INTRODUCTION TO STATISTICS

COURSE NOTES

Prepared by R. Hassad, , PhD, MPH, CStat

- Statistics is a set of tools for summarizing and making sense of data
- Statistics is about measurement (hence it is quantitative or numerical)
- Statistics can be viewed as data management
- And in order to effectively manage data, we must know:
  - ✓ **What it is**
  - ✓ **How it is represented**
  - ✓ **What are the types of data**
  - ✓ **Why they are different, and**
  - ✓ **How they are used for decision-making**

<u>Note</u> that increasingly the word "data" is being used with a plural verb: data are, data were, etc. You may say an item or element of data is -------. For improved readability I have used either is/are for these notes.

Data is a collection of variables (about a particular domain or area of interest), which broadly speaking, can be either NUMERICAL (quantitative) or CATEGORICAL (qualitative). We can say that a variable is the smallest unit of data.

What exactly is a variable? A variable is any characteristic, quality or factor that can change. Knowing the specific type of variable/data that we are working with allows us to perform and generate appropriate statistical analyses and graphical formats.

For practical purposes there are two types of numerical variables/data:

1. Continuous – which we get from measuring, and the values can have a fraction of a unit, for example; a body temperature of 101.6 degrees Fahrenheit.
2. Discrete - which we get from counting, and the values are whole numbers only, for example; the number of students in this class is 22.

Also, for practical purposes, there are two types of categorical (or qualitative) variables/data:

1. Nominal – which refers to unordered categories, for example; gender, ethnicity, and marital status. The categories of these variables do not indicate or represent that one is higher or lower than the other. The category is simply a name (note that Nominal

comes from the the Latin word "nomen" which means "name").

2. Ordinal – which refers to ordered categories, for example; ranks in the military, severity of diseases (mild, moderate, severe), height (short, medium, tall), and stages of college life (freshman, sophomore, junior, senior). The categories of these variables have a particular rank or order, and hence one can be higher or lower than the other (unlike nominal variables/data).

Note that categorical variables/data are sometimes referred to as discrete, given that the categories are separate and generally different from the others, and this is what the word discrete means.

IMPORTANT: Types of data are also referred to as scales of measurement, and are classified as follows:

Scale	Type of Data
Ratio	Quantitative/Numerical
Interval	
Ordinal	Qualitative/Categorical
Nominal	

NOTE that for practical purposes, ratio and interval-scaled data can be classified as numerical.

NOTE: There is a key difference between variables with respect to what they represent, and how they are measured.

Some characteristics are basic, objective, and easy to measure, for example; age, weight, height, ethnicity, and marital status. Each of these is UNIDIMESNIONAL (has one dimension or aspect), and if we want to ascertain any of these from our participants, a single (that is, one) question or item will okay, for example: What was your age last birthday? What is your gender?

On the other hand, characteristics such as, IQ (intelligence quotient), self-esteem, depression, personality, and anxiety are MULTIDIMENSIONAL (have more than one dimensions or aspects) and complex, and hence more challenging and difficult to measure. Such characteristics or factors are referred to as CONSTRUCTS, because they are built from, or are composed of many variables or items.

In order to measure a construct we must use a scale or instrument that is reliable and valid.

What is a scale: A scale is a set of inter-related items intended to tap or measure a particular concept or construct. See for example the Rosenberg self-esteem scale (http://www.yorku.ca/rokada/psyctest/rosenbrg.pdf).

Another commonly used scale in psychology and the behavior sciences is the BDI (Beck's Depression Inventory (http://www.ibogaine.desk.nl/graphics/3639b1c_23.pdf)

NOTE: When we begin to perform data analysis, I will introduce you to statistical techniques/methods, which will be either descriptive (related to the sample) or inferential (related to the population).

Coming Attractions!!: In the next assignment we will use the data given for assignment # 1 (which are real) to perform basic descriptive statistical analyses, in order to summarize (using numerical measures and graphs) the distribution of each variable given in the table.