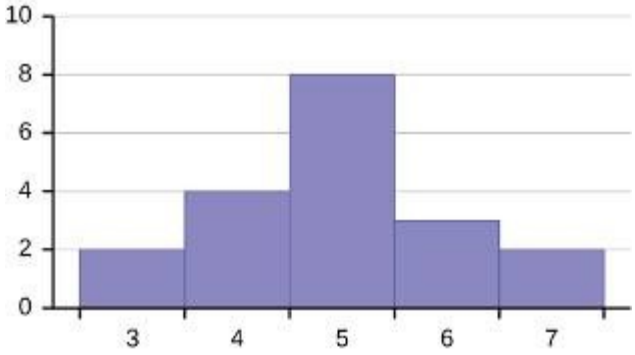


Figure 2.37

Solution No, the mean is 4.9 and the median is five. The distribution, though close to symmetrical, is not perfectly symmetrical, so the mean and median will differ slightly.

Exercise 61. *Describe the shape of this distribution.*

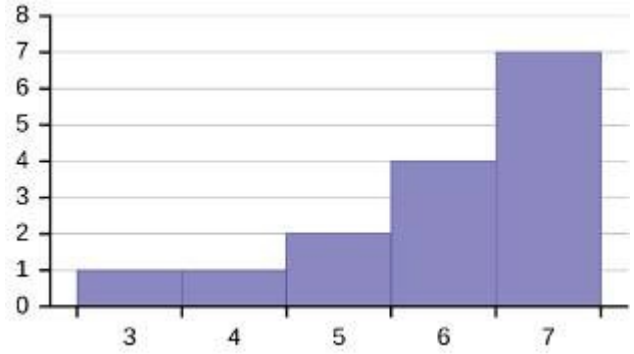


Figure 2.38

Solution The distribution is skewed left because it looks pulled out to the left.

Exercise 62. *Describe the relationship between the mode and the median of this distribution.*

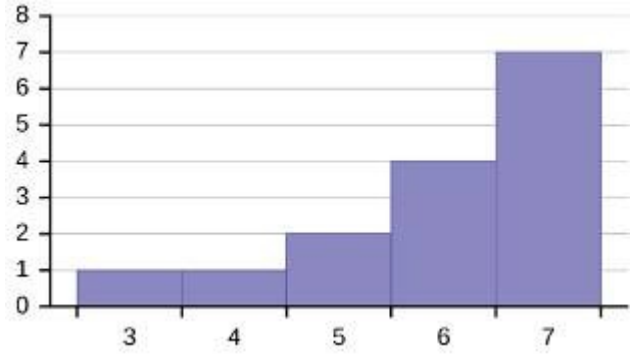


Figure 2.39

Solution The mode is seven and is the greater than the median, which is six.

Exercise 63. *Describe the relationship between the mean and the median of this distribution.*

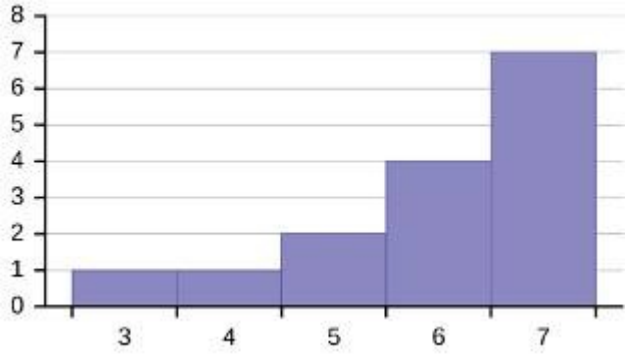


Figure 2.40

Solution The mean and the median are both six.

Exercise 64. *The mean and median for the data are the same.*
3; 4; 5; 5; 6; 6; 6; 6; 7; 7; 7; 7; 7; 7; 7
Is the data perfectly symmetrical? Why or why not?

Solution No, even though the mean and the median are the same, the mode lies to the right of the median. The data are skewed to the left.

Exercise 65. *Which is the greatest, the mean, the mode, or the median of the data set?*
11; 11; 12; 12; 12; 12; 13; 15; 17; 22; 22; 22

Solution The mode is 12, the median is 12.5, and the mean is 15.1. The mean is the largest.

Exercise 66. *Which is the least, the mean, the mode, and the median of the data set?*
56; 56; 56; 58; 59; 60; 62; 64; 64; 65; 67

Solution The mode is 56, the median is 60, and the mean is 60.6. The mode is the least value.

Exercise 67. *Of the three measures, which tends to reflect skewing the most, the mean, the mode, or the median? Why?*

Solution The mean tends to reflect skewing the most because it is affected the most by outliers.

Exercise 68. *In a perfectly symmetrical distribution, when would the mode be different from the mean and median?*

Solution The mode would be different if the distribution were bimodal, which means there would be two modes, one on either side of the mean and median.

Exercise 69. *The following data are the distances between 20 retail stores and a large distribution center. The distances are in miles. 29; 37; 38; 40; 58; 67; 68; 69; 76; 86; 87; 95; 96; 96; 99; 106; 112; 127; 145; 150*

Use a graphing calculator or computer to find the standard deviation and round to the nearest tenth.

Solution $s = 34.5$

Exercise 70. *The following data are the distances between 20 retail stores and a large distribution center. The distances are in miles. 29; 37; 38; 40; 58; 67; 68; 69; 76; 86; 87; 95; 96; 96; 99; 106; 112; 127; 145; 150*

Find the value that is one standard deviation below the mean.

Solution Mean = 29 + 37 + 38 + 40 + 58 + 67 + 68 + 69 + 76 + 86 + 87 + 95 + 96 + 96 + 99 + 106 + 112 + 127 + 145 + 150 $= 1681, \frac{1681}{20} = 84.05$ $\bar{x}$ -1(s) = 84.05 – 1(34.5) = 49.55

Exercise 71. *Two baseball players, Fredo and Karl, on different teams wanted to find out who had the higher batting average when compared to his team. Which baseball player had the higher batting average when compared to his team?*

Baseball Player	***Batting Average***	***Team Batting Average***	***Team Standard Deviation***
Fredo	*0.158*	*0.166*	*0.012*
Karl	*0.177*	*0.189*	*0.015*

Table 2.57

Solution For Fredo: $z = \frac{0.158 - 0.166}{0.012} = -0.67$ For Karl: $z = \frac{0.177 - 0.189}{0.015} = -0.8$ Fredo's z-score of -0.67 is higher than Karl's z-score of -0.8. For batting average, higher values are better, so Fredo has a better batting average compared to his team.

Exercise 72. *Use **Table 2.57** to find the value that is three standard deviations:*
a. above the mean.
b. below the mean.

Solution a. 0.202
b. 0.144

Exercise 73. *Find the standard deviation for the following frequency tables using the formula. Check the calculations with the TI 83/84.*

a.

Grade	***Frequency***
49.5–59.5	*2*
59.5–69.5	*3*
69.5–79.5	*8*
79.5–89.5	*12*
89.5–99.5	*5*

Table 2.58

b.

Daily Low Temperature	***Frequency***
49.5 – 59.5	*53*
59.5 – 69.5	*32*
69.5 – 79.5	*15*
79.5 – 89.5	*1*
89.5 – 99.5	*0*

Table 2.59

c.

Points per Game	***Frequency***
49.5 – 59.5	*14*
59.5 – 69.5	*32*

69.5 – 79.5	*32*
79.5 – 89.5	*23*
89.5 – 99.5	*2*

Table 2.60

Solution

a. $s_x = \sqrt{\frac{\sum fm^2}{n} - \bar{x}^2} = \sqrt{\frac{193157.45}{30} - 79.5^2} = 10.88$

b. $s_x = \sqrt{\frac{\sum fm^2}{n} - \bar{x}^2} = \sqrt{\frac{380945.3}{101} - 60.94^2} = 7.62$

c. $s_x = \sqrt{\frac{\sum fm^2}{n} - \bar{x}^2} = \sqrt{\frac{440051.5}{86} - 70.66^2} = 11.14$

Exercise 74. *Student grades on a chemistry exam were: 77, 78, 76, 81, 86, 51, 79, 82, 84, 99*
a. Construct a stem-and-leaf plot of the data.
b. Are there any potential outliers? If so, which scores are they? Why do you consider them outliers?

Solution a.

Stem	Leaf
5	1
6	
7	6 7 8 9
8	1 2 4 6
9	9

b. 51, 99
51 is 25 points below the nearest exam value, and 99 is 13 points above the nearest exam value. These two scores are clearly separate from the rest of the data. They are outliers.

Exercise 75. *Table 2.61 contains the 2010 obesity rates in U.S. states and Washington, DC.*

State	***Percent (%)***	***State***	***Percent (%)***	***State***	***Percent (%)***
Alabama	*32.2*	*Kentucky*	*31.3*	*North Dakota*	*27.2*
Alaska	*24.5*	*Louisiana*	*31.0*	*Ohio*	*29.2*
Arizona	*24.3*	*Maine*	*26.8*	*Oklahoma*	*30.4*
Arkansas	*30.1*	*Maryland*	*27.1*	*Oregon*	*26.8*
California	*24.0*	*Massachusett*	*23.0*	*Pennsylvan*	*28.6*

		s		*ia*	
Colorado	*21.0*	*Michigan*	*30.9*	*Rhode Island*	*25.5*
Connecticut	*22.5*	*Minnesota*	*24.8*	*South Carolina*	*31.5*
Delaware	*28.0*	*Mississippi*	*34.0*	*South Dakota*	*27.3*
Washington, DC	*22.2*	*Missouri*	*30.5*	*Tennessee*	*30.8*
Florida	*26.6*	*Montana*	*23.0*	*Texas*	*31.0*
Georgia	*29.6*	*Nebraska*	*26.9*	*Utah*	*22.5*
Hawaii	*22.7*	*Nevada*	*22.4*	*Vermont*	*23.2*
Idaho	*26.5*	*New Hampshire*	*25.0*	*Virginia*	*26.0*
Illinois	*28.2*	*New Jersey*	*23.8*	*Washington*	*25.5*
Indiana	*29.6*	*New Mexico*	*25.1*	*West Virginia*	*32.5*
Iowa	*28.4*	*New York*	*23.9*	*Wisconsin*	*26.3*
Kansas	*29.4*	*North Carolina*	*27.8*	*Wyoming*	*25.1*

Table 2.61

a. Use a random number generator to randomly pick eight states. Construct a bar graph of the obesity rates of those eight states.
b. Construct a bar graph for all the states beginning with the letter "A."
c. Construct a bar graph for all the states beginning with the letter "M."

