2.1
Describing Location in a Distribution

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.

Example 1: Use the scores on Mr. Pryor's first statistics test to find the percentiles for the following students:

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

a) Norman, who earned a 72.

b) Katie, who scored 93.

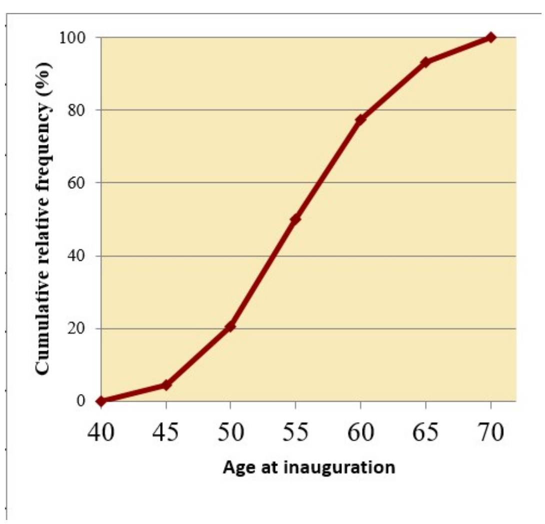
6	7
7	2334
7	5777899
8	00123334
8	569
9	03

c) The two students who earned scores of 80.

Two students scored an 80 on Mr. Pryor's first test. Because 12 of the 25 scores in the class were less than 80, these two students are at the 48th percentile.

A **cumulative relative frequency** graph (or **ogive**) displays the cumulative relative frequency of each class of a frequency distribution.

Age of First 44 Presidents When They Were Inaugurated		
Age	Frequency	
40-44	2	
45-49	7	
50-54	13	
55-59	12	
60-64	7	
65-69	3	



Interpreting cumulative relative frequency graphs

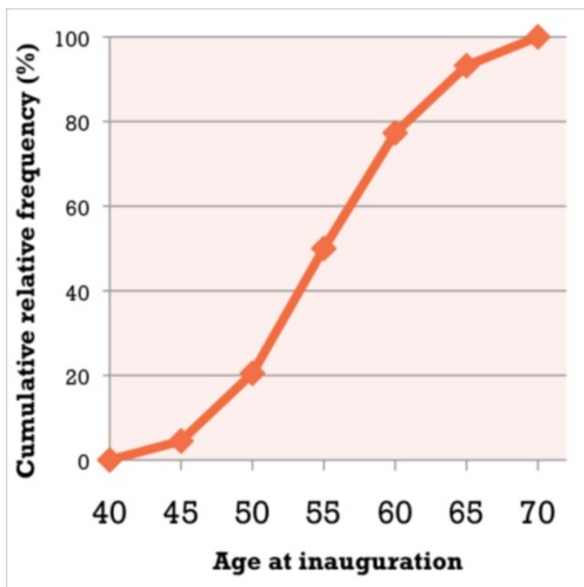
Example 2: Use the graph to the right to answer the following questions.

a) Was Barack Obama, who was inaugurated at age 47, unusually young?

about 11% of the presidents were younger than him.

b) Estimate and interpret the 65th percentile of the distribution.

Age 58 corresponds with the 65th percentile ranking, meaning 65% of the presidents are younger than 58 years of age.



Measuring Position: *z*-Scores

Converting observations from original values to standard deviation units is known as **standardizing**. To standardize a value, subtract the mean of the distribution and then divide by the standard deviation.

Definition:

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

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

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

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

Example 3: Use the information in stemplot to find the standardized scores (*z*-scores) for each of the following students in Mr. Pryor's class. Interpret each value in context.

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

mean = 80 standard deviation = 6.07

a) Katie, who scored 93.

b) Norman, who earned a 72.

Using z-scores for Comparison

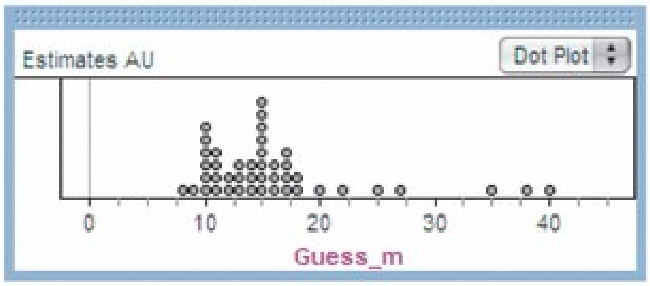
We can use z-scores to compare the position of individuals in different distributions.

