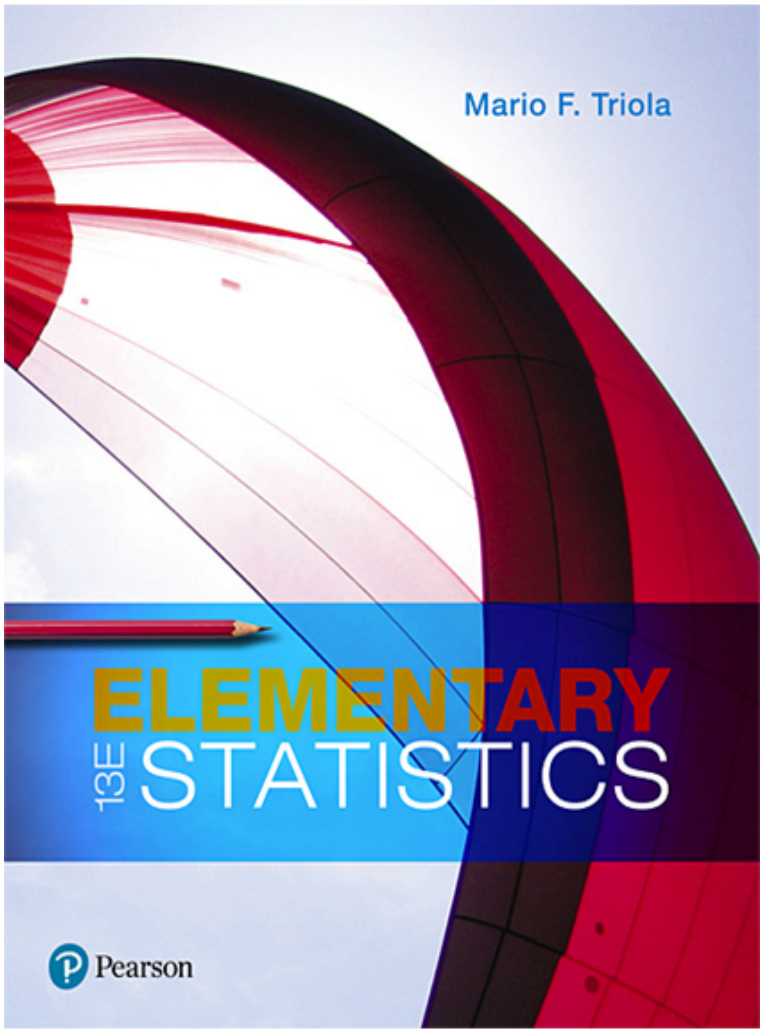
Elementary Statistics

Thirteenth 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 and Boxplots

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 (or Arithmetic Mean)

- Mean (or Arithmetic 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.

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.

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.

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

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

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

Example: Mean (1 of 2)

Data Set 32 “Airport Data Speeds” in Appendix B includes measures of data speeds of smartphones from four different carriers. Find the mean of the first five data speeds for Verizon: 38.5, 55.6, 22.4, 14.1, and 23.1 (all in megabits per second, or Mbps).

Example: Mean (2 of 2)

Solution

$$\bar{x} = \frac{\Sigma x}{n} = \frac{38.5+55.6+22.4+14.1+23.1}{5}$$

$$= \frac{153.7}{5}$$

$$= 30.74 \text{ Mbps}$$

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.

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 and Notation of the Median

The median of a sample is sometimes denoted by $\tilde{x}$ (pronounced "*x*-tilde") or ***M*** or Med.

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 five data speeds for Verizon: 38.5, 55.6, 22.4, 14.1, and 23.1 (all in megabits per second, or Mbps).

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:

14.1 22.4 23.1 38.5 55.6

Because there are 5 data values, the number of data values is an odd number (5), so the median is the number located in the exact middle of the sorted list, which is 23.1 Mbps.

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

Repeat of the previous example after including the sixth data speed of 24.5 Mbps. That is, find the median of these data speeds: 38.5, 55.6, 22.4, 14.1, 23.1, 24.5 (all in Mbps).