Solution

a. Example solution for using the random number generator for the TI-84+ to generate a simple random sample of 8 states. Instructions are as follows.
Number the entries in the table 1–51 (Includes Washington, DC; Numbered vertically)
Press MATH
Arrow over to PRB
Press 5:randInt(
Enter 51,1,8)
Eight numbers are generated (use the right arrow key to scroll through the numbers). The numbers correspond to the numbered states (for this example: {47 21 9 23 51 13 25 4}. If any numbers are repeated, generate a different number by using 5:randInt(51,1)). Here, the states (and Washington DC) are {Arkansas, Washington DC, Idaho, Maryland, Michigan, Mississippi, Virginia, Wyoming}.
Corresponding percents are {30.1, 22.2, 26.5, 27.1, 30.9, 34.0, 26.0, 25.1}.

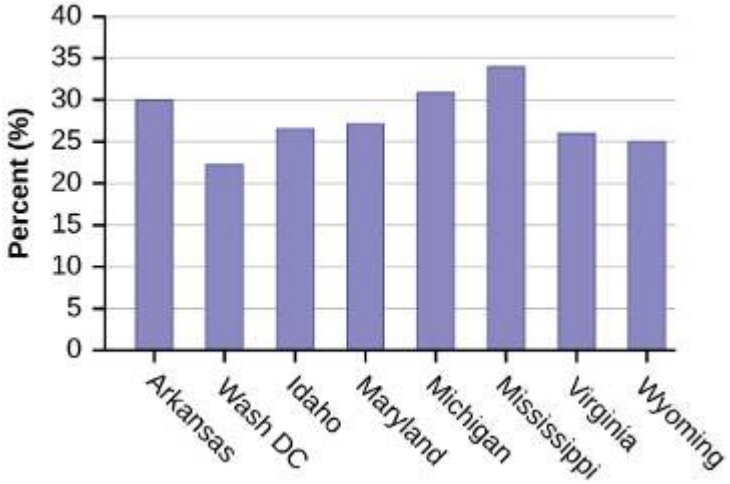
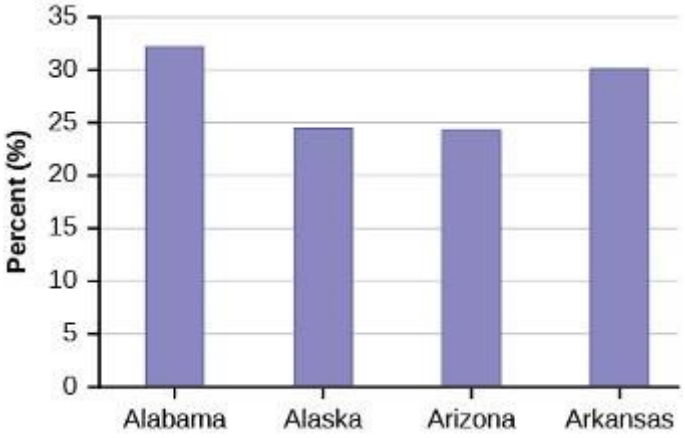
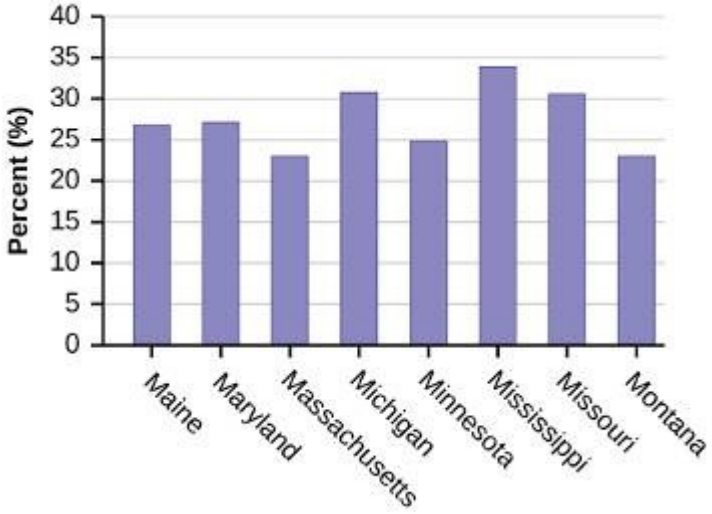


Figure 2.58



b.

Figure 2.59



c.

Figure 2.60

Exercise 76. *Suppose that three book publishers were interested in the number of fiction paperbacks adult consumers purchase per month. Each publisher conducted a survey. In the survey, adult consumers were asked the number of fiction paperbacks they had purchased the previous month. The results are as follows:*

# of books	***Freq.***	***Rel. Freq.***
0	*10*	
1	*12*	
2	*16*	
3	*12*	
4	*8*	
5	*6*	
6	*2*	
8	*2*	

Table 2.62 Publisher A

# of books	***Freq.***	***Rel. Freq.***
0	*18*	
1	*24*	
2	*24*	
3	*22*	
4	*15*	
5	*10*	

7	*5*	
9	*1*	

Table 2.63 Publisher B

# of books	***Freq.***	***Rel. Freq.***
0–1	*20*	
2–3	*35*	
4–5	*12*	
6–7	*2*	
8–9	*1*	

Table 2.64 Publisher C

a. Find the relative frequencies for each survey. Write them in the charts.
b. Using either a graphing calculator, computer, or by hand, use the frequency column to construct a histogram for each publisher's survey. For Publishers A and B, make bar widths of one. For Publisher C, make bar widths of two.
c. In complete sentences, give two reasons why the graphs for Publishers A and B are not identical.
d. Would you have expected the graph for Publisher C to look like the other two graphs? Why or why not?
e. Make new histograms for Publisher A and Publisher B. This time, make bar widths of two.
f. Now, compare the graph for Publisher C to the new graphs for Publishers A and B. Are the graphs more similar or more different? Explain your answer.

Solution

# of books	**Frequency**	**Relative Frequency**
0	10	0.15
1	12	0.18
2	16	0.24
3	12	0.18
4	8	0.12
5	6	0.09

6	2	0.03
8	2	0.03

Table 2.76 Publisher A

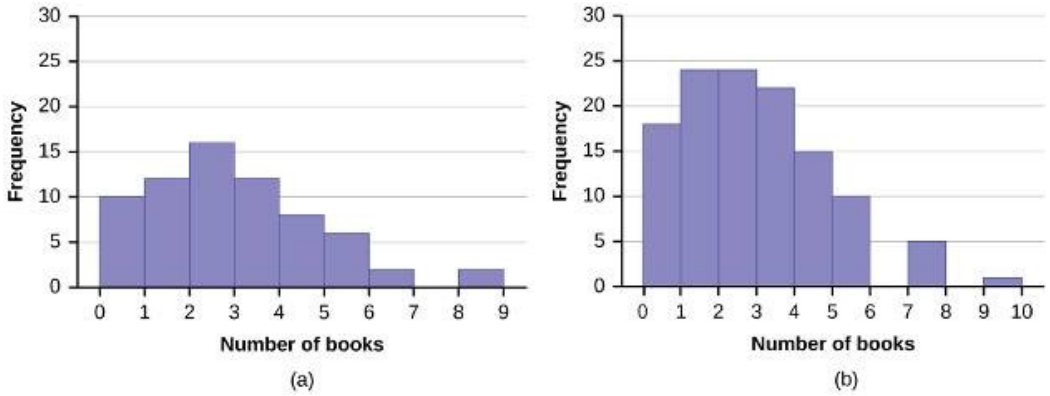
# of books	**Frequency**	**Relative Frequency**
0	18	0.15
1	24	0.2
2	24	0.2
3	22	0.18
4	15	0.13
5	10	0.08
7	5	0.04
9	1	0.01

Table 2.77 Publisher B

# of books	**Frequency**	**Relative Frequency**
0–1	20	0.29
2–3	35	0.5
4–5	12	0.17
6–7	2	0.03
8–9	1	0.01

Table 2.78 Publisher C

a. See Table 2.76, Table 2.77, and Table 2.78.
b. Check student's solution.
c. Answers will vary. Possible answers include: The different publishers sold different numbers of books. Different customers were included in each survey with different reading habits. The graphs have different scales. The publishers have different catalogs with different titles and different prices.
d. I would expect the same general trend, but I do expect the graph for Publisher C to have a different look. By changing the width of the class intervals, you change the number of bars and the height of each bar.
e. Check student's solution.



f. The graphs are more similar now that they have the same widths for class intervals. There is a clear pattern common to all three graphs. However, the graphs display different values for each class interval, and there are a different number of classes in the graph for Publisher A.

Exercise 77. *Often, cruise ships conduct all on-board transactions, with the exception of gambling, on a cashless basis. At the end of the cruise, guests pay one bill that covers all onboard transactions. Suppose that 60 single travelers and 70 couples were surveyed as to their on-board bills for a seven-day cruise from Los Angeles to the Mexican Riviera. Following is a summary of the bills for each group.*

Amount($)	***Frequency***	***Rel. Frequency***
51–100	*5*	
101–150	*10*	
151–200	*15*	
201–250	*15*	
251–300	*15*	
301–350	*5*	

Table 2.65 Singles

Amount($)	***Frequency***	***Rel. Frequency***
100–150	*5*	
201–250	*5*	
251–300	*5*	
301–350	*5*	
351–400	*10*	
401–450	*10*	
451–500	*10*	
501–550	*10*	
551–600	*5*	
601–650	*5*	

Table 2.66 Couples

a. Fill in the relative frequency for each group.

b. Construct a histogram for the singles group. Scale the x-axis by $50 widths. Use relative frequency on the y-axis.

c. Construct a histogram for the couples group. Scale the x-axis by $50 widths. Use relative frequency on the y-axis.

d. Compare the two graphs:

i. List two similarities between the graphs.

ii. List two differences between the graphs.

iii. Overall, are the graphs more similar or different?

e. Construct a new graph for the couples by hand. Since each couple is paying for two individuals, instead of scaling the x-axis by $50, scale it by $100. Use relative frequency on the y-axis.

f. Compare the graph for the singles with the new graph for the couples:

i. List two similarities between the graphs.

ii. Overall, are the graphs more similar or different?

g. How did scaling the couples graph differently change the way you compared it to the singles graph?

h. Based on the graphs, do you think that individuals spend the same amount, more or less, as singles as they do person by person as a couple? Explain why in one or two complete sentences.

