BIOS 6222: Biostatistics II

Analysis of Categorical Data

Instructor: Qingzhao Yu

Outline

- Response Variables and Explanatory Variables
- Variable Measurements
- Sampling Framework
- Review of Discrete Distributions

Response vs. Explanatory Variables

Explanatory variables are the variables that affect the responses. In a given problem, how one makes the distinction depends on the study design and the scientific goals of the investigation. The distinction is clear in regression context.

Variable Measurements

- Qualitative
  - Nominal Scale: Categorizes the data into distinct groups. Examples: sex(male/female) and race(Black/Caucasian/ Native American/Asian/others).
  - Ordinal Scale: Categorizes the data into groups and order the groups. Examples: pain(none/moderate/severe) and socio-economic status(low/middle/high).

Variable Measurements(Cont'd)

- Quantitative
  - Interval Scale: Categorize, order, and quantify comparisons between pairs of measurements. Require a unit of measurement and an arbitrary origin. Example: temperature measured in ${}^0F$.
  - Ratio Scale: Categorize, order, and quantify comparisons between pairs of measurements, and quantify comparisons between measurements. Require a unit of measurement and an absolute origin. Example: length.

Variable Measurements(Cont'd)

- Discrete: The variables can assume either a finite or countable number of values.
- Continuous: The variables can assume any value in some intervals.
- A discrete variable can be either qualitative (pain) or quantitative (# of accidents).
  - Example 1: arthrit.sas
  - Example 2: respire.sas
- A continuous variable is quantitative.

Sampling Framework

- Historical Data: the study population has a geographic or circumstantial definition. Assumes no randomization.
- Experimental data: drawn from studies that involve the random allocation of subjects to different treatments of one sort or another.
- Sample Survey Data: subjects are randomly chosen from a larger study population.

Review of Discrete Distributions

- Binomial
  - With sample size n
  - Mean is $n\pi$ and variance is $n\pi(1-\pi)$

$$P(Y=y|n)=\frac{n!}{y!(n-y)!}\pi^y(1-\pi)^{n-y}, y=0,1,\ldots,n$$

- Poisson
  - with mean μ and variance μ

$$P(Y=y)=\frac{\mu^y}{y!}e^{-\mu}, y=0,1,2,\ldots$$

- Negative Binomial Distribution and Multinomial Distribution