Section 2.1 Summary

- Two ways of describing an individual's location within a distribution are **percentiles** and **z-scores.** An observation's percentile is the percent of the distribution that is below the value of that observation. To standardize any observation x, subtract the mean of the distribution and then divide the difference by the standard deviation. The resulting z-score

$$z = \frac{x - \text{mean}}{\text{standard deviation}}$$

  says how many standard deviations x lies above or below the distribution mean. We can also use percentiles and z-scores to compare the location of individuals in different distributions.
- A **cumulative relative frequency graph** allows us to examine location within a distribution. Cumulative relative frequency graphs begin by grouping the observations into equal-width classes (much like the process of making a histogram). The completed graph shows the accumulating percent of observations as you move through the classes in increasing order.
- It is common to **transform data,** especially when changing units of measurement. When you add a constant a to all the values in a data set, measures of center (median and mean) and location (quartiles and percentiles) increase by a. Measures of spread do not change. When you multiply all the values in a data set by a positive constant b, measures of center, location, and spread are multiplied by b. Neither of these transformations changes the shape of the distribution.

Section 2.1 Exercises

1. **Shoes** How many pairs of shoes do students have? Do girls have more shoes than boys? Here are data from a random sample of 20 female and 20 male students at a large high school:

Female:	50	26	26	31	57	19	24	22	23	38
	13	50	13	34	23	30	49	13	15	51
Male:	14	7	6	5	12	38	8	7	10	10
	10	11	4	5	22	7	5	10	35	7

(a) Find and interpret the percentile in the female distribution for the girl with 22 pairs of shoes.

(b) Find and interpret the percentile in the male distribution for the boy with 22 pairs of shoes.

(c) Who is more unusual: the girl with 22 pairs of shoes or the boy with 22 pairs of shoes? Explain.

2. **Old folks** Here is a stemplot of the percents of residents aged 65 and older in the 50 states:

```
 7 | 0
 8 | 8
 9 | 8
10 | 019
11 | 16777
12 | 01122456778999
13 | 0001223344455689
14 | 023568
15 | 24
16 | 9
```

Key: 15|2 means 15.2% of this state's residents are 65 or older

(a) Find and interpret the percentile for Colorado, where 10.1% of the residents are aged 65 and older.

(b) Find and interpret the percentile for Rhode Island, where 13.9% of the residents are aged 65 and older.

(c) Which of these two states is more unusual? Explain.

3. **Math test** Josh just got the results of the statewide Algebra 2 test: his score is at the 60th percentile. When Josh gets home, he tells his parents that he got 60 percent of the questions correct on the state test. Explain what's wrong with Josh's interpretation.

4. **Blood pressure** Larry came home very excited after a visit to his doctor. He announced proudly to his wife, "My doctor says my blood pressure is at the 90th percentile among men like me. That means I'm better off than about 90% of similar men." How should his wife, who is a statistician, respond to Larry's statement?

5. **Growth charts** We used an online growth chart to find percentiles for the height and weight of a 16-year-old girl who is 66 inches tall and weighs 118 pounds. According to the chart, this girl is at the 48th percentile for weight and the 78th percentile for height. Explain what these values mean in plain English.

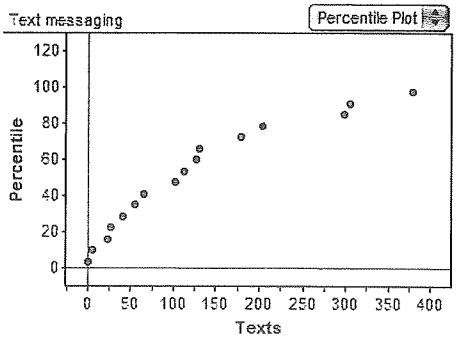
6. **Run fast** Peter is a star runner on the track team. In the league championship meet, Peter records a time that would fall at the 80th percentile of all his race times that season. But his performance places him at the 50th percentile in the league championship meet. Explain how this is possible. (Remember that lower times are better in this case!)

Exercises 7 and 8 involve a new type of graph called a ***percentile plot****. Each point gives the value of the variable being measured and the corresponding percentile for one individual in the data set.*

7. **Text me** The percentile plot below shows the distribution of text messages sent and received in a two-day period by a random sample of 16 females from a large high school.

(a) Describe the student represented by the highlighted point.

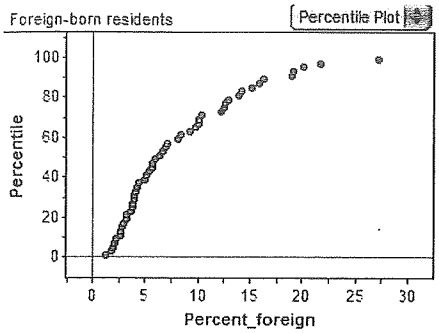
(b) Use the graph to estimate the median number of texts. Explain your method.