Solution

Amount($)	Frequency	Relative Frequency
51–100	5	0.08
101–150	10	0.17
151–200	15	0.25
151–200	15	0.25
251–300	10	0.17
301–350	5	0.08

Table 2.86 Singles

Amount($)	Frequency	Relative Frequency
100–150	5	0.07
201–250	5	0.07
251–300	5	0.07
301–350	5	0.07
351–400	10	0.14
401–450	10	0.14
451–500	10	0.14
501–550	10	0.14
551–600	5	0.07
601–650	5	0.07

Table 2.87 Couples

a. See Table 2.86 and Table 2.87.

b. In the following histogram data values that fall on the right boundary are counted in the class interval, while values that fall on the left boundary are not counted (with the exception of the first interval where both boundary values are included).

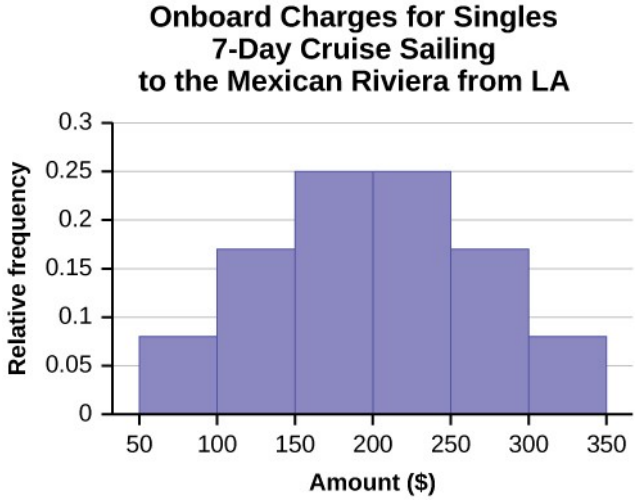


Figure 2.61

c. In the following histogram, the data values that fall on the right boundary are counted in the class interval, while values that fall on the left boundary are not counted (with the exception of the first interval where values on both boundaries are included).

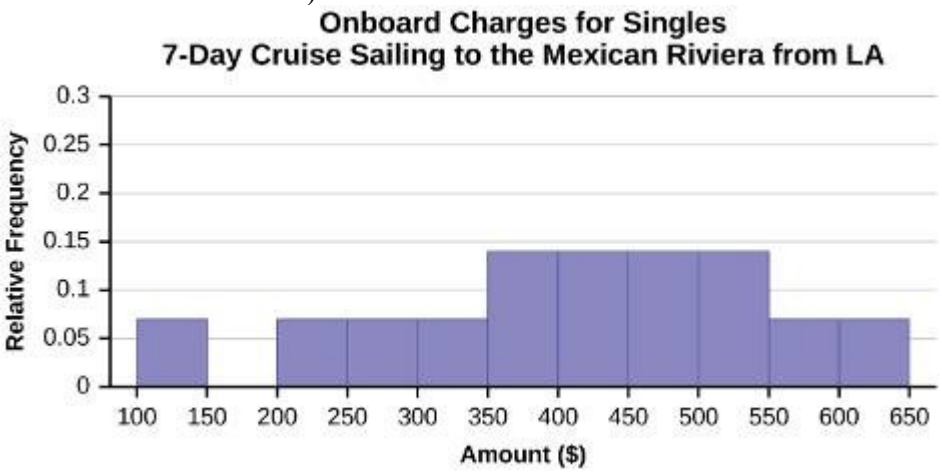


Figure 2.62

d. Compare the two graphs:
i. Answers may vary. Possible answers include:
▪ Both graphs have a single peak.
▪ Both graphs use class intervals with width equal to $50.
ii. Answers may vary. Possible answers include:
▪ The couples graph has a class interval with no values.
▪ It takes almost twice as many class intervals to display the data for couples.
iii. Answers may vary. Possible answers include: The graphs are more similar than different because the overall patterns for the graphs are the same.
e. Check student's solution.
f. Compare the graph for the Singles with the new graph for the Couples:
i .▪ Both graphs have a single peak.
▪ Both graphs display 6 class intervals.
▪ Both graphs show the same general pattern.
ii. Answers may vary. Possible answers include: Although the width of the class intervals for couples is double that of the class intervals for singles, the graphs are more similar than they are different.
g. Answers may vary. Possible answers include: You are able to compare the graphs interval by interval. It is easier to compare the overall patterns with the new scale on the Couples graph. Because a couple represents two individuals, the new scale leads to a more accurate comparison.
h. Answers may vary. Possible answers include: Based on the histograms, it seems that spending does not vary much from singles to individuals who are part of a couple. The overall patterns are the same. The range of spending for couples is approximately double the range for individuals.

Exercise 78. *Twenty-five randomly selected students were asked the number of movies they watched the previous week. The results are as follows.*

# of movies	***Frequency***	***Relative Frequency***	***Cumulative Relative Frequency***
0	*5*		
1	*9*		
2	*6*		
3	*4*		
4	*1*		

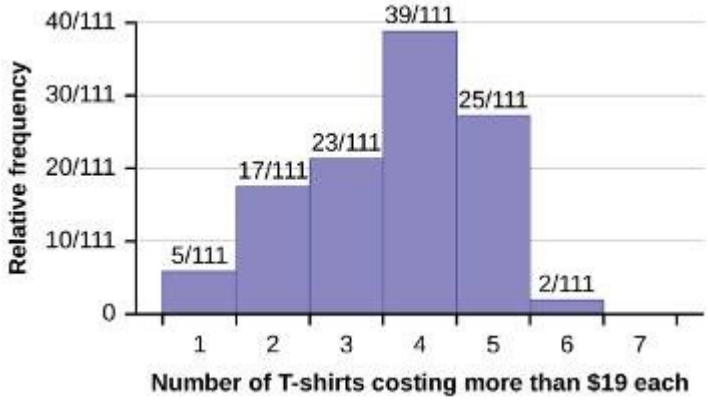
Table 2.67

a. Construct a histogram of the data.
b. Complete the columns of the chart.

Solution a. Answers will vary. One possible histogram is shown:
In this histogram, data values that fall on the left boundary of the class interval are counted in that interval. Data values that fall on the right boundary of an interval are not counted in the interval.
b.

# of Movies	**Frequency**	**Relative Frequency**	**Cumulative Relative Frequency**
0	4	0.2	0.2
1	9	0.36	0.56
2	6	0.24	0.8
3	4	0.16	0.96
4	4	0.04	1

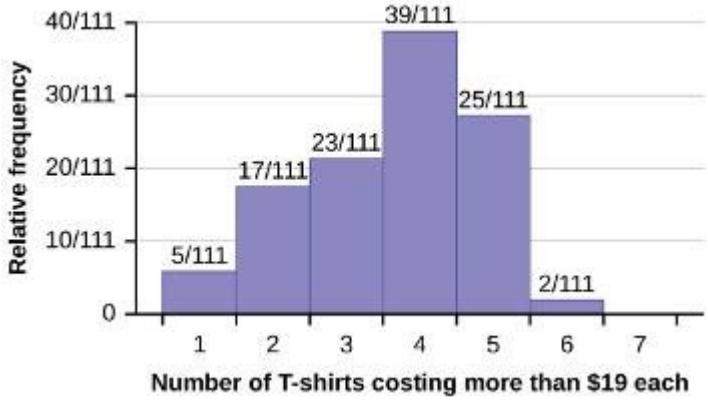
Exercise 79. *Suppose one hundred eleven people who shopped in a special t-shirt store were asked the number of t-shirts they own costing more than $19 each.*



The percentage of people who own at most three t-shirts costing more than $19 each is approximately:
a. 21
b. 59
c. 41
d. Cannot be determined

Solution C

Exercise 80. *Suppose one hundred eleven people who shopped in a special t-shirt store were asked the number of t-shirts they own costing more than $19 each.*



If the data were collected by asking the first 111 people who entered the store, then the type of sampling is:
a. cluster
b. simple random
c. stratified
d. convenience

Solution d

Exercise 81. *Following are the 2010 obesity rates by U.S. states and Washington, DC.*

State	***Percent (%)***	***State***	***Percent (%)***	***State***	***Percent (%)***
Alabama	*32.2*	*Kentucky*	*31.3*	*North Dakota*	*27.2*
Alaska	*24.5*	*Louisiana*	*31.0*	*Ohio*	*29.2*
Arizona	*24.3*	*Maine*	*26.8*	*Oklahoma*	*30.4*
Arkansas	*30.1*	*Maryland*	*27.1*	*Oregon*	*26.8*
California	*24.0*	*Massachusetts*	*23.0*	*Pennsylvania*	*28.6*
Colorado	*21.0*	*Michigan*	*30.9*	*Rhode Island*	*25.5*
Connecticut	*22.5*	*Minnesota*	*24.8*	*South Carolina*	*31.5*
Delaware	*28.0*	*Mississippi*	*34.0*	*South Dakota*	*27.3*
Washington, DC	*22.2*	*Missouri*	*30.5*	*Tennessee*	*30.8*
Florida	*26.6*	*Montana*	*23.0*	*Texas*	*31.0*
Georgia	*29.6*	*Nebraska*	*26.9*	*Utah*	*22.5*
Hawaii	*22.7*	*Nevada*	*22.4*	*Vermont*	*23.2*
Idaho	*26.5*	*New Hampshire*	*25.0*	*Virginia*	*26.0*
Illinois	*28.2*	*New Jersey*	*23.8*	*Washington*	*25.5*
Indiana	*29.6*	*New Mexico*	*25.1*	*West Virginia*	*32.5*
Iowa	*28.4*	*New York*	*23.9*	*Wisconsin*	*26.3*
Kansas	*29.4*	*North Carolina*	*27.8*	*Wyoming*	*25.1*

Table 2.68

Construct a bar graph of obesity rates of your state and the four states closest to your state. Hint: Label the x-axis with the states.

