

3.E: Descriptive Statistics (Optional Exercises)

2.4: Measures of the Location of the Data

Q 2.4.1

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

- Based upon this information, give two reasons why the black median age could be lower than the white median age.
- Does the lower median age for blacks necessarily mean that blacks die younger than whites? Why or why not?
- 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?

Q 2.4.2

Six hundred adult Americans were asked by telephone poll, "What do you think constitutes a middle-class income?" The results are in Table. 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

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

S 2.4.2

- $1 - (0.02 + 0.09 + 0.19 + 0.26 + 0.18 + 0.17 + 0.02 + 0.01) = 0.06$
- $0.19 + 0.26 + 0.18 = 0.63$
- Check student's solution.
- 40th percentile will fall between 30,000 and 40,000
80th percentile will fall between 50,000 and 75,000
- Check student's solution.

Q 2.4.3

Given the following box plot:


 This is a horizontal boxplot graphed over a number line from 0 to 13. The first whisker extends from the smallest value, 0, to the first quartile, 2. The box begins at the first quartile and extends to third quartile, 12. A vertical, dashed line is drawn at median, 10. The second whisker extends from the third quartile to largest value, 13.

Figure 2.4.1.

- which quarter has the smallest spread of data? What is that spread?
- which quarter has the largest spread of data? What is that spread?
- find the interquartile range (*IQR*).
- are there more data in the interval 5–10 or in the interval 10–13? How do you know this?
- which interval has the fewest data in it? How do you know this?
 - 0–2
 - 2–4
 - 10–12
 - 12–13
 - need more information

Q 2.4.4

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


 A box plot with values from 0 to 105, with Q1 at 17, M at 33, and Q3 at 50.

Figure 2.4.2.

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

S 2.4.4

- 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%.
- 62.4%

2.5: Box Plots

Q 2.5.1

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.


 This shows three boxplots graphed over a number line from 0 to 11. The boxplots match the supplied data, and compare the countries' results. The China boxplot has a single whisker from 0 to 5. The Germany box plot's median is equal to the third quartile, so there is a dashed line at right edge of box. The America boxplot does not have a left whisker.

Figure 2.5.1.

- In complete sentences, describe what the shape of each box plot implies about the distribution of the data collected.
- Have more Americans or more Germans surveyed been to over eight foreign countries?
- 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?

Q 2.5.2

Given the following box plot, answer the questions.


 This is a boxplot graphed over a number line from 0 to 150. There is no first, or left, whisker. The box starts at the first quartile, 0, and ends at the third quartile, 80. A vertical, dashed line marks the median, 20. The second whisker extends the third quartile to the largest value, 150.

Figure 2.5.2.

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

S 2.5.2

- 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.
- 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.

Q 2.5.3

Given the following box plots, answer the questions.

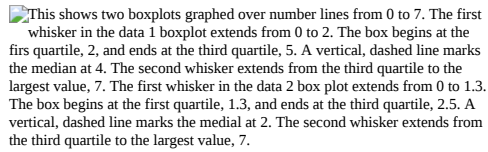
This shows two boxplots graphed over number lines from 0 to 7. The first whisker in the data 1 boxplot extends from 0 to 2. The box begins at the first quartile, 2, and ends at the third quartile, 5. A vertical, dashed line marks the median at 4. The second whisker extends from the third quartile to the largest value, 7. The first whisker in the data 2 box plot extends from 0 to 1.3. The box begins at the first quartile, 1.3, and ends at the third quartile, 2.5. A vertical, dashed line marks the median at 2. The second whisker extends from the third quartile to the largest value, 7.

Figure 2.5.3.

- In complete sentences, explain why each statement is false.
 - Data 1** has more data values above two than **Data 2** has above two.
 - The data sets cannot have the same mode.
 - For **Data 1**, there are more data values below four than there are above four.
- For which group, Data 1 or Data 2, is the value of “7” more likely to be an outlier? Explain why in complete sentences.

Q 2.5.4

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.

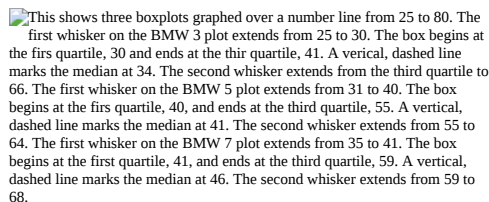
This shows three boxplots graphed over a number line from 25 to 80. The first whisker on the BMW 3 plot extends from 25 to 30. The box begins at the first quartile, 30 and ends at the third quartile, 41. A vertical, dashed line marks the median at 34. The second whisker extends from the third quartile to 66. The first whisker on the BMW 5 plot extends from 31 to 40. The box begins at the first quartile, 40, and ends at the third quartile, 55. A vertical, dashed line marks the median at 41. The second whisker extends from 55 to 64. The first whisker on the BMW 7 plot extends from 35 to 41. The box begins at the first quartile, 41, and ends at the third quartile, 59. A vertical, dashed line marks the median at 46. The second whisker extends from 59 to 68.

Figure 2.5.4.

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

S 2.5.4

- 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%.
- The BMW 3 series is most likely to have an outlier. It has the longest whisker.
- 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.
- The second quarter has the smallest spread. There seems to be only a three-year difference between the first quartile and the median.
- The third quarter has the largest spread. There seems to be approximately a 14-year difference between the median and the third quartile.
- $IQR \sim 17$ years
- 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.

Q 2.5.5

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

Construct a box plot of the data.

2.6: Measures of the Center of the Data

Q 2.6.1

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

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

Q 2.6.2

Table 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

Percent of Underweight Children	Number of Countries
37.8–43.25	6
43.25–48.7	1

S 2.6.2

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

2.7: Skewness and the Mean, Median, and Mode

Q 2.7.1

The median age of the U.S. population in 1980 was 30.0 years. In 1991, the median age was 33.1 years.

- What does it mean for the median age to rise?
- Give two reasons why the median age could rise.
- For the median age to rise, is the actual number of children less in 1991 than it was in 1980? Why or why not?

2.8: Measures of the Spread of the Data

Use the following information to answer the next nine exercises: 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.

- $\mu = 1000$ FTES
- median = 1,014 FTES
- $\sigma = 474$ FTES
- first quartile = 528.5 FTES
- third quartile = 1,447.5 FTES
- $n = 29$ years

Q 2.8.1

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.

S 2.8.1

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.

Q 2.8.2

75% of all years have an FTES:

- at or below: _____
- at or above: _____

Q 2.8.3

The population standard deviation = _____

S 2.8.3

474 FTES

Q 2.8.4

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

Q 2.8.5

What is the *IQR*? What does the *IQR* represent?

S 2.8.5

919

Q 2.8.6

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 FTES	1,585	1,690	1,735	1,935	2,021	1,890

Q 2.8.7

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

S 2.8.7

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

Q 2.8.8

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.

Q 2.8.9

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?

S 2.8.10

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

Q 2.8.11

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

Q 2.8.12

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.

S 2.8.12

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.

Q 2.8.13

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.

- Why is Kenji considered a better runner than Nedda, even though Nedda ran faster than he?
- Who is the fastest runner with respect to his or her class? Explain why.

Q 2.8.14

The most obese countries in the world have obesity rates that range from 11.4% to 74.6%. This data is summarized in [Table 14](#).

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

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.

S 2.8.14

- $\bar{x} = 23.32$
- 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 have an unusually high percentage of obese people.

Q 2.8.15

[Table](#) 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

Percent of Underweight Children	Number of Countries
37.8–43.25	6
43.25–48.7	1

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.

This page titled [3.E: Descriptive Statistics \(Optional Exercises\)](#) 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.