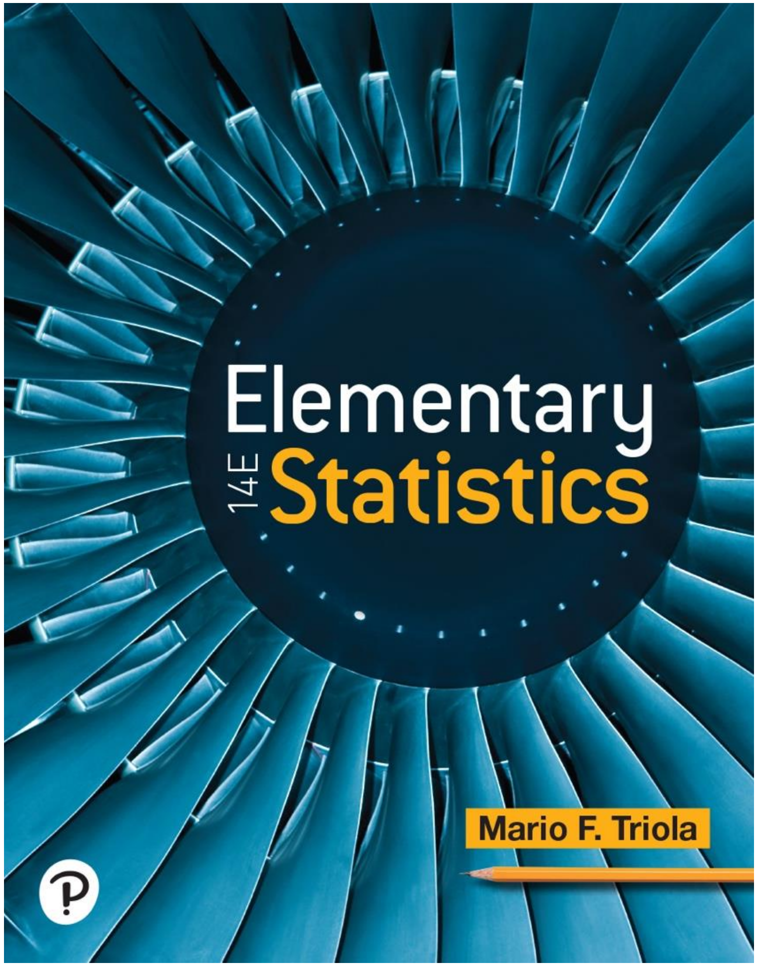
Elementary Statistics

Fourteenth Edition



Chapter 3

Describing, Exploring, and Comparing Data

Describing, Exploring, and Comparing Data

3-1 Measures of Center

3-2 Measures of Variation

3-3 Measures of Relative Standing

Key Concept

The focus of this section is to obtain a value that measures the **center** of a data set. In particular, we present measures of center, including **mean** and **median**. Our objective here is not only to find the value of each measure of center, but also to interpret those values.

Measure of Center

- Measure of Center
  – A **measure of center** is a value at the center or middle of a data set.

Mean

- Mean
  - The **mean** (or **arithmetic mean**) of a set of data is the measure of center found by adding all of the data values and dividing the total by the number of data values.

Mean

- Caution
  - Never use the term **average** when referring to a measure of center. The word **average** is often used for the mean, but it is sometimes used for other measures of center.
- The term **average** is not used by statisticians.
- The term **average** is not used by the statistics community or professional journals.

Important Properties of the Mean

- Sample means drawn from the same population tend to vary less than other measures of center.
- The mean of a data set uses every data value.
- A disadvantage of the mean is that just one extreme value (outlier) can change the value of the mean substantially. (Using the following definition, we say that the mean is not **resistant.**)

Resistant

- Resistant
  - A statistic is **resistant** if the presence of extreme values (outliers) does not cause it to change very much.