Solution Answers will vary.

Exercise 82. *The median age for U.S. blacks currently is 30.9 years; for U.S. whites it is 42.3 years.*

a. Based upon this information, give two reasons why the black median age could be lower than the white median age.

b. Does the lower median age for blacks necessarily mean that blacks die younger than whites? Why or why not?

c. How might it be possible for blacks and whites to die at approximately the

same age, but for the median age for whites to be higher?

Solution

a. There are several reasons that this difference could be observed. The lower black median age could be related to the greater number of births in the black minority group than in the whites. Another reason could be the aging of the non-Hispanic white group to a post child-bearing age and thus the general aging of this subgroup of the population. In general, differences in group medians are due to the distribution of ages and the observed range of the ages in the two different population groups.

b. No, the lower median age for blacks is most likely a result of the increase in minority births. Many other factors are involved in the timing of death between these two population groups (e.g., socioeconomic status, access to health care, etc). If the comparison is to be made only on the variable of age, the lower age is a balance of increased minority births and aging white population.
c. Even with the same range of the data on age for these two groups, the distribution of the data between the two groups may be very different.

Exercise 83.

Six hundred adult Americans were asked by telephone poll, "What do you think constitutes a middle-class income?" The results are in Table 2.69. Also, include left endpoint, but not the right endpoint.

Salary ($)	***Relative Frequency***
< 20,000	*0.02*
20,000–25,000	*0.09*
25,000–30,000	*0.19*
30,000–40,000	*0.26*
40,000–50,000	*0.18*
50,000–75,000	*0.17*

75,000–99,999	*0.02*
100,000+	*0.01*

Table 2.69

a. What percentage of the survey answered "not sure"?
b. What percentage think that middle-class is from $25,000 to $50,000?
c. Construct a histogram of the data.
i. Should all bars have the same width, based on the data? Why or why not?
ii. How should the <20,000 and the 100,000+ intervals be handled? Why?
d. Find the 40^{th} and 80^{th} percentiles
e. Construct a bar graph of the data

Solution
a. 1 – (0.02+0.09+0.19+0.26+0.18+0.17+0.02+0.01) = 0.06
b. 0.19+0.26+0.18 = 0.63
c. Check student's solution.
d. 40th percentile will fall between 30,000 and 40,000
80th percentile will fall between 50,000 and 75,000
e. Check student's solution.

Exercise 84.
Given the following box plot:

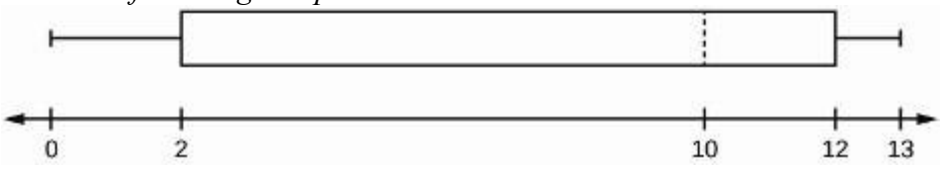


Figure 2.41

a. which quarter has the smallest spread of data? What is that spread?
b. which quarter has the largest spread of data? What is that spread?
c. find the interquartile range (IQR).
d. are there more data in the interval 5–10 or in the interval 10–13? How do you know this?
e. which interval has the fewest data in it? How do you know this?
i. 0–2
ii. 2–4
iii. 10–12
iv. 12–13
v. need more information

Solution
a. the last 25% (Q_3 to max), spread from 12–13
b. the second 25% (Q_1–Q_2), spread from 2–10
c. $IQR = Q_3 – Q_1 = 12 – 2 = 10$

d. There are more data in the interval 10–13. This interval holds 50% of the distribution of data. There is only 25% in the interval 2–10 therefore the interval 5–10 will hold less than 25%.
e. The interval from 2–4 has the fewest proportion of the data as it is less than 25%, it is a subset of the interval 2–10. All of the other intervals listed hold 25% of the data.

Exercise 85. *The following box plot shows the U.S. population for 1990, the latest available year.*

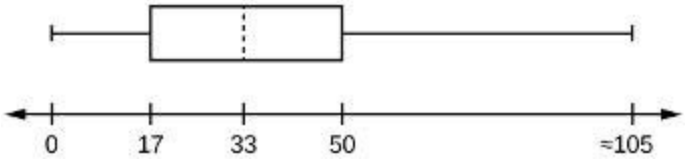


Figure 2.42
a. Are there fewer or more children (age 17 and under) than senior citizens (age 65 and over)? How do you know?
b. 12.6% are age 65 and over. Approximately what percentage of the population are working age adults (above age 17 to age 65)?

Solution a. more children; The left whisker shows that 25% of the population are children 17 and younger. The right whisker shows that 25% of the population are adults 50 and older, so adults 65 and over represent less than 25%.
b. 62.4%

Exercise 86. *In a survey of 20-year-olds in China, Germany, and the United States, people were asked the number of foreign countries they had visited in their lifetime. The following box plots display the results.*

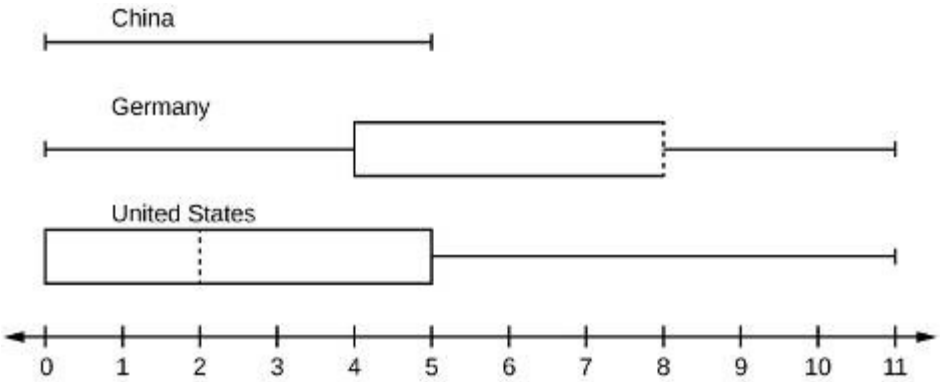


Figure 2.43
a. In complete sentences, describe what the shape of each box plot implies about the distribution of the data collected.
b. Have more Americans or more Germans surveyed been to over eight foreign countries?
c. Compare the three box plots. What do they imply about the foreign travel of

20-year-old residents of the three countries when compared to each other?

Solution

a. Seventy-five percent of Chinese participants have never visited a foreign country. The only variability occurs in the top 25%, so this data is skewed to the right. The lower 50% of data for Germans is more spread out than the top 50%, so the German box plot is skewed to the left. The data for U.S. participants is skewed to the right.
b. At least 50% of Germans surveyed have been to eight or more foreign countries. Fewer than 25% of U.S. participants have been to eight or more foreign countries. A greater proportion of Germans have traveled to eight or more foreign countries. This is reasonable because Germany is in close proximity to many more foreign countries than the United States.
c. According to the box plots, 20-year-olds in China rarely travel outside their home country. Germans travel a great deal. The greatest variability is seen in the U.S. where 25% of participants have never traveled outside the country, yet 25% have traveled to at least 5 foreign countries.

Exercise 87.

Given the following box plot, answer the questions.

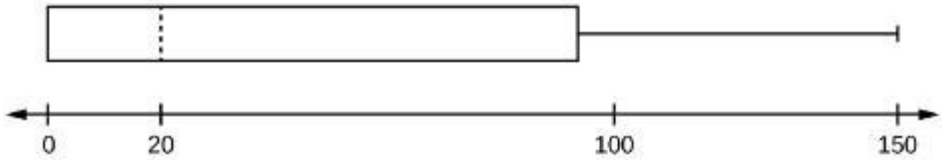


Figure 2.44

a. Think of an example (in words) where the data might fit into the above box plot. In 2–5 sentences, write down the example.
b. What does it mean to have the first and second quartiles so close together, while the second to third quartiles are far apart?

Solution

a. Answers will vary. Possible answer: State University conducted a survey to see how involved its students are in community service. The box plot shows the number of community service hours logged by participants over the past year.
b. Because the first and second quartiles are close, the data in this quarter is very similar. There is not much variation in the values. The data in the third quarter is much more variable, or spread out. This is clear because the second quartile is so far away from the third quartile.

Exercise 88.

Given the following box plots, answer the questions.

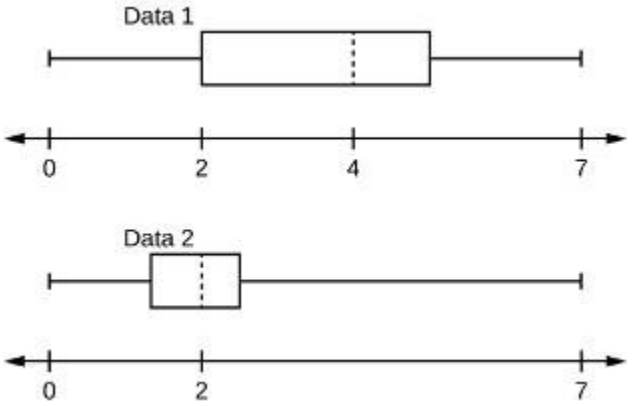


Figure 2.45

a. In complete sentences, explain why each statement is false.

*i. **Data 1** has more data values above two than **Data 2** has above two.*

ii. The data sets cannot have the same mode.

*iii. For **Data 1**, there are more data values below four than there are above four.*

b. For which group, Data 1 or Data 2, is the value of "7" more likely to be an outlier? Explain why in complete sentences.

