Statistics 514: Design of Experiments

Topic 2 Summary

Review – Concepts

- *Random Variables*
  - Have distribution function
  - Summaries – mean ($\mathsf{E}()$), variance ($\mathsf{Var}()$)
  - Independence – implies uncorrelated
- *Sample Summaries*
  - Samples can be thought of as collections of RV's.
  - Sample mean – unbiased with variance σ^2/n; affected by CLT
  - Sample variance – unbiased with limiting distribution χ^2

$$S^2 = \frac{1}{n-1}\sum_i X_i^2 - \frac{1}{n(n-1)}\left(\sum_i X_i\right)^2$$

More Basics

- Setup – Assume model and estimate/infer parameters from data
  - Based on understanding as random variable
  - Get *estimate* with actual sample
  - Want nice *estimator* properties – low variance, bias
  - Dividing by *degrees of freedom* takes bias away
- Hypothesis Testing – is data "unusual" for null model?

Important Distributions

- Associated with tests of particular parameters
- Generates null distribution (is data unusual?)
- Often used to generate other distributions
- Often derived using *Central Limit Theorem*
  - Helps with model validity for large samples

Examples

- *Normal distribution* – building block of other distributions
- *Chisquare distribution* – used for modeling variance
- *t-distribution* – used for testing mean
- *F-distribution* – testing using ratio of sums of squares

Most distributions (in this class) have "non-central" form, primarily used for power analysis.

- *Noncentral* χ^2 – distribution of mean square under alternative
- *Noncentral* F – distribution of F-ratio under alternative
- Working with each distribution in tables/SAS is an important skill.

Hypothesis Testing Mechanics

- Use cdf's to find p-values
  - 1-sided ($\theta > \theta_0$): $P(X > u)$
  - 2-sided ($\theta \neq \theta_0$): $2(1 - cdf(|u|))$
- p-values may be too small for SAS to compute numerically.
- Use quantiles ($1 - \alpha$ or $1 - \alpha/2$) to find rejection regions for hypothesis tests.