8. **Foreign-born residents** The following percentile plot shows the distribution of the percent of foreign-born residents in the 50 states.

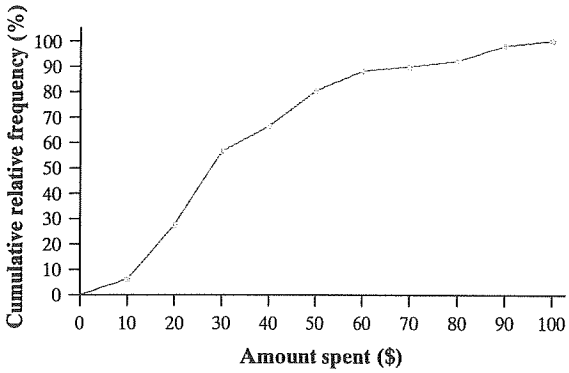
(a) The highlighted point is for Maryland. Describe what the graph tells you about this state.

(b) Use the graph to estimate the 30th percentile of the distribution. Explain your method.



9. **Shopping spree** The figure below is a cumulative relative frequency graph of the amount spent by 50 consecutive grocery shoppers at a store.

pg 88

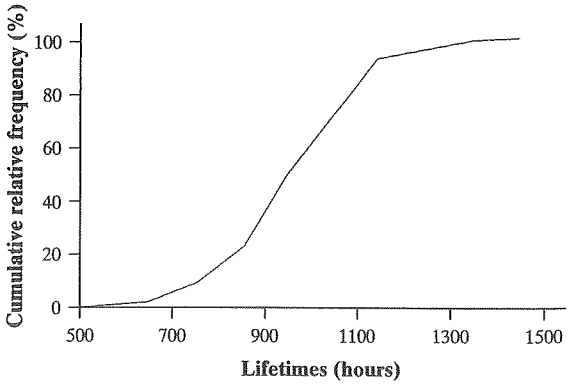


(a) Estimate the interquartile range of this distribution. Show your method.

(b) What is the percentile for the shopper who spent $19.50?

(c) Draw the histogram that corresponds to this graph.

10. **Light it up!** The graph below is a cumulative relative frequency graph showing the lifetimes (in hours) of 200 lamps.[4]



(a) Estimate the 60th percentile of this distribution. Show your method.

(b) What is the percentile for a lamp that lasted 900 hours?

(c) Draw a histogram that corresponds to this graph.

11. **SAT versus ACT** Eleanor scores 680 on the SAT Mathematics test. The distribution of SAT scores is symmetric and single-peaked, with mean 500 and standard deviation 100. Gerald takes the American College Testing (ACT) Mathematics test and scores 27. ACT scores also follow a symmetric, single-peaked distribution—but with mean 18 and standard deviation 6. Find the standardized scores for both students. Assuming that both tests measure the same kind of ability, who has the higher score?

pg 91

12. **Comparing batting averages** Three landmarks of baseball achievement are Ty Cobb's batting average of 0.420 in 1911, Ted Williams's 0.406 in 1941, and George Brett's 0.390 in 1980. These batting averages cannot be compared directly because the distribution of major league batting averages has changed over the years. The distributions are quite symmetric, except for outliers such as Cobb, Williams, and Brett. While the mean batting average has been held roughly constant by rule changes and the balance between hitting and pitching, the standard deviation has dropped over time. Here are the facts:

Decade	Mean	Standard deviation
1910s	0.266	0.0371
1940s	0.267	0.0326
1970s	0.261	0.0317

Find the standardized scores for Cobb, Williams, and Brett. Who was the best hitter?[5]

13. **Measuring bone density** Individuals with low bone density have a high risk of broken bones (fractures). Physicians who are concerned about low bone density (osteoporosis) in patients can refer them for specialized testing. Currently, the most common method for testing bone density is dual-energy X-ray absorptiometry (DEXA). A patient who undergoes a DEXA test usually gets bone density results in grams per square centimeter (g/cm^2) and in standardized units.

Judy, who is 25 years old, has her bone density measured using DEXA. Her results indicate a bone density in the hip of 948 g/cm^2 and a standardized score of $z = -1.45$. In the reference population of 25-year-old women like Judy, the mean bone density in the hip is 956 g/cm^2.[6]

(a) Judy has not taken a statistics class in a few years. Explain to her in simple language what the standardized score tells her about her bone density.