Example 4: The day after receiving her statistics test result of 86 from Mr. Pryor, Jenny earned an 82 on Mr. Goldstone's chemistry test. At first, she was disappointed. Then Mr. Goldstone told the class that the distribution of scores was fairly symmetric with a mean of 76 and a standard deviation of 4. On which test did Jenny perform better relative to the rest of her class?


STOP

15 15 16 16 16 17 17 17 17 18 18 20 22
25 27 35 38 40

Figure 2.4 includes a dotplot of the data and some numerical summaries.



	n	$\bar{x}$	s_x	Min	Q_1	Med	Q_3	Max	IQR	Range
Guess	44	16.02	7.14	8	11	15	17	40	6	32

Center: The median guess was 15 meters and the mean guess was about 16 meters. Due to the clear skewness and potential outliers, the median is a better choice for summarizing the "typical" guess.

S

C

Spread: Because Q_1 = 11, about 25% of the students estimated the width of the room to be fewer than 11 meters. The 75th percentile of the distribution is at about Q_3 = 17. The *IQR* of 6 meters describes the spread of the middle 50% of students' guesses. The standard deviation tells us that the typical distance of students' guesses from the mean was about 7 meters. Because *sx* is not resistant to extreme values, we prefer the *IQR* to describe the variability of this distribution.

S

Outliers: By the 1.5 × *IQR* rule, values greater than 17 + 9 = 26 meters or less than 11 − 9 = 2 meters are identified as outliers. So the four highest guesses—which are 27, 35, 38, and 40 meters–are outliers.

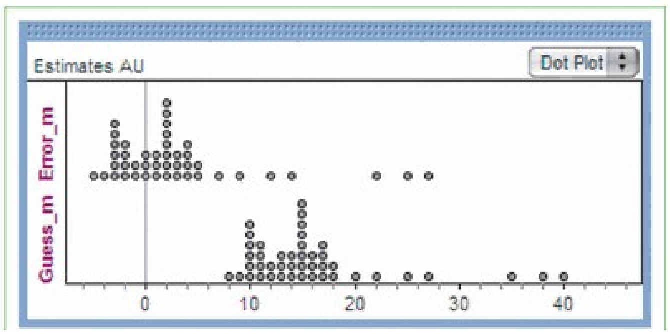
O

Example 6 Estimating Room Width

Effect of subtracting a constant

Let's see how accurate your predictions were (you did make predictions, right?). **Figure 2.5** shows dotplots of students' original guesses and their errors on the same scale. We can see that the original distribution of guesses has been shifted to the left.

error = guess – 13   room was 13 meters



What do you notice about both dot plots?

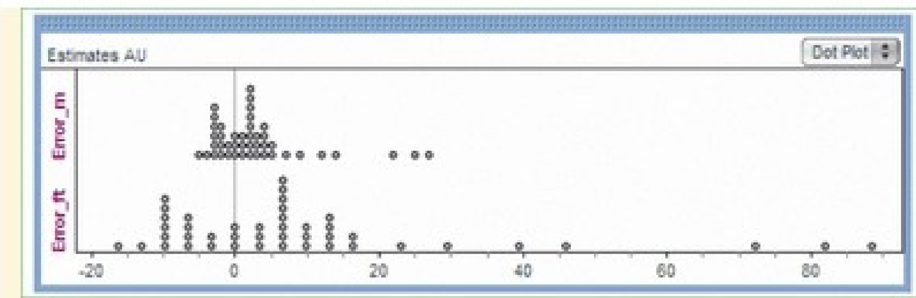
What do you think it did to the mean, median, and interquartile range?

EFFECT OF ADDING (OR SUBTRACTING) A CONSTANT

- As the example shows, subtracting the same positive number from each value in a data set shifts the distribution to the left by that number. Adding a positive constant to each data value would shift the distribution to the right by that constant.

