Measures of Central Tendency & Variation

Finding the Measures of Central Tendency

Example 1: Find the mean median, and mode for the data. Deer at a feeder each hour: 3, 0, 2, 0, 1, 2, 4	Now you Try: Example 2: Find the mean, median, and mode of the data set. $\{6, 9, 3, 8\}$

A box-and-whisker plot shows the spread of data set. It displays the 5 key points; the ______________ and ______________ values, the ______________, and the __________ and _____________ ______________.

The quartiles are the medians of the lower and upper halves of the data set. If there are an odd number of data, do not include the median in either half.

The interquartile range, or IQR, is the difference between the 1^{st} and 3^{rd} quartiles, or Q3-Q1. It represents the middle 50% of the data.

Example 3: Make a box-and-whisker plot of the data. Find the interquartile range. $\{6, 8, 7, 5, 10, 6, 9, 8, 4\}$	Example 4: Make a box-and-whisker plot of the data. Find the interquartile range. $\{13, 14, 18, 13, 12, 17, 15, 12, 13, 19, 11, 14, 14, 18, 22, 23\}$

The data sets $\{19, 20, 21\}$ and $\{0, 20, 40\}$ have the same mean and median, but the sets are very different. The way the data are spread out from the mean or median is important to the study of statistics.

A measure of variation is a value that describes the spread of a data set. The most commonly used measures of variation are the range, the interquartile range, the variance, and the standard deviation.

Symbols commonly used to represent **mean** is _____ or ____ ______. The symbol for **standard deviation** is the lowercase Greek letter ___________, _____. The **variance**, denoted by _______, is the average of the squared differences from the mean. **Standard deviation**, denoted by _____, is the square root of the variance and is one of the most common and useful measures of variation.

Low standard deviation indicate data that are clustered near the measures of central tendency, whereas high standard deviations indicate data that are spread out form the center.

Finding Variance and Standard Deviation
Step 1: Find the mean of the data $\bar{X}$.
Step 2: Find the difference between the mean and each data value and square it.
Step 3: Find the variance, σ^2, by adding the squares of all of the differences from the mean and dividing by the number of data values.
Step 4: Find the standard deviation, σ , by taking the square root of the variance.

Standard Deviation

Example 5: Find the mean and standard deviation for the data set of the number of people getting on and off a bus for several stops. $\{6, 8, 7, 5, 10, 6, 9, 8, 4\}$	Example 6: Find the mean and standard deviation for the data set of the number of elevator stops for several rides. $\{0, 3, 1, 1, 0, 5, 1, 0, 3, 0\}$

An **outlier** is an extreme value that is much less than or much greater than the other data values. Outliers have a strong effect on the mean and standard deviation. If an outlier is the result of measurement error or represents data from the wrong population it is usually removed.

There are different ways to determine whether a value is an outlier. One is to look for data values that are more than 3 standard deviations from the mean.

Example 7: in the 2003-2004 American League Championship Series, the New Your Yankees scored the following numbers of runs against the Boston Red Sox: 2, 6, 4, 2, 4, 6, 6, 10, 3, 19, 4, 4, 2, 3. Identify the outlier, and describe how if affects the mean and standard deviation.

Try It: Use the data set for 1-3
$\{9, 4, 7, 8, 5, 8, 24, 5\}$

1. Make a box-and-whisker plot of the data. Find the interquartile range.

2. Find the variance and the standard deviation of the data set.

3. Are there any outliers in the date set?