DATA ANALYSIS FOR PLANNING AND MANAGEMENT II

COURSE: UPP 589
CALL NO.: 22738
PREREQUISITES: UPP 503 or a graduate equivalent.
CLASS MEETS ON: 12:30-3:30 PM on Thursdays
LOCATION: 2113 ADH
DURATION: January 12 to May 4, 2006
SPRING BREAK: March 20-24, 2006
FINAL EXAM WEEK: May 1-May 5, 2006

Instructor: Dr. Piyushimita Thakuriah (Vonu)
Email: *vonu-pt@uic.edu* • **Phone:** (312) 355-0447
Course Web Site: Login to *http://blackboard.uic.edu*

DESCRIPTION OF THE COURSE

The objective of the course is to introduce a series of statistical methods that will enable urban planning and policy students to: (i) formulate empirical problems for the purpose of making planning/policy decisions (ii) determine an appropriate method for estimating parameters (iii) conduct exploratory data analysis (iv) develop a model that is appropriate to the problem and the data and finally (v) to diagnose problems with the model and to validate the model.

The broad framework of regression and related methods will be introduced using policy and planning-related examples as case studies. The instruction will emphasize upon the exploratory nature of empirical planning and policy analysis. To this end, the course will focus heavily on exploratory data analysis and model diagnostics.

Students will work on a number of planning and policy-related examples that will highlight the methods. For this purpose, a number of datasets will be distributed in class, using which, students will be able to develop statistical models using, in many cases, real-life examples.

Specific topics will include (i) linear and nonlinear regression (ii) qualitative and limited dependent variable models (iii) categorical data analysis and (iv) spatial analysis. The estimation techniques of (i) least squares and (ii) maximum likelihood will be covered. The relationships among different classes of models will be shown. We will begin the course with a review of random variables, probability distributions and matrices. The course will involve heavy usage of SAS, the Statistical Analysis System and Stata. Most techniques that we will cover will be implemented in the computer via appropriate examples. SAS is available for purchase at a heavily discounted price by students; go to http://www.uic.edu/depts/accc/index.html.

Do not be concerned about the focus given to two softwares; the windows interface of all the software I just mentioned will allow you to move easily from one software to another in estimating models. I/O issues are different and I will address data import/export.

Course requirements are two take-home assignments, a class presentation and a final term paper.

METHOD OF INSTRUCTION

The course has four high-level requirements:

1) Understanding statistical concepts,
2) Analysis of data,
3) Interpretation of statistical modeling results for a non-technical audience and
4) Technical writing for an applied policy analysis or planning journal.

All data analysis work will be implemented in a computer using SAS or Stata. However, in order to formulate the model and to interpret the results, a thorough understanding of the underlying concepts will be required. The instructor will present the statistical concepts in class, based on the various readings and using applied planning and policy examples. Students will be given a series of handouts developed for the course, on the use of the software for different end uses, such as obtaining estimates of parameters, testing, diagnostics and so on. Class discussion will focus on integrating the concepts with the application problems. The application problems will center around homework assignments given every week (on which the student will NOT be graded; it is for self-learning purposes); these assignments will be discussed in class so that the reasons behind ``why the computer spits out certain sets of numbers'' become clear to the students. In addition, class assignments are given in the second half of each class to review the concepts and material presented in the first half of the class. I talk extensively about empirical planning and policy problems. Students are expected to discuss in class modeling issues that are relevant to their own paper. Overall, these conversations are intended to bring up empirical modeling issues that are relevant to our profession.

Once students are more or less familiar with the software, we will focus our attention on the contextual material. Please bring to every session of the class:

1) Your textbooks and other material.
2) Computer printouts from the assignments.

