Goals

Goals

1. Learn about the 3 most common measures of **central tendency**: mean, median and mode.

Goals

1. Learn about the 3 most common measures of **central tendency**: mean, median and mode.

2. Learn about the differences between these concepts.

Central Tendency

Central Tendency

- If we can only use **one value** to represent the data set (a pretty crude representation) what should we use?

Central Tendency

- If we can only use **one value** to represent the data set (a pretty crude representation) what should we use?
- Often the “center” of the data set will be used.

Central Tendency

- If we can only use **one value** to represent the data set (a pretty crude representation) what should we use?
- Often the “center” of the data set will be used.
- But the “center” of a data set could mean different things.

Central Tendency

- If we can only use **one value** to represent the data set (a pretty crude representation) what should we use?
- Often the “center” of the data set will be used.
- But the “center” of a data set could mean different things.
- **Mean**, **median** and **mode** are the 3 most common measures of central tendency.

Mean

Mean

The mean (aka average, expected value) of a data set is:

$$\text{Mean} = \frac{\text{sum of the data values}}{\text{size of the data set}}$$

Mean

The mean (aka average, expected value) of a data set is:

$$\text{Mean} = \frac{\text{sum of the data values}}{\text{size of the data set}}$$

- The mean of X is usually denoted by $\overline{x}$ or μ_x.

Mean

The mean (aka average, expected value) of a data set is:

$$\text{Mean} = \frac{\text{sum of the data values}}{\text{size of the data set}}$$

- The mean of X is usually denoted by $\overline{x}$ or μ_x.
- Symbolically $\overline{x} = \dfrac{\sum x}{n}$

Mean

The mean (aka average, expected value) of a data set is:

$$\text{Mean} = \frac{\text{sum of the data values}}{\text{size of the data set}}$$

- The mean of X is usually denoted by $\overline{x}$ or μ_x.
- Symbolically $\overline{x} = \dfrac{\sum x}{n}$
- Intuitively, $\overline{x}$ is the "average" values of the x's. However, note that $\overline{x}$ needs not be a value of the data (e.g. the mean value of rolling a fair die.)

Mean

Mean

22	14	27	27	27
3	29	30	15	27
18	18	14	28	20
23	15	22	25	19
18	17	26	17	22
23	27	14	25	24

Table: Test 1 scores for class A

- Use **Count** and **SUM** in Excel, or
- Use **AVERAGE**.

Mean

22	14	27	27	27
3	29	30	15	27
18	18	14	28	20
23	15	22	25	19
18	17	26	17	22
23	27	14	25	24

Table: Test 1 scores for class A

- Use **Count** and **SUM** in Excel, or
- Use **AVERAGE**.
- $n = 30$, $\sum x = 636$. And so $\overline{x} = 636/30 = 21.2$.

Median

The median of a data set, roughly speaking, is the middle value when the values are listed in ascending order.

Median

The median of a data set, roughly speaking, is the middle value when the values are listed in ascending order.

(lower) half of the data points $\leq$ *median* $\leq$ *(upper) half of the data points.*

Median

The median of a data set, roughly speaking, is the middle value when the values are listed in ascending order.

$$\textit{(lower) half of the data points} \leq \textit{median} \leq \textit{(upper) half of the data points.}$$

To find the median:

Median

The median of a data set, roughly speaking, is the middle value when the values are listed in ascending order.

$$\textit{(lower) half of the data points} \leq \textit{median} \leq \textit{(upper) half of the data points.}$$

To find the median:

1. order the data points: $x_1 \leq x_2 \leq \cdots \leq x_n$

Median

The median of a data set, roughly speaking, is the middle value when the values are listed in ascending order.

(lower) half of the data points $\leq$ *median* $\leq$ *(upper) half of the data points.*

To find the median:

1. order the data points: $x_1 \leq x_2 \leq \cdots \leq x_n$
2. The median = the middle data point if n is odd and is the average of the two data points that are closest to the middle if n is even.

Median

The median of a data set, roughly speaking, is the middle value when the values are listed in ascending order.

(lower) half of the data points $\leq$ *median* $\leq$ *(upper) half of the data points.*

To find the median:

1. order the data points: $x_1 \leq x_2 \leq \cdots \leq x_n$
2. The median = the middle data point if n is odd and is the average of the two data points that are closest to the middle if n is even.
3. E.g. if $n = 15$, the middle position is the 8-th position hence x_8 is the median. If $n = 16$, then the median is the average of x_8 and x_9, i.e. $(x_8 + x_9)/2$.

