12-1 Descriptive Statistics

Objective:

I can describe a distribution by its shape, outliers, center, and spread.

Vocabulary:

Population: Set of all

Sample: A subset of the population

Parameter: Measures of a population

-Use $\mu = population\ \ mean$

$\sigma =$ *population standard deviation*

U.S.

Statistics: Measures of a sample

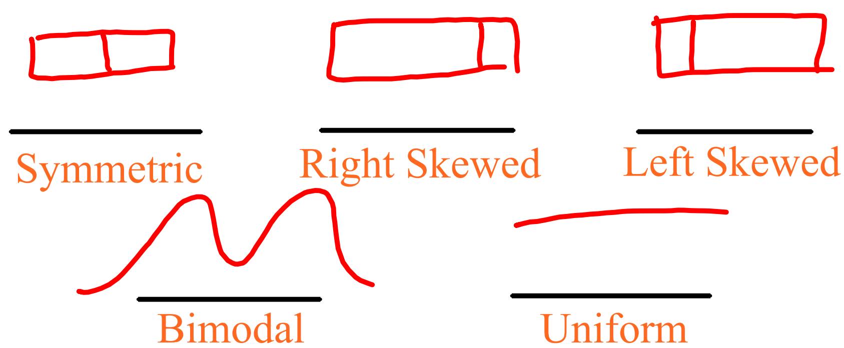
-Use $\bar{x} = sample\ \ mean$

$s =$ *sample standard deviation*

Utah

"Remember your S.O.C.S"

1. Symmetry:



2. Outliers: Data far away from the rest of the data. Formula to come ...

3. Center: Measures of central tendency:

   1. Mean - arithmetic average of the data

   2. Median - Middle value when placed in order, or average of the two middle values

   3. Mode - Most frequently occurring value(s)

4. Standard deviation: Measure of the variability in the data

Mean - Median - Mode ?

The average on the test was an 84 -

The average test score puts you in the middle of the class -

The average American student starts college at 18-

Find the mean, median, and mode for the following set of data:

12, 14, 10, 1, 9, 13, 17, 14, 16

1, 9, 10, 12, 13, 14, 14, 16, 17

$\bar{x} = 11.\bar{7}$

Med = 13

Mode = 14

Is there an outlier for the following set? If so, find the mean, median, and mode without the outlier and describe how it affects the data.

Test scores from a class: 70, 70, 75, 75, 90, 70, 80, 85, 65, 95, 70, 85, 90, 70, 20

With 20

$\bar{x} = 74$

Med = 75

W/o 20

$\bar{x} = 78$

Med = 75

changed by outliers

The salaries of the LA Lakers (who makes more than a million a year) for the 2013-2014 season

Kobe Bryant: $30,453,805
Pau Gasol: $19,285,850
Steve Nash: $9,300,500
Jordan Hill: $3,563,600
Chris Kaman: $3,183,000
Jodie Meeks: $1,550,000
MarShon Brooks: $1,210,080
Nick Young: $1,106,942
Jordan Farmar: $1,106,942
Chris Duhon: $1,500,000

Mean:

~ $7 million

Median:

~ 2.3 million

Mode:

1,106,942

~~**Range:**~~

Why do we have all of these measures?

Example: On a cul-de-sac, you have 5 houses built for:

$200,000, $200,000, $200,000, $200,000, $1,200,000

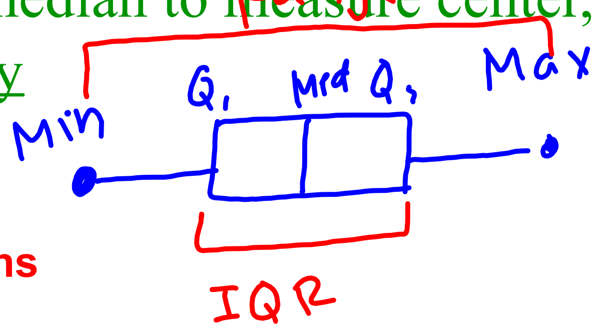
Find the median and the mean? Which one is a better measure?

Mean: 400,000
Med: 200,000

Find the standard deviation: Weights in grams of 30 loon chicks

79.5	87.5	88.5	89.2	91.6	84.5	82.1	82.3	85.1	89.8
84.0	84.8	88.2	88.2	82.9	89.8	89.2	94.1	88.0	91.1
91.8	87.0	87.7	88.0	85.4	94.4	91.3	86.3	85.7	86.0

Spread: When we use the median to measure center, we use <u>5-Number Summary</u>



<u>Range</u> = maximum - minimum

Quartiles split the data into **fourths**

Min = lowest

<u>First Quartile</u> (Q_1) = the median of the lower half of the data

<u>Second Quartile</u> = the median

<u>Third Quartile</u> (Q_3) = the median of the upper half of the data

Max = highest

<u>Interquartile Range</u> (IQR) measures the spread between Q_1 and Q_3

$$IQR = Q_3 - Q_1$$

<u>Five number summary</u> = {minimum, Q_1, median, Q_3, maximum}

Find the five number summary for the male and female life expectancies in South American nations and compare. Then draw its boxplot.

