Biostatistics Workshop
March 16-18, 2011
Introduction

ANWAR AHMAD, PHD, MBA, MCIS

Biostatistics

- **Biostatistics** or **biometry** is the application of statistics to a wide range of topics in biology. It has particular applications to medicine and to agriculture.

Statistics

- Statistics is the study of methods and procedures for collecting, classifying, summarizing, and analyzing data and for making scientific inferences from such data.

Descriptive Statistics

- Abstraction of various properties of sets of observations using graphic, tabular or numerical devices.
- Frequency
- Typical or usual value, central tendency
- Variability, dispersion
- Relationship between the two or more variables.

Descriptive Statistics

- No conclusion is drawn
- Prelude to conclusion
- Help determine the conclusion

Inferential Statistics

- Intelligence gathering out of piles of information
- Leading to conclusion based on incomplete information
- Generalization, Inductive Reasoning
- Ex. Mendelian Law of Inheritance
- Generalizing the college student demographics to the US pop based on JSU

Inferential Statistics

- Clinical trial data on medication efficacy and extrapolating to the entire population

Statistical Inference

- Inductive reasoning to infer about a population based upon a sample
- Statistical inference is concerned with the procedures whereby such generalizations or inductions can be made
- Biostatistician alone may not be in a position to exclusively infer

Statistical Observations

- Investigators are concerned with subjects or other experimental units
- Statisticians are concerned with numbers

Statistical Observations

- Blood pressure
- Egg production
- Soil fertility
- Behavioral charectistics
- ABO blood group

Variables

- The charectistics observable on the unit
- The charectistic is called a variable because it can vary
- Charectistics that do not vary are called constants

Variables

- A characteristic or an attribute of the statistical unit/entity of interest with values that are numeric (in the case of a **quantitative variable**) or non-numeric (in the case of a **qualitative variable**).

Variables

- The standard notation for a variable is X in the case of a univariate analysis, X and Y in the case of a bivariate analysis, or X, Y and Z in the case of a three-variable multivariate analysis.

Variables

- Several kinds of variables
- Qualitative versus quantitative
- Discrete versus Continuous

Variables

- Dependent vs. independent variables
- Independent variables are those that are manipulated whereas dependent variables are only measured or registered

Variables-Measurement scales

- Variables may be classified as
- (a) nominal,
- (b) ordinal,
- (c) interval or
- (d) ratio.

Nominal variables

- Nominal variables allow for only qualitative classification.
- That is, they can be measured only in terms of whether the individual items belong to some distinctively different categories, but we cannot quantify or even rank order those categories.

Nominal variables

- For example, two individuals are different in terms of variable A (e.g., they are of different race),
- Which one "has more" of the quality represented by the variable?
- Typical examples of nominal variables are gender, race, color, city, etc.

Ordinal variables

- Ordinal variables order the items in terms of "less and more" of the quality represented by the variable
- A typical example of an ordinal variable is the socioeconomic status of families

Ordinal variables

- For example, upper-middle is higher than middle but how much higher?, 18% higher
- Nominal measurement provides less information than ordinal measurement, "how much less" it does not inform

Interval variables

a. Interval variables not only rank order the items that are measured, but also to quantify and compare the sizes of differences between them
b. For example, temperature, as measured in degrees Fahrenheit or Celsius, constitutes an interval scale
c. A temperature of 40 degrees is higher than a temperature of 30 degrees, and that an increase from 20 to 40 degrees is twice as much as an increase from 30 to 40 degrees

Ratio variables

a. Ratio variables are similar to interval variables; in addition to all the properties of interval variables, they feature an identifiable absolute zero point, thus they allow for statements such as x is two times more than y.
b. Typical examples of ratio scales are measures of time or space. For example, as the Kelvin temperature scale is a ratio scale, not only can we say that a temperature of 200 degrees is higher than one of 100 degrees, we can correctly state that it is twice as high. Interval scales do not have the ratio property.
c. Most statistical data analysis procedures do not distinguish between the interval and ratio properties of the measurement scales.

Statistical terminologies

- Study or investigation: An organized scientific undertaking with a defined set of purposes or objectives

Statistical terminologies

- Survey: A study to assess conditions as they exist in nature, altering them as little as possible
- E.g.. Higher stroke rates among minority population
- Higher breast cancer incidence among majority population

Statistical terminologies

- Experiment: A study that alters existing conditions in a defined manner in order to assess the effect of one or more “treatments”
- Effect of dietary intake on blood pressure and cholesterol

Statistical terminologies

- Unit: That smallest object or individual that can be investigated, the source of the basic information
- Experimental units
- Sampling units

Statistical terminologies

- Design: The detailed specification of the procedures whereby information will be obtained

Statistical terminologies

- Population or universe: A very large, infinite, group of units concerning which scientific inferences are to be made
- All college going students in the US
- All milking cows in Europe
- All sheep in Australia

Statistical terminologies

- Parameter: A characteristic of population
- Average resting heart rate
- Milk production
- GPA of senior class
- Crop yield

Statistical terminologies

- Sample: A subset of units in the underlying population or universe
- The sample provides the actual numerical information used in making inferences about the population

Statistical terminologies

- Random sampling
- Stratified sampling
- Convenient sampling

Statistical terminologies

- Statistic: A charectistics of a sample, used for making inferences about parameter

Statistical terminologies

- Analysis: The procedures for summarizing and extracting numerical information on the variables observed on the units selected for study and making inferences from these data

Statistical terminologies

- Statistical inference: A conclusion about a population on the basis of information contained in a sample
- Exact knowledge of this population can never be found unless the entire population is sampled
- Statistical inference utilize the laws of probability in arriving at conclusions