General Biostatistics

Part 10

General Biostatistics

Summary

Marie Diener-West, Ph.D.
Department of Biostatistics
Johns Hopkins University
Bloomberg School of Public Health

Outline

- "Biostatistics"
- Descriptive statistics
- Statistical inference
- Common statistical methods
- Sample size and power
- Simple linear regression

Biostatistics

- Uses the tools of probability and probability distributions to quantify the certainty/uncertainty associated with observations.
- Provides methods for describing and summarizing natural phenomena
- Assesses the strength of evidence for or against an hypothesis

Descriptive Statistics

- Exploratory data analysis
  - Stem and leaf displays
  - Box and whiskers plots
- Graphical displays and tables
- Summarizing data
  - Measures of central tendency
  - Measures of spread

Statistical Inference

- Populations and samples
  - A parameter = a summary measure computed from a population
  - A statistic = a summary measure computed from a sample
- Statistical inference is a conclusion about the unknown truth (the population) based on the data (the sample).

Methods for Statistical Inference

- Based on theoretical sampling distributions of sample statistics
- Estimation - does not require an hypothesis
  - sample statistic $\pm$ (z or t) standard error of the sample statistic
  - the precision of the estimate increases and the width of the CI decreases with increasing sample size.

Methods for Statistical Inference

- Hypothesis testing - requires Ho and Ha
  - minimize errors
  - designate magnitude of a clinically relevant difference; set Ho and Ha
  - maximize sample size
  - rejection of Ho allows one to conclude that Ha appears to be true.

Common Statistical Methods

- Comparing:
  - Means of 2 groups - Z or t test
  - Pair-matched data - t test
  - Proportions of 2 groups - Z test or χ^2 test
  - Means of 3 or more groups - ANOVA for continuous outcomes; Bonferroni adjustment for pairwise comparisons
  - Proportions of 3 or more groups - χ^2 for dichotomous outcomes

Sample Size and Power

- Components of sample size
  - Type I and II errors
  - Estimated variance
  - Clinically relevant difference of interest
- Sample size increases with:
  - Smaller assumed errors
  - Large variance
  - Smaller clinically relevant difference

Sample Size and Power

- Power is related to sample size and the magnitude of the clinically relevant difference
  - Power to detect a difference of interest increases with increasing sample size
  - Power to detect a difference of interest decrease with increasing variability

Simple Linear Regression

- Correlation analysis describes the strength and direction of a linear relationship between 2 variables
  - r=+1, perfect positive correlation
  - r =0, no correlation
  - r=-1, perfect negative correlation
- Regression analysis provides a prediction of 1 variable based on another variable

What's next?

- Multiple linear regression analysis
  - Linear relationship between a continuous response variable and multiple predictor variables
- Multiple logistic regression analysis
  - Relationship between a dichotomous variable and multiple predictor variables

What's next?

- Cox regression analysis
  - The risk of a response based on multiple predictor variables and time to event