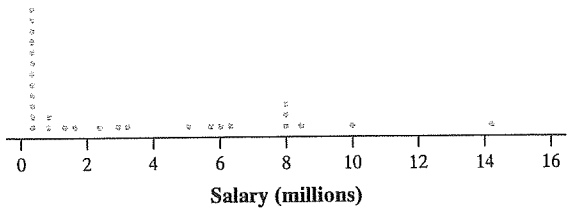
(b) Use the information provided to calculate the standard deviation of bone density in the reference population.

14. **Comparing bone density** Refer to the previous exercise. One of Judy's friends, Mary, has the bone density in her hip measured using DEXA. Mary is 35 years old. Her bone density is also reported as 948 g/cm^2, but her standardized score is $z = 0.50$. The mean bone density in the hip for the reference population of 35-year-old women is 944 grams/cm^2.

(a) Whose bones are healthier—Judy's or Mary's? Justify your answer.

(b) Calculate the standard deviation of the bone density in Mary's reference population. How does this compare with your answer to Exercise 13(b)? Are you surprised?

Exercises 15 and 16 refer to the dotplot and summary statistics of salaries for players on the World Champion 2008 Philadelphia Phillies baseball team.[7]



Variable	n	Mean	Std. dev.	Min	Q_1	Med	Q_3	Max
Salary	29	3388617	3767484	390000	440000	1400000	6000000	14250000

15. **Baseball salaries** Brad Lidge played a crucial role as the Phillies' "closer," pitching the end of many games throughout the season. Lidge's salary for the 2008 season was $6,350,000.

pg 90

(a) Find the percentile corresponding to Lidge's salary. Explain what this value means.

(b) Find the z-score corresponding to Lidge's salary. Explain what this value means.

16. **Baseball salaries** Did Ryan Madson, who was paid $1,400,000, have a high salary or a low salary compared with the rest of the team? Justify your answer by calculating and interpreting Madson's percentile and z-score.

17. The scores on Ms. Martin's statistics quiz had a mean of 12 and a standard deviation of 3. Ms. Martin wants to transform the scores to have a mean of 75 and a standard deviation of 12. What transformations should she apply to each test score? Explain.

18. Mr. Olsen uses an unusual grading system in his class. After each test, he transforms the scores to have a mean of 0 and a standard deviation of 1. Mr. Olsen then assigns a grade to each student based on the transformed score. On his most recent test, the class's scores had a mean of 68 and a standard deviation of 15. What transformations should he apply to each test score? Explain.

19. **Tall or short?** Mr. Walker measures the heights (in inches) of the students in one of his classes. He uses a computer to calculate the following numerical summaries:

pg 95

Mean	Std. dev.	Min	Q_1	Med	Q_3	Max
69.188	3.20	61.5	67.75	69.5	71	74.5

Next, Mr. Walker has his entire class stand on their chairs, which are 18 inches off the ground. Then he measures the distance from the top of each student's head to the floor.

(a) Find the mean and median of these measurements. Show your work.

(b) Find the standard deviation and *IQR* of these measurements. Show your work.

20. **Teacher raises** A school system employs teachers at salaries between $28,000 and $60,000. The teachers' union and the school board are negotiating the form of next year's increase in the salary schedule.

(a) If every teacher is given a flat $1000 raise, what will this do to the mean salary? To the median salary? Explain your answers.

(b) What would a flat $1000 raise do to the extremes and quartiles of the salary distribution? To the standard deviation of teachers' salaries? Explain your answers.

21. **Tall or short?** Refer to Exercise 19. Mr. Walker converts his students' original heights from inches to feet.

(a) Find the mean and median of the students' heights in feet. Show your work.

(b) Find the standard deviation and *IQR* of the students' heights in feet. Show your work.

22. **Teacher raises** Refer to Exercise 20. If each teacher receives a 5% raise instead of a flat $1000 raise, the amount of the raise will vary from $1400 to $3000, depending on the present salary.

(a) What will this do to the mean salary? To the median salary? Explain your answers.

(b) Will a 5% raise increase the *IQR*? Will it increase the standard deviation? Explain your answers.

23. **Cool pool?** Coach Ferguson uses a thermometer to measure the temperature (in degrees Celsius) at 20 different locations in the school swimming pool. An analysis of the data yields a mean of 25°C and a standard deviation of 2°C. Find the mean and standard deviation of the temperature readings in degrees Fahrenheit (recall that °F = (9/5)°C + 32).

24. **Measure up** Clarence measures the diameter of each tennis ball in a bag with a standard ruler. Unfortunately, he uses the ruler incorrectly so that each of his measurements is 0.2 inches too large. Clarence's data had a mean of 3.2 inches and a standard deviation of 0.1 inches. Find the mean and standard deviation of the corrected measurements in centimeters (recall that 1 inch = 2.54 cm).

Multiple choice: Select the best answer for Exercises 25 to 30.