  - does not change the shape of the distribution or measures of spread (range, *IQR*, standard deviation).

Effect of multiplying or dividing by a constant: Because our group of Australian students is having some difficulty with the metric system, it may not be helpful to tell them that their guesses tended to be about 2 to 3 meters too high. Let's convert the error data to feet before we report back to them. There are roughly 3.28 feet in a meter. So for the student whose error was −5 meters, that translates to



	n	$\bar{x}$	s_x	Min	Q_1	Med	Q_3	Max	IQR	Range
Error (m)	44	3.02	7.14	−5	−2	2	4	27	6	32
Error (ft)	44	9.91	23.43	−16.4	−6.56	6.56	13.12	88.56	19.68	104.96

Are the graphs the same when you multiply?

When the errors were measured in meters, the median was 2 and the mean was 3.02. For the transformed error data in feet, the median is 6.56 and the mean is 9.91. Can you see that the measures of center were multiplied by 3.28? That makes sense. If we multiply all the observations by 3.28, then the mean and median should also be multiplied by 3.28.

What about the spread? Multiplying each observation by 3.28 increases the variability of the distribution. By how much? You guessed it–by a factor of 3.28. The numerical summaries in **Figure 2.6** show that the standard deviation, the interquartile range, and the range have been multiplied by 3.28.

EFFECT OF MULTIPLYING (OR DIVIDING) BY A CONSTANT

Multiplying (or dividing) each observation by the same positive 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

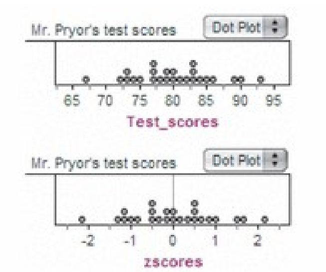
- does not change the shape of the distribution.

It is not common to multiply (or divide) each observation in a data set by a *negative* number *b*. Doing so would multiply (or divide) the measures of spread by the *absolute value* of *b*. We can't have a negative amount of variability! Multiplying or dividing by a negative numbe would also affect the shape of the distribution as all values would be reflected over the *y* axis.

Connecting transformations and z-scores

$$z = \frac{x - \text{mean}}{\text{st.dev.}}$$

- **Shape:** The shape of the distribution of *z*-scores would be the same as the shape of the original distribution of test scores. Neither subtracting a constant nor dividing by a constant would change the shape of the graph. The dotplots confirm that the combination of these two transformations does not affect the shape.
- **Center:** Subtracting 80 from each data value would also reduce the mean by 80. Because the mean of the original distribution was 80, the mean of the transformed data would be 0. Dividing each of these new data values by 6.07 would also divide the mean by 6.07. But because the mean is now 0, dividing by 6.07 would leave the mean at 0. That is, the mean of the *z*-score distribution would be 0.
- **Spread:** The spread of the distribution would not be affected by subtracting 80 from each observation. However, dividing each data value by 6.07 would also divide our common measures of spread by 6.07. The standard deviation of the distribution of *z*-scores would therefore be 6.07/6.07 = 1.



The Minitab computer output below confirms the result: *If we standardize every observation in a distribution, the resulting set of z-scores has mean 0 and standard deviation 1.*

Descriptive Statistics: Test scores, z-scores

Variable	n	Mean	StDev	Minimum	Q_1	Median	Q_3	Maximum
Test scores	25	80.00	6.07	67.00	76.00	80.00	83.50	93.00
z-scores	25	0.00	1.00	−2.14	−0.66	0.00	0.58	2.14

CHECK YOUR UNDERSTANDING

The figure below shows a dotplot of the height distribution for Mrs. Navard's class, along with summary statistics from computer output.

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.

Show Answer

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?

Show Answer

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.

Show Answer

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.

Hide Answer

Correct Answer
Shape will not change. However, it will multiply the center (mean, median) and spread (range, *IQR*, standard deviation) by 2.54.

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?

Hide Answer

Correct Answer
Shape and spread will not change. It will, however, add 6 inches to the center (mean, median).

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.

Hide Answer

Correct Answer
Shape will not change. However, it will change the mean to 0 and the standard deviation to 1.