**Worksheet : Measures of Central Tendency**

We use statistical terms to describe something about a set of data points. With a specific data set, it is often important to know the measurements around which the observations tend to cluster. Three measures of the "center" of the data are the mean, the median, and the mode. For Gaussian distribution, the measures of central tendency are the same value. In other words, the mean=median=mode.

**Mean** ( $\overbar{x}$ )

The mean, also called the arithmetic mean or the average, is the sum of all the data points divided by the number of points. The average is the most common way of calculating central tendency.

Example: For the data set containing 7 numbers {2, 5, 9, 3, 5, 7, 4}, the mean is calculated as:

2+5+9+3+5+7+4 = 35/7 = 5 is the mean

7

The properties of the mean are:

* easy to calculate
* only one exists for any data set
* affected by all observations, and strongly affected by outliers

**Median**

The median of a data set is the value of the middle point, when they are arranged in order.

Using the previous data set and arranging from lowest to highest {2, 3, 4, 5, 5, 7, 9}, we can determine the median by crossing off the lowest and highest values, then the next lowest and next highest value. Continue crossing off values from both ends until only one value, the middle value, remains {2, 3, 4, 5, 5, 7, 9}. For this data set, the median is 5.

If there is an even number of points, average the two middle values.

Example: For the following data set containing 6 numbers, {2, 3, 4, 5, 7, 9}, we can determine the median as follows: 2, 3, 4, 5, 7, 9. For this data set, two numbers, 4 and 5, lie at the center. To determine the median for this data set, we would take the average of 4 and 5 as follows:

4+5 = 9/2 = 4.5. The median for this data set is 4.5.

2

The properties of the median are:

* always exists for a set of data
* unique
* not strongly affected by extreme values
* corresponds to the 50th percentile

**Mode**

The mode is the value that occurs most frequently in a data set. There can be more than one mode, if there are two or more measurements that are tied for occurring most frequently. In cases where two numbers occur most frequently, the distribution of data would then be classified as bimodal (having two modes).

For the data set, {2, 5, 9, 3, 5, 7, 4}, all numbers occur only once except the number 5; it occurs twice or more frequently than the other numbers. Therefore, the mode for this data set is 5.

The properties of the mode are:

* requires no calculation
* not necessarily unique
* very insensitive to extreme values
* may not be close to the center of the distribution

**Practice Calculations**

Calculate the mean, median, and mode for the following data sets:

 **Mean Median Mode**

Data Set #1 {2, 2, 2, 2, 42, 2, 2, 2, 2, 2} \_\_\_\_\_ \_\_\_\_\_\_ \_\_\_\_\_

**Statistics are meaningful only in context** – the value of 42 may or may NOT be an outlier

Data Set #2 {9, 2, 3, 4, 11, 5, 8, 6, 7, 5} \_\_\_\_\_ \_\_\_\_\_\_\_ \_\_\_\_\_

Data Set #3 {6, 6, 6, 6, 6, 6, 6, 6, 6, 6} \_\_\_\_\_ \_\_\_\_\_\_\_ \_\_\_\_\_

You randomly select 20 sodium specimens submitted to you laboratory to track turn-around time for the day. The results, in minutes, are as follows:

 {45 48 41 49 102 44 43 141 44 46 43 43 45 49 41 42 40 43 48 43}

What is the mean for these 20 samples? \_\_\_\_\_\_\_\_\_

What is the median for these 20 samples? \_\_\_\_\_\_\_\_\_

What is the mode for these 20 samples? \_\_\_\_\_\_\_\_\_\_

Bring this worksheet with you to the workshop. We will review the *practice calculations* together as a class.

Please note, even though Data Set #3’s mean = median = mode, it is not a Gaussian distribution. In class, we will discover other qualities that must be present for a distribution to be Gaussian.