Measures of Center

Colby Community College

Definition

A **measure of center** is a value at the center or middle of a data set.

Definition

A **measure of center** is a value at the center or middle of a data set.

Definition

The **mean** of a set of data is the measure of center found by adding all the data values and dividing by the total number of data values.

Definition

A **measure of center** is a value at the center or middle of a data set.

Definition

The **mean** of a set of data is the measure of center found by adding all the data values and dividing by the total number of data values.

Properties of the Mean

- Sample means drawn from the same population tend to vary less than other measures of center.

Definition

A **measure of center** is a value at the center or middle of a data set.

Definition

The **mean** of a set of data is the measure of center found by adding all the data values and dividing by the total number of data values.

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.

Definition

A **measure of center** is a value at the center or middle of a data set.

Definition

The **mean** of a set of data is the measure of center found by adding all the data values and dividing by the total number of data values.

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.

Definition

A **measure of center** is a value at the center or middle of a data set.

Definition

The **mean** of a set of data is the measure of center found by adding all the data values and dividing by the total number of data values.

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.

Definition

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

Notation

Sample statisttics are usually represented by English letters, such as $\bar{x}$, and population parameters are usually represented by Greek letters, such as μ.

$\sum$	denotes the sum of a set of data values.
x	is 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.
$\bar{x} = \dfrac{\sum x}{n}$	is the mean of a set of sample values.
$\mu = \dfrac{\sum x}{N}$	is the mean of all values in a population.

Example 1

Data set 32 "Airport Data Speeds" in Appendix B includes measures of data speeds of smartphones from four different carriers. The table contains five data speeds, in megabits per second (Mbps), from the data set.

Verizon	38.5	55.6	22.4	14.1	23.1

Example 1

Data set 32 "Airport Data Speeds" in Appendix B includes measures of data speeds of smartphones from four different carriers. The table contains five data speeds, in megabits per second (Mbps), from the data set.

Verizon	38.5	55.6	22.4	14.1	23.1

The mean is

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

Example 1

Data set 32 "Airport Data Speeds" in Appendix B includes measures of data speeds of smartphones from four different carriers. The table contains five data speeds, in megabits per second (Mbps), from the data set.

Verizon	38.5	55.6	22.4	14.1	23.1

The mean is

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

Example 1

Data set 32 "Airport Data Speeds" in Appendix B includes measures of data speeds of smartphones from four different carriers. The table contains five data speeds, in megabits per second (Mbps), from the data set.

Verizon	38.5	55.6	22.4	14.1	23.1

The mean is

$$\bar{x} = \frac{\sum x}{n} = \frac{38.5 + 55.6 + 22.4 + 14.1 + 23.1}{5} = \frac{153.7}{5} = 30.74 \text{ Mbps}$$

Example 2

There are at least two ways to obtain the mean class size.

Consider a large university:

- If we take the numbers of students in all 737 classes, we get a mean of 40 students.
- If we compile a lists of the class sizes for each for each student and use this list, we could get a mean class size of 147.

What explains this discrepancy?

Example 2

There are at least two ways to obtain the mean class size.

Consider a large university:

- If we take the numbers of students in all 737 classes, we get a mean of 40 students.
- If we compile a lists of the class sizes for each for each student and use this list, we could get a mean class size of 147.

What explains this discrepancy?

At a large school, most students are in large classes, while there are few students in small classes.

Definition

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 magnitude.

Definition

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 magnitude.

Properties

- The median does not change by large amounts when we include just a few extreme value. (The median is resistant.)

Definition

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 magnitude.

Properties

- The median does not change by large amounts when we include just a few extreme value. (The median is resistant.)
- The median does not directly use every data value.

Definition

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 magnitude.

Properties

- The median does not change by large amounts when we include just a few extreme value. (The median is resistant.)
- The median does not directly use every data value.

Notation

The median of a sample is denoted $\tilde{x}$.

Definition

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 magnitude.

Properties

- The median does not change by large amounts when we include just a few extreme value. (The median is resistant.)
- The median does not directly use every data value.

