2.S: Descriptive Statistics (Solutions)

40.

a.

Table 1.19

# flossing per week	Frequency	Relative frequency	Cumulative relative frequency
0	27	0.4500	0.4500
1	18	0.3000	0.7500
3	11	0.1833	0.9333
6	3	0.0500	0.9833
7	1	0.0167	1

b. 5.00%
c. 93.33%

42.

The sum of the travel times is 1,173.1. Divide the sum by 50 to calculate the mean value: 23.462. Because each state's travel time was measured to the nearest tenth, round this calculation to the nearest hundredth: 23.46.

44.

b

1.

Figure 2.*S*. 26

3.

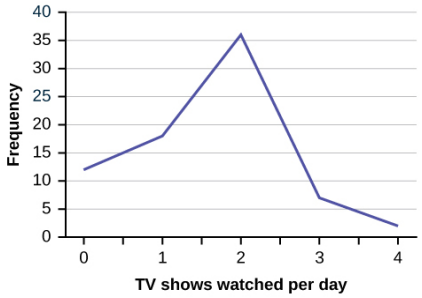

Figure 2.*S*. 27

5.

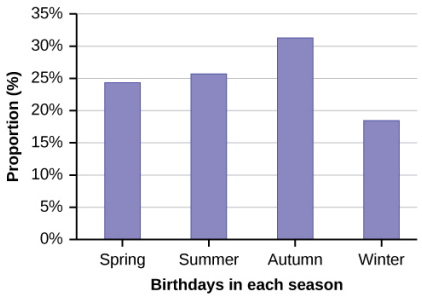

Figure 2.S.28

7.

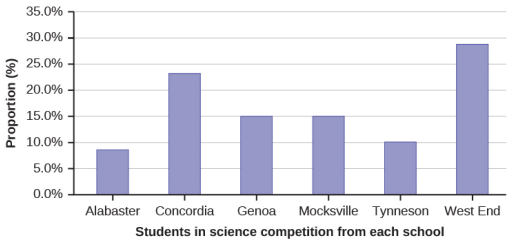

Figure 2.S.29

9.

65

11.

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.

13.

Answers will vary. One possible histogram is shown:

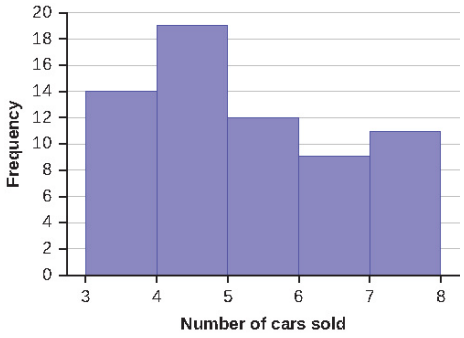

Figure 2.S.30

15.

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.

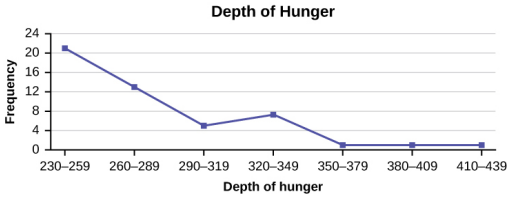

Figure 2.S.31

17.

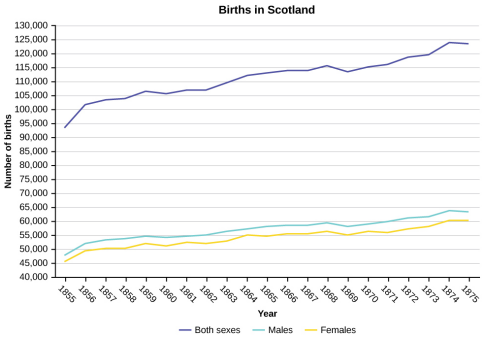

Figure 2.S. 32

19.

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

21.

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.

23.

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).

25.

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.

27.

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.

29.

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.

31.

4

33.

$6 - 4 = 2$

35.

6

37.

Mean: $16 + 17 + 19 + 20 + 20 + 21 + 23 + 24 + 25 + 25 + 25 + 26 + 26 + 27 + 27 + 27 + 28 + 29 + 30 + 32 + 33 + 33 + 34 + 35 + 37 + 39 + 40 = 738$;

$\frac{738}{27} = 27.33$

39.

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

41.

4

44.

39.48 in.

45.

$21,574

46.

15.98 ounces

47.

81.56

48.

4 hours

49.

2.01 inches

50.

18.25

51.

10

52.

14.15

53.

14

54.

14.78

55.

44%

56.

100%

57.

6%

58.

33%

59.

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.

61.

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.

63.

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

65.

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

67.