Median

For instance, the test scores in ascending order:

$$3, 14, 14, 14, 15, 15, 17, 17, 18, 18,$$
$$18, 19, 20, 22, 22, 22, 23, 23, 24, 25,$$
$$25, 26, 27, 27, 27, 27, 27, 28, 29, 30$$

Median

For instance, the test scores in ascending order:

$$3, 14, 14, 14, 15, 15, 17, 17, 18, 18,$$
$$18, 19, 20, 22, 22, 22, 23, 23, 24, 25,$$
$$25, 26, 27, 27, 27, 27, 27, 28, 29, 30$$

- Since there are 30 data points (an even number) the median is the average of the 15th and the 16th data point, i.e. $(22 + 22)/2 = 22$.

Median

For instance, the test scores in ascending order:

$$3, 14, 14, 14, 15, 15, 17, 17, 18, 18,$$
$$18, 19, 20, 22, 22, 22, 23, 23, 24, 25,$$
$$25, 26, 27, 27, 27, 27, 27, 28, 29, 30$$

- Since there are 30 data points (an even number) the median is the average of the 15th and the 16th data point, i.e. $(22 + 22)/2 = 22$.
- The median of a data set, just like the mean, **needs not be** the value of any data.

Median

For instance, the test scores in ascending order:

$$3, 14, 14, 14, 15, 15, 17, 17, 18, 18,$$
$$18, 19, 20, 22, 22, 23, 23, 23, 24, 25,$$
$$25, 26, 27, 27, 27, 27, 27, 28, 29, 30$$

- Since there are 30 data points (an even number) the median is the average of the 15th and the 16th data point, i.e. $(22 + 22)/2 = 22$.
- The median of a data set, just like the mean, **needs not be** the value of any data. E.g. if the 16th scores is 23 instead of 22,

Median

For instance, the test scores in ascending order:

$$3, 14, 14, 14, 15, 15, 17, 17, 18, 18,$$
$$18, 19, 20, 22, 22, 23, 23, 23, 24, 25,$$
$$25, 26, 27, 27, 27, 27, 27, 28, 29, 30$$

- Since there are 30 data points (an even number) the median is the average of the 15th and the 16th data point, i.e. $(22 + 22)/2 = 22$.
- The median of a data set, just like the mean, **needs not be** the value of any data. E.g. if the 16th scores is 23 instead of 22, the median would then be $(22 + 23)/2 = 22.5$.

Mode

Mode

A mode is a data value with the **highest frequency**.

Mode

A mode is a data value with the **highest frequency**.

By convention, the data set has **no mode** if **every value** has the **same frequency**.

Mode

A mode is a data value with the **highest frequency**.

By convention, the data set has **no mode** if **every value** has the **same frequency**.

- 27 is the only mode of the test scores, it appears 5 times.

Mode

A mode is a data value with the **highest frequency**.

By convention, the data set has **no mode** if **every value** has the **same frequency**.

- 27 is the only mode of the test scores, it appears 5 times.

- The data set

  {cow, cow, cow, sheep, sheep, duck, duck, duck, chicken, pig, pig}

Mode

A mode is a data value with the **highest frequency**.

By convention, the data set has **no mode** if **every value** has the **same frequency**.

- 27 is the only mode of the test scores, it appears 5 times.

- The data set

  {cow, cow, cow, sheep, sheep, duck, duck, duck, chicken, pig, pig}

  has two modes, namely cow and duck.

Mode

A mode is a data value with the **highest frequency**.

By convention, the data set has **no mode** if **every value** has the **same frequency**.

- 27 is the only mode of the test scores, it appears 5 times.

- The data set

  {cow, cow, cow, sheep, sheep, duck, duck, duck, chicken, pig, pig}

  has two modes, namely cow and duck. Such a data set is called bimodal.

Mode

A mode is a data value with the **highest frequency**.

By convention, the data set has **no mode** if **every value** has the **same frequency**.

- 27 is the only mode of the test scores, it appears 5 times.

- The data set

  {cow, cow, cow, sheep, sheep, duck, duck, duck, chicken, pig, pig}

  has two modes, namely cow and duck. Such a data set is called bimodal. In general, a data set is multi-modal if it has more than one mode.