Section 1.2
Measures of Center

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

Measure of center include: Mean, median and mode

Mean is denoted with the Greek letter μ when referring to the population mean, and with the symbol $\bar{x}$ (read "*x*-bar") when referring to the sample mean.

In any case, we find the mean by adding up all the values and dividing by the number of values.

Mean Formula

$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{x_1 + x_2 + ... + x_n}{n}$, where $x_1, x_2, ..., x_n$ are values and *n* is the number of values.

The **median** is found by putting all values in order from least to greatest and finding the middle value. If two middle data values exist, add those up and divide by 2.

We will be using a free program called R to do a lot of different type of problems in this course. To download R, you should watch the following video at:
http://www.math.uh.edu/~bekki/RStudioInstall/RStudioInstall.html

It has many commands, some of which we'll learn throughout the course. We will also be able to use it as a regular calculator.

Example 1: Suppose your sample data set consists of the values: 24, 20, 22, 20, 21, and 19. Find the mean and the median.

This data set is not too large so we can either do it using R as a calculator (to calculate the mean) or use its commands to find both the mean and median. Let's use its commands, so we'll first need to enter the data set as follows:

- To enter a data set in R, use the command: c( )
- The cursor will then appear inside the parenthesis and you'll enter the data set, separating each number with a comma. Lastly, hit enter.
- Many times we will want to give a name to the data set, so the command would be: "you pick a name for the data set"=c( )
- Then enter the data set as described above.

- Either way you choose to enter the data, if you made a mistake entering the data, to correct without entering the whole set of data again, hit the up arrow on the keypad to recall any previously entered commands.

- To find the mean of a data set, use the command: mean("name")
- To find the median of a data set, use the command: median("name")
- If you need to see the sorted data, use the command: sort("name")

Now, let's write out the commands to calculate the mean and median.

Commands: Results:

Both the mean and the median can be used to describe the center of a data set. It is up to the statistician to decide which would give the best description of center. Median is resistant to extreme values (outliers) in the data set. Mean is NOT robust against extreme values. The mean is pulled away from the center of the distribution toward any extreme values.

> *For example: We have 100 employees at a company whose salaries are $40,000 each. One other employee's salary is $1,000,000. The mean is:*
>
> $$\frac{100*40000+1000000}{101}=\$49,504.95$$
>
> *Is this a good measure of the center of this data?*
>
> *No! The median would be the best descriptor. In fact, when there are these types of outliers the median is the best way to describe the center.*

The **mode** is the value that occurs most often. It is used as a description of center for categorical data. When working with categorical data, mean and median do not make sense to use to describe the data set. The mode can be used to describe categorical data as well as quantitative data.

In R, the mode function simply indicates whether the data set is numeric or categorical, so we'll use the sort command to find the mode of a data set.

Example 2: Ten employees were asked how many cups of coffee they drank during the work day. The values reported were 6, 3, 1, 2, 3, 1, 2, 3, 2, and 0. What is the mode of the data?

Commands:
cups=c(6,3,1,2,3,1,2,3,2,0)
sort(cups)

Result of the sort:
0 1 1 2 2 2 3 3 3 6

The mode is:

Example 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.

Using R as a calculator, we'll find the mean of the combined group.

Command: The mean is:

We can use the same idea used in Example 3 to determine whether certain statements are valid or invalid. For example: A report states that at least 98% of the company's employees earn less than the mean salary. Is this report valid?

Answer: Yes. For example, let's say 75 employees are paid \$45,000 per year and an executive is paid \$800,000 per year.

$$\frac{75*45000+800000}{76}=\$54,934.21$$

And $75/76 \approx 98.7\%$ which states that more than 98% of the company's employees earn less than the mean salary.

Example 4: 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

Find the mean, median and mode for the data.
Commands: Results: