Building Parametric Statistical Models

IPSA Summer School in Sao Paulo Brazil

January 26-29 2016

Dr. Randy Stevenson

randystevenson@rice.edu

Overview: This week-long course will focus on general strategies for building statistical models and introduce maximum likelihood as a method to estimate their unknown parameters. In this week, only limited attention will be paid to estimating measures of uncertainty. Two examples of the general model building strategy that we will examine in depth are the normal-linear model (regression) and the Bernoulli-logistic model (logit), which is often appropriate when modelling a dichotomous dependent variable.

Each day of the course will involve about three hours of lecture time with breaks, then lunch, then three to four hours of exercises and hands on instruction in analysis using STATA (with some initial data exercises by hand or in excel). Data sets for assignments will be provided.

We will move at a pace set by you. It is more important that you understand the material, than we cover many topics. I usually find that repeated presentations of some of material, as well as study outside of class, is necessary.

Course Prerequisites: No previous experience in statistical modelling or mathematics is required. That said, we will not avoid the mathematics that are necessary to understand statistical modeling at a level required become a good applied data analyst. I will teach all the math required (which is really less than most people think!).

Course Readings: Lecture Notes will be available for download.

Background reading: Gary King Unifying Political Methodology: The Likelihood Theory of Statistical Inference" Michigan University Press. 1998

http://www.amazon.com/Unifying-Political-Methodology-Likelihood-Statistical/dp/0472085549

Topic 1: Overview of building a statistical model and probability distributions

- Course Overview
- Understanding Parametric Statistical Models
- Introduction to Probability Distributions
- Examples: The Bernoulli-Logistic (Logit) and Normal-Linear-Homoscedastic Models ("Regression")

Topic 2: Introduction to Maximum Likelihood Estimation

- From a Parametric Statistical Model to a Likelihood Function
- The Likelihood Principle
- Graphical, Analytical, and Numerical Approaches to Maximizing the Likelihood Function
- Tools in Stata and Excel for finding Maximum Likelihood Estimates “by hand”

Topic 3: Strategies for Building a Statistical Model: The Stochastic Component

- Building the Stochastic Component of a Statistical Model
    - The shallow way
    - From “first principals”
- Examples
    - Deriving the Bernoulli-Logistic Model from first principles
    - Binomial-Logistic Model
    - Double Truncated Normal Model

Topic 4: Strategies for Building a Statistical Model: The Systematic Component

- Moments of a Probability Distribution
- The Linear-aggregator function
- Functions in the Systematic Component
- Choosing Covariates to include in the Systematic Component
- Nominal, Polynomial, and Interactive Covariates

Topic 5: Interpreting Results

- Using the systematic component to produce predicted values and substantive effects
- The problems of Non-Linearity and Non-additivity for interpretation
- How not to (accidentally) lie about your results
- Choosing “typical” cases
- Averaging over the sample
- Programming for Presenting Results