Measures of Central Tendancy:

- Mean
- Median
- Mode

These give a general location for the "middle" of the data

✓

Measures of Variability:

- Range
- Interquartile Range
- Standard Deviation

These give an idea of how spread out the data is and how much variation there is amongst 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 smaller range would indicate not only that the max and min are close together, but, therefore, all the rest of the data is also relatively close together.

Interquartile Range:

Upper Quartile - Lower Quartile

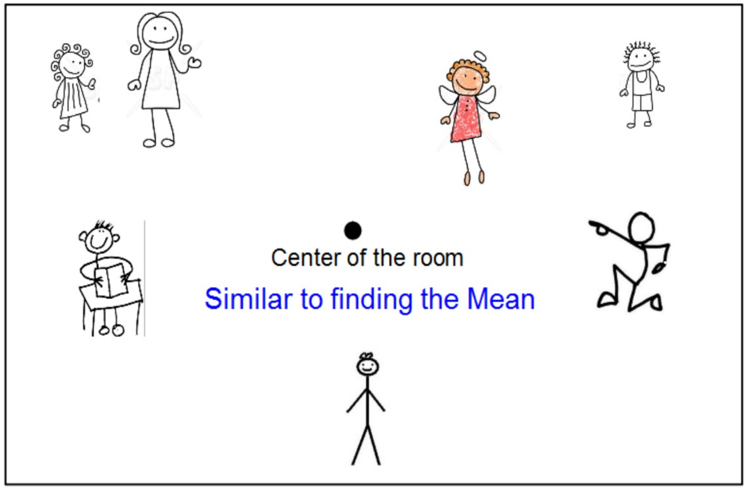
Gives a measure of how spread out the middle 50% is

Similar to Range it doesn't tell the whole story because it is found using only 2 data values.

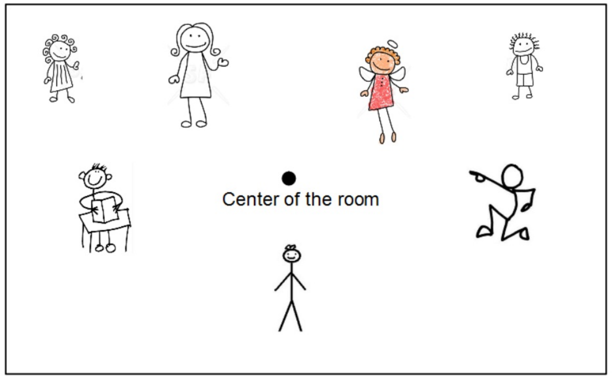
Standard Deviation:

- 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
- Most of the time it's 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 measure spread out people are in a room.

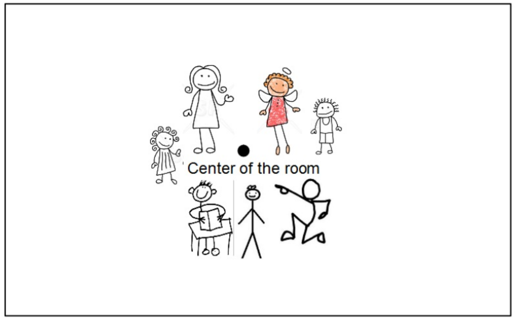


Standard Deviation is similar to the average distance each person is from the center of the room



Large or small Standard Deviation?

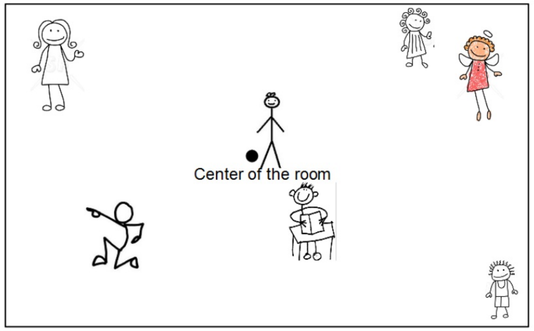
Is there a little or a lot of variation in the data set?



Small: They are all "pretty" close to the center of the room and all 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 Sigma

Standard Deviation Formula:

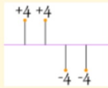
$$\sigma = \sqrt{\frac{\sum(x-\overline{x})^2}{n}}$$

1. Find the mean $\overline{x}$
2. Find the difference between each value & the mean $x-\overline{x}$
3. Square the difference $(x-\overline{x})^2$
4. Find the sum of these squares $\sum(x-\overline{x})^2$
5. Find the mean of these squares $\frac{\sum(x-\overline{x})^2}{n}$
6. Take the square root. $\sqrt{\frac{\sum(x-\overline{x})^2}{n}}$

***Footnote: Why *square* the differences?**

If we just added up the differences from the mean ... the negatives would cancel the positives:

Mean = 0



$$\frac{4+4-4-4}{4} = 0$$

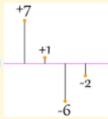
So that won't work. How about we use absolute values?