Formula

$$\text{Mean} = \frac{\sum x}{n} = \frac{\text{sum of all data values}}{\text{number of data values}}$$

If the data are a *sample* from a population, the mean is denoted by $\bar{x}$ (pronounced "*x*-bar"); if the data are the entire population, the mean is denoted by μ (lowercase Greek mu).

Notation (1 of 2)

$\sum$ denotes the **sum** of a set of data values.

x is the variable usually used to represent the individual data values.

n represents the number of data values in a **sample.**

N represents the number of data values in a **population.**

Notation (2 of 2)

$\bar{x}$ is pronounced "x-bar" and is the mean of a set of **sample** values.

$$\bar{x} = \frac{\sum x}{n}$$

μ is pronounced "mu" and is the mean of all values in a **population**.

$$\mu = \frac{\sum x}{N}$$

Example: Mean (1 of 2)

Data Set 33 “Disney World Wait Times” in Appendix B includes wait times (minutes) for six popular rides. Find the mean of the first eleven wait times for “Space Mountain” at 10 AM:

50 25 75 35 50 25 30 50 45 25 20

Example: Mean (2 of 2)

Solution

$$\bar{x} = \frac{\sum x}{n} = \frac{50 + 25 + 75 + 35 + 50 + 25 + 30 + 50 + 45 + 25 + 20}{11}$$

$$= \frac{430}{11} = 39.1 \text{ min}$$

The mean of the wait times for “Space Mountain” is 39.1 minutes.

Median

- Median
  - The **median** of a data set is the measure of center that is the **middle value** when the original data values are arranged in order of increasing (or decreasing) magnitude.

Important Properties of the Median

- The median does not change by large amounts when we include just a few extreme values, so the median is a **resistant** measure of center.
- The median does not directly use every data value. (For example, if the largest value is changed to a much larger value, the median does not change.)

Calculation of the Median

To find the median, first **sort** the values (arrange them in order) and then follow one of these two procedures:

1. If the number of data values is **odd**, the median is the number located in the exact middle of the sorted list.
2. If the number of data values is **even,** the median is found by computing the mean of the two middle numbers in the sorted list.

Example: Median with an Odd Number of Data Values (1 of 2)

Find the median of the first eleven wait times (mins) for "Space Mountain" at 10 AM:

50 25 75 35 50 25 30 50 45 25 20

Example: Median with an Odd Number of Data Values (2 of 2)

Solution

First sort the data values by arranging them in ascending order, as shown below:

20 25 25 25 30 35 45 50 50 50 75

Because the number of data values is an odd number (11), the median is the data value that is in the exact middle of the sorted list, which is 35.0 minutes.

Note the median, 35.0, is different from the mean, 39.1.

Example: Median with an Even Number of Data Values (1 of 2)

Repeat of the previous example after including the twelfth wait time for “Space Mountain” at 10 AM. That is, find the median of these wait times (minutes):

50 25 75 35 50 25 30 50 45 25 20 50

Example: Median with an Even Number of Data Values (2 of 2)

Solution

First arrange the values in ascending order:

20 25 25 25 30 35 45 50 50 50 50 75

Because the number of data values is an even number (12), the median is found by computing the mean of the two data values in the middle of the sorted list, which are 35 and 45.

$$\frac{(35 + 45)}{2} = 40.0 \text{ minutes}$$

Mode

- Mode
  - The **mode** of a data set is the value(s) that occur(s) with the greatest frequency.

Important Properties of the Mode

- The mode can be found with qualitative data.
- A data set can have no mode or one mode or multiple modes.

Finding the Mode

A data set can have one mode, more than one mode, or no mode.

- When two data values occur with the same greatest frequency, each one is a mode, and the data set is said to be **bimodal**.
- When more than two data values occur with the same greatest frequency, each is a mode, and the data set is said to be **multimodal**.
- When no data value is repeated, we say that there is **no mode**.

Example: Mode

Find the mode of the first eleven wait times for “Tower of Terror” at 10 AM.