Notation

The median of a sample is denoted $\tilde{x}$.

Procedure

1. Sort the values.
2.
   - If the number of data values is odd, the median is the number located in the exact middle of the sorted list.
   - 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 3

Data set 32 "Airport Data Speeds" in Appendix B includes measures of data speeds of smartphones from four different carriers. The table contains five data speeds, in megabits per second (Mbps), for Verizon.

38.5	55.6	22.4	14.1	23.1

Example 3

Data set 32 "Airport Data Speeds" in Appendix B includes measures of data speeds of smartphones from four different carriers. The table contains five data speeds, in megabits per second (Mbps), for Verizon.

38.5	55.6	22.4	14.1	23.1

First sort the data values.

14.1	22.4	23.1	38.5	55.6

Example 3

Data set 32 "Airport Data Speeds" in Appendix B includes measures of data speeds of smartphones from four different carriers. The table contains five data speeds, in megabits per second (Mbps), for Verizon.

38.5	55.6	22.4	14.1	23.1

First sort the data values.

14.1	22.4	23.1	38.5	55.6

We have 5 data values so the median is 23.1 Mbps.

Example 3

Data set 32 “Airport Data Speeds” in Appendix B includes measures of data speeds of smartphones from four different carriers. The table contains five data speeds, in megabits per second (Mbps), for Verizon.

38.5	55.6	22.4	14.1	23.1

First sort the data values.

14.1	22.4	23.1	38.5	55.6

We have 5 data values so the median is 23.1 Mbps.

Note

This different than the mean 30.74 Mbps.

Example 4

Data set 32 "Airport Data Speeds" in Appendix B includes measures of data speeds of smartphones from four different carriers. The table contains six data speeds, in megabits per second (Mbps), for Verizon.

38.5	55.6	22.4	14.1	23.1	24.5

Example 4

Data set 32 "Airport Data Speeds" in Appendix B includes measures of data speeds of smartphones from four different carriers. The table contains six data speeds, in megabits per second (Mbps), for Verizon.

38.5	55.6	22.4	14.1	23.1	24.5

First sort the data values.

14.1	22.4	23.1	24.5	38.5	55.6

Example 4

Data set 32 "Airport Data Speeds" in Appendix B includes measures of data speeds of smartphones from four different carriers. The table contains six data speeds, in megabits per second (Mbps), for Verizon.

38.5	55.6	22.4	14.1	23.1	24.5

First sort the data values.

14.1	22.4	23.1	24.5	38.5	55.6

We have 6 data values so the median is $\dfrac{23.1 + 24.5}{2} = 23.80$ Mbps.

Definition

The **mode** of a data set is the values that occur with the greatest frequency.

Definition

The **mode** of a data set is the values that occur with the greatest frequency.

Properties

- The mode can be found with categorical data.

Definition

The **mode** of a data set is the values that occur with the greatest frequency.

Properties

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

Definition

The **mode** of a data set is the values that occur with the greatest frequency.

Properties

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

Procedure

- When two data values occur with the same greatest frequency, each one is a mode and the data set is **bimodal**.

Definition

The **mode** of a data set is the values that occur with the greatest frequency.

Properties

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

Procedure

- When two data values occur with the same greatest frequency, each one is a mode and the data set is **bimodal**.
- When more than two data values occur with the same greatest frequency, each is a mode and the data set is **multimodal**.

Definition

The **mode** of a data set is the values that occur with the greatest frequency.

Properties

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

Procedure

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

Example 5

Data set 32 "Airport Data Speeds" in Appendix B includes measures of data speeds of smartphones from four different carriers. The table contains five data speeds, in megabits per second (Mbps), from the data set.

Sprint	0.2	0.3	0.3	0.3	0.6	0.6	1.2

Example 5

Data set 32 "Airport Data Speeds" in Appendix B includes measures of data speeds of smartphones from four different carriers. The table contains five data speeds, in megabits per second (Mbps), from the data set.

Sprint	0.2	0.3	0.3	0.3	0.6	0.6	1.2

The mode is 0.3 Mbps since it occurs most often.