Solution

a. i. There is no information given about the number of data values in the data sets. It is true that a larger proportion of values in Data Set 1 are greater than two. Data Set 2 may be a much larger set, however, and may actually have a greater number of data values greater than two.

ii. The box plots do not show individual data values. We have no information about the modes of the data sets. It is possible that the data sets do have the same mode.

iii. As in (i), there is a greater proportion of values below four in Data Set 2 than in Data Set 1. However, this does not allow us to compare the number of values in each interval.

b. The value "7" is more likely to be an outlier in Data Set 2, where more than 75% of the data falls below three.

Exercise 89.

A survey was conducted of 130 purchasers of new BMW 3 series cars, 130 purchasers of new BMW 5 series cars, and 130 purchasers of new BMW 7 series cars. In it, people were asked the age they were when they purchased their car. The following box plots display the results.

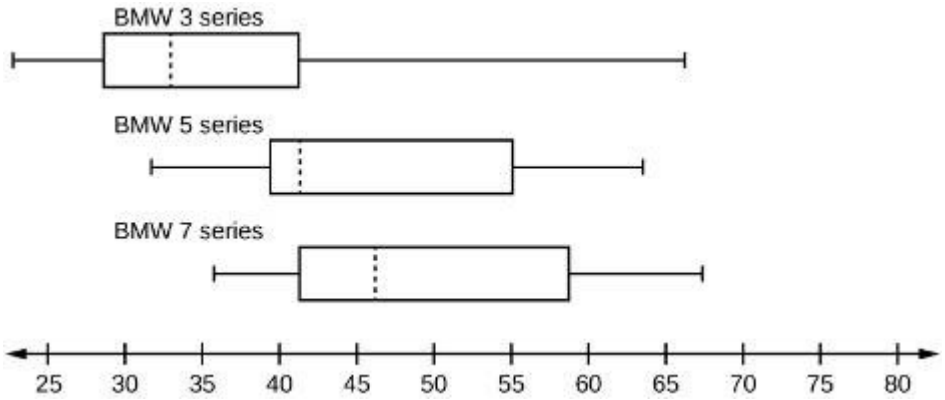


Figure 2.46

a. In complete sentences, describe what the shape of each box plot implies about the distribution of the data collected for that car series.
b. Which group is most likely to have an outlier? Explain how you determined that.
c. Compare the three box plots. What do they imply about the age of purchasing a BMW from the series when compared to each other?
d. Look at the BMW 5 series. Which quarter has the smallest spread of data? What is the spread?
e. Look at the BMW 5 series. Which quarter has the largest spread of data? What is the spread?
f. Look at the BMW 5 series. Estimate the interquartile range (IQR).
g. Look at the BMW 5 series. Are there more data in the interval 31 to 38 or in the interval 45 to 55? How do you know this?
h. Look at the BMW 5 series. Which interval has the fewest data in it? How do you know this?
i. 31–35
ii. 38–41
iii. 41–64

Solution

a. Each box plot is spread out more in the greater values. Each plot is skewed to the right, so the ages of the top 50% of buyers are more variable than the ages of the lower 50%.
b. The BMW 3 series is most likely to have an outlier. It has the longest whisker.
c. Comparing the median ages, younger people tend to buy the BMW 3 series, while older people tend to buy the BMW 7 series. However, this is not a rule, because there is so much variability in each data set.
d. The second quarter has the smallest spread. There seems to be only a three-year difference between the first quartile and the median.
e. The third quarter has the largest spread. There seems to be approximately a 14-year difference between the median and the third quartile.
f. $IQR \sim 17$ years
g. There is not enough information to tell. Each interval lies within a quarter, so

we cannot tell exactly where the data in that quarter is concentrated.
h. The interval from 31 to 35 years has the fewest data values. Twenty-five percent of the values fall in the interval 38 to 41, and 25% fall between 41 and 64. Since 25% of values fall between 31 and 38, we know that fewer than 25% fall between 31 and 35.

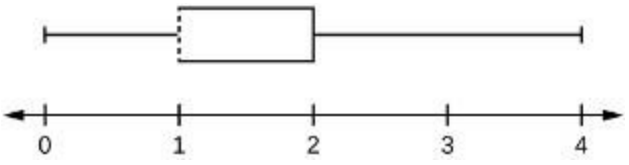
Exercise 90. *Twenty-five randomly selected students were asked the number of movies they watched the previous week. The results are as follows:*

# of movies	***Frequency***
0	*5*
1	*9*
2	*6*
3	*4*
4	*1*

Table 2.70

Construct a box plot of the data.

Solution



Exercise 91. *The most obese countries in the world have obesity rates that range from 11.4% to 74.6%. This data is summarized in the following table.*

Percent of Population Obese	***Number of Countries***

11.4–20.45	*29*
20.45–29.45	*13*
29.45–38.45	*4*
38.45–47.45	*0*
47.45–56.45	*2*
56.45–65.45	*1*
65.45–74.45	*0*
74.45–83.45	*1*

Table 2.71

a. What is the best estimate of the average obesity percentage for these countries?
b. The United States has an average obesity rate of 33.9%. Is this rate above average or below?
c. How does the United States compare to other countries?

Solution

$$\bar{x} = \frac{\sum fm}{\sum f} = \frac{1165.78}{50} = 23.32$$

a. The sample mean,
b. Since the average for the United Sates is 33.9%, compared to the average obesity rate of 23.32%, the United States has an obesity rate 10.58% higher than average.
c. The United States has an obesity rate higher than 84% of the countries listed.

Exercise 92. *Table 2.72 gives the percent of children under five considered to be underweight. What is the best estimate for the mean percentage of underweight children?*

Percent of Underweight Children	***Number of Countries***

16–21.45	*23*
21.45–26.9	*4*
26.9–32.35	*9*
32.35–37.8	*7*
37.8–43.25	*6*
43.25–48.7	*1*

Table 2.72

Solution

The mean percentage, $\overline{x} = \frac{1328.65}{50} = 26.75$

Exercise 93. *The median age of the U.S. population in 1980 was 30.0 years. In 1991, the median age was 33.1 years.*
a. What does it mean for the median age to rise?
b. Give two reasons why the median age could rise.
c. For the median age to rise, is the actual number of children less in 1991 than it was in 1980? Why or why not?

Solution

a. A rising median age in a population usually indicates an aging population.
b. 1. The median age could rise due to a decrease in births over time.
2. This trend could also be observed if there were an increased lifespan and thus a decreased number of deaths.
c. Not necessarily but that could be the case. In 1980, 50% of the population was 30 years old or less. In 1991, 50% of the population was 33.1 years old or less. On July 1, 1980, the population was 227.23 million. On July 1, 1991, the population was 252.98 million. Since half of the population in 1991 was larger than that in 1980 and the median was greater, there could have been fewer children. But, then again, there could have been the same or more children.

Exercise 94. *The population parameters below describe the full-time equivalent number of students (FTES) each year at Lake Tahoe Community College from 1976–1977 through 2004–2005.*

- *μ = 1000 FTES*
- *median = 1,014 FTES*

- *σ = 474 FTES*
- *first quartile = 528.5 FTES*
- *third quartile = 1,447.5 FTES*
- *n = 29 years*

A sample of 11 years is taken. About how many are expected to have a FTES of 1014 or above? Explain how you determined your answer.

Solution The median value is the middle value in the ordered list of data values. The median value of a set of 11 will be the 6th number in order. Six years will have totals at or below the median.

Exercise 95. *The population parameters below describe the full-time equivalent number of students (FTES) each year at Lake Tahoe Community College from 1976–1977 through 2004–2005.*

- *μ = 1000 FTES*
- *median = 1,014 FTES*
- *σ = 474 FTES*
- *first quartile = 528.5 FTES*
- *third quartile = 1,447.5 FTES*
- *n = 29 years*

75% of all years have an FTES:

a. at or below: _____
b. at or above: _____

Solution a. 1,447.5
b. 528.5

Exercise 96. *The population parameters below describe the full-time equivalent number of students (FTES) each year at Lake Tahoe Community College from 1976–1977 through 2004–2005.*

- *μ = 1000 FTES*
- *median = 1,014 FTES*
- *σ = 474 FTES*
- *first quartile = 528.5 FTES*
- *third quartile = 1,447.5 FTES*
- *n = 29 years*

The population standard deviation = _____

Solution 474 FTES

Exercise 97. *The population parameters below describe the full-time equivalent number of students (FTES) each year at Lake Tahoe Community College from 1976–1977*

through 2004–2005.

- *μ = 1000 FTES*
- *median = 1,014 FTES*
- *σ = 474 FTES*
- *first quartile = 528.5 FTES*
- *third quartile = 1,447.5 FTES*
- *n = 29 years*

What percent of the FTES were from 528.5 to 1447.5? How do you know?

Solution These values are the first and third quartiles, so 50% of the FTES are within this range.

Exercise 98. *The population parameters below describe the full-time equivalent number of students (FTES) each year at Lake Tahoe Community College from 1976–1977 through 2004–2005.*

- *μ = 1000 FTES*
- *median = 1,014 FTES*
- *σ = 474 FTES*
- *first quartile = 528.5 FTES*
- *third quartile = 1,447.5 FTES*
- *n = 29 years*

What is the IQR? What does the IQR represent?

Solution 919

Exercise 99. *The population parameters below describe the full-time equivalent number of students (FTES) each year at Lake Tahoe Community College from 1976–1977 through 2004–2005.*

- *μ = 1000 FTES*
- *median = 1,014 FTES*
- *σ = 474 FTES*
- *first quartile = 528.5 FTES*
- *third quartile = 1,447.5 FTES*
- *n = 29 years*

How many standard deviations away from the mean is the median?
Additional Information: The population FTES for 2005–2006 through 2010–2011 was given in an updated report. The data are reported here.

