Wednesday, June 3, 2020

Analyzing Data:

- Measures of Central Tendancy
- Measures of Variability

Some Statistics Vocabulary:

Measures of Central Tendency (the 3 M's):

- Mean
- Median
- Mode

These statistics give an indication of where the "middle" of the data is.

Analyzing Data

Measures of Central Tendency:

- Mean $\frac{\text{Sum of data}}{\text{\# of data items}}$ Symbol for Mean: $\overline{x}$ "x bar"
- Median The middle # or the mean of the middle two #'s (#'s must be in order!)
- Mode The # or #'s that occur the most often

What is the mode of this set of data?

41, 47, 46, 47, 39, 41, 39, 46

There is NO MODE because all #'s appear the same amount of times.
In other words, no # or #'s appear more than the other #'s.

What is the mode of this set of data?

13, 17, 21, 17, 13, 21,13

Mode = 13

13 appears three times which is more often than any other number.

What is the Mode of this data set:

14, 19, 32, 16, 19, 11, 16

Mode = 16 & 19

These #'s appear twice each when all other #'s only appear once each.

Given a set of data, how many Modes could there be?

- None
- One
- Many

Find the Mean, Median, and Mode of this set of data.

2,13, 27, 19, 21, 8, 14, 25, 15, 19

Mean:

$$\overline{x} = \frac{\text{Sum of \#'s}}{\text{\#'s of items}} = \frac{163}{10} = 16.3$$

Median

2,13, 27, 19, 21, 8, 14, 25, 15, 19

First, put the #'s in order:

2, 8, 13, 14, 15, 19, 19, 21, 25, 27

Since there is an even # of data items the Median is the average of the middle two.

$$\text{Median} = \frac{15 + 19}{2} = 17$$

Mode:

2,13, 27, 19, 21, 8, 14, 25, 15, 19

It's helpful but not necessary to have the #'s in order.

2, 8, 13, 14, 15, 19, 19, 21, 25, 27

Mode = 19

There are two 19's when every other number only appears once each.

Measures of Variability:

- Range
- Standard Deviation

These statistics give an indication of how spread out, or, how much variation there is in the data.

Range: Max Value - Min Value

Gives a measure of the Spread in a data set

Range by itself doesn't describe the whole data set because it is found using only 2 data values.

Which would be more significant?
A small range OR A large range?

- A large range tells us that the Max and Min are far apart but it doesn't give us any indication how spread out the rest of the data is.
- A small range tells us that the Max and Min are close which would also mean that the remaining data must also be close together.
  Therefore, all the data must be close to each other.

Find the range of the data set used earlier:

2,13, 27, 19, 21, 8, 14, 25, 15, 32

It's helpful but not necessary to have the #'s in order.

2, 8, 13, 14, 15, 19, 19, 21, 25, 27

Range = Max - Min = 27 - 2 = 25

Since range only tells us how spread out the Max and Min are we need another statistic to give us an indication of how spread out the entire data set is.

Standard Deviation:

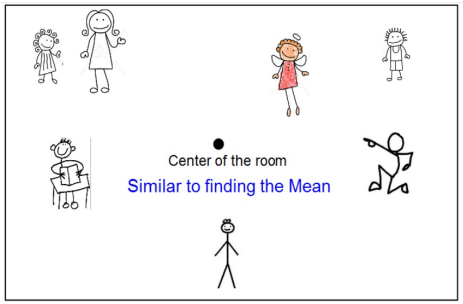
A measure of how much variation there is in a set of data.

Used by itself it doesn't tell you that much about a data set

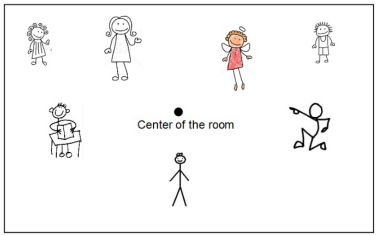
Best used to compare sets of data

Standard Deviation is a measure of how far on average each data value is from the mean.

Bigger Standard Deviation means more variation

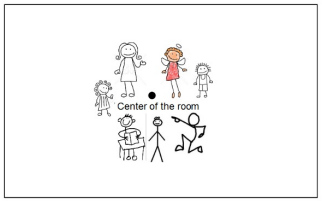


Standard Deviation is similar to the average distance each person is from the center of the room



Large or small Standard Deviation?

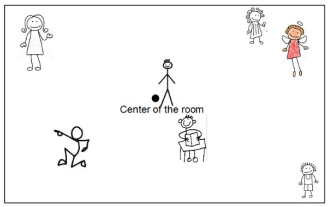
Is there a little or a lot of variation in the data set?



Small: They are all "pretty" close to the center of the room and all are about the same distance from the center.

Large or small Standard Deviation?

Is there a little or a lot of variation in the data set?



Larger: Their distances from the center of the room vary more and are for the most part further away than the previous picture.

Symbol for Standard Deviation: σ Lower Case Greek letter Sigma

Standard Deviation Formula:

$$\sigma = \sqrt{\frac{\sum(x-\bar{x})^2}{n}}$$

Using this data: 5, 6, 7, 9, 13, 15, 20, 23, 31, 40

$\sigma = 11.04$

What this number tells us is that the average distance from the mean of the entire data set is 11.04.

Standard Deviation: Commonly used to compare sets of data.

Which set of data has more variation?

Set 1: 95, 100, 105, 110, 115, 120, 125, 130

$\sigma = 11.456$

Set 2: 26, 27, 37, 39, 44, 50, 58, 61

$\sigma = 12.224$

The greater the Std Dev the more variation there is in a set of data.

Therefore, Set 2 has more variation.

Which set of data has more variation?

Set A: 12, 17, 22, 27, 32, 37, 42, 47, 52, 57

$\sigma = 14.36$

Set B: 85, 78, 79, 83, 81, 84, 86, 75, 82, 81

$\sigma = 3.2$

Set A has more variation due to a larger Std Dev.

You can now do problems 7 and 8 on Practice #30

This practice will be due on Sunday, June 7 by 10:00 pm