

We have 3 ways of measuring the center of a distribution:

1. The mean or average of our data
2. The median
3. The mode

Lets look at these three a bit more closely.

### **The Mean:**

The mean or the average of our data set is best used as a measure of center if our data is approximately symmetric and doesn't contain outliers.

The formula for the mean is  $\frac{\sum_{i=1}^n x_i}{n}$ . This formula tells us to add up all the values in our dataset and then divide by the number of values we added.

The mean is very sensitive to outliers. The mean will always get pulled in the direction of the largest outlier

### **The Median:**

The median is best used as a measure of center when outliers are present in the data. This is because the median will not be affected by extremely small or large observations.

The median is the data points where 50% of the observations are above and below that datapoint. To find the position of the median from a dataset with n observations we have to consider two separate cases:

1. The dataset has an odd number of observations:
  - a. If this is the case then we find the position of the median by taking  $\frac{n+1}{2}$
2. The dataset has an even number of observations:
  - a. In this case we find the middle two observations and average them.

Example of the median:

Take this data set as an example:

12	6	7	14	8	18	2	1	2	22	15
----	---	---	----	---	----	---	---	---	----	----

The first step we need to take for finding the median is to first order our dataset:

1	2	2	6	7	8	12	14	15	18	22
---	---	---	---	---	---	----	----	----	----	----

Now since n=11 in this case we can find the median by taking  $n+1/2$ , which in this case

becomes 6. So we count to the 6th observation in our dataset which happens to be the number 8.

1	2	2	6	7	8	12	14	15	18	22
---	---	---	---	---	---	----	----	----	----	----

If instead we had 10 observations we would average the two values in the middle

1	2	2	6	7	8	12	14	15	18
---	---	---	---	---	---	----	----	----	----

This would make our median 7.5.

### **The mode:**

This measure of center is best used when analyzing categorical datasets. The number is the number, range of numbers, or category that occurs the most frequently.

The mode is also very resistant to outliers since it relies on which observation occurs the most and not the actual value of the observation.