Year	*2005–06*	*2006–07*	*2007–08*	*2008–09*	*2009–10*	*2010–11*
Total	*1,585*	*1,690*	*1,735*	*1,935*	*2,021*	*1,890*

FTES						

Table 2.73

Solution 0.03

Exercise 100. *The population parameters below describe the full-time equivalent number of students (FTES) each year at Lake Tahoe Community College from 1976–1977 through 2004–2005.*

- *μ = 1000 FTES*
- *median = 1,014 FTES*
- *σ = 474 FTES*
- *first quartile = 528.5 FTES*
- *third quartile = 1,447.5 FTES*
- *n = 29 years*

Additional Information: The population FTES for 2005–2006 through 2010–2011 was given in an updated report. The data are reported here.

Year	*2005–06*	*2006–07*	*2007–08*	*2008–09*	*2009–10*	*2010–11*
Total FTES	*1,585*	*1,690*	*1,735*	*1,935*	*2,021*	*1,890*

Calculate the mean, median, standard deviation, the first quartile, the third quartile and the IQR. Round to one decimal place.

Solution
- mean = 1,809.3
- median = 1,812.5
- standard deviation = 151.2
- first quartile = 1,690
- third quartile = 1,935
- IQR = 245

Exercise 101. *The population parameters below describe the full-time equivalent number of students (FTES) each year at Lake Tahoe Community College from 1976–1977 through 2004–2005.*

- *μ = 1000 FTES*
- *median = 1,014 FTES*
- *σ = 474 FTES*
- *first quartile = 528.5 FTES*
- *third quartile = 1,447.5 FTES*
- *n = 29 years*

Additional Information: The population FTES for 2005–2006 through 2010–2011 was given in an updated report. The data are reported here.

Year	*2005–06*	*2006–07*	*2007–08*	*2008–09*	*2009–10*	*2010–11*
Total FTES	*1,585*	*1,690*	*1,735*	*1,935*	*2,021*	*1,890*

What additional information is needed to construct a box plot for the FTES for 2005–2006 through 2010–2011 and a box plot for the FTES for 1976–1977 through 2004–2005?

Solution The maximum and minimum values for the data.

Exercise 102. *The population parameters below describe the full-time equivalent number of students (FTES) each year at Lake Tahoe Community College from 1976–1977 through 2004–2005.*

- *μ = 1000 FTES*
- *median = 1,014 FTES*
- *σ = 474 FTES*
- *first quartile = 528.5 FTES*
- *third quartile = 1,447.5 FTES*
- *n = 29 years*

Additional Information: The population FTES for 2005–2006 through 2010–2011 was given in an updated report. The data are reported here.

Year	*2005–06*	*2006–07*	*2007–08*	*2008–09*	*2009–10*	*2010–11*
Total FTES	*1,585*	*1,690*	*1,735*	*1,935*	*2,021*	*1,890*

Compare the IQR for the FTES for 1976–77 through 2004–2005 with the IQR for the FTES for 2005-2006 through 2010–2011. Why do you suppose the IQRs are so different?

Solution Hint: Think about the number of years covered by each time period and what happened to higher education during those periods.
Check student's solution.

Exercise 103. *Three students were applying to the same graduate school. They came from schools with different grading systems. Which student had the best GPA when compared to other students at his school? Explain how you determined your answer.*

Student	***GPA***	***School Average GPA***	***School Standard Deviation***
Thuy	*2.7*	*3.2*	*0.8*
Vichet	*87*	*75*	*20*
Kamala	*8.6*	*8*	*0.4*

Table 2.74

Solution Kamala

Exercise 104. *A music school has budgeted to purchase three musical instruments. They plan to purchase a piano costing $3,000, a guitar costing $550, and a drum set costing $600. The mean cost for a piano is $4,000 with a standard deviation of $2,500. The mean cost for a guitar is $500 with a standard deviation of $200. The mean cost for drums is $700 with a standard deviation of $100. Which cost is the lowest, when compared to other instruments of the same type? Which cost is the highest when compared to other instruments of the same type. Justify your answer.*

Solution For pianos, the cost of the piano is 0.4 standard deviations BELOW the mean. For guitars, the cost of the guitar is 0.25 standard deviations ABOVE the mean. For drums, the cost of the drum set is 1.0 standard deviations BELOW the mean. Of the three, the drums cost the lowest in comparison to the cost of other instruments of the same type. The guitar costs the most in comparison to the cost of other instruments of the same type.

Exercise 105. *An elementary school class ran one mile with a mean of 11 minutes and a standard deviation of three minutes. Rachel, a student in the class, ran one mile in eight minutes. A junior high school class ran one mile with a mean of nine minutes and a standard deviation of two minutes. Kenji, a student in the class, ran 1 mile in 8.5 minutes. A high school class ran one mile with a mean of seven minutes and a standard deviation of four minutes. Nedda, a student in the class, ran one mile in eight minutes.*

a. Why is Kenji considered a better runner than Nedda, even though Nedda ran faster than he?

b. Who is the fastest runner with respect to his or her class? Explain why.

Solution a. Kenji is considered a better runner than Nedda because Kenji's time for one mile was 0.25 standard deviations faster (below, $z = -0.25$) than the mean of his class and Nedda's time as 0.25 standard deviations slower (above, $z = +0.25$)

than her class.
b. Rachel was the fastest runner with respect to her class as she had a time that was one standard deviation faster (below, z = –1.00) than her class.

Exercise 106. *The most obese countries in the world have obesity rates that range from 11.4% to 74.6%. This data is summarized in* ***Table 2.75****.*

Percent of Population Obese	***Number of Countries***
11.4–20.45	*29*
20.45–29.45	*13*
29.45–38.45	*4*
38.45–47.45	*0*
47.45–56.45	*2*
56.45–65.45	*1*
65.45–74.45	*0*
74.45–83.45	*1*

Table 2.75

What is the best estimate of the average obesity percentage for these countries? What is the standard deviation for the listed obesity rates? The United States has an average obesity rate of 33.9%. Is this rate above average or below? How "unusual" is the United States' obesity rate compared to the average rate? Explain.

Solution
- $\bar{x} = 23.32$
- Using the TI 83/84, we obtain a standard deviation of: $s_x = 12.95$.
- The obesity rate of the United States is 10.58% higher than the average obesity rate.
- Since the standard deviation is 12.95, we see that 23.32 + 12.95 = 36.27 is the obesity percentage that is one standard deviation from the mean. The United States obesity rate is slightly less than one standard deviation from the mean. Therefore, we can assume that the United States, while 34% obese, does not hav e an unusually high percentage of obese people.

Exercise 107. *Table 2.76 gives the percent of children under five considered to be underweight.*

Percent of Underweight Children	***Number of Countries***
16–21.45	*23*
21.45–26.9	*4*
26.9–32.35	*9*
32.35–37.8	*7*
37.8–43.25	*6*

43.25–48.7	*1*

Table 2.76

What is the best estimate for the mean percentage of underweight children? What is the standard deviation? Which interval(s) could be considered unusual? Explain.

Solution

- $\bar{x} = 26.57$
- Using the TI 83/84, we obtain a standard deviation of: $s_x = 8.54$
- Unusual values typically fall into ranges of: $\pm 2sx$

This means that unusual values are more than two standard deviations away from the mean. Therefore any values greater than 43.65 or less than 9.49 could be considered unusual. We see from the table that no class is less than 16%. We can then conclude that there are no unusually low percentages. On the other hand, we see there is one country in the highest class, 43.25–48.7. Since this class midpoint is greater than the high unusual value, we can assume this country has an unusually high percentage of underweight children.

Exercise 108.

Santa Clara County, CA, has approximately 27,873 Japanese-Americans. Their ages are as follows:

Age Group	***Percent of Community***
0–17	*18.9*
18–24	*8.0*
25–34	*22.8*
35–44	*15.0*
45–54	*13.1*
55–64	*11.9*
65+	*10.3*

Table 2.77

a. Construct a histogram of the Japanese-American community in Santa Clara County, CA. The bars will ***not*** *be the same width for this example. Why not? What impact does this have on the reliability of the graph?*
b. What percentage of the community is under age 35?
c. Which box plot most resembles the information above?

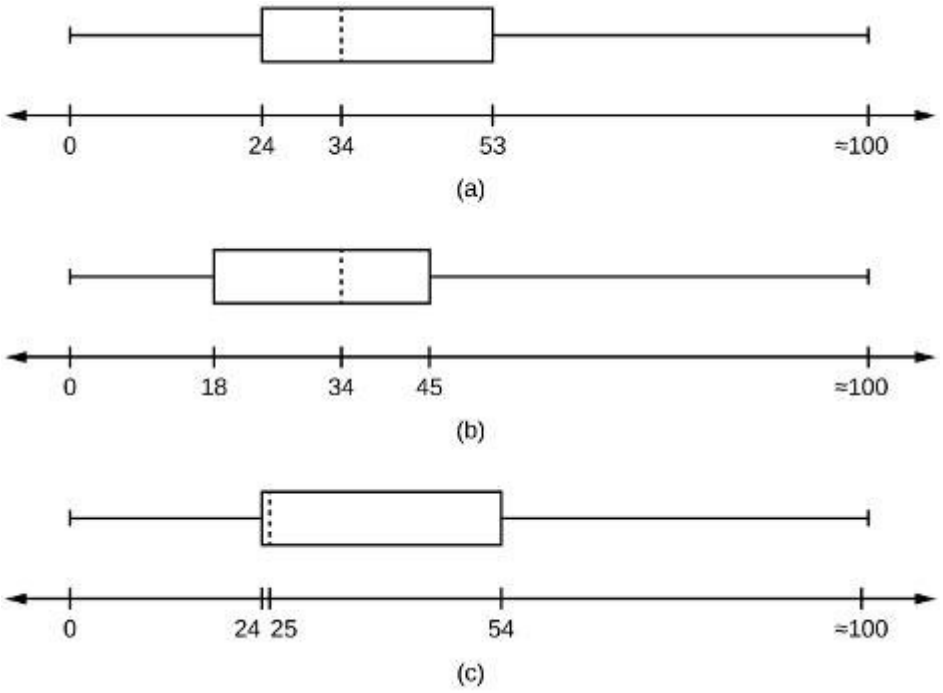


Figure 2.47

Solution

a. For histogram, check student's solution. In this histogram, values that fall on the left boundary are counted in the interval, while values that fall on the right boundary are not. Note that the width of the intervals varies. The data was recorded in classes of varying width. Pay close attention to the scale labels of the graph because the histogram can be misleading when the intervals have different widths.

b. 49.7% of the community is under the age of 35.
c. Based on the information in the table, graph (a) most closely represents the data.