25. Jorge's score on Exam 1 in his statistics class was at the 64th percentile of the scores for all students. His score falls

(a) between the minimum and the first quartile.

(b) between the first quartile and the median.

(c) between the median and the third quartile.

(d) between the third quartile and the maximum.

(e) at the mean score for all students.

26. When Sam goes to a restaurant, he always tips the server $2 plus 10% of the cost of the meal. If Sam's distribution of meal costs has a mean of $9 and a standard deviation of $3, what are the mean and standard deviation of the distribution of his tips?

(a) $2.90, $0.30

(b) $2.90, $2.30

(c) $9.00, $3.00

(d) $11.00, $2.00

(e) $2.00, $0.90

27. Scores on the ACT college entrance exam follow a bell-shaped distribution with mean 18 and standard deviation 6. Wayne's standardized score on the ACT was −0.5. What was Wayne's actual ACT score?

(a) 5.5 (b) 12 (c) 15 (d) 17.5 (e) 21

28. George has an average bowling score of 180 and bowls in a league where the average for all bowlers is 150 and the standard deviation is 20. Bill has an average bowling score of 190 and bowls in a league where the average is 160 and the standard deviation is 15. Who ranks higher in his own league, George or Bill?

(a) Bill, because his 190 is higher than George's 180.

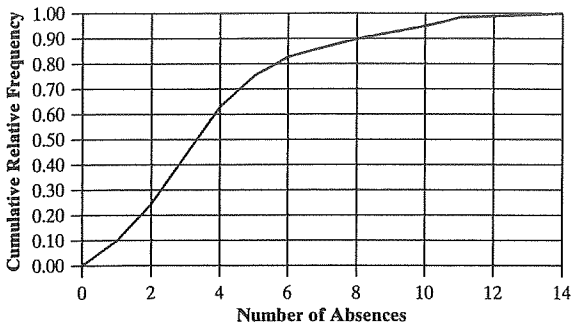
(b) Bill, because his standardized score is higher than George's.

(c) Bill and George have the same rank in their leagues, because both are 30 pins above the mean.

(d) George, because his standardized score is higher than Bill's.

(e) George, because the standard deviation of bowling scores is higher in his league.

Exercises 29 and 30 refer to the following setting. The number of absences during the fall semester was recorded for each student in a large elementary school. The distribution of absences is displayed in the following cumulative relative frequency graph.



29. What is the interquartile range (*IQR*) for the distribution of absences?

(a) 1 (c) 3 (e) 14

(b) 2 (d) 5

30. If the distribution of absences was displayed in a histogram, what would be the best description of the histogram's shape?

(a) Symmetric

(b) Uniform

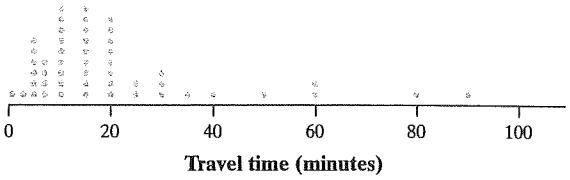
(c) Skewed left

(d) Skewed right

(e) Cannot be determined

Exercises 31 and 32 refer to the following setting. We used CensusAtSchool's Random Data Selector to choose a sample of 50 Canadian students who completed a survey in a recent year.

31. **Travel time** (1.2) The dotplot below displays data on students' responses to the question "How long does it usually take you to travel to school?" Describe the shape, center, and spread of the distribution. Are there any outliers?



32. **Lefties** (1.1) Students were asked, "Are you right-handed, left-handed, or ambidextrous?" The responses are shown below (R = right-handed; L = left-handed; A = ambidextrous).

R	R	R	R	R	R	R	R	R	R	R	L	R	R
R	R	R	R	R	R	R	R	R	R	R	R	R	A
R	R	R	R	A	R	R	L	R	R	R	R	L	A
R	R	R	R	R	R	R	R						

(a) Make an appropriate graph to display these data.

(b) Over 10,000 Canadian high school students took the CensusAtSchool survey that year. What percent of this population would you estimate is left-handed? Justify your answer.

2.2 Density Curves and Normal Distributions

WHAT YOU WILL LEARN By the end of the section, you should be able to:

- Estimate the relative locations of the median and mean on a density curve.
- Use the 68–95–99.7 rule to estimate areas (proportions of values) in a Normal distribution.
- Use Table A or technology to find (i) the proportion of *z*-values in a specified interval, or (ii) a *z*-score from a percentile in the standard Normal distribution.
- Use Table A or technology to find (i) the proportion of values in a specified interval, or (ii) the value that corresponds to a given percentile in any Normal distribution.
- Determine whether a distribution of data is approximately Normal from graphical and numerical evidence.