Statistics

What are statistics?

Statistics are numbers that describe a set of data. A set of data is any set of numbers that describe something. They can be grades on school tests, temperatures in Seattle throughout the year or numbers that come up on dice that you throw.

For example, let's say we rolled 2 dice 11 times and we wrote down their totals. Here they are:

6 4 10 8 8 7 2 3 9 8 12

Minimum, Maximum & Range

The **maximum** is the largest number. In our set of data, the maximum is 12.

The **minimum** is the smallest number. In our set of data, the minimum is 2.

The **range** of a set of data is the maximum (highest) minus the minimum (lowest) number.
In our set of data, the range is 12 - 2 = 10.

Mean

When people talk about finding the average of a set of numbers, they are usually talking about finding the mean. To find the mean, you add all of the numbers in the set and divide by how many numbers are in the set.

The average (mean) of the 11 numbers in our set is:

$\frac{6 + 4 + 10 + 8 + 8 + 7 + 2 + 3 + 9 + 8 + 12}{11}$ = $\frac{77}{11}$ = 7

Mode

The mode is the most common number in the set. It is the one that occurs most often. In our set, the mode is 8 because there are 3 of them and only one each of the other numbers.

6 4 10 **8 8** 7 2 3 9 **8** 12

It is possible for a set of data values to have more than one mode. If there are two data values that occur most frequently, we say that the set of data values is **bimodal**. If there is no data value or data values that occur most frequently, we say that the set of data values has no mode.

Median

The median is the number in the middle. You can find the middle number by crossing out the lowest number, then the highest, then the next lowest, then the next highest and so on until you have only one number left. That last number is the *median*.

An easier way is to reorganize the numbers in your set from smallest to largest, then count in from both sides to find the number in the middle. If you have two numbers in the middle (because you have an even number of numbers in your set), the median is the average of those two numbers.

Let’s reorganize our numbers to find the median:

~~2~~ ~~3~~ ~~4~~ ~~6~~ ~~7~~ **8** ~~8~~ ~~8~~ ~~9~~ ~~10~~ ~~12~~

The median number is 8.

But Which One Do I Use?

The mean, median and mode of a data set are collectively known as **measures of central tendency** because they focus on where the data is centered or clustered.  To analyze data using the mean, median and mode, we need to use the most appropriate measure of central tendency.  In general:

* The mean is useful for predicting future results when there are no extreme values in the data set.

For example, on his last 8 math quizzes, Joe received scores of 85, 95, 92, 88, 99, 95, 86 and 96. His average (mean) for the past 8 quizzes is 92.

* The median may be more useful than the mean when there are extreme values in the data set as it is not affected by the extreme values.

Let’s look at the scores from Joe’s last 9 math quizzes. He wasn’t feeling well for the first one, and only scored 20. If we take the mean (average) of the 9 scores (20, 85, 95, 92, 88, 99, 95, 86 and 96), our result is 84. If we take the median, our result is 92. Which result more accurately predicts how well Joe will score on future tests?

* The mode is useful when the most common item, characteristic or value of a data set is required. We can even find the mode of a set that doesn’t contain numbers.

Let’s look at Joe’s sock drawer to try to figure out what his favorite type of sock looks like. He has 12 white socks with purple polka dots, 4 blue socks with yellow stripes, 10 green and orange argyle socks and 1 black sock. The mode of Joe’s sock drawer (or most popular sock) is “white sock with purple polka dots”.