Exercise 109.

Javier and Ercilia are supervisors at a shopping mall. Each was given the task of estimating the mean distance that shoppers live from the mall. They each randomly surveyed 100 shoppers. The samples yielded the following information.

	Javier	*Javier*

$\bar{x}$	*6.0 miles*	*6.0 miles*
s	*4.0 miles*	*7.0 miles*

Table 2.78

a. How can you determine which survey was correct ?
b. Explain what the difference in the results of the surveys implies about the data.
c. If the two histograms depict the distribution of values for each supervisor, which one depicts Ercilia's sample? How do you know?

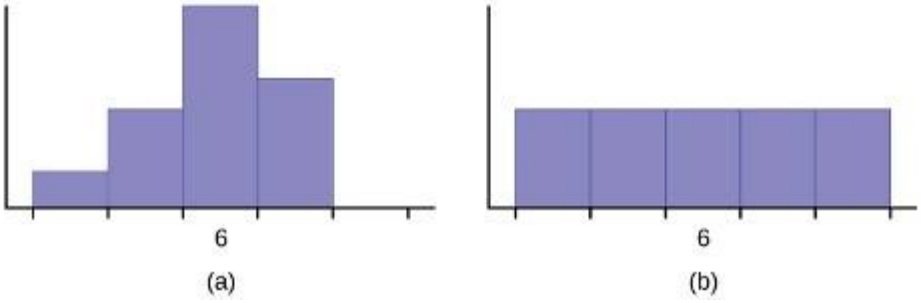


Figure 2.48

d. If the two box plots depict the distribution of values for each supervisor, which one depicts Ercilia's sample? How do you know?

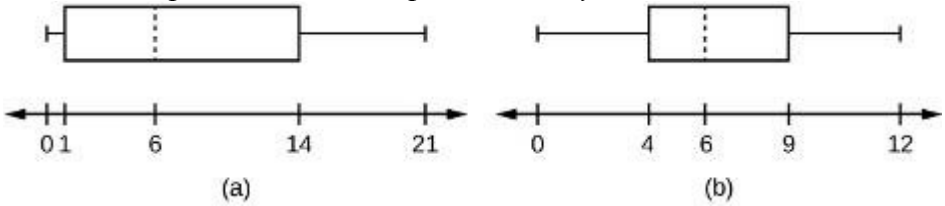


Figure 2.49

Solution

a. There is no way to determine from these numbers which survey was correct. Both Javier and Ercilia found approximately the same mean distance that shoppers live from the mall so this may be a good estimate. More samples of the same size or a much larger sample would help to obtain a better estimate of the mean.
b. The difference in the results shows that Javier's sample had a smaller standard deviation than Ercilia's sample. This shows that, although the mean distance appears to be around 6 miles, there is probably some variability in the distances that people live from the mall.
c. (ii) there appears to be a larger range of values
d. (i) the range of the entire box plot and variability shown in the box would be appropriate for the larger standard deviation in Ercilia's sample

Exercise

We are interested in the number of years students in a particular elementary

110. *statistics class have lived in California. The information in the following table is from the entire section.*

Number of years	***Frequency***	***Number of years***	***Frequency***
7	*1*	*22*	*1*
14	*3*	*23*	*1*
15	*1*	*26*	*1*
18	*1*	*40*	*2*
19	*4*	*42*	*2*
20	*3*		
			Total = 20

Table 2.79

What is the IQR?

a. 8

b. 11

c. 15

d. 35

Solution a

Exercise 111. *We are interested in the number of years students in a particular elementary statistics class have lived in California. The information in the following table is from the entire section.*

Number of years	***Frequency***	***Number of years***	***Frequency***
7	*1*	*22*	*1*
14	*3*	*23*	*1*
15	*1*	*26*	*1*
18	*1*	*40*	*2*
19	*4*	*42*	*2*

20	*3*		
			Total = 20

Table 2.79

What is the mode?
a. 19
b. 19.5
c. 14 and 20
d. 22.65

Solution a

Exercise 112. *We are interested in the number of years students in a particular elementary statistics class have lived in California. The information in the following table is from the entire section.*

Number of years	***Frequency***	***Number of years***	***Frequency***
7	*1*	*22*	*1*
14	*3*	*23*	*1*
15	*1*	*26*	*1*
18	*1*	*40*	*2*
19	*4*	*42*	*2*
20	*3*		
			Total = 20

Table 2.79

Is this a sample or the entire population?
a. sample
b. entire population
c. neither

Solution b

Exercise 113. *Twenty-five randomly selected students were asked the number of movies they watched the previous week. The results are as follows:*

# of movies	***Frequency***
0	*5*
1	*9*
2	*6*
3	*4*
4	*1*

Table 2.80

a. Find the sample mean $\bar{x}$.

b. Find the approximate sample standard deviation, s.

Solution

a. 1.48

b. 1.12

Exercise 114.

Forty randomly selected students were asked the number of pairs of sneakers they owned. Let X = the number of pairs of sneakers owned. The results are as follows:

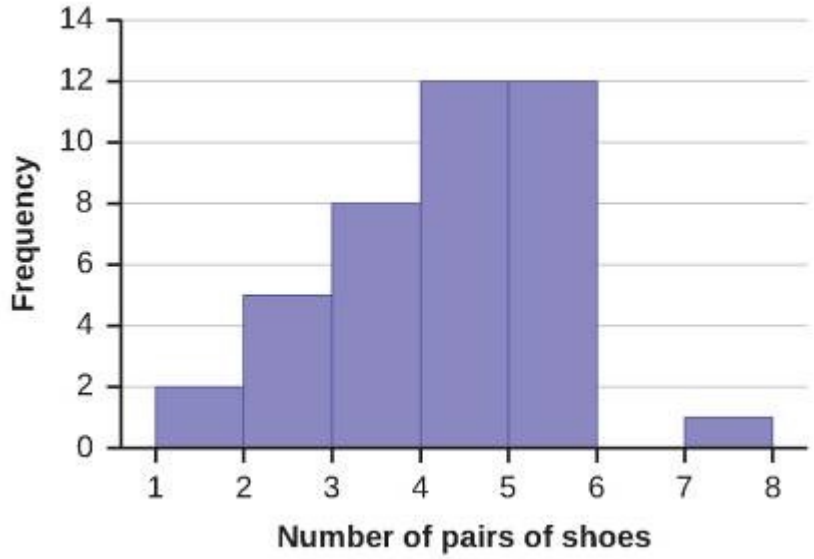
X	***Frequency***
1	*2*
2	*5*
3	*8*

4	*12*
5	*12*
6	*0*
7	*1*

Table 2.81

a. Find the sample mean $\bar{x}$
b. Find the sample standard deviation, s
c. Construct a histogram of the data.
d. Complete the columns of the chart.
e. Find the first quartile.
f. Find the median.
g. Find the third quartile.
h. Construct a box plot of the data.
i. What percent of the students owned at least five pairs?
j. Find the 40^{th} *percentile.*
k. Find the 90^{th} *percentile.*
l. Construct a line graph of the data
m. Construct a stemplot of the data

Solution a. 3.78
b. 1.29



c.

d.

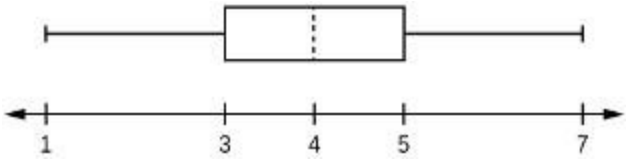
X	**Frequency**	**Relative Frequency**	**Cumulative Relative Frequency**
1	7	0.05	0.050
2	5	0.125	0.175
3	8	0.2	0.375
4	12	0.3	0.675
5	12	0.3	0.975
6	0	0	0.975
7	1	0.025	1.000

e. 3

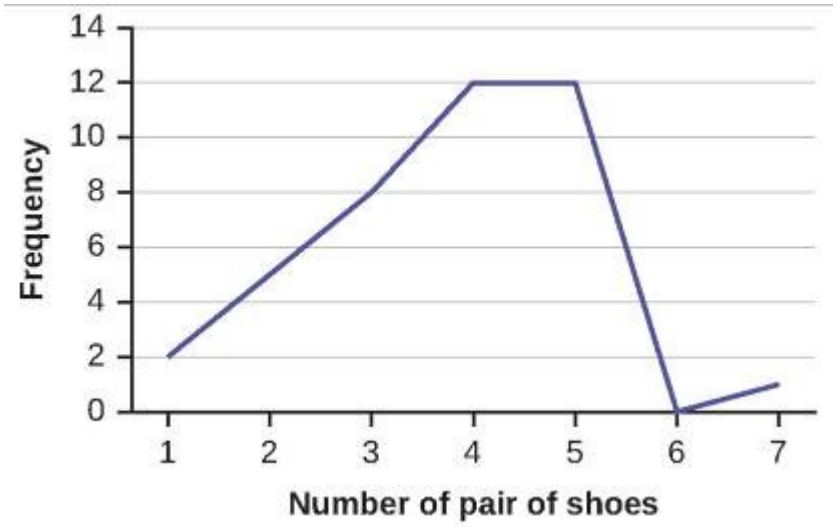
f. 4

g. 5

h.



i. 32.5%
j. 4
k. 5



l.
m.

Leaf	Stem
1	0 0
2	0 0 0 0 0
3	0 0 0 0 0 0 0 0
4	0 0 0 0 0 0 0 0 0 0 0 0
5	0 0 0 0 0 0 0 0 0 0 0 0
6	
7	0

Exercise 115. *Following are the published weights (in pounds) of all of the team members of the San Francisco 49ers from a previous year.*
177; 205; 210; 210; 232; 205; 185; 185; 178; 210; 206; 212; 184; 174; 185; 242; 188; 212; 215; 247; 241; 223; 220; 260; 245; 259; 278; 270; 280; 295; 275; 285; 290; 272; 273; 280; 285; 286; 200; 215; 185; 230; 250; 241; 190; 260; 250; 302; 265; 290; 276; 228; 265

a. Organize the data from smallest to largest value.
b. Find the median.
c. Find the first quartile.
d. Find the third quartile.
e. Construct a box plot of the data.
f. The middle 50% of the weights are from _____ to _____.
g. If our population were all professional football players, would the above data be a sample of weights or the population of weights? Why?

