2.1 **Learning Objectives** I will be able to…

- Find and interpret the percentile of an individual value within a distribution of data.
- Estimate percentiles and individual values using a cumulative relative frequency graph.
- Find and interpret the standardized score (*z*-score) of an individual value within a distribution of data.

Measuring Position: Percentiles

One way to describe the location of a value in a distribution is to tell what percent of observations are less than it.

The ***p*th percentile** of a distribution is the value with *p* percent of the observations less than it.

Example

Jenny earned a score of 86 on her test. How did she perform relative to the rest of the class?

6	7
7	2334
7	5777899
8	00123334
8	569
9	03

21

n=25

Her score was greater than 21 of the 25 observations. Since 21 of the 25, or 84%, of the scores are below hers, Jenny is at the 84th percentile in the class's test score distribution.

Percentile is 84th

Wins in Major League Baseball

The stemplot below shows the number of wins for each of the 30 Major League Baseball teams in 2012.

Key: 6|1 represents a team with 61 wins.

5	5
6	14
6	6899
7	234
7	569
8	113
8	56889
9	033444
9	578

Problem: Find the percentiles for the following teams:

(a) The Minnesota Twins, who won 66 games.

(b) The Washington Nationals, who won 98 games.

(c) The Texas Rangers and Baltimore Orioles, who both won 93 games.

2.1.2 Cumulative Relative Frequency Graphs

There are some interesting graphs that can be made with percentiles. One of the most common graphs starts with a frequency table for a quantitative variable. For instance, the frequency table in the margin summarizes the ages of the first 44 U.S. presidents when they took office.

Some people refer to cumulative relative frequency graphs as "ogives" (pronounced "o-jives").

Let's expand this table to include columns for relative frequency, cumulative frequency, and cumulative relative frequency.

- To get the values in the *relative frequency* column, divide the count in each class by 44, the total number of presidents. Multiply by 100 to convert to a percent.
- To fill in the *cumulative frequency* column, add the counts in the frequency column for the current class and all classes with smaller values of the variable.
- For the *cumulative relative frequency* column, divide the entries in the cumulative frequency column by 44, the total number of individuals. Multiply by 100 to convert to a percent.

Age	Frequency	Relative freq. (%)	cumulative freq.	cumulative relative freq (%)
40–44	2	4.5	2	4.5
45–49	7	15.9	9	20.5
50–54	13	29.5	22	50
55–59	12	27.3	34	77.3
60–64	7	15.9	41	93.2
65–69	3	6.8	44	100

To make a **cumulative relative frequency graph,** we plot a point corresponding to the cumulative relative frequency in each class at the smallest value of the *next* class. For example, for the 40 to 44 class, we plot a point at a height of 4.5% above the age value of 45. This means that 4.5% of presidents were inaugurated *before* they were 45 years old. (In other words, age 45 is the 4.5th percentile of the inauguration age distribution.)

It is customary to start a cumulative relative frequency graph with a point at a height of 0% at the smallest value of the first class (in this case, 40). The last point we plot should be at a height of 100%. We connect consecutive points with a line segment to form the graph. **Figure 2.1** shows the completed cumulative relative frequency graph.

Here's an example that shows how to interpret a cumulative relative frequency graph.

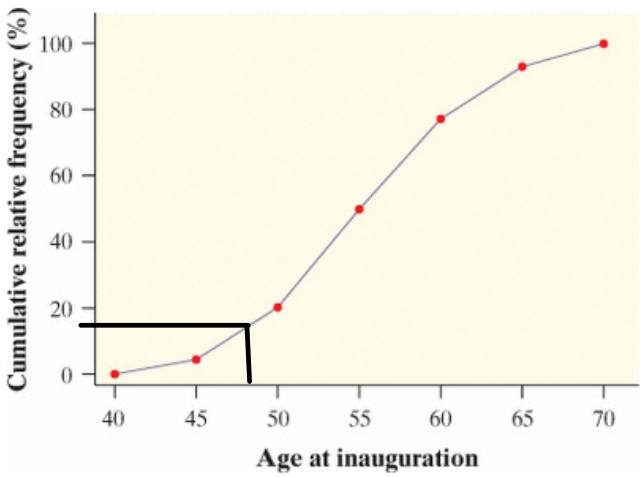


Figure 2.1 Cumulative relative frequency graph for the ages of U.S. presidents at inauguration.

