

Princeton University
Woodrow Wilson School
Graduate Program

Fall Term – 2002

WWS 507C
Quantitative Analysis: Advanced

Prof. Josh Goldstein
257 Wallace Hall
Ext 8-5513
josh@princeton.edu

The aim of this course is to introduce students to statistical methods and their applications in social science and public policy. The course is divided into four sections. The introductory section covers descriptive statistics, including statistical summary measures and the graphical display of data. The second section covers the probability theory that forms the basis of statistical inference. The third section covers sampling, estimation and inference. The final section covers linear regression. Through the course, examples from many fields will be discussed to demonstrate the wide applicability of quantitative analysis and to introduce statistical methods.

Students are expected to be familiar with algebra and calculus. We will also use matrix notation for our discussion to multivariate regression. Students who would like a course with less math should take WWS 507b. Although mathematical proofs will not be the focus of the course, formal analysis will be used often to show the inter-relatedness of concepts and to generalize specific examples.

Grades will be based on weekly exercises (20%), two projects (30%), a mid-term exam (20%) and a final (30%). In addition to this graded work, students are also required to participate in the joint-class survey exercise (with WWS 507b) which will involve approximately two hours for a telephone poll on the New Jersey senate race. Exams will allow students to bring in one page of notes and will require the use of a calculator that handles square roots, logarithms, and the exponential function.

Because the only way to learn statistics is by practicing it, the exercises and projects are a critical part of the class. A total of 10 problem sets will be handed out. Students can choose not to hand in one of the problem sets. The problem sets consist of a wide array of questions ranging from formal “drills” to practical applications. Students are encouraged to collaborate on problem solutions and to discuss alternative solution strategies. However, each student is required to hand in his or her own solutions. Precepts will meet once a week. Problem sets will typically be due in precept. Late problem sets will not be accepted. Precepts will review problem set solutions as well as answer other questions.

The two projects are a chance to work with real world data to help answer policy questions. The first project will be an analysis of racial profiling on the New Jersey turnpike and will involve analysis of speeding data gathered by automatic cameras. The official analysis of the camera data finds that Black drivers were more likely to speed

Weeks 2 and part of 3 Introduction to Probability

Lecture 3 Probability defined and combinatorics

- What are the chances?
- The binomial theorem
- Discrimination in workplace
- The birthday problem: the high likelihood of coincidences

Lecture 4 Conditional Probability

- Statistical independence
- Bayes Theorem
- Monte's dilemma
- ID'ing terrorists
- Sensitive item questionnaires

Lecture 5 Random Variables

- What is a random variable?
- Density and Distribution Functions
- Moments (and properties)
- Common Discrete and Continuous Distributions

Weeks 3, 4, 5 and 6 Estimation and Inference from Samples

Lecture 6 Why sample and how to design a sample

- Simple random sampling
- Law of large numbers
- China vs. Rhode Island
- Finite sampling correction (mentioned)

Lecture 7* Guest Lecture on Survey Methodology

- Designing a survey
- Sampling with the telephone
- Our Torricelli-Forester poll

*May be outside of normal class time

Lecture 8 Normality

- Central Limit Theorem stated
- CLT demonstrated
- Properties of the Normal distribution

Lecture 9 Confidence

- From normality to confidence intervals
- Distinction between randomness of estimate and non-randomness of parameter

** Week of election survey **

Lecture 10 Hypothesis testing
Defining the alternative
p-values
One-sided and two-sided test
Z-test

Lecture 11 Comparing two means
Types of errors
t-test
Definition of F-test

Lecture 12 Properties of Estimators
Unbiasedness
Efficiency
Mean Squared Error
Maximum Likelihood
Capture-Recapture estimator
Geometric waiting time example

MID-TERM EXAM (may be outside of class)

MID-TERM BREAK

Weeks 7 through 12

Multivariate Analysis (Regression)

Lecture 12 Contingency tables and goodness of fit
What is a cross-tabulation?
Independence
Chi-squared tests

Lecture 13 Modeling tabular data
Two-way modeling
Additive and multiplicative models
Introduction to the idea of control
Infant mortality by age and race
Income group by sex and occupation

Lecture 14 Regression formula
Interpretation
Uses of regression
Endogeneity
Causality
Prediction

"Control" in observational studies

Lecture 15 How to fit a straight line

OLS and LAD

Estimation of slope and intercept

Lecture 16 The uncertainty of regression estimates

Illustration of parameter estimate variability

Derivation of standard errors

The importance (and unimportance) of statistical significance

Lecture 17 OLS Assumptions

Illustrate heteroskedasticity

Outliers

Non-linearity

Gauss-Markov theorem

Lecture 18 Extensions to multiple regression

Matrix algebra formulation

of multiple regression

Lecture 19 Including and interpreting covariates

Dummy variables

Interactions

Lecture 20 The importance of covariates

Omitted variable bias

Including irrelevant variables

Multicollinearity

Lecture 21 Experimental design vs. regression

Discussion of Freedman's critique

Lecture 22 Yes-no outcome regressions

The linear probability model

The logistic alternative

Note: If possible, one or two lectures in the second half of the semester will be spent discussing the projects.