TEXTS AND READINGS

While you are free to buy all of the books below, selected chapters will be made available, as needed, from:

1) Freund, R. J. and W. J. Wilson (2002). *Statistical Methods*. Academic Press. Revised Edition. Selected chapters.
2) Sen, A. K. and M. Srivastava (1990*). Regression Analysis: Theory, Methods and Applications*. Springer-Verlag, New York. Selected chapters.
3) Gujarati, D.N. (2003). *Basic Econometrics*. McGraw-Hall. Chapters 14, 15 and 16 *will be photocopied for class.*
4) Freund, R. J. and R. C. Ramon. (2000). *SAS(r) System for Regression*. John Wiley and Sons, Inc.
5) Wooldridge, J. (2002). *Econometric Analysis of Cross-Section and Panel Data*. MIT Press.
6) Fotheringham, S., C. Brunsdon and M. Charlton (2000). *Quantitative Geography*. Sage Publications.
7) Online journal articles.
8) Online data sources.

Course Requirements

Coursework will include:

1) Assignments that will count towards the final grade: Two take-home assignments during the semester, one class presentation and one final term paper that will be due the last week of class.
2) Assignments that will NOT count towards the final grade: (i) Homework will be assigned every week; however, the student will not be graded on the weekly home-work - these will be given so that students can get adequate practice. Failure to do these ungraded assignments will make the students fall far behind the class, such that the actual graded assignments may suffer. It is essential that the student make an effort to learn SAS and Stata rather well within the first few weeks, so that they can concentrate on the problems better later on. (ii) Class assignments: Student discussions based on class assignments given in the second half of each class to review the concepts and material presented in the first half of the class.
3) Readings each week of statistical concepts.

All assignments are to be submitted only to the Blackboard Digital Drop Box. It is essential that students use a mathematical typesetting language or an Equation Editor to be able to write down the math symbols in an electronic file for electronic submission of the home-work assignments. Further, the assignment is due in the Digital Drop Box by 12:30 PM of the due date. Late submissions will automatically mean a grade reduction of 15%.

Grading Scheme

Assignment I: 20%
Assignment II: 20%
Class presentation: 25%
Final paper: 35%

SYLLABUS

Weeks 1 through 3:

Introductory Material and Review.

1) Data modeling and its role in planning and policy analysis.
2) Review of probability, linear algebra, matrix manipulations and fundamental concepts in statistics.
3) Introduction to SAS, Stata, exploratory data analysis.
4) Introduction to simple and multiple regression.

Weeks 4 to 8:

Multiple regression and regression diagnostics.

1) Model fitting and measures of fit.
2) Gauss-Markov Condition.
3) Testing and General Linear Hypothesis.
4) Indicator Variables; Analysis of Variance.
5) Heteroscedasticity and weighting.
6) Outlier Detection.

7) Polynomial Models, splines, broken line regression, transformations including Box-Cox/Box-Tidwell transformations
8) Nonlinear Least Squares
9) Multicollinearity.
10) Principal Components, Factor Analysis and ANOVA.

Weeks 9 to 14:

Qualitative, Limited Dependent Variable and Count models.
1) Introduction to Maximum Likelihood (ML)
2) Differences between Least Squares (LS) and ML
3) Generalized linear model.
4) Introduction to latent variable modeling.
5) Binary logit and probit models.
6) Multinomial logit model
7) Count data - what's wrong with using least squares
8) Weighted least squares for Poisson data
9) Direct maximization of likelihood function
10) Log-linear model for Poisson data
11) Reiteratively Weighted Least Squares.
12) Negative binomial model
13) Truncated and censored data. The Tobit model.
14) Calculation of marginal effects and elasticities.
15) Categorical Data Analysis.
16) Correlated Errors
17) Temporal and spatial dependencies: effects on data modeling.
18) Introduction to time series and spatial analysis.

For a detailed week-by-week break down of lectures, readings and schedule of assignments, please consult Table 1.

GUIDELINES FOR FINAL PAPER

The final term paper is intended to be a publication quality paper for an applied planning or policy analysis journal. In that context, the paper should be structured as a journal article on an empirical problem, with a very focused set of research questions, which is investigated using (publicly-available) data, a statistical model and on the basis of which the research questions are interpreted. You are required to have a non-technical Introduction section and a non-technical Interpretation and Summary of Results section.

