Quantitative Analysis – International and Public Affairs
U6500 Sections 3, 4, 5
Fall 2010 Syllabus

Professor Alan Yang asy2@columbia.edu
Office Hours: Wednesday 1:30-3:30 (1309 IAB)

Section 3: Tuesday	***11:00-12:50 p.m.***	***(410 IAB)***
Labs: Thursday	6:10 p.m.-8:00 p.m.	
Friday:	1:00 p.m.-2:50 p.m.	

Section 4: Tuesday	***2:10-4:00 p.m.***	***(410 IAB)***
Labs: Thursday:	9:00 a.m.-10:50 a.m.	
Friday:	3:00 p.m.-4:50 p.m.	

Section 5: Wednesday	***11:00-12:50 p.m.***	***(411 IAB)***
Labs: Thursday:	2:10 p.m.-4:00 p.m.	
Friday:	9:00 a.m.-10:50 a.m.	

Course Description:

This course introduces students to the fundamentals of statistical analysis. We will examine the principles and basic methods for analyzing quantitative data, with a focus on applications to problems in public policy, management, and the social sciences. We will begin with simple statistical techniques for describing and summarizing data and build toward the use of more sophisticated techniques for drawing inferences from data and making predictions about the social world.

The course will assume that students have little mathematical background beyond high school algebra. Students will be trained on STATA, which is supported in the SIPA Computer lab. This powerful statistical package is frequently used to manage and analyze quantitative data in many organizational/institutional contexts. A practical mastery of a major statistical package will be an important proficiency for many of you down the road. All students are *strongly* encouraged to enroll in the SIPA *Computer Skills Course*. You can obtain more information about the course and your lab sticker at the SIPA lab, which is located on the 5th floor of IAB.

The formal mathematical foundation of statistics is downplayed; students who expect to make extensive and customized use of advanced statistical methods may be better served by a different course. This course also offers less practice in writing research papers using quantitative analysis than some courses (e.g., Political Science 4910). Most MPA students, however, should benefit from our emphasis on generating and interpreting statistical results in many different practical contexts.

Requirements:

Class attendance is required. Lectures will sometimes cover matters related directly to the homework assignments that are not covered fully in the assigned readings.

Students are required to review and obtain any relevant material (e.g., weekly handouts) in advance of each class by going to *Courseworks* at https://courseworks.columbia.edu. This site will include all course materials including: the syllabus, weekly class handouts, homework assignments, solution sets, midterm and final exam review sheets, and information on data as well as downloadable datasets.

Students are required to come to class having *already completed* the assigned readings for that class. The purpose of this requirement is to ensure that our twelve classes (not counting an in-class midterm) can focus on learning how to bring to bring statistical concepts and methods to life in an applied context. Class will be conducted in a manner that assumes this advance preparation has been done.

Students are required to attend a weekly lab session in addition to the regular lecture. These labs will be important supplements to each lecture, where concepts and methods will be reviewed and students will receive direction and support as they learn STATA.

Problem sets will be assigned at least a week in advance of their due dates. ***Late problem sets will not be accepted for credit***. You are encouraged to be actively engaged in the completion of every problem set since hands-on work (computer-based or otherwise) is essential to fully absorbing the material presented in this course. *Problem sets will be done in groups of two.*

Grading:

The three components to the final course grade will include weekly problem sets (30%), a midterm exam (35%), and a final exam (35%). In "borderline" cases, the quality of your class attendance and participation will weigh heavily in determination of a final grade.

SIPA Computer Lab Policy 2010-2011

The SIPA computer lab accommodates a maximum of 44 students per session. All students taking classes or attending recitations in the computer lab must adhere to this limit. Additional students will not be allowed to share computer stations, sit on the floor, or sit in the back of the room. Instructors, TAs, and computer lab staff will enforce this policy.

SIPA students must secure guest IDs in order to gain access to the SIPA computer lab during class sessions. Guest IDs will not be available during the first week of classes, but shortly thereafter they will be provided to each registered non-SIPA student. Guest ID's will allow non-SIPA students to access the SIPA computer lab *during class time only*. Non-SIPA students will obtain their guest IDs from the SIPA Consultant Desk in 510 IAB.

Non-SIPA students who wish to use the SIPA computer lab *outside of regular class/recitation time* must pay $160 per semester (payable by check or cash in 510 IAB). Non-SIPA students who choose not to pay this fee should consult their course instructor or TA and the IT office at their own school about any special software required for the course. SIPA IT is not equipped to provide technical support to non-SIPA students who have not paid the $160 per semester fee.

