Cumulative Frequency Distributions

GOAL Construct and interpret cumulative frequency histograms, and compute and interpret percentiles and quartiles.

You have already seen frequency distributions. A cumulative frequency distribution describes the frequency of all data values that are less than or equal to a given value. You can display a cumulative frequency distribution in a cumulative frequency table or in a cumulative frequency histogram.

EXAMPLE 1 Making a Cumulative Frequency Histogram

The table shows the time, in minutes, that it took 50 visitors to find their way through a cornfield maze. Make a cumulative frequency histogram of the data.

Time to Complete Maze

Time, x (min)	Frequency
$0 < x \le 10$	2
$10 < x \le 20$	8
$20 < x \le 30$	15
$30 < x \le 40$	14
$40 < x \le 50$	6
$50 < x \le 60$	5

SOLUTION

Step 1 Make a cumulative frequency table. Add a third column to the given table to show cumulative frequencies. The cumulative frequency of any row is the sum of the frequency of that row and the cumulative frequency of the row before it.

Time to Complete Maze

Time, x (min)	Frequency	Cumulative frequency
$0 < x \le 10$	2	2
$10 < x \le 20$	8	$8 + 2 = 10$
$20 < x \le 30$	15	$15 + 10 = 25$
$30 < x \le 40$	14	$14 + 25 = 39$
$40 < x \le 50$	6	$6 + 39 = 45$
$50 < x \le 60$	5	$5 + 45 = 50$

Step 2 Make a cumulative frequency histogram. Use the values in the cumulative frequency column of the table to determine the heights of the bars. In the cumulative frequency histograms in this lesson, the value shown on the right end of each interval is included in that interval, and the value on the left end is not.

Time to Complete Maze
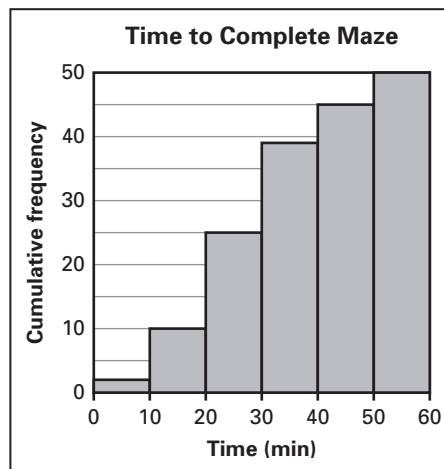
✓ **CHECK** **Example 1**

1. The table shows the scores of 32 students on a history exam. Make a cumulative frequency histogram of the data.

Score	50–59	60–69	70–79	80–89	90–99
Frequency	4	4	7	9	8

In Example 1, it is easy to see that 100% of the data is less than or equal to 60 minutes. You can also ask what percent of the data is less than or equal to other values. This leads to the idea of a *percentile*.

The **nth percentile** is the value of a variable for which n% of the items in a data set are less than or equal to that value. For example, if a veterinarian reports that the weight of a puppy is at the 40th percentile for puppies of the same breed and age, then 40% of all puppies of the same breed and age weigh the same as or less than that particular puppy.

Recall that you have already used *quartiles* to make box-and-whisker plots. The lower quartile (or first quartile) is the 25th percentile. The median (or second quartile) is the 50th percentile. The upper quartile (or third quartile) is the 75th percentile.

EXAMPLE **2** **Interpreting Percentiles and Quartiles**

On a quiz taken by 40 students, a score of 85 is the 60th percentile and a score of 90 is the upper quartile.

a. How many students scored 85 or less?

b. How many students scored greater than 90?

SOLUTION

a. Since 85 is the 60th percentile, 60% of the students scored 85 or less.

$0.6 \times 40 = 24$ *Find 60% of 40, the total number of students.*

24 students scored 85 or less.

b. Since 90 is the upper quartile, 75% of the students scored 90 or less.

$0.75 \times 40 = 30$ *Find 75% of 40, the total number of students.*

30 students scored 90 or less. Therefore, 10 students scored greater than 90.

✓ **CHECK** **Example 2**

A biologist weighs 60 sea otters at a marine sanctuary. A weight of 47 pounds is the lower quartile and a weight of 71 pounds is the 80th percentile.

2. How many otters weigh more than 47 pounds?

3. How many otters weigh 71 pounds or less?

You can use a frequency table to estimate percentiles. Note that the techniques used in Examples 3 and 4 tend to overestimate percentiles for small sets of data. For larger data sets the methods result in more accurate estimates.

