Percentile: A number that represents the percent of data below a certain score.

If you took a test and were told that you were in the 80th percentile that means:

- Your score was better than 80% of the other scores
- 80% of the other scores were below yours.

Use this list of numbers.

23, 25, 27, 28, 39, 39, 40, 42, 42, 43, 44, 49, 51, 55, 57

39 is at what percentile? $\frac{4}{15} = 26.7^{th}$ %-tile

there are 4 data values below 39.

55 is at what percentile? $\frac{13}{15} = 86.7^{th}$ %-tile

There are 13 data values below 55

What number is at the 60th percentile? $(.60)(15) = 9$

43 43 has 9 data values below it.

57, 63, 65, 66, 68, 68, 70, 73, 73, 75, 80, 86, 87, 93, 98

Using our definition of Percentile:

Could your score be at the 100th percentile?

No, if you had the highest score then you can't be below yourself.

Could your score be at the 0th percentile?

Yes, it's possible that nobody scored below you

57 is at what percentile? $\longrightarrow$ 0^{th} %-tile

98 is at what percentile? $\longrightarrow$ $\frac{14}{15} = 93.3^{rd}$ %-tile

Hwk #28

Sec 12-3 Due tomorrow.

Pages 664-665

Problems 1, 2, 8-11, 16-18

Measures of Central Tendancy:

- Mean
- Median
- Mode

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

Measures of Variability:

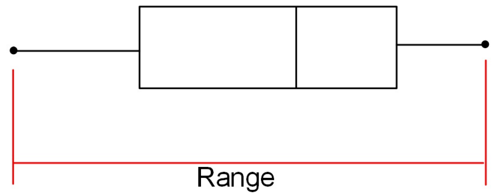
- Range
- Interquartile Range
- Standard Deviation

These give an measure of how spread out the data is and how much variation there is amongst the data

How can you find the Range of a set of data using a Box-and-Whisker plot?

Range = Upper Extreme - Lower Extreme

Max - Min



Range is a measure of the "spread" of the data.

Which tells you more about a data set?

- A large Range or
- A small Range

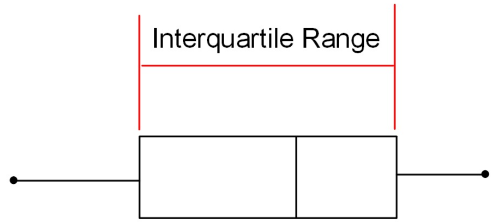
A large Range: Tells you that the Max and Min are far apart but it doesn't give much info about the remaining data values.

A small Range: Tells you that all the data are tightly packed together.

Range only uses two data values.

How can you find the Interquartile Range of a set of data using a Box-and-Whisker plot?

Interquartile Range = Upper Quartile - Lower Quartile



Interquartile Range:

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:

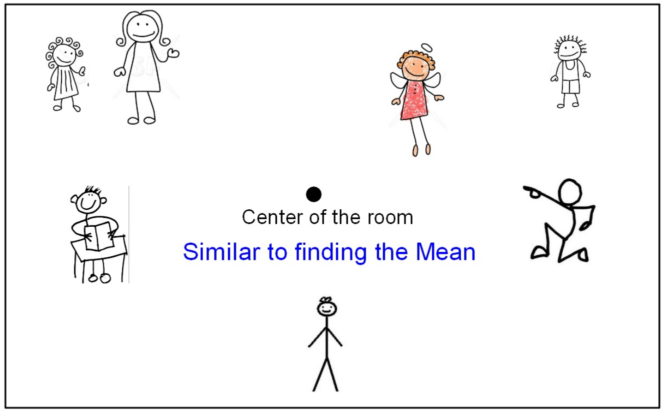
A measure of how much variation there is in a set of data.

