CHAPTER 2
Modeling Distributions of Data
2.1 Describing Location in a Distribution

The Practice of Statistics, 5th Edition
Starnes, Tabor, Yates, Moore

Bedford Freeman Worth Publishers

Describing Location in a Distribution

Learning Objectives

After this section, you should 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.
- DESCRIBE the effect of adding, subtracting, multiplying by, or dividing by a constant on the shape, center, and spread of a distribution of data.

Quartiles give a measure of position of data elements in a distribution. Percentiles are another way of indicating where a data element falls in a data set relative to others in the set.

Percentiles divide the data set into 100 equal sized parts.

Example: If the weight of a 6-month old child is in the 95 percentile, it means that the child weighs more than 95% of all 6-month old infants.

In any data set

The 25th percentile = ___________

The ________ percentile = the median

The ________ percentile = the mean

The ________ percentile = Q_3

The ________ percentile = the maximum

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

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 test score distribution.

Cumulative Relative Frequency Graphs

Age of the first 44 presidents when they took office:

See activity on separate handout

Measuring Position: z-scores

For Example STANDARDIZING SKIING TIMES

The men's super combined skiing event debuted in the 2010 Winter Olympics in Vancouver. It consists of two races: a downhill and a slalom. Times for the two events are added together, and the skier with the lowest total time wins. At Vancouver, the mean slalom time was 52.67 seconds with a standard deviation of 1.614 seconds. The mean downhill time was 116.26 seconds with a standard deviation of 1.914 seconds. Bode Miller of the United States, who won the gold medal with a combined time of 164.92 seconds, skied the slalom in 51.01 seconds and the downhill in 113.91 seconds.

On which part of the race did Bodie do better compared to the competition?

Using the Standard Deviation as a Ruler:

Over and over in this course, we will ask questions such as "how far from the mean is this value" or "how different are these 2 statistics".

The answer in every case will be to measure in standard deviations.

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**.

Calculate the z-scores for Bodie's races:

Home Run Kings alternative activity

Transforming Data

How wide in meters is our classroom?

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.

Example:

4, 8, 10, 10, 12, 14 ⇨ 14, 18, 20, 20, 22, 24

Mean
Median
Q_1, Q_3
Range
IQR
Std Dev

Adding the same number, *n*, to every observation

- adds *n* to the
- but does not change

Multiplying (or dividing) each observation by the same number, m:

Example:

4, 8, 10, 10, 12, 14 ⇨ 16, 32, 40, 40, 48, 56

Mean
Median
Q_1, Q_3
Range
IQR
Std Dev

Multiplying (or dividing) each observation by the same number, m

- multiplies (divides)
- multiplies (divides)
-

Let's convert our room estimates into feet. There are approximately 3.28 feet per meter.

Check your understanding:

The figure below shows a dotplot of the height distribution for Mrs. Navard's class, alon 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.

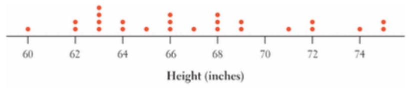
Show Answer

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

Describing Location in a Distribution

Section Summary

In this section, we learned how 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.
- DESCRIBE the effect of adding, subtracting, multiplying by, or dividing by a constant on the shape, center, and spread of a distribution of data.