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

69.

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

71.

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

73.

The mean and the median are both six.

75.

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

77.

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

79.

$s = 34.5$

81.

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.

83.

a. $s_x = \sqrt{\frac{\sum fm^2}{n} - \bar{x}^2} = \sqrt{\frac{193157.45}{30} - 79.5^2} = 10.88$
b. $s_x = \sqrt{\frac{\sum fm^2}{n} - \bar{x}^2} = \sqrt{\frac{38045.3}{101} - 60.94^2} = 7.62$
c. $s_x = \sqrt{\frac{\sum fm^2}{n} - \bar{x}^2} = \sqrt{\frac{440051.5}{86} - 70.66^2} = 11.14$

84.

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 $\{30.1, 22.2, 26.5, 27.1, 30.9, 34.0, 26.0, 25.1\}$

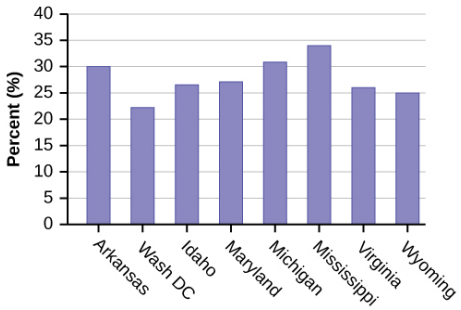

Figure 2.S. 33

b.

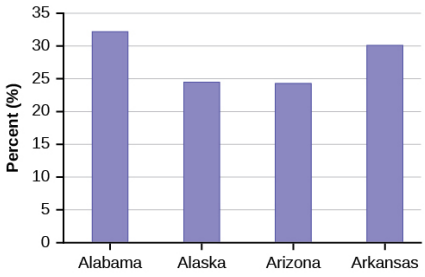

Figure 2.S. 34

c.

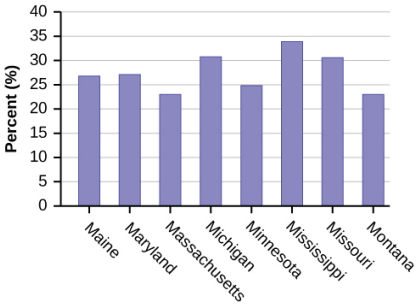

Figure 2.S. 35

86.

Amount($)	Frequency	Relative frequency
51–100	5	0.08
101–150	10	0.17
151–200	15	0.25
201–250	15	0.25
251–300	10	0.17
301–350	5	0.08

Table2.87 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

Amount($)	Frequency	Relative frequency
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

Table2.88 Couples

a. See Table $2.S.87$ and Table $2.S.88$.
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).

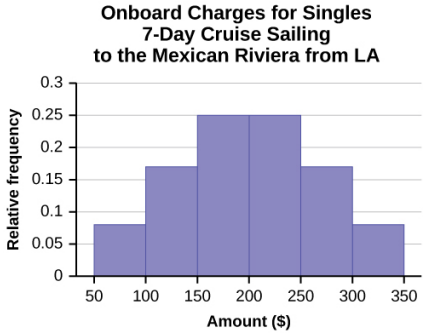

Figure $2.S.36$

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).

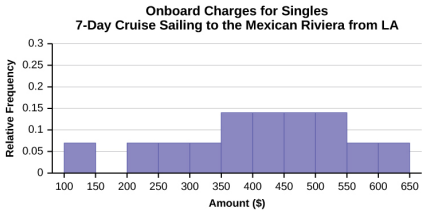

Figure $2.S.37$

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.

88.

c

90.

Answers will vary.

92.

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. 40^{th} percentile will fall between 30,000 and 40,000

   80^{th} percentile will fall between 50,000 and 75,000

e. Check student's solution.

94.

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

95.

a. Yes
b. The sample is 0.5 higher.

96.

a. 20
b. No

97.

51

98.

a. 42
b. 99

99.

$10.19

100.

17%

101.

$30,772.48

102.

4.4%

103.

7.24%

104.

-1.27%

106.

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.

108.

474 FTES

110.

919

112.

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

113.

Hint: Think about the number of years covered by each time period and what happened to higher education during those periods.

115.

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.

117.

- $\overline{x} = 23.32$
- 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.

120.

a

122.

b

123.

a. 1.48
b. 1.12

125.

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

127.

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

129.

Table 2.*S*. 89

a.

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

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

131.

a

This page titled 2.S: Descriptive Statistics (Solutions) is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by OpenStax via source content that was edited to the style and standards of the LibreTexts platform.

- **2.15: Solutions** by OpenStax is licensed CC BY 4.0. Original source: https://openstax.org/details/books/introductory-business-statistics.