11-1 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 μ = *population mean*

σ = *population standard deviation*

Statistics: Measures of a sample

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

s = *sample standard deviation*

"Remember your S.O.C.S"

1. S_____:

_____ Symmetric _____ Right Skewed _____ Left Skewed

_____ Bimodal _____ Uniform

2. O_____: Data far away from the rest of the data.

3. C_____: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. S_____: 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, mode and standard deviation for the following set of data:

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

Use your calculator to find the mean, median, mode and standard deviation of the following. Is there an outlier?

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

11-1

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:

Median:

Mode:

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

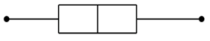
Find the median and the mean? Which one is a better measure?

Spread: When we use the median to measure center, we use 5-Number Summary

Quartiles split the data into **fourths**

Five number summary = {min, Q_1, median, Q_3,max}

Min= Lowest value



First Quartile (Q_1) = the median of the lower half of the data

Second Quartile = the median

Third Quartile (Q_3) = the median of the upper half of the data

Max= Highest Value

Range = maximum - minimum

Interquartile Range (IQR) measures the spread between Q_1 and Q_3

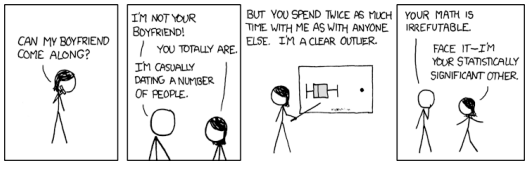
$IQR = Q_3 - Q_1$

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.

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}



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 outlier if it is more than 1.5 X IQR below Q_1 or above Q_3

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

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