1. Why does the graph get very steep at age 50?

majority of Presidents were inagurated in their 50's

2. What percent of presidents were between the ages 55 and 59 at inauguration?

Using table 27%

Using graph approx 25%

3. Was Barack Obama, first inaugurated at age 47, unusually young?

Yes, approx 10-15% were 47 or younger

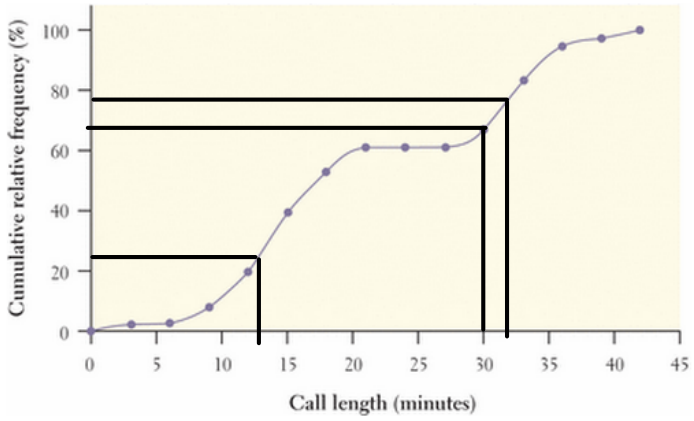
4. Estimate and interpret the 65th percentile.

65% of presidents were inagurated at about 57-58 years or less young

1. *Multiple choice: Select the best answer.* Mark receives a score report detailing his performance on a statewide test. On the math section, Mark earned a raw score of 39, which placed him at the 68th percentile. This means that

(a) Mark did better than about ~~39%~~ of the students who took the test.
(b) Mark did worse than about ~~39%~~ of the students who took the test.
(c) Mark did better than about 68% of the students who took the test.
(d) Mark did worse than about 68% of the students who took the test.
(e) Mark got fewer than half of the questions correct on this test.

2. Mrs. Munson is concerned about how her daughter's height and weight compare with those of other girls of the same age. She uses an online calculator to determine that her daughter is at the 87th percentile for weight and the 67th percentile for height. Explain to Mrs. Munson what this means.



Questions 3 and 4 relate to the following setting. The graph displays the cumulative relative frequency of the lengths of phone calls made from the mathematics department office at Gabalot High last month.

3. About what percent of calls lasted less than 30 minutes? 30 minutes or more?

4. Estimate Q_1, Q_3, and the *IQR* of the distribution.

$Q_1 = 13$ min $Q_3 = 32$ min

IQR = 19 min

Measuring Position: z-Scores

A z-score tells us how many standard deviations from the mean an observation falls, and in what direction.

If x is an observation from a distribution that has known mean and standard deviation, the **standardized score** of x is:

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

A standardized score is often called a **z-score**.

Example

$\bar{x} = 80$

Jenny earned a score of 86 on her test. The class mean is 80 and the standard deviation is 6.07. What is her standardized score?

$S_x = 6.07$

$$z = \frac{x - \text{mean}}{\text{standard deviation}} = \frac{86 - 80}{6.07} = 0.99$$

Wins in Major League Baseball
In 2012, the mean number of wins for teams in Major League Baseball was 81 with a standard deviation of 11.9 wins.
Problem: Find and interpret the *z*-scores for the following teams.
(a) The New York Yankees, with 95 wins.
(b) The New York Mets, with 74 wins.

Yankees

$$z = \frac{95-81}{11.9} = 1.18$$

Yankees are 1.18 standard deviations above the mean # of wins.

Mets

$$z = \frac{74-81}{11.9} = -.59$$

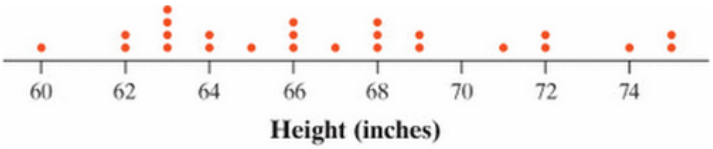
Mets are .59 std dev below the mean # of wins.

Home run kings
The single-season home run record for Major League Baseball has been set just three times since Babe Ruth hit 60 home runs in 1927. Roger Maris hit 61 in 1961, Mark McGwire hit 70 in 1998,and Barry Bonds hit 73 in 2001. In an absolute sense, Barry Bonds had the best performance of these four players, because he hit the most home runs in a single season. However, in a relative sense, this may not be true. Baseball historians suggest that hitting a home run has been easier in some eras than others. This is due to many factors, including quality of batters, quality of pitchers, hardness of the baseball, dimensions of ballparks, and possible use of performance-enhancing drugs. To make a fair comparison, we should see how these performances rate relative to those of other hitters during the same year.
Problem: Compute the standardized scores for each performance using the information in the table.
Which player had the most outstanding performance relative to his peers?

Year	Player	HR	Mean	SD	z
1927	Babe Ruth	60	7.2	9.7	5.44
1961	Roger Maris	61	18.8	13.4	3.16
1998	Mark McGwire	70	20.7	12.7	3.87
2001	Barry Bonds	73	21.4	13.2	3.91

All are exceptional relative to their season, but Babe is the single season champ!