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

Solution

First arrange the values in ascending order:

14.1 22.4 23.1 24.5 38.5 55.6

Because the number of data values is an even number (6), the median is found by computing the mean of the two middle numbers, which are 23.1 and 24.5.

$$\text{Median} = \frac{23.1 + 24.5}{2} = \frac{47.6}{2} = 23.80 \text{ Mbps}$$

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 these Sprint data speeds (in Mbps):

0.2 0.3 0.3 0.3 0.6 0.6 1.2

Solution

The mode is 0.3 Mbps, because it is the data speed occurring most often (three times).

Other Mode Examples

Two modes: The data speeds (Mbps) of 0.3, 0.3, 0.6, 4.0, and 4.0 have two modes: 0.3 Mbps and 4.0 Mbps.

No mode: The data speeds (Mbps) of 0.3, 1.1, 2.4, 4.0, and 5.0 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 these Verizon data speeds: 38.5, 55.6, 22.4, 14.1, and 23.1 (all in Mbps)

Solution

The midrange is found as follows:

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

$$= \frac{55.6 + 14.1}{2}$$

$$= 34.85 \text{ Mbps}$$

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.
- We should also think about the sampling method used to collect the data.

Example: Critical Thinking and Measures of Center (1 of 2)

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 2)

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

b. Ranks of selected national universities of Harvard, Yale, Duke, Dartmouth, and Brown (from **U.S. News & World Report**): 2, 3, 7, 10, 14. The ranks reflect an ordering, but they don’t measure or count anything.

Calculating the Mean from a Frequency Distribution

- Mean from a Frequency Distribution
  - First multiply each frequency and class midpoint; then add the products.

$$\bar{x} = \frac{\sum (f \cdot x)}{\sum f}$$ (Result is an approximation.)

Sum of frequencies (equal to *n*)

Example: Computing the Mean from a Frequency Distribution (1 of 2)

The first two columns of the table shown here are the same as the frequency distribution of Table 2-2 from Chapter 2. Use the frequency distribution in the first two columns to find the mean.

Time (seconds)	Frequency f	Class Midpoint x	$f \cdot x$
75 – 124	11	99.5	1094.5
125 – 174	24	149.5	3588.0
175 – 224	10	199.5	1995.0
225 – 274	3	249.5	748.5
275 – 324	2	299.5	599.0
Totals:	$\sum f = 50$		$\sum(f \cdot x) = 8025.0$

Example: Computing the Mean from a Frequency Distribution (2 of 2)

Solution

When working with data summarized in a frequency distribution, we make calculations possible by pretending that all sample values in each class are equal to the class midpoint.

$$\bar{x} = \frac{\sum (f \cdot x)}{\sum f} = \frac{8025.0}{50} = 160.5 \text{ seconds}$$

The result of x = 160.5 seconds is an **approximation** because it is based on the use of class midpoint values instead of the original list of service times.

Calculating a Weighted Mean

- Weighted Mean
  - When different *x* data values are assigned different weights *w*, we can compute a weighted mean.

$$\bar{x} = \frac{\Sigma(w \cdot x)}{\Sigma w}$$

Example: Computing Grade-Point Average (1 of 4)

In her first semester of college, a student of the author took five courses. Her final grades, along with the number of credits for each course, were A (3 credits), A (4 credits), B (3 credits), C (3 credits), and F (1 credit).

The grading system assigns quality points to letter grades as follows: A = 4; B = 3; C = 2; D = 1; F = 0. Compute her grade-point average.

Example: Computing Grade-Point Average (2 of 4)

Solution

- Use the numbers of credits as weights: w = 3, 4, 3, 3, 1.
- Replace the letter grades of A, A, B, C, and F with the corresponding quality points: x = 4, 4, 3, 2, 0.

Example: Computing Grade-Point Average (3 of 4)

Solution

$$\bar{x} = \frac{\Sigma(w \cdot x)}{\Sigma w}$$

$$= \frac{(3 \times 4) + (4 \times 4) + (3 \times 3) + (3 \times 2) + (1 \times 0)}{3 + 4 + 3 + 3 + 1}$$

$$= \frac{43}{14} = 3.07$$

Example: Computing Grade-Point Average (4 of 4)

Solution

The result is a first-semester grade-point average of 3.07. (In using the preceding round-off rule, the result should be rounded to 3.1, but it is common to round grade-point averages to two decimal places.)