Academic Integrity Statement:

The School of International & Public Affairs does not tolerate cheating and/or plagiarism in any form. Those students who violate the Code of Academic & Professional Conduct will be subject to the Dean's Disciplinary Procedures. Click here to view the Code of Academic & Professional Conduct online.
http://sipa.columbia.edu/resources_services/student_affairs/academic_policies/deans_discipline_policy.html

Please familiarize yourself with the proper methods of citation and attribution. The School provides some useful resources online; we strongly encourage you to familiarize yourself with these various styles before conducting your research:
http://sipa.columbia.edu/resources_services/student_affairs/academic_policies/code_of_conduct.html

Violations of the Code of Academic & Professional Conduct should be reported to the Associate Dean for Student Affairs.

Readings:

The *required* and *recommended* textbooks may be purchased at Bookculture (536 West 112th Street).

Required Texts:
D. Moore and G. McCabe. *Introduction to the Practice of Statistics.* 6th Ed (2009)
  W.H. Freeman and Company
  ISBN: 9781429216227
M. Lewis-Beck, *Applied Regression.* (1980)
  SAGE
  ISBN: 9780803914940

Recommended:
Lawrence C. Hamilton. *Statistics with STATA: Updated for Version 10*
  Thomson-Brooks/Cole
  ISBN: 9780495557869

Supplemental:

Wonnacott & Wonnacott also covers much of the same material as Moore & McCabe. Students may want to supplement the weekly readings for some units. .
T. Wonnacott and R.Wonnacott, *Introductory Statistics.* 5th Ed (1990)
(Chapters 1-4, 6, 8, 9, 11-14, 17)

Units on research methods/design and sampling will refer to the readings in Hoyle, et al.
R. Hoyle, M. Harris, C. Judd, *Research Methods in Social Relations.* 7th Ed (2002)

C. Achen, *Interpreting and Using Regression.* (1982)
(For Multiple Regression Units)

COURSE OUTLINE

Session 1:   Introduction to Statistics & Research Design
September 7 (Sect. 3 & 4)/September 8 (Sect. 5)

- How can we ask and answer meaningful empirical questions? Testing Theories (Concepts and Variables, Unit of Analysis, Independent and Dependent Variables)
- Randomized Experiments and Causality, Observational Studies
- Sampling Design (Random Samples, Probability Sampling), Surveys and Questionnaires

STATA: Introduction to STATA (Log files, Data Editor, Recoding Variables)

Required Readings:
M & M: Chapters 3.1-3.2 (Research Design)
        Supplemental: Hoyle, et al.: Chapter 8, 9, 11

Courseworks: Session 1 Handout, Homework 1, Introduction to STATA Handout

Session 2:   Exploratory Data Analysis—Single Variable
September 14 (Sect. 3 & 4)/September 15 (Sect. 5)

- Classification of Variables, Frequency Distributions, Graphical Displays (Bar Charts, Pie Charts, Histograms)
- Measures of Central Tendency (Mean, Median, Mode), Measures of Dispersion (Quartiles, Boxplots, Variance, Standard Deviation)
- Density curves, Standard Normal Distribution (Z Table, Normal Quantile Plots)

STATA: Frequency tables, Charts and Graphs, Summary Statistics

Required Readings:
M & M: 1.1-1.3

Courseworks: Session 2 Handout, Homework 2, General Social Survey Online Codebook Handout

Session 3:   Exploratory Data Analysis—Relationships between variables
September 21 (Sect. 3 & 4)/September 22 (Sect. 5)

- Association between categorical explanatory variable and quantitative response variable
- Association between 2 categorical variables
- Association between 2 quantitative variables: Scatterplots, Correlation

STATA: Means tables, Crosstabs, scatterplots

Required Readings:
M & M: 2.1-2.2

Courseworks: Session 3 Handout, Homework 3

Session 4:   Exploratory Data Analysis—Simple Regression
September 28 (Sect. 3 & 4)/September 29 (Sect. 5)

- Simple Linear Regression (Least Squares Criterion, Calculating and Interpreting regression coefficients)
- Predicted Values and Residuals, R-square, Outliers, Lurking Variables, Causation
-
STATA: Simple Regression, residual analysis

Required Readings:
M & M: Chapter 2.3-2.4

Courseworks: Session 4 Handout, Homework 4

Session 5:   Concepts and Applications in Probability
October 5 (Sect. 3 & 4)/October 6 (Sect. 5)