EXAMPLE 3   Estimating Percentiles

The 12 players on a college basketball team reported their heights to the nearest inch. The table shows the frequencies. Estimate the percentile for a height of 78 inches.

Height (in.)	Frequency
75	1
76	2
77	4
78	3
79	2

SOLUTION

Step 1 Find the cumulative frequency for 78 inches.

$1 + 2 + 4 + 3 = 10$ *Add the frequencies for heights less than or equal to 78 inches.*

Step 2 Estimate the percentile for 78 inches.

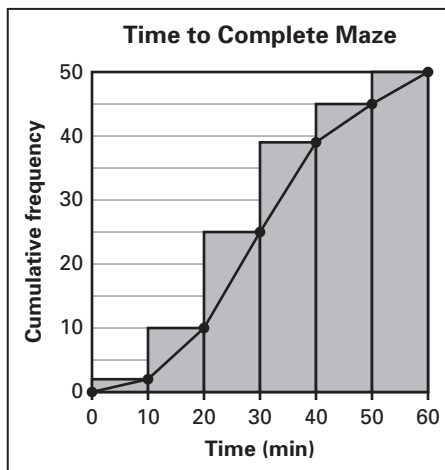$\dfrac{10}{12} = 0.8\overline{3}$ *Divide the cumulative frequency by the total number of players.*

$\approx 83\%$ *Round to the nearest percent.*

A height of 78 inches is approximately the 83rd percentile.

CHECK Example 3

4. Using the data from Example 3, estimate the percentile for a height of 77 inches.

A cumulative frequency graph is like a cumulative frequency histogram, but uses line segments instead of bars. You can make a cumulative frequency graph from a cumulative frequency histogram by drawing line segments that connect the upper right corners of the histogram's bars. The figure at the right shows a cumulative frequency graph superimposed on the cumulative frequency histogram from Example 1.


Time to Complete Maze

Cumulative frequency graphs are usually shown using *cumulative relative frequencies*. To find the cumulative relative frequency, divide the cumulative frequency by the total number of data values and express the result as a percent. The table on the next page shows the cumulative relative frequencies for the table from Example 1.

Notice that the cumulative relative frequency in the last row of the table is 100%. This makes sense because 100% of the data is less than or equal to 60 minutes.

Time to Complete Maze			
Time, x (min)	Frequency	Cumulative frequency	Cumulative relative frequency
$0 < x \leq 10$	2	2	$\frac{2}{50} = 0.04 = 4\%$
$10 < x \leq 20$	8	10	$\frac{10}{50} = 0.2 = 20\%$
$20 < x \leq 30$	15	25	$\frac{25}{50} = 0.5 = 50\%$
$30 < x \leq 40$	14	39	$\frac{39}{50} = 0.78 = 78\%$
$40 < x \leq 50$	6	45	$\frac{45}{50} = 0.9 = 90\%$
$50 < x \leq 60$	5	50	$\frac{50}{50} = 1 = 100\%$

The vertical axis of a cumulative frequency graph can be relabeled with cumulative relative frequencies without redrawing the graph. Now the graph makes it easy to estimate quartiles and percentiles. For example, to estimate the median (second quartile) for the maze data, find 50% on the vertical axis, move across to the graph, then move down to find the corresponding value on the horizontal axis. The median time to complete the maze is about 30 minutes.

Time to Complete Maze
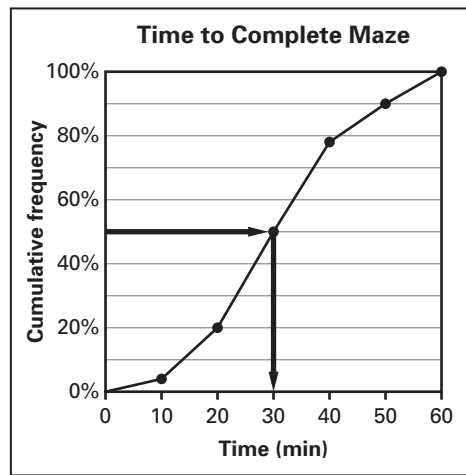
EXAMPLE 4 Reading a Cumulative Frequency Graph

A survey was conducted of 200 households to find out how long the primary television set is on during a typical day. The cumulative frequency graph shows the results. What is the 60th percentile? Interpret your answer.