Heights of students in my Algebra 2 class:



Variable	n	$\bar{x}$	s_x	Min	Q_1	Med	Q_3	Max
Height	25	67	4.29	60	63	66	69	75

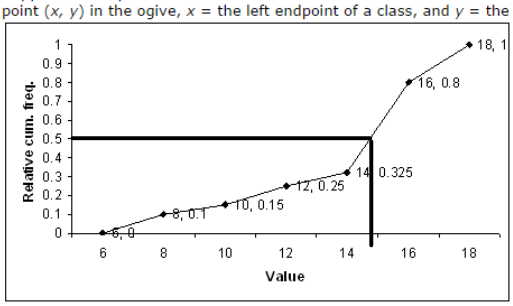
HW 2.1 Part 1
#1-15 odd
SKIP 7

1. Lynette, a student in the class, is 65 inches tall. Find and interpret her z-score.

2. Another student in the class, Brent, is 74 inches tall. How tall is Brent compared with the rest of the class? Give appropriate numerical evidence to support your answer.

3. Brent is a member of the school's basketball team. The mean height of the players on the team is 76 inches. Brent's height translates to a z-score of -0.85 in the team's height distribution. What is the standard deviation of the team members' heights?

1. Suppose that a particular set of observations has the cumulative relative frequency graph (ogive) shown below. Recall that for each point (x, y) in the ogive, x = the left endpoint of a class, and y = the percentage of observations in all classes below that class.



The median of the distribution is located
- A. in the class from 6 up to but not including 8.
- B. in the class from 14 up to but not including 16.
- C. in the class from 16 up to but not including 18.

2. Scores on the American College Testing (ACT) college entrance exam follow a Normal distribution with mean 18 and standard deviation 6. Lisa's standardized score on the ACT was z = -0.7. What was her actual ACT score?
- A. 4.2
- B. 13.8
- C. 22.2

3. An office uses two brands of fluorescent light bulbs in its overhead light fixtures. From past experience, it is known that Brand A bulbs have a mean life length of 3000 hours and a standard deviation of 200 hours, while Brand B bulbs have a mean life length of 2700 hours and a standard deviation of 250 hours. Which bulb has a longer life relative to its brand, a Brand A bulb that lasts 3150 hours or a Brand B bulb that lasts 2850 hours?
- A. The Brand A bulb has a longer life relative to its brand.
- B. The Brand B bulb has a longer life relative to its brand.
- C. The two bulbs have equally long lives.

Without talking to anyone else, write down your estimate for how wide (side to side) the room is in meters (round to nearest m).

_____

2.1 Part 2 Transforming Data

Describe the effect of adding, subtracting, multiplying by, or dividing by a constant on the shape, center, and spread of a distribution of data.

Analyze the distribution of the estimates.

The actual width of the room is about 11 m (10.765m).

Let's analyze the distribution of the errors.

error = guess - 11

Effect of Adding or Subtracting a constant

Transforming Data

Transforming converts the original observations from the original units of measurements to another scale. Transformations can affect the shape, center, and spread of a distribution.

Effect of Adding (or Subtracting) a Constant

Adding the same number *a* to (subtracting *a* from) each observation:

- adds *a* to (subtracts *a* from) measures of center and location (mean, median, quartiles, percentiles), but
- Does not change the shape of the distribution or measures of spread (range, *IQR*, standard deviation).

Effect of Multiplying or Dividing by a constant

We aren't used to working with meter, let's convert the errors to feet.

Let's analyze the distribution of the errors in feet.

error (m) = error (ft) * 3.28

There is about 3.28 feet in 1 meter

Transforming Data

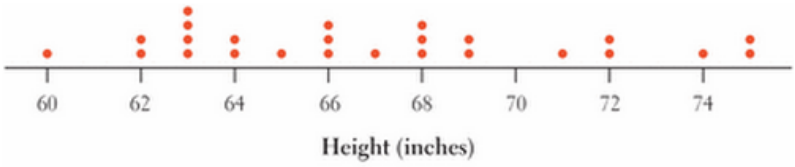
Transforming converts the original observations from the original units of measurements to another scale. Transformations can affect the shape, center, and spread of a distribution.

Effect of Multiplying (or Dividing) by a Constant

Multiplying (or dividing) each observation by the same number *b*:

- multiplies (divides) measures of center and location (mean, median, quartiles, percentiles) by *b*
- multiplies (divides) measures of spread (range, *IQR*, standard deviation) by $|b|$, but
- does not change the shape of the distribution

Taxi cabs
In 2010, taxi cabs in New York City charged an initial fee of $2.50 plus $2 per mile.
In equation form, *fare* = 2.50 + 2(*miles*).
At the end of a month, a businessman collects all his taxi cab receipts and calculates some numerical summaries.
The mean fare he paid was $15.45, with a standard deviation of $10.20.
Problem: What are the mean and standard deviation of the lengths of his cab rides in miles?



Variable	n	$\bar{x}$	s_x	Min	Q_1	Med	Q_3	Max
Height	25	67	4.29	60	63	66	69	75

1. Suppose that you convert the class's heights from inches to centimeters (1 inch = 2.54 cm). Describe the effect this will have on the shape, center, and spread of the distribution.

2. If Mrs. Navard had the entire class stand on a 6-inch-high platform and then had the students measure the distance from the top of their heads to the ground, how would the shape, center, and spread of this distribution compare with the original height distribution?

3. Now suppose that you convert the class's heights to *z*-scores. What would be the shape, center, and spread of this distribution? Explain.