VERY IMPORTANT: You will be required to present your paper topic to class, along with (i) your research questions (ii) data source and (iii) preliminary analysis of data on March 31 (or April 7). So, it is important that you start working on your research paper almost in the second week of the semester. In the past, students who set aside time to work on their paper in the last month or so of the semester were unable to turn in good papers.

The following gives a (non-sequential) set of activities to guide you to develop your final paper.

1) Select an area/topic you would like to work on (for example, factors which affect housing prices, effect of demographics on travel time to work, effect of the regional economy on migration to an area and so on are just a few of many, many policy and planning questions).

2) Look for publicly available data (on the web or from government agencies) that would allow you to empirically estimate your model.
3) Select research questions in your chosen topic.
4) Download data; conduct exploratory data analysis.
5) Formulate research questions.
6) Develop a statistical model that would allow you to develop inferences from data regarding your research questions.
7) Extensively diagnose and interpret results; you may have to do this several times to make sure your results make sense.
8) Write paper; edit paper.

Format of paper:

1) Title page with Name of student, social security number, email address, total number of pages (including cover page) and assignment number in a cover page.
2) Abstract (not to exceed 250 words).
3) A set of 5 keywords.
4) List of tables.
5) List of figures.

6) Main body of paper. Section and subsection heavily.

    (a) Introduction section.
    (b) Data Source and Description section.
    (c) Research Question or Research Objectives section.
    (d) Model section.
    (e) Model Estimation (including model fit and diagnostics) section.
    (f) Results section.
    (g) Section on Interpretation of Results in the context of Research Questions.
    (h) Summary section.
    (i) Directions for Future Research section.

7) References

Table 1: Weekly break-down of course materials, readings and assignments.

Week	Date	Topic	Readings
1	Jan. 12	Introduction to data analysis and statistics and course requirements. Relation of course material to research design.	1) Government Accounting Office. *Quantitative Data Analysis. An Introduction*. Read Chapters 1, 6 and 7 of this publication. Online in Blackboard. 2) Freund, R. J. and W. J. Wilson: *Statistical Methods.* Chapter 1. Data and Statistics. Online in Blackboard.
2	Jan. 19	Review material (fundamentals of probability, linear algebra) and introduction to Ordinary Least Squares.	1) Freund, R. J. and W. J. Wilson: *Statistical Methods.* Chapter 2. Probability and Sampling Distributions **(**Read Sections 2.2, 2.3, 2.4 and 2.5). 2) Document on matrices. *Introduction to Matrices for Linear Regression*. matrices_spring04.pdf 3) Gujarati. Chapter 1. 4) Sen and Srivastava. Chapter 1.
3	Jan. 26	Introduction to software by guest lecturer, Yihua Liao	1) SAS 2) Stata
4	Feb. 2	Introduction to Multiple Regression.	1) Freund and Wilson. Chapter 8. 2) Sen and Srivastava. Chapter 2.
5	Feb. 9	Multiple Regression. Indicator Variables, Interaction terms. Review Class I	1) Gujarati. Chapter 9.
6	Feb. 16	Multiple Regression Diagnostics: Outlier detection, and multicollinearity. Weighted Least Squares.	1) Sen and Srivastava. Chapters 6, 8 and 10.
7	Feb. 23	Guest lecturer. Discussion of journal articles.	
8	Mar. 2	Polynomial models, splines, transformation, non-linear LS.	1) Sen and Srivastava. Chapter 9.
9	Mar. 9	Review Class II: Introduction to Maximum Likelihood	1) Gujarati. Chapter 15.
	Mar. 9	First home-work assignment due	
10	Mar. 16	Binary logit and probit models	1) Gujarati. Chapter 15.
	Mar. 23	Spring break	

11	Mar. 30	Multinomial logit models. Introduction to models for Count data (Poisson regression and Negative Binomial)	1) Gujarati. Chapter 15.
12	Apr. 6	Study session.	
	Apr. 6	Second home-work assignment due	
13	Apr. 13	Spatial patterns and spatial autoregressive models	1) Fotheringham, Brundson and Charlton. Chapter 7.
14	Apr 20	Student presentations of final paper.	
15	Apr 27	Review Class III and final papers due	