Television Survey

Television Survey
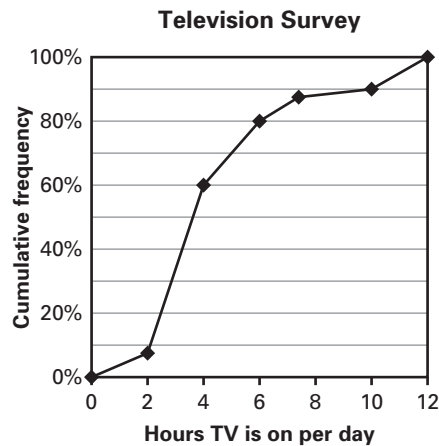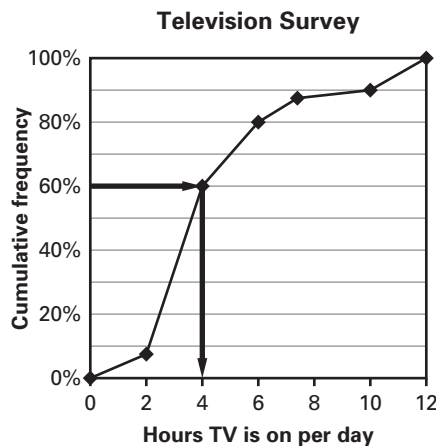
SOLUTION

Find 60% on the vertical axis. Move across to the graph and then move down to find the corresponding value on the horizontal axis. The 60th percentile is approximately 4 hours.

This means that the primary television is on for 4 hours or less in 60%, or 120, of the households that were surveyed.

5. What is the 80th percentile for the television survey data in Example 4? Interpret your answer.

EXERCISES

1. A census taker collected information about the number of people under the age of 18 in a new housing development. Use the data in the table to make a cumulative frequency histogram.

| Age	0–2	3–5	6–8	9–11	12–14	15–17
Frequency	6	4	8	12	5	5

2. An editor at a travel magazine surveyed 60 families who took weekend road trips. The table shows the number of hours the families spent driving. Make a cumulative frequency histogram of the data.

Time Spent Driving	
Time, x (h)	Frequency
$0 < x \le 5$	25
$5 < x \le 10$	8
$10 < x \le 15$	15
$15 < x \le 20$	12

In Exercises 3–6, use the following information.

The *shot put* is a track and field event in which athletes throw a heavy metal ball as far as possible. The cumulative frequency histogram shows the distances for a shot put event in which each athlete got one throw.

Shot Put Distances

3. How many athletes participated in the event?

4. How many of the throws had a distance of 12 meters or less?

5. What percent of the throws went farther than 4 meters?

6. How many of the throws had a distance greater than 16 meters but less than or equal to 20 meters? How do you know?
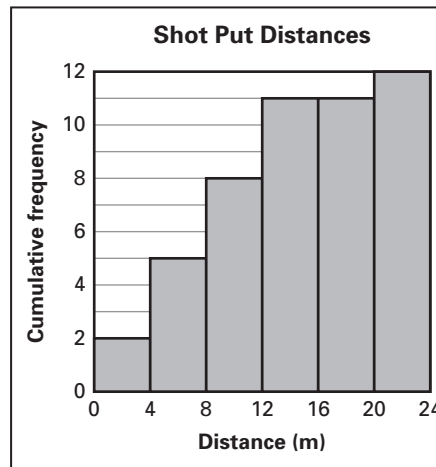
In Exercises 7–10, use the following information.

A *bar exam* is a test that is used to determine if a prospective lawyer is qualified to practice law. The table shows results from the 2009 California bar exam.

Percentile	5	15	25	35	45	55	65	75	85	95
Score	1155	1252	1300	1335	1364	1388	1429	1465	1509	1605

7. What percent of the candidates scored 1465 or less?

8. What percent of the candidates scored greater than 1364?

9. What is the lower quartile? the upper quartile?

10. Is it possible to determine the median score from the table? Why or why not?

In Exercises 11–14, use the following information.

A zoologist recorded the lengths of 80 snakes at a zoo's reptile house. A length of 0.4 meter is the lower quartile, a length of 0.9 meter is the 65th percentile, and a length of 2.1 meters is the 90th percentile.

11. How many snakes are 0.4 meter long or less?

12. How many snakes are longer than 0.4 meter?

13. How many snakes are 0.9 meter long or less?

14. How many snakes are longer than 0.9 meter but no longer than 2.1 meters?

In Exercises 15–17, use the following information.

A Web site lets users rate laptop computers on a scale of 1 to 5. The table shows the ratings for one of the computers.

15. Estimate the percentile for a rating of 3.

16. Estimate the percentile for a rating of 4.

17. Explain how you can find the percentile for a rating of 5 without doing any calculations.

Laptop Computer Ratings	
Rating	**Frequency**
1	14
2	20
3	35
4	38
5	21

In Exercises 18–22, use the following information.