Mean = 0

$$\frac{|4|+|4|+|-4|+|-4|}{4} = \frac{4+4+4+4}{4} = 4$$

That looks good (and is the Mean Deviation), but what about this case:

Mean = 0

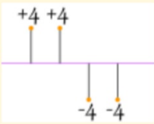


$$\frac{|7|+|1|+|-6|+|-2|}{4} = \frac{7+1+6+2}{4} = 4$$

Oh No! It also gives a value of 4, Even though the differences are more spread out!

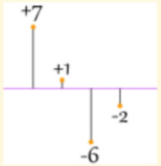
So let us try squaring each difference (and taking the square root at the end):

Mean = 0



$$\sqrt{\frac{4^2 + 4^2 + 4^2 + 4^2}{4}} = \sqrt{\frac{64}{4}} = 4$$

Mean = 0



$$\sqrt{\frac{7^2 + 1^2 + 6^2 + 2^2}{4}} = \sqrt{\frac{90}{4}} = 4.74...$$

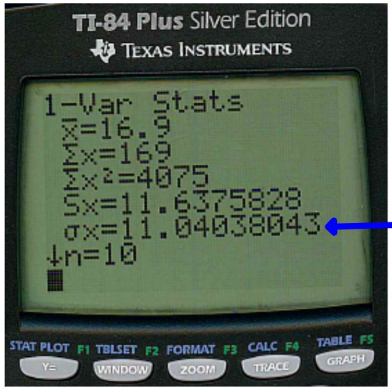
That is nice! The Standard Deviation is bigger when the differences are more spread out ... just what we want!

Using this set of numbers: 5, 6, 7, 9, 13, 15, 20, 23, 31, 40

To find the Standard Deviation using the graphing calculator:

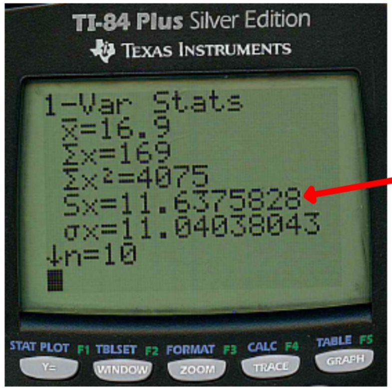
1. Enter the numbers into a List
2. Perform 1-Var Stats
3. Look for σx

σx = 11.04



σ

Population Standard Deviation: Uses all data values



This is called the Sample Standard Deviation and is used when you only have a sample of data, not all of it (population).

Using Excel to find Standard Deviation

	A	B	C
1		5	15
2		6	20
3		7	23
4		9	31
5		13	40
6			
7			11.0404

=stdevp(B1:C5)

=stdevp(B1:C5)

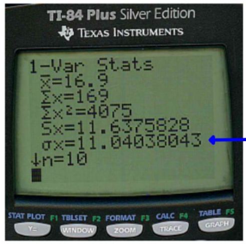
p stands for Population which means you are using ALL the data.

Population Standard Deviation:

This is the kind of standard deviation we will be using since we will have ALL of the data to work with (the whole population).

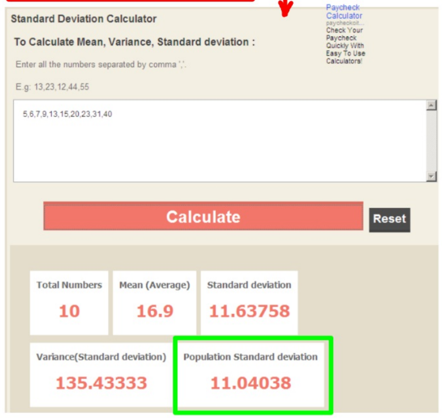
Use the formula on page 669: $\sigma = \sqrt{\dfrac{\Sigma(x - \overline{x})^2}{n}}$

This will match what you get when using the graphing calculator:



These will also match what you get when:

Using the link from my Blog:



Using Excel:

	A	B	C
1		5	15
2		6	20
3		7	23
4		9	31
5		13	40
6			
7			11.0404

=stdevp(B1:C5)

Standard Deviation Calculator Link on my blog:

Population Standard Deviation - uses all of the data values

Standard Deviation:
Mostly used to compare two sets of data

Which set of data has more variation?

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

σ = 11.456

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

σ = 12.224

The greater the Standard Deviation the more variation there is in the 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

$\sigma_x = 14.36$

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

$\sigma_x = 3.2$

Due to it's greater Standard Deviation, Set A has more variation.

Which set of data has more variation?

Set 1: 5,6,8,10,13,15,19

$\sigma = 4.703$

Set 2: 48,50,51,53,56,57,60

$\sigma = 3.959$

Due to it's greater Standard Deviation, Set A has more variation.