## Are You Outlying?

Outliers affect how we view data. They can seriously skew data and make us misinterpret conclusions. Let's take a summer camp for children as an example.

On the y-axis, the manager of the summer camp will list the frequency. The frequency is the number of times a student of a certain age attends the summer camp. On the x-axis, the manager will list the age of the children. Let's say the first distribution looks like this.
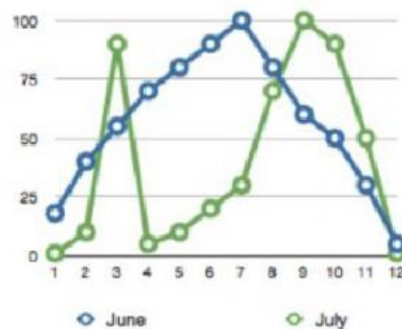
Age of Children at Summer Camp

June's data looks smooth with the center at approximately 6 or 7. This signifies that the mean and the median are somewhere between 6 and 7.

The spread of the data is consistent with a significant number of children on either side of the center. There isn't a high degree of cluster around the center meaning that many children of different ages attend the camp. This isn't a specialty camp for only elementary school aged children nor is it primarily for toddlers.

July's data is quite different. It appears like there is a shift to older age groups with the center at approximately 9 or 10 with a tighter spread. This means that most children at the camp are now in the upper elementary school years.

However, there is a large amount of 3 year olds that seemingly comes from nowhere! All of the sudden ~90 3 year olds show up at the camp in July!

Are You Outlying?

This data point is an example of an outlier. Outliers are significant because they affect both the mean and standard deviation. Although most children at the camp are on the older side of the spectrum, 8 years old to 11 years old, the high number of 3 year olds will skew this number and make the mean appear to be much smaller than would be practical to interpret.

For example the mean might show the average age is 7 years of age, and the manager might schedule activities for mostly 7 year olds when in fact most students are between 8 and 11. You can see why outliers make a huge difference to data!
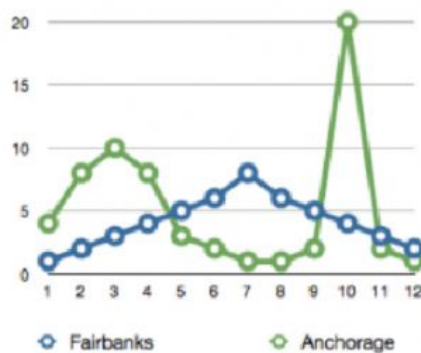
Are You Outlying?

Directions:

Statistics help us understand data and make better decisions. The mean and median both describe the center of the data. The standard deviation and interquartile range describe the spread of the data. Outliers skew the standard deviation and mean and can be seen by their drastic impact on the graphs.

1.) Cities continually measure the populations in order to prepare to provide these new citizens services and collect tax revenue. The data below shows the number of people moving to Fairbanks and Anchorage Alaska in the year 2013.

Below is the data distribution for the months of January through December.

Which month appears to be the outlier for which city? Why? Use a complete sentence

_____

_____

_____

Are You Outlying?

2.) In the problem above, which city appears to have a center of spring? Why? Use a complete sentence.

_____

_____

_____

3.) How would the outlier negatively affect the city planners as they plan to build services for people moving to their city? Refer to center and spread in your answer. Use 2 complete sentences.

_____

_____

_____

_____

_____