35 35 20 50 95 75 45 50 30 35 30

Solution

Sort the list to make it easier to find values that occur more than once:

20 <u>30 30</u> <u>35 35 35</u> 45 <u>50 50</u> 75 95

The mode is 35 minutes, because it is the value occurring most often (three times).

Other Mode Examples

Two modes: The wait times (mins) of 30, 30, 50, 50, and 75 have two modes: 30 mins and 50 mins.

No mode: The wait times (mins) of 20, 30, 35, 50, and 75 have no mode because no value is repeated.

Midrange

- Midrange
  - The **midrange** of a data set is the measure of center that is the value midway between the maximum and minimum values in the original data set. It is found by adding the maximum data value to the minimum data value and then dividing the sum by 2, as in the following formula:

$$\text{Midrange} = \frac{\text{maximum data value} + \text{minimum data value}}{2}$$

Important Properties of the Midrange (1 of 2)

Because the midrange uses only the maximum and minimum values, it is very sensitive to those extremes so the midrange is not **resistant.**

Important Properties of the Midrange (2 of 2)

- In practice, the midrange is rarely used, but it has three redeeming features:
  1. The midrange is very easy to compute.
  2. The midrange helps reinforce the very important point that there are several different ways to define the center of a data set.
  3. The value of the midrange is sometimes used incorrectly for the median, so confusion can be reduced by clearly defining the midrange along with the median.

Example: Midrange

Find the midrange of the first eleven wait times (mins) for “Space Mountain” at 10 AM:

50 25 75 35 50 25 30 50 45 25 20

Solution

The midrange is found as follows:

$$\text{Midrange} = \frac{\text{maximum data value} + \text{minimum data value}}{2}$$

$$= \frac{75 + 20}{2} = 47.5 \text{ min}$$

The midrange is 47.5 minutes.

Round-Off Rules for Measures of Center

- For the mean, median, and midrange, carry one more decimal place than is present in the original set of values.
- For the mode, leave the value as is without rounding (because values of the mode are the same as some of the original data values).

Critical Thinking

- We can always calculate measures of center from a sample of numbers, but we should always think about whether it makes sense to do that.

Example: Critical Thinking and Measures of Center (1 of 5)

See each of the following illustrating situations in which the mean and median are **not** meaningful statistics.

a. Zip codes of the Gateway Arch in St. Louis, White House, Air Force division of the Pentagon, Empire State Building, and Statue of Liberty: 63102, 20500, 20330, 10118, 10004. (The zip codes don’t measure or count anything. The numbers are just labels for geographic locations.)

Example: Critical Thinking and Measures of Center (2 of 5)

See each of the following illustrating situations in which the mean and median are **not** meaningful statistics.

b. Ranks of selected medical schools of Harvard, Johns Hopkins, New York University, Stanford University, and Duke University: 1, 2, 3, 4 10. (The ranks reflect an ordering, but they don’t measure or count anything.)

Example: Critical Thinking and Measures of Center (3 of 5)

See each of the following illustrating situations in which the mean and median are **not** meaningful statistics.

c. Numbers on the jerseys of the starting offense for the New England Patriots when they won Super Bowl LIII: 12, 26, 46, 15, 11, 87, 77, 62, 60, 69, 61. (The numbers on the football jerseys don’t measure or count anything; they are just substitutes for names.)

Example: Critical Thinking and Measures of Center (4 of 5)

See each of the following illustrating situations in which the mean and median are **not** meaningful statistics.

d. Top 5 annual compensation of chief executive officers (in millions of dollars): 513.3, 256.0, 146.6, 141.7, 130.7. (Such “top 5” or “top 10” lists include data that are not at all representative of the larger population.)

Example: Critical Thinking and Measures of Center (5 of 5)

See each of the following illustrating situations in which the mean and median are **not** meaningful statistics.

e. The 50 mean ages computed from the means in each of the 50 states. (If you calculate the mean of those 50 values, the result is not the mean age of people in the entire United States. The population sizes of the 50 different states must be taken into account.)