MATH 3307
Lesson 2

One question we want to answer about data is about its location, particularly the location of its center.

- Mean –

  Calculated (or arithmetic average). This is the best measure of center, when the data set does not contain outliers.

  $\sum_{all\ i} \frac{x_i}{n}$

  Symbols for mean: $\bar{x}$ *vs.* μ

  $\bar{x}$: "x-bar": sample mean

  μ : "mu": population mean

- Median –

  Center-most value. Located in the middle of your data set. (arrange your data from smallest to largest and then select the middle data point (odd-n) or the mean of the two middle data points (even-n).

- Mode –

  Most often occurring data point. There can be zero, one, two, or three modes.

Examples:

1. Twelve babies spoke for the first time at the following ages (in months):

8 9 10 11 12 13 15 15 18 20 20 26

a. What is the mean of the data?

```
> mean(baby)
[1] 14.75
```

b. What is the median of the data?

```
> median(baby)
[1] 14
```

To Copy: 8,9,10,11,12,13,15,15,18,20,20,26

R Studio Commands:

To assign to a list: assign(“name”,c(1,2,3,4,5))

mean(name)

median(name)

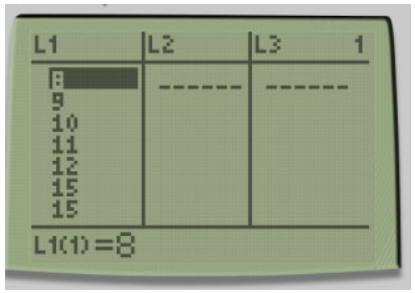
min(name)

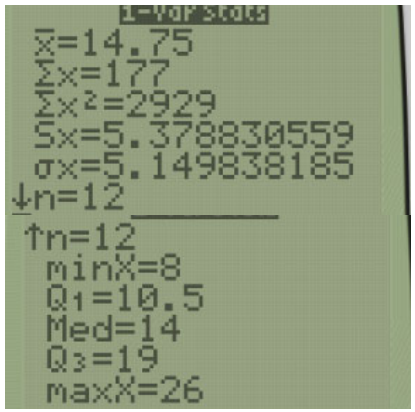
max(name)

sort(name)

length(name)

```
> assign("baby",c(8,9,10,11,12,13,15,15,18,20,20,26))
> baby
 [1]  8  9 10 11 12 13 15 15 18 20 20 26
> baby
 [1]  8  9 10 11 12 13 15 15 18 20 20 26
> mean(baby)
[1] 14.75
> median(baby)
[1] 14
> min(baby)
[1] 8
> max(baby)
[1] 26
> length(baby)
[1] 12
> sort(baby)
 [1]  8  9 10 11 12 13 15 15 18 20 20 26
```





TI-83/84 Calculator Commands

To assign values to a list: STAT → (option 1): Edit

Fill in the values to the L1

(to clear a list, scroll up to highlight list name, press CLEAR, scroll down)

To find measures:

STAT → (right arrow) Calc →

(option 1): 1-Var Stats

List: list name (usually L1)

FreqList: >>Blank<<

$\bar{x}$: Mean
Sx: Standard Deviation
n: size of list
minX: Minimum Value
Q1: First Quartile
Med: Median Value
Q3: Third Quartile
maxX: Maximum Value

2. Here are the weights (in pounds) of 20 steers on an experimental feed diet:

174 142 131 145 175 150 176 151 110 162

133 163 135 178 178 154 166 146 156 167

a. What is the mean of the data?

```
> mean(cow)
[1] 154.6
```

b. What is the median of the data?

```
> median(cow)
[1] 155
```

c. What is the mode of the data? 178

```
> sort(cow)
 [1] 110 131 133 135 142 145 146 150 151 154 156 162 163 166 167 174 175 176 178 178
```

Based on this group of data, the mean and median would be a good representation of the center.
The mode, since it is on one end of the data set, would not be.

To copy:
174,142,131,145,175,150,176,151,110,162,133,163,135,178,178,154,166,146,156,167

3. The test scores of a class of 20 students have a mean of 71.6 and the test scores of another class of 14 students have a mean of 78.4. Find the mean of the combined group.

This is a weighted average. (one class average carries more weight that the other)

$$\frac{(\bar{x}_1)(n_1) + (\bar{x}_2)(n_2)}{n_1 + n_2}$$

```
(71.6*20+78.4*14)/(20+14)
L] 74.4
```

Whenever your numerator and/or denominator have more than one term added or subtracted, enclose the entire numerator/denominator in parenthesis.

4. Explain why the conclusion drawn is not valid:

A businesswoman calculates that the median cost of the five business trips that she took in a month is $600 and concludes that the total cost must have been $3000.

Thought behind this reasoning: middle value is $600, there are 5 trips, 600*5 = 3000

Issue with this reasoning: the median is not affected by outliers in your data set. Meaning, if there was an outlier in these five trips, this calculation is no longer accurate.

Example 1: 400 500 600 700 800
These five data points have a median of 600, and they have a total of: 3000

Example 2: 400 500 600 700 1200
These five data points still have a median of 600, but now they have a total of: 3400

Since both data sets have the same median but different totals, this is not a valid method of reasoning.