Percentile:

A number that represents the percent of data that falls <u>below</u> a given value.

Use this set of data:

23, 39, 45, 46, 49, 50, 51, 51, 51, 68, 68, 71, 73, 82, 88, 89, 92, 95, 98, 99

6. 95 is at what percentile? $\frac{17}{20}$ → 85 %-tile

7. 51 is at what percentile? $\frac{6}{20}$ 30th %-tile

8. What number is at the 20th percentile? $(.20)(20) = 4$ 49 there are 4 numbers below 49

9. What number is at the 65th percentile? $(.65)(20) = 13$ 82 there are 13 numbers below 82

Using our definition of Percentile:

Could your score be at the 100th percentile?

No, 100% of the scores couldn't be below yours because you are one of the scores. (You can't be below yourself!)

Could your score be at the 0th percentile?

Yes, if your's was the lowest score then no scores are below yours.

You can now finish Hwk #25:

Sec 12-3 Due tomorrow

Pages 664-665

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

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
1. Where the middle 50% lies
2. 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:

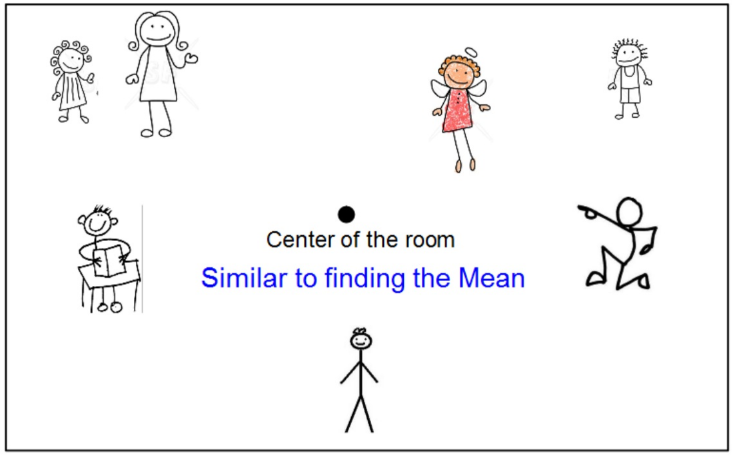
A measure of how much variation there is in a set of data.

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

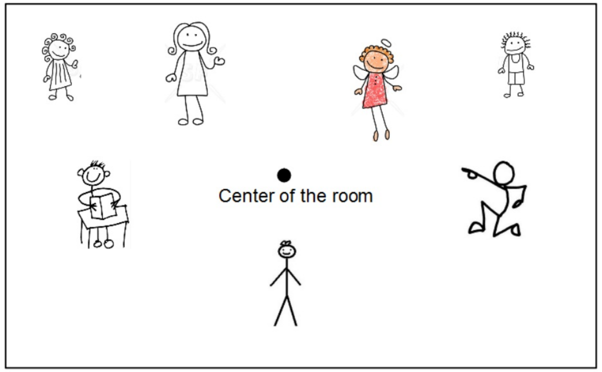
Best used to compare sets of data

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

Bigger Standard Deviation means more variation

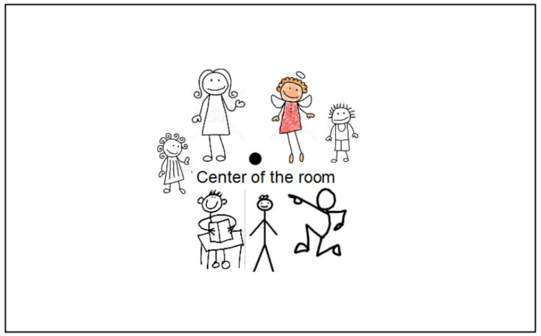


Standard Deviation is similar to the average distance each person is from the center of the room



Large or small Standard Deviation?

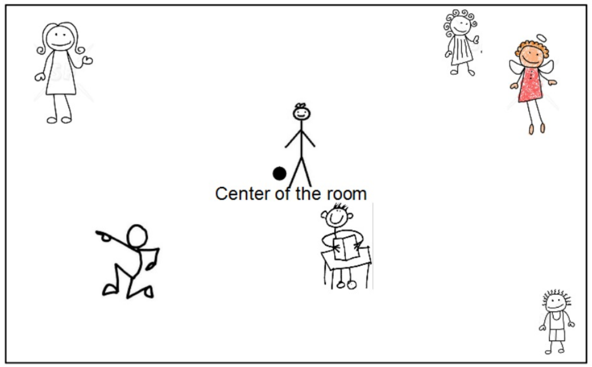
Is there a little or a lot of variation in the data set?