A standardized test was given to a group of students. The cumulative frequency graph shows the results.
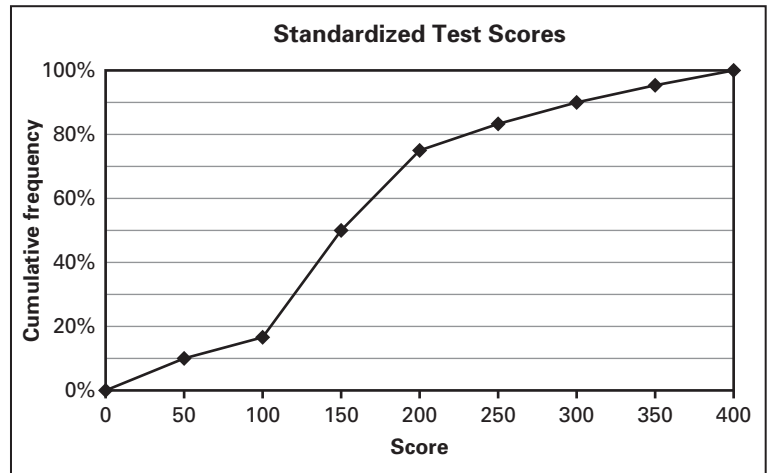
18. Estimate the median test score.

19. What is the 90th percentile? What does this tell you?

20. Estimate the upper quartile.

21. Suppose 120 students took the test. How many students got a score of 50 or less?

22. A student took the test and got a score of 250. Estimate the student's percentile.

Standardized Test Scores

**23.** Percentiles are useful in describing the spread of data. Consider the two data sets shown below:

Data Set A: 21, 21, 23, 37, 52, 67, 74, 81, 82, 87
Data Set B: 19, 34, 42, 53, 54, 60, 62, 70, 75, 83

a. Estimate the 40th percentile and the 60th percentile for data set A.

b. Estimate the 40th percentile and the 60th percentile for data set B.

c. The difference between the 60th percentile and the 40th percentile is one way to describe the amount of spread of the middle of a data set. Use this difference to compare the two data sets.

In Exercises 24–28, use the following information. Determine whether each statement about the film festival is true or false.

The organizer of a film festival collects data on the lengths of all 50 films being shown at the festival. The cumulative frequency graph shows the data.

Lengths of Films
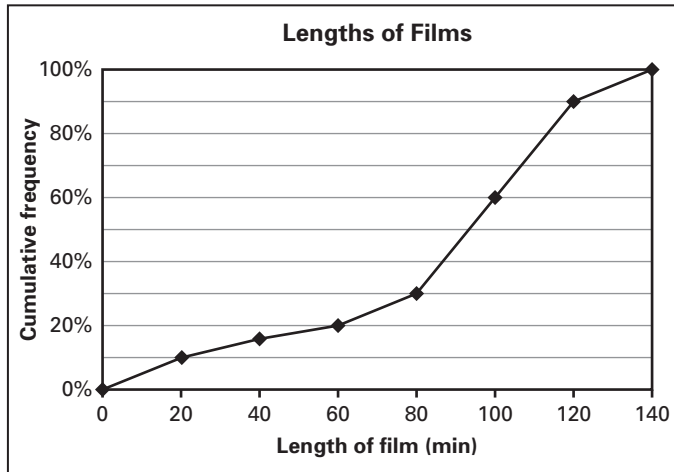
24. A length of 60 minutes is the 20th percentile.

25. The 60th percentile is about 100 minutes.

26. There are 10 films that are longer than 120 minutes.

27. There are 30 films that are at most 80 minutes long.

28. A program of all the short films (films that last 20 minutes or less) consists of 5 films.

In Exercises 29–31, use the following information.

A *decile* is one of the values that divides an ordered data set into 10 equal parts. Thus, 10% of the data are less than or equal to the 1st decile, 20% of the data are less than or equal to the 2nd decile, and so on.

29. The 320 employees of a company were surveyed about their commute times to the office. A commute time of 15 minutes is the 4th decile. How many employees commute for 15 minutes or less?

30. A geologist found the mass of 130 quartz samples. A mass of 550 milligrams is the 6th decile. How many of the samples have a mass greater than 550 milligrams?

31. On a Spanish exam, a score of 90 is the 8th decile and a score of 95 is the 9th decile. What percent of the scores are greater than 90 but less than or equal to 95?