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

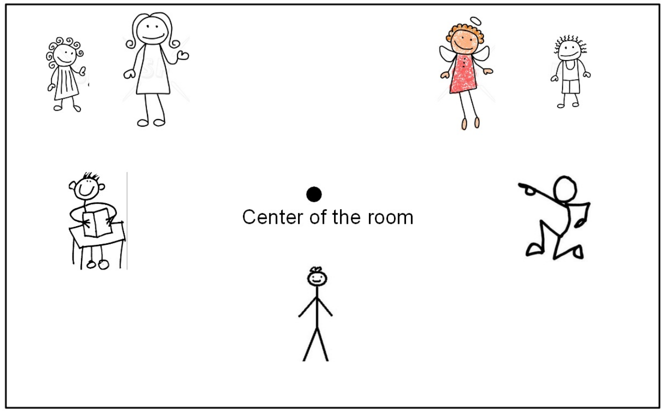
Best used to compare sets of data

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

Bigger Standard Deviation means more variation



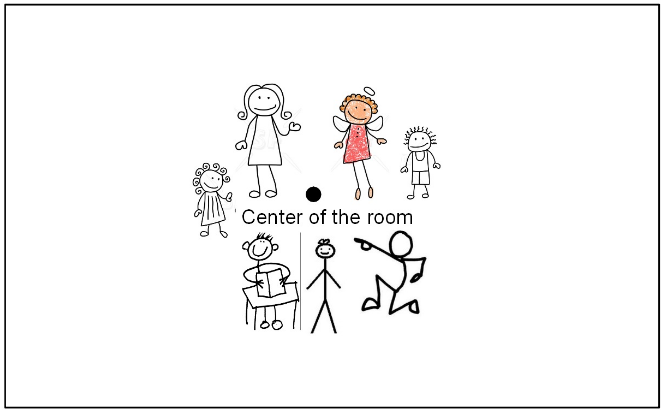
Standard Deviation is similar to the average distance each person is from the center of the room



Large or small Standard Deviation?

Is there a little or a lot of variation in their distances from the center of the room?

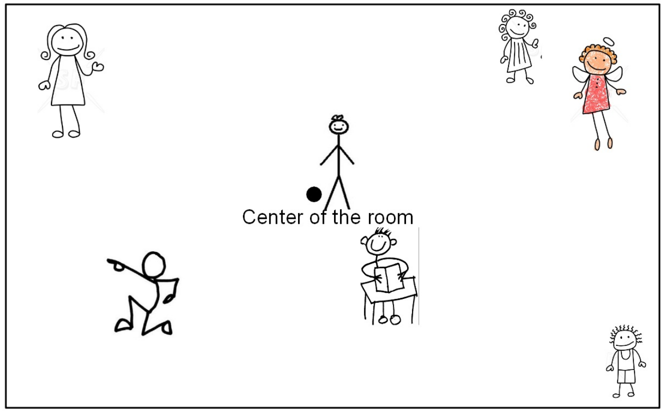
Smaller Standard Deviation



Large or small Standard Deviation?

Is there a little or a lot of variation in their distances from the center of the room?

Larger Standard Deviation



Symbol for Standard Deviation: σ Lower case Sigma

Standard Deviation Formula:

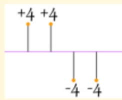
$$\sigma = \sqrt{\frac{\sum (x - \overline{x})^2}{n}}$$

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:

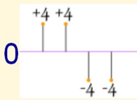
Mean = 0



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

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

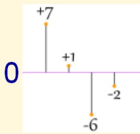
Mean = 0



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

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

Mean = 0

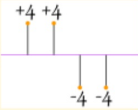


$\frac{|7| + |1| + |-6| + |-2|}{4} = \frac{7 + 1 + 6 + 2}{4} = 4$

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

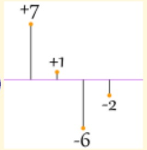
So let us try squaring each difference (and taking the square root at the end):

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!

Use the calculator to find the Standard Deviation:

15, 23, 18, 14, 33, 16, 18, 19, 22, 19

Enter the numbers in a list

Calculater 1: 1-Var Stats

$\sigma_x = 5.18$

Population Std Dev

Using Excel to find Standard Deviation.

	15		
	23		
	18		
	14		
	33		
	16		
	18		
	19		
	22		
	19		
=STDEVP(D3:D12)			

Population Standard Deviation

On Excel you can also find the

Mean =AVERAGE(

Median =MEDIAN(

Mode =MODE(

Use an online standard deviation calculator.

Chose Population Standard Deviation
when given the choice.

Which set of data has more variation?

Find the Standard Deviation for each set.

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 its Standard Deviation is bigger

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

$18 - 12.5 = 5.5$
$\div 2.1$
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:

$\overline{x} = 35.5$

$\sigma = 1.5$

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

a) x = 46

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

b) x = 39

$\frac{39-35.5}{1.5} = 2.3$

c) x = 28

$\frac{28-35.5}{1.5} = -5$

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

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

$x = 87.08$

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.

$5.2(2.6) = \frac{95-\overline{x}}{5.2}(5.2)$

$13.52 = 95-\overline{x}$

$+81.48 = +\overline{x}$