A **box plot** (sometimes called box and whisker plot) is a graph that depicts the five number summary of a data set.

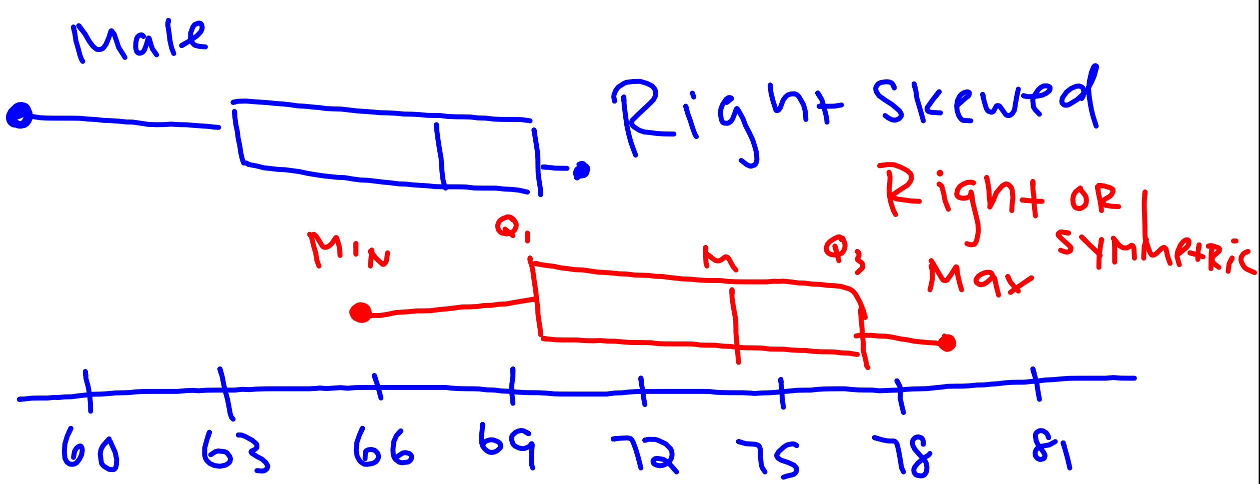
Q1

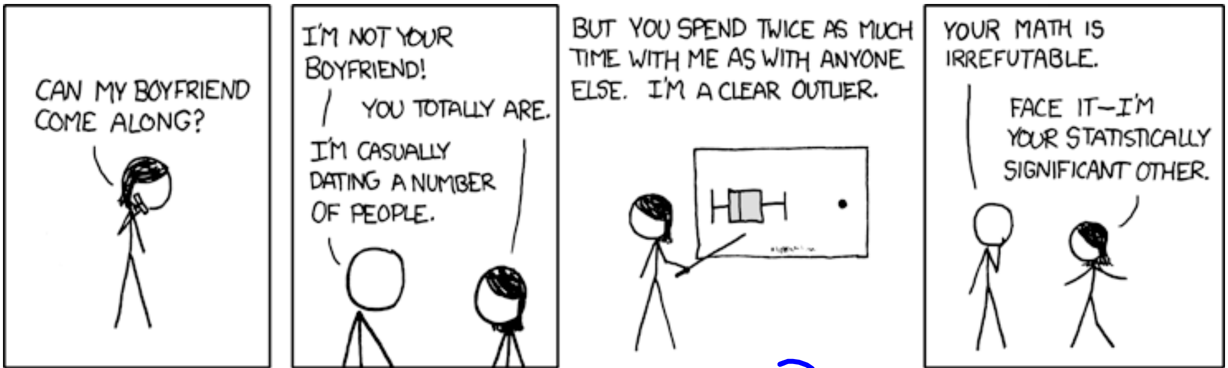
5#: {66.2, 70.25, 74.5, 77.7, 79.4}

females: {66.2, 66.7, 67.7, 72.8, 74.3, 74.4, 74.6, 76.5, 76.6, 78.8, 79.0, 79.4}

males: {59.0, 60.5, 61.5, 66.7, 67.9, 68.5, 69.0, 70.3, 71.4, 71.9, 72.1, 72.6}

5#: {59, 64.1, 68.75, 71.65, 72.6}





OUTLIER?

Box and Whisker plots allow us to get a good visual of outliers: a number that makes one of the whiskers noticeably longer than the box:

RULE OF THUMB: a number is considered an <u>outlier</u> if it is more than 1.5 X IQR below Q_1 or above Q_3

1. Find IQR
2. Multiply by 1.5
3. Add to Q3
4. Subtract from Q1

Outlier Range

Is 61 an outlier in Roger Maris's home run data?

Q_1 Q_3

Five number summary = {5, 11, 19.5, 30.5, 61}

1. Find IQR = 30.5 - 11 = 19.5

2. IQR x 1.5 = 29.25

3. Add to 30.5+29.25 = 59.75

4. Subtract from 11-29.25 = -18.25

61 is an outlier