- Basic Definitions, Compound Events, Joint Probabilities, Conditional Probabilities
- Statistical Independence, Contingency Probability Tables
- Random Variables (Discrete versus Continuous), Mean and Variance of Random Variables, Probability Distributions

STATA: Contingency probability tables (crosstabs)

Required Readings:
M & M: Chapters 4.1-4.5 & 2.5 (Contingency Probability Tables)

Courseworks: Session 5 Handout, Homework 5

Session 6:   Sampling Distributions
October 12 (Sect. 3 & 4)/October 13 (Sect. 5)

- Sampling Distribution for Counts (Calculating Binomial Probabilities), Sampling Distribution for proportions, Normal approximation
- Sampling Distribution of the Mean, Central Limit Theorem.

Required Readings:
M & M: Chapters 3.3 & 5.1-5.2

Courseworks: Session 6 Handout, Pseudo Homework 6, Midterm Review Practice Problems, Midterm Review Key Topics Handout

Session 7:   MIDTERM EXAM (In-Class)
October 19 (Sect. 3 & 4)/October 20 (Sect. 5)

Session 8:   Basics of Statistical Inference
October 26 (Section 3 & 4)/October 27 (Sect. 5)

- Student's t-distribution, Confidence Interval around a single Mean
- Hypothesis testing (Null and Alternative Hypotheses, One versus two-sided tests, p-values)

STATA: One-Sample t-test

Required Readings:
M & M: Chapters 6.1-6.3 & 7.1

Courseworks: Session 8 Handout, Homework 7

Session 9:   Inference for two sample means, a single proportion, and two proportions
November 3 (Sect. 5)/November 9 (Sect. 3 & 4)
Note: No class on Tuesday November 2 for Election Day Holiday

- Inference for Difference in Means (Independent samples, Pooled Variance), Confidence Interval around a Difference in Means, Hypothesis testing
- Inference for Single Proportion (Confidence Interval around a proportion, Hypothesis testing)
- Inference for Difference between two proportions (Confidence Interval around difference in two proportions, Hypothesis testing)

STATA: Independent Samples t-test, Test for a single proportion, Two-sample test of proportion

Required Readings:
M & M: Chapters 7.2 (Difference in Means) & 8.1-8.2 (Inference for proportions)

Courseworks: Session 9 Handout, Homework 8

Session 10:   Contingency Tables & Inference for Simple Linear Regression
November 10 (Sect. 5)/November 16 (Sect. 3 & 4)

- Contingency Tables Reviewed, Chi-Square test of statistical independence
- Simple Linear Regression, Goodness of fit, Gauss-Markov Assumptions
- Hypothesis testing of slope coefficient (Standard error of b-coefficient, Confidence Intervals, p-values)

STATA: Crosstabs, Chi-Square, Simple regression estimates, assess statistical significance, calculate predicted values and residuals, goodness of fit

Required Readings:
M & M: Chapters 2.5 (review) 9.1-9.2 (Chi-Square) & 10.1-10.2 (Inference for Simple Regression)
Lewis-Beck: pp. 9-47

Courseworks: Session 10 Handout, Homework 9

Session 11:   Multiple Linear Regression I
November 17 (Sect. 5)/November 23 (Sect. 3 & 4)
Note: No Class on Wednesday November 25 for Thanksgiving Holiday

- Review Randomized Experiments versus Observational Studies, Confounding Factors
- Fitting the model, Check of regression assumptions, Model fit (F-test, R-square)

STATA: Multiple regression estimates, use residuals to check regression assumptions

Required Readings:
M & M: Chapters 11.1-11.2
Lewis-Beck: pp. 47-54

Courseworks: Session 11 Handout, Homework 10

Session 12:   Multiple Linear Regression II
November 30 (Sect. 3 & 4)/December 1 (Sect. 5)

- Dummy Independent Variables (Same slopes, different intercepts)
- Interaction terms (Different slopes, different intercepts)
- Categorical Independent Variables

STATA: Creating and using dummy independent variables, creating and using interaction terms

Required Readings:
Lewis Beck: pp. 54-74

Courseworks: Session 12 Handout, Pseudo Homework 11, Final Review Key Topics Handout, Final Exam Practice Problems

Session 13:   FINAL EXAM (In-Class)
December 7 (Sect. 3 & 4)/December 8 (Sect. 5)