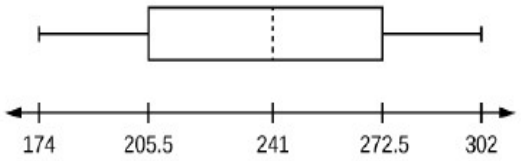
h. If our population included every team member who ever played for the San Francisco 49ers, would the above data be a sample of weights or the population of weights? Why?

i. Assume the population was the San Francisco 49ers. Find:
i. the population mean, μ.
ii. the population standard deviation, σ.
iii. the weight that is two standard deviations below the mean.
iv. When Steve Young, quarterback, played football, he weighed 205 pounds. How many standard deviations above or below the mean was he?

j. That same year, the mean weight for the Dallas Cowboys was 240.08 pounds with a standard deviation of 44.38 pounds. Emmit Smith weighed in at 209 pounds. With respect to his team, who was lighter, Smith or Young? How did you determine your answer?

Solution

a. 174; 177; 178; 184; 185; 185; 185; 185; 188; 190; 200; 205; 205; 206; 210; 210; 210; 212; 212; 215; 215; 220; 223; 228; 230; 232; 241; 241; 242; 245; 247; 250; 250; 259; 260; 260; 265; 265; 270; 272; 273; 275; 276; 278; 280; 280; 285; 285; 286; 290; 290; 295; 302
b. 241
c. 205.5
d. 272.5



e.
f. 205.5, 272.5
g. sample
h. population
i. i. 236.34
ii. 37.50
iii. 161.34
iv. 0.84 std. dev. below the mean
j. Young

Exercise 116.

One hundred teachers attended a seminar on mathematical problem solving. The attitudes of a representative sample of 12 of the teachers were measured before and after the seminar. A positive number for change in attitude indicates that a teacher's attitude toward math became more positive. The 12 change scores are as follows
3; 8; –1; 2; 0; 5; –3; 1; –1; 6; 5; –2
a. What is the mean change score?
b. What is the standard deviation for this population?

c. What is the median change score?
d. Find the change score that is 2.2 standard deviations below the mean.

Solution
a. +1.92
b. +3.50
c. +1.50
d. –5.78

Exercise 117. *Refer to Figure 2.50 determine which of the following are true and which are false. Explain your solution to each part in complete sentences.*

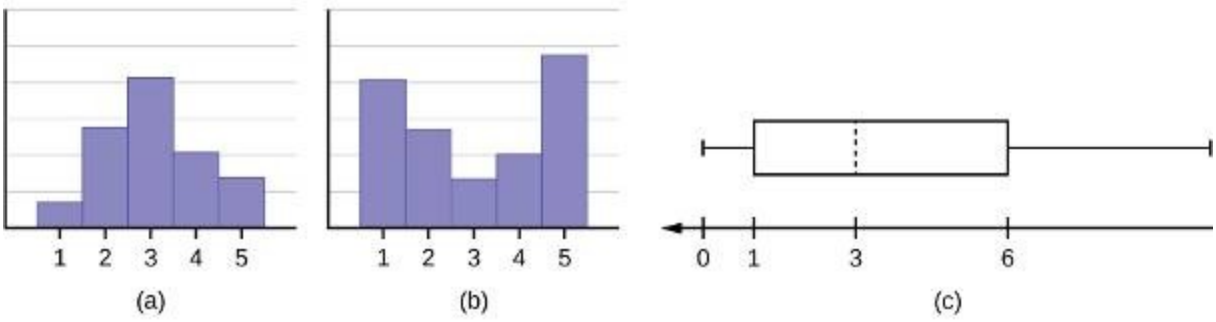


Figure 2.50
a. The medians for all three graphs are the same.
b. We cannot determine if any of the means for the three graphs is different.
c. The standard deviation for graph b is larger than the standard deviation for graph a.
d. We cannot determine if any of the third quartiles for the three graphs is different.

Solution
a. True
b. True
c. True
d. False

Exercise 118. *In a recent issue of the IEEE Spectrum, 84 engineering conferences were announced. Four conferences lasted two days. Thirty-six lasted three days. Eighteen lasted four days. Nineteen lasted five days. Four lasted six days. One lasted seven days. One lasted eight days. One lasted nine days. Let X = the length (in days) of an engineering conference.*

a. Organize the data in a chart.
b. Find the median, the first quartile, and the third quartile.
c. Find the 65^{th} percentile.
d. Find the 10^{th} percentile.

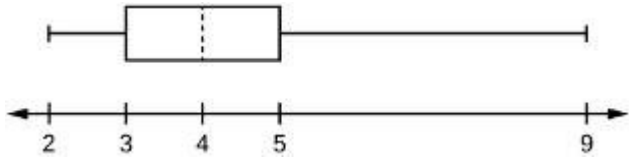
e. Construct a box plot of the data.
f. The middle 50% of the conferences last from _____ days to _____ days.
g. Calculate the sample mean of days of engineering conferences.
h. Calculate the sample standard deviation of days of engineering conferences.
i. Find the mode.
j. If you were planning an engineering conference, which would you choose as the length of the conference: mean; median; or mode? Explain why you made that choice.
k. Give two reasons why you think that three to five days seem to be popular lengths of engineering conferences.

Solution

a.

Number of Conferences	Days lasted
4	2
36	3
18	4
19	5
4	6
1	7
1	8
1	9

b. 4,3,5
c. 4
d. 3
e.



f. 3,5
g. 3.94
h. 1.28
i. 3
j. mode
k. Answers will vary.

Exercise 119.

A survey of enrollment at 35 community colleges across the United States yielded the following figures: 6414; 1550; 2109; 9350; 21828; 4300; 5944; 5722; 2825; 2044; 5481; 5200; 5853; 2750; 10012; 6357; 27000; 9414; 7681; 3200; 17500; 9200; 7380; 18314; 6557; 13713; 17768; 7493; 2771; 2861; 1263; 7285; 28165; 5080; 11622

a. Organize the data into a chart with five intervals of equal width. Label the

two columns "Enrollment" and "Frequency."
b. Construct a histogram of the data.
c. If you were to build a new community college, which piece of information would be more valuable: the mode or the mean?
d. Calculate the sample mean.
e. Calculate the sample standard deviation.
f. A school with an enrollment of 8000 would be how many standard deviations away from the mean?

Solution

a.

Enrollment	Frequency
1000-5000	10
5000-10000	16
10000-15000	3
15000-20000	3
20000-25000	1
25000-30000	2

Table 2.88

b. Check student's solution.
c. mode
d. 8628.74
e. 6943.88
f. –0.09

Exercise 120.

X = the number of days per week that 100 clients use a particular exercise facility.

x	***Frequency***
0	*3*
1	*12*
2	*33*
3	*28*
4	*11*
5	*9*
6	*4*

Table 2.82

The 80th percentile is _____
a. 5
b. 80
c. 3

d. 4

Solution d

Exercise 121. *X = the number of days per week that 100 clients use a particular exercise facility.*

x	***Frequency***
0	*3*
1	*12*
2	*33*
3	*28*
4	*11*
5	*9*
6	*4*

Table 2.82

The number that is 1.5 standard deviations BELOW the mean is approximately

a. 0.7
b. 4.8
c. –2.8
d. Cannot be determined

Solution a

Exercise 122. *Suppose that a publisher conducted a survey asking adult consumers the number of fiction paperback books they had purchased in the previous month. The results are summarized in the Table 2.83.*

# of books	***Freq.***	***Rel. Freq.***
0	*18*	
1	*24*	
2	*24*	
3	*22*	
4	*15*	
5	*10*	
7	*5*	

9	*1*	

Table 2.83

a. Are there any outliers in the data? Use an appropriate numerical test involving the IQR to identify outliers, if any, and clearly state your conclusion.
b. If a data value is identified as an outlier, what should be done about it?
c. Are any data values further than two standard deviations away from the mean? In some situations, statisticians may use this criteria to identify data values that are unusual, compared to the other data values. (Note that this criteria is most appropriate to use for data that is mound-shaped and symmetric, rather than for skewed data.)
d. Do parts a and c of this problem give the same answer?
e. Examine the shape of the data. Which part, a or c, of this question gives a more appropriate result for this data?
f. Based on the shape of the data which is the most appropriate measure of center for this data: mean, median or mode?

Solution

a. $IQR = 4 - 1 = 3$; $Q^1 - 1.5(IQR) = 1 - 1.5(3) = -3.5$; $Q_3 + 1.5(IQR) = 4 + 1.5(3) = 8.5$. The data value nine is greater than 8.5. The purchase of nine books in one month is an outlier.
b. The outlier should be investigated to see if there is an error or some other problem in the data; then a decision whether to include or exclude it should be made based on the particular situation. If it was a correct value then the data value should remain in the data set. If there is a problem with this data value, then it should be corrected or removed from the data. For example: If the data were recorded incorrectly (perhaps a nine was miscoded and the correct value was six), then the data should be corrected. If it was an error but the correct value is not known it must be removed from the data set.
c. $\bar{x} - 2s = 2.45 - 2(1.88) = -1.31$; $\bar{x} + 2s = 2.45 + 2(1.88) = 6.21$. Using this method, the five data values of seven books purchased and the one data value of nine books purchased would be considered unusual.
d. No, part a identifies only the value of nine to be an outlier but part c identifies both seven and nine.
e. The data is skewed to the right. It would be more appropriate to use the method involving the *IQR* in part a, identifying only the one value of nine books purchased as an outlier. Note that part c remarks that identifying unusual data values by using the criteria of being further than two standard deviations away from the mean is most appropriate when the data are mound-shaped and symmetric.
f. The data are skewed to the right. For skewed data it is more appropriate to use the median as a measure of center.