**Statistics Basics**

The most common basic statistics terms you’ll come across are the mean, mode and median. These are all what are known as “Measures of Central Tendency.” Also important in this early chapter of statistics is the shape of a distribution. This tells us something about how data is spread out around the mean or median. Perhaps the most common distribution you’ll see is the normal distribution, sometimes called a bell curve. Heights, weights, and many other things found in nature tend to be shaped like this:



**Example of a Normal Distribution**

A normal distribution is an arrangement of a data set in which most values cluster in the middle of the range and the rest taper off symmetrically toward either extreme.

Height is one simple example of something that follows a normal distribution pattern: Most people are of average height, the numbers of people that are taller and shorter than average are fairly equal and a very small (and still roughly equivalent) number of people are either extremely tall or extremely short.

****

****

**Measure of Central Tendency**

****

****

**Mean, median, Mode and Range**

Mean, median, and Mode are three kinds of "**averages**". The "**mean**" is the "**average**" you're used to, where you add up all the numbers and then divide by the number of numbers. The "**median**" is the "**middle**" value in the list of numbers. To find the median, your numbers have to be listed in numerical order from smallest to largest, so you may have to rewrite your list before you can find the median. The "**mode**" is the value that occurs most often. If no number in the list is repeated, then there is no mode for the list. The "**range**" of a list a numbers is just the difference between the largest and smallest values.

**Exercise**

Find the mean, median, mode, and range for the following list of values:
13, 18, 13, 14, 13, 16, 14, 21, 13

The mean is the usual average, so I'll add and then divide:
(13 + 18 + 13 + 14 + 13 + 16 + 14 + 21 + 13) ÷ 9 = 15

Note that the mean, in this case, isn't a value from the original list. This is a common result. You should not assume that your mean will be one of your original numbers.

The median is the middle value, so first I'll have to rewrite the list in numerical order:
13, 13, 13, 13, 14, 14, 16, 18, 21

There are nine numbers in the list, so the middle one will be the (9 + 1) ÷ 2 = 10 ÷ 2 = 5th number:
13, 13, 13, 13, 14, 14, 16, 18, 21

So the median is 14.
The mode is the number that is repeated more often than any other, so 13 is the mode.
The largest value in the list is 21, and the smallest is 13, so the range is 21 – 13 = 8.

**Mean**: 15
 **Median**: 14
 **Mode**: 13
 **Range**: 8

**What is an Average?**
The word “average” is used in everyday life to describe where the middle number of a data set is. It’s the typical number you would expect to find in a series of numbers. In statistics, the average is called the “arithmetic mean,” usually just shortened to the mean. Both the average and the mean use the same formula:

**Examples**

**Example 1:** You earned €129, €139, €155 and €176 over the last 4 weeks. What is your average pay?

Step 1: Add up all of the numbers in the set. €129 + €139 + €155 + €176 = €599.

Step 2: Divide Step 1 by the total number of items in the set. There are 4 items in the set, so €599 / 4 = €149.75.

**Example 2:** You have semester grades of B, C, D, A, B and B. What is your average grade?

Step 1: Add up all of the numbers in the set. You have grades here, so you need to convert them on a 4.0 scale:
B = 3.0
C = 2.0
D = 1.0
A = 4.0
B = 3.0
B = 3.0
So we have: 3.0 + 2.0 + 1.0 + 4.0 + 3.0 + 3.0 = 16.0.

Step 2: Divide Step 1 by the total number of items in the set. There are 6 items in the set, so 16.0/6 = 2.66.

**Bias in Statistics**

Bias is the tendency of a statistic to overestimate or underestimate a parameter. To understand the difference between a statistic and a parameter, see this article. Bias can seep into your results for a slew of reasons including sampling or measurement errors, or unrepresentative samples.

Sampling error is the tendency for a statistic not to exactly match the population. Error doesn’t necessarily mean that a mistake was made in your sampling; Sampling Variability could be a more accurate name. For example, let’s say you have a population in the United States with an average height of 5 feet 9 inches. If you take a sample, even a fairly sizable sample of say, 10,000 people, it’s unlikely that you’ll get exactly 5 feet 9 inches. You might get very close, perhaps to within a fraction of an inch. If you repeat the experiment, you might get another very close result. For example, in experiment 1 you might get 5 feet 8.9 inches and in experiment 2 you might get 5 feet 9.1 inches. The tendency for statistics to get very close, but not exactly right, is called sampling error. Note: If the statistic is unbiased, the average of all statistics from all samples will average the true population parameter.

**Standard Deviation**

Standard Deviation a quantity expressing by how much the members of a group differ from the mean value for the group.

****

**Formula for calculating Standard Deviation from a grouped frequency table.**



**Variance**

The sample variance, s2, is used to calculate how varied a sample is. A sample is a select number of items taken from a population. For example, if you are measuring American people’s weights, it wouldn’t be feasible (from either a time or a monetary standpoint) for you to measure the weights of every person in the population. The solution is to take a sample of the population, say 1000 people, and use that sample size to estimate the actual weights of the whole population. The variance helps you to figure out how spread out your weights are.



Grouped frequency table