Definition

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.

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

Definition

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.

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

Properties

- The midrange is very sensitive to extreme values.

Definition

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.

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

Properties

- The midrange is very sensitive to extreme values.
- The midrange is very easy to compute.

Example 6

Data set 32 "Airport Data Speeds" in Appendix B includes measures of data speeds of smartphones from four different carriers. The table contains five data speeds, in megabits per second (Mbps), from the data set.

Verizon	38.5	55.6	22.4	14.1	23.1

Example 6

Data set 32 "Airport Data Speeds" in Appendix B includes measures of data speeds of smartphones from four different carriers. The table contains five data speeds, in megabits per second (Mbps), from the data set.

Verizon	38.5	55.6	22.4	14.1	23.1

The midrange is

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

Example 6

Data set 32 "Airport Data Speeds" in Appendix B includes measures of data speeds of smartphones from four different carriers. The table contains five data speeds, in megabits per second (Mbps), from the data set.

Verizon	38.5	55.6	22.4	14.1	23.1

The midrange is

$$\text{Midrange} = \frac{\text{minimum} + \text{maximum}}{2} = \frac{14.1 + 55.6}{2}$$

Example 6

Data set 32 "Airport Data Speeds" in Appendix B includes measures of data speeds of smartphones from four different carriers. The table contains five data speeds, in megabits per second (Mbps), from the data set.

Verizon	38.5	55.6	22.4	14.1	23.1

The midrange is

$$\text{Midrange} = \frac{\text{minimum} + \text{maximum}}{2} = \frac{14.1 + 55.6}{2} = 34.85 \text{ Mbps}$$

Rounding Measures of Center

- For the mean, median, and midrange, carry one more decimal place than is present for the original set of values.
- For the mode, leave the value as is without rounding.

Rounding Measures of Center

- For the mean, median, and midrange, carry one more decimal place than is present for the original set of values.
- For the mode, leave the value as is without rounding.

Note

Only round the final answer, not intermediate values that occur during calculations.

Rounding Measures of Center

- For the mean, median, and midrange, carry one more decimal place than is present for the original set of values.
- For the mode, leave the value as is without rounding.

Note

Only round the final answer, not intermediate values that occur during calculations.

Critical Thinking

We can always calculate measures of center from a sample, but we need to think about whether it makes sense to do so.

Example 7

Consider the following zip codes:

Location	Zipcode
Gateway Arch in St. Louis	63102
Whitehouse	20500
Air Force division of the Pentagon	20330
Empire State Building	10118
Statue of Liberty	10004

Example 7

Consider the following zip codes:

Location	Zipcode
Gateway Arch in St. Louis	63102
Whitehouse	20500
Air Force division of the Pentagon	20330
Empire State Building	10118
Statue of Liberty	10004

The data are categorical. We can calculate the mean and median, but it would be meaningless.

Example 8

U.S. News & World Report ranks national universities:

University	**Ranking**
Harvard	2
Yale	3
Duke	7
Dartmouth	10
Brown	14

Example 8

U.S. News & World Report ranks national universities:

University	Ranking
Harvard	2
Yale	3
Duke	7
Dartmouth	10
Brown	14

While the ranks reflect an ordering, they don't count anything. The mean and median would be meaningless.

Example 9

The top five incomes, in millions of dollars, of chief executive officers in 2018 are given below. (USA Today)

Name	**Company**	**Income**
Brian Duperreault	AIG	43.1
Dirk Van de Put	Mondelez International	42.4
Mark V. Hurd	Oracle	40.8
Safra A. Catz	Oracle	40.7
Robert A. Iger	Disney	36.3

Example 9

The top five incomes, in millions of dollars, of chief executive officers in 2018 are given below. (USA Today)

Name	**Company**	**Income**
Brian Duperreault	AIG	43.1
Dirk Van de Put	Mondelez International	42.4
Mark V. Hurd	Oracle	40.8
Safra A. Catz	Oracle	40.7
Robert A. Iger	Disney	36.3

The data values are numerical, but the sample is not representative of the population at large.