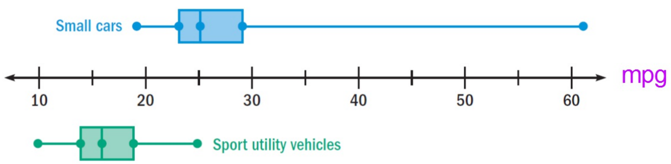
Which vehicles get better gas mileage? Give a reason.



Small cars get better gas mileage because...

- 50% of them get 25mpg or better when no SUV gets higher than 25mpg
- All of them get 18mpg or better when only 25% of SUV's get 18mpg or better

On a standardized test your score was reported to be in the 90th percentile.

What does this mean?

Your score was better than 90% of all those who took the test.

Percentile:

A number that represents the percent of data that falls below a given value.

If you tested at the 85th percentile that means that you scored higher than 85% of those taking the test.

Or you could say that 85% of those testing ended up below your score.

Or you could say that only 15% of those testing scored higher than you.

12, 9, 8, 15, 20, 3, 17, 9, 10, 14

First put the data in order!

3, 8, 9, 9, 10, 12, 14, 15, 17, 20

1. 17 is at what percentile? $\frac{8}{10}$ → 80th
2. What number is at the 40th percentile? 10
3. 9 is at what percentile? $\frac{2}{10}$ → 20th

24,28,29,32,33,38,38,39,41,43,44,56,57,60,68

1. What percentile is 38 at? $\frac{5}{15} = 33^{rd}$

2. What value is at the 80th percentile? $.80(15) = 12$

57

Could you score at the 100th percentile?

Not using this definition of percentile.
You can't score better than 100% of all those who took the test.
(you can't score better than youself!)

You can now finish Hwk #18 Sec 12-3

Pages 664-665

Problems 1, 2, 9-11, 14, 16-18

Due tomorrow

In a shipment of 20 computers, 3 are defective. Three computers are randomly selected and tested. What is the probability that all three are defective if the first and second ones are not replaced after being tested?

$\frac{3}{20} \cdot \frac{2}{19} \cdot \frac{1}{18}$

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 given and 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?

Interquartile Range:

Upper Quartile - Lower Quartile

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

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

Standard Deviation:

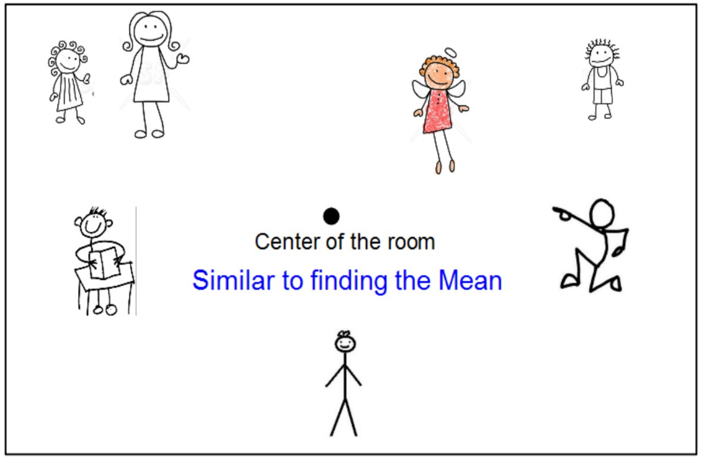
A measure of how much variation there is in a set of data.

