# BASIC STATISTICS for SPSS

Prepared by R. Hassad, MPH, PhD (October, 2008)

**Statistics** is a structured and logical process of collecting, organizing, analyzing, interpreting and presenting data, based on specific objectives, for the purpose of decision-making.

WHEN we interpret statistics, we must not focus on the numerical values only, but more importantly on the **CONTEXT** of the data, in particular, what the data represent.

There are basically two classifications of statistical methods, and these are:

(1) Descriptive Statistics – which provide basic information about characteristics of the sample,
(2) Inferential Statistics – which allow us to estimate measurements about the population using information from the sample, in other words, we can generalize from the sample to the population using inferential statistical techniques.

**The rest of these notes will address DESCRIPTIVE STATISTICS.**

When we perform descriptive statistics, we must comment on the CENTER, SHAPE, and SPREAD of the distribution of the variable(s) being analyzed.

**CENTER** refers to measures of central tendency, and these are:

(1) MEAN (the typical value or score)
(2) MEDIAN (at least 50% of the sample will have this value or higher/lower)
(3) MODE (the value with the highest frequency)

**Note** that all three measures are called averages, but each tells us something different about the distribution, and when used together, we can get a more comprehensive understanding of the data, for decision-making.

However, the measures of central tendency alone are NOT adequate to describe a distribution. We MUST also comment on the **SPREAD**.

NOTE that if the distribution is skewed (such as when outliers are present), then the MEAN becomes unreliable, and in this case the MEDIAN is the preferred measure of central tendency.

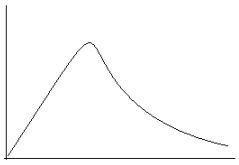
**SPREAD** refers to measures of dispersion, and these are:

(1) STANDARD DEVIATION (the average deviation of the values from the mean)
(2) VARIANCE (the square of the standard deviation)
(3) RANGE (the difference between the lowest and highest values)

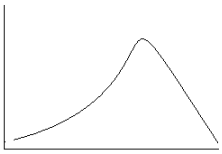
**NOTE** that measures of central tendency and measures of dispersion are numerical values, which may not give us the true picture of our data, therefore, in addition to these, we MUST examine the **SHAPE** of the distribution using an appropriate graph such as a line graph, bar chart, histogram, box plot, or stem and leaf plot Some common shapes are:

1. NORMAL or BELL-SHAPED: This shape has one peak or mode (and hence is also called unimodal). It is symmetric and balanced, and the MEAN, MEDIAN and MODE are EQUAL. Such a distribution can indicate that the majority of the values are close to the mean, and in general, the sample performed alike (that is, a homogeneous distribution).

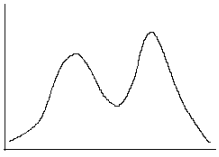
2. POSITIVELY-SKEWED (NON-NORMAL): The right tail of the curve/distribution is extended indicating the presence of outliers. **An outlier** is an extreme value that does not fit the body of the data/distribution. An extreme value can be either higher or lower than the mean. In this case the outlier(s) is/are higher than the mean, the reason the curve is skewed to the right (the side with the higher values)



3. NEGATIVELY-SKEWED (NON-NORMAL): The left tail of the curve/distribution is extended indicating the presence of outliers. An outlier is an extreme value that does not fit the body of the data/distribution. An extreme value can be either higher or lower than the mean. In this case the outlier(s) is/are lower than the mean, the reason the curve is skewed to the left (the side with the lower values)



4. BIMODAL (NON-NORMAL): Note that there are two peaks or modes, hence bimodal, and this indicates that there are two different subgroups within the distribution. It is often important to note this pattern, as each subgroup may have to be treated differently.



**NOTE:** If there are three or more peaks, and each peak represents a meaningful subgroup, then the curve/distribution is referred to as MULTIMODAL.

## WHAT'S NEXT?

In the next document (course notes), I will use this information on statistics to show you how to interpret the data on "number of sex partners", and write a brief practical report.