Statistics

January 13, 2010

Jobayer Hossain, Ph.D. & Tim Bunnell, Ph.D.

Nemours Bioinformatics Core Facility

Overview

- Class goals
  - Master basic statistical concepts
  - Learn analytic techniques & when to apply them
  - Learn how to interpret analysis results
  - Develop familiarity with statistical package SPSS and related tools
  - Gain understanding that will transfer to a broad range of other statistics tool

Overview

- Class structure
  - 8 sessions
  - 1.5 hours per session
  - Several homework assignments
- Class website
  - http://www.medsci.udel.edu/open/StatClass/January2010

Statistics Package for the Social Science (SPSS)

- SPSS can import data from almost any source to generate reports, plots of distributions and trends, descriptive statistics, and complex statistical analyzes
- Opening SPSS through Citrix server
  - Click on MetaFrame Presentation Server on the desktop of your computer
  - Click on SPSS v17 and it will open SPSS for you
- Open data file to SPSS
  - File → Open and look for data source; follow steps on the window wizard

Statistics Package for the Social Science (SPSS)

- Data Editor: Spreadsheet-like window which contains the data to be analyzed
- The data editor has two views:
  - Data View: contains the data
  - Variable View: Displays variable information of the stored dataset
- Data Manipulation: Can be done either by writing syntax or by using drop down menu in the SPSS data view window
- Menus Edit, View, Data, and Transform are important for data management, and Analyze and Graphs are important for data analysis

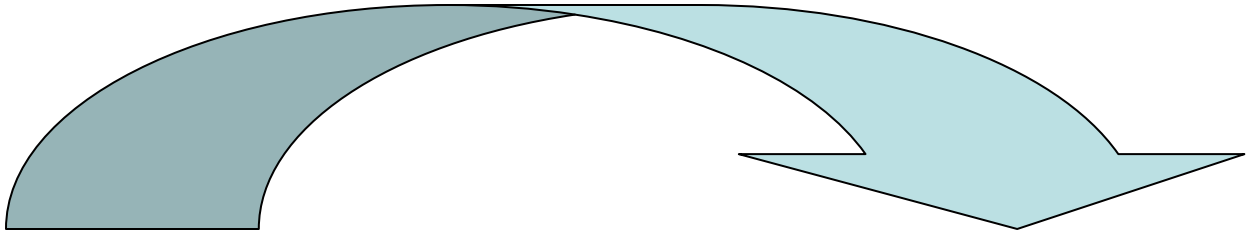
Statistics

- Science of data collection, summarization, analysis and interpretation.
- Descriptive versus Inferential Statistics:
  - Descriptive Statistic: Data descriptions (summarization) such as center, variability and shape for quantitative variables (e.g. age) and number (frequency) and percentage for categorical variables (e.g. gender, race etc).
  - Inferential Statistic : Drawing conclusions that extend beyond the sample studied, allowing for prediction.

Statistical Description of Data

- Statistics describes a numeric set of data by its
  - Central tendency (mean, median, mode etc)
  - Variability (standard deviation, range etc)
  - Shape (skewness, kurtosis etc)
- Statistics describes a categorical set of data by
  - Frequency, percentage or proportion of each category

Statistical Inference



sample population

- Statistical inference is the process by which we acquire information about populations from samples.
- Two types of estimates for making inferences:
  - *Point estimation.*
  - *Interval estimate.*

Population and Sample

- **Population**: The entire collection of individuals or measurements about which information is desired.
- **Sample**: A subset of the population selected for study.
  - Primary objective is to create a subset of population whose center, spread and shape are representative of the population.

Parameter v.s. Statistic

- **Parameter**:
  - Any statistical characteristic of a *population*.
  - Population mean, population median, population standard deviation are examples of parameters.
  - Parameters describe the distribution of a population
  - Parameters are fixed and their exact values are usually unknown

Parameter v.s. Statistic

- **Statistic**:
  - Any statistical characteristic of a *sample*.
  - Sample mean, sample median, sample standard deviation are some examples of statistics.
  - Statistics ***estimate*** parameters of the population
  - The value of a statistic is determined (known exactly) from a sample, but different samples can yeild different values of statistics.
  - Thus, statistics are used for making inferences about population parameters

Parameter v.s. Statistic

- Statistical Issue
  - Estimate a population parameter using a sample statistic.
  - E.g., the sample *mean* is an estimate of the population *mean*.