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

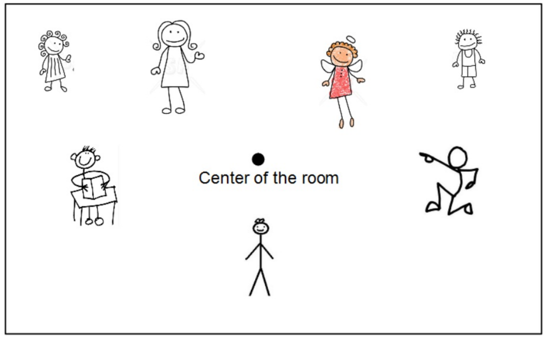
Best used to compare sets of data

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

Bigger Standard Deviation means more variation

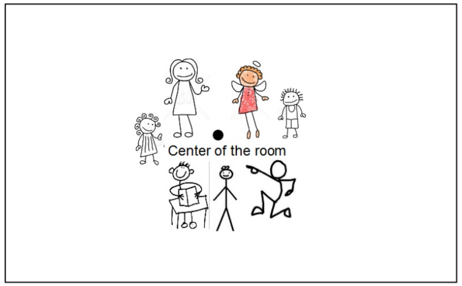


Standard Deviation is similar to the average distance each person is from the center of the room



Large or small Standard Deviation?

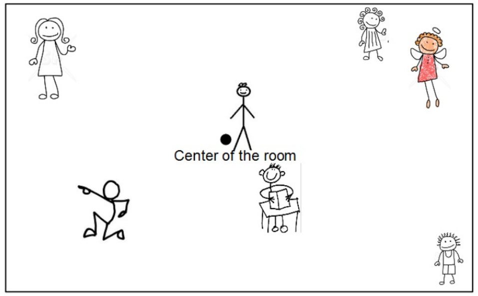
Is there a little or a lot of variation in the data set?



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

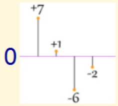
$\frac{4 + 4 - 4 - 4}{4} = 0$

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

Mean = 0 

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

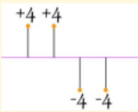
That looks good (and is the Mean Deviation), but what about this case:

Mean = 0 

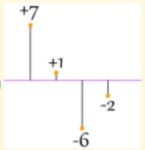
$\frac{|7| + |1| + |-6| + |-2|}{4} = \frac{7 + 1 + 6 + 2}{4} = 4$

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!

Standard Deviation Demonstration:

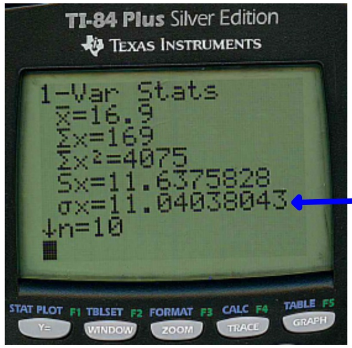
http://www.stat.rice.edu/~stat280/Applets/stddev.html

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

Find the Standard Deviation rounded to the nearest hundredth.

$$\sigma = 11.04$$

Using a Graphing Calculator



σ

Population Standard Deviation: Uses all data values

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.

Standard Deviation Calculator Link on my blog:

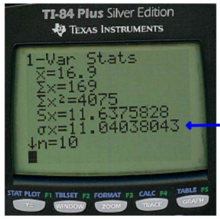
Population Standard Deviation - uses all of the data values

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).

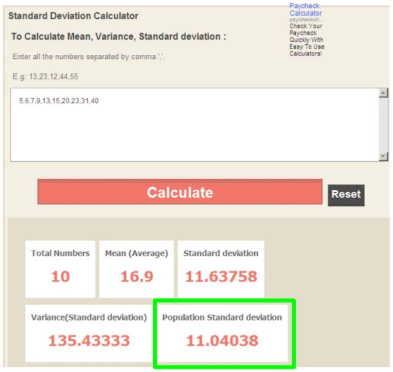
Use the formula on page 669: $\sigma = \sqrt{\frac{\Sigma(x - \overline{x})^2}{n}}$

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:

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

$\sigma = 11.456$

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

$\sigma = 12.224$

The greater the Standard Deviation the more variation there is in the set of data.

Set #2 has more variation because it has a larger Standard Deviation

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

$\sigma_x = 3.2$

Set A has more variation because it has a larger Standard Deviation

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

$\sigma = 3.959$

Set #1 has more variation because it has a larger Standard Deviation