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?



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

Symbol for Standard Deviation: σ Lower case Sigma

Standard Deviation Formula:

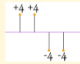
$$\sigma = \sqrt{\frac{\sum (x - \overline{x})^2}{n}}$$

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:

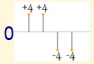
Mean = 0



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

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

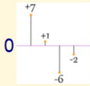
Mean = 0



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

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

Mean = 0

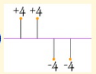


$$\frac{|7| + |1| + |-6| + |-2|}{4} = \frac{7 + 1 + 6 + 2}{4} = 4$$

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

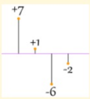
So let us try squaring each difference (and taking the square root at the end):

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:

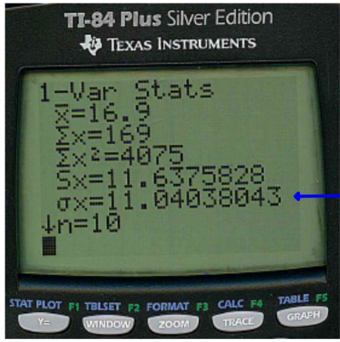
http://www.stat.tamu.edu/~west/ph/stddev.html

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

Find the Standard Deviation.

$\sigma = 11.04$

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:

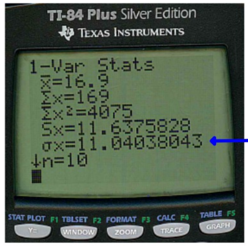
Population Standard Deviation - uses all of the data values

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).

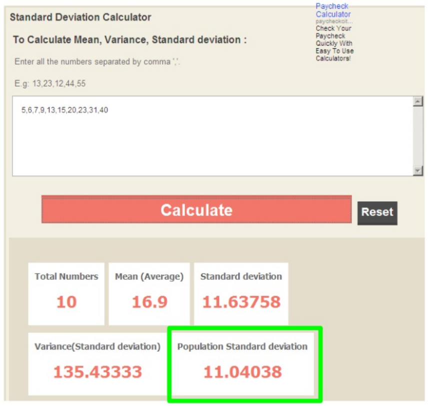
Use the formula on page 669: $\sigma = \sqrt{\dfrac{\Sigma(x-\overline{x})^2}{n}}$

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:

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

$\sigma = 11.4$

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

$\sigma = 12.2$

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

Law of Large Numbers:

The variation in a set of data decreases as the sample size increases.

In general, the larger the data set the smaller the standard deviation.

Z-scores:

The number of Standard Deviations a value is from the mean.

Given the following statistics for a set of data:

$\bar{x} = 12.5$

$\sigma_x = 2.1$

Find the z-score for the data value x=18

$18 - 12.5 = \frac{5.5}{2.1}$

$z = 2.6$

Z-score Formula:

$$z = \frac{x - \bar{x}}{\sigma}$$

x = Data value

$\bar{x}$ = mean

σ = Standard Deviation

Use the following statistics of a set of data:

$\bar{x} = 35.5$

$\sigma = 1.5$

Find the z-score for each data value to the nearest tenth.

a) x = 46

$\frac{46 - 35.5}{1.5}$ $\frac{10.5}{1.5} = 7$

b) x = 39

$z = 2.3$

Quiz scores:

12, 13, 15, 17, 17, 21, 22, 22, 22, 31

$\bar{x} = 19.2$ $\sigma = 5.3$

1. If your quiz was 21 find your z-score. $z = .34$

2. If your quiz was 13 find your z-score. $z = -1.2$

The mean on a test was 82.4 and the standard deviation was 3.6. Find your score on the test if you had a z-score of 1.3

$$\frac{x - 82.4}{3.6} = 1.3$$

$$x = 87.08$$

The standard deviation on a test was 5.2. Your score of 95 gave a z-score of 2.6. Find the mean on the test.

$$\frac{x - \bar{x}}{\sigma} = z$$

$$5.2 \cdot \frac{95 - \bar{x}}{5.2} = 2.6\,(5.2)$$

$$95 - \bar{x} = 13.52$$

-95 $\quad -95$

$$\frac{-\bar{x}}{-1} = \frac{-81.48}{-1}$$

$$\boxed{81.48}$$