Master in Political and Social Sciences, UPF

Techniques of Statistical Analysis II, winter 2007

Dr. Jorge Rodriguez (office 20: 1E16; tel: 2239)
jorge.rodriguez@upf.edu

Subject teaching plan

1. Introduction

Techniques of Statistical Analysis II is an optional course that aims to provide an introduction to econometrics and advanced techniques of statistical analysis for the social sciences. The course builds on the previous course on Techniques of Statistical Analysis I to deepen students' knowledge of linear regression, with special emphasis on correct model specification and general lineal models. The course also provides an introduction to logistic and multinomial regression and event history analysis. The course avoids a formalistic-mathematical presentation of its contents and focuses instead on the basic logic behind the techniques. The lectures will combine a theoretical presentation of the subjects with hand-on applications using SPSS and STATA.

2. Prerequisites

Techniques of Statistical Analysis I

3. Competences to be achieved in the subject

- General competences (Instrumental, Interpersonal and Systemic)

Instrumental Competences
Ability to analyse and synthesise
Planning and management of time
Basic computer competence and ability of using statistical software like SPSS and STATA
Information management abilities (ability to search and analyse information coming from a variety of sources)

Interpersonal Competences
Critical and self-critical ability
Team work
Ability to work in a interdisciplinary team
Ability to communicate with people that are not experts in the subject

Systemic Competences
Research abilities
Ability to work autonomously
Ability to generate new ideas (creativity)
Design and management of projects

- Specific competences
  a. Knowledge of the basics of linear regression analysis
  b. An understanding of the assumptions underlying classical linear regression models, and of the importance for testing for, and applying remedial measures to, deviations from such assumptions
  c. Knowledge of regression techniques for categorical dependent variables and rates
  d. Interpretation skills, as crystallized into the ability to describe results from intermediate statistical analyses

4. Assessment

- Assessment will be based on the 0-100 scale
  - ✓ 0-50 – Fail
  - ✓ 51-60 – Pass
  - ✓ 61-70 – Good
  - ✓ 71 to high – Outstanding
- 2 take-home tests at weeks 4 (due on week 6) and 9 (due a week after classes end), with an emphasis on problem solving and interpretations of results using real data. Each will count 30% of the final mark.
- 1 final in-class test combining true-false and multiple-choice questions with short problems worth 40% of the final mark
- Every week, students will be given sets of problems/exercises to complete at home (due in class on the following week). The problems will not be marked, i.e., but for every 2 set not returned on time, 5 marks will be deducted from the final mark

5. Contents

- There will be five basic content blocks, although much of the time will be spent on block 2.

1. **Simple & multiple regression revisited**
   Regression assumptions. Deterministic & probabilistic models; metric, significance & strength of linear relations; OLS and ML estimation; parameter interpretation; t & F tests, r and R^2; R^2 decomposition; F increment tests.
2. **Deviations from regression assumptions**
   Sources of bias (miss-specification): wrong regressors, endogeneity, non-linearities and parameter inconsistency. Some tests and remedial measures for miss-specification. Sources of inefficiency: heteroskedasticity, autocorrelation, & multicollinearity. Brief introduction to testing and correcting inefficiency: generalized (weighted) least squares.
3. **Regression applications: ANOVA, ANCOVA, & non-linear regression**
   Regression with dummy variables (ANOVA revisited); regression with nominal & continuous independent variables (ANCOVA). Non-linear functions: quadratic & logarithmic functions; linear transformations; generalized linear models.
4. **Regression for dichotomous dependent variables**
   Logit & probit models; logistic regression: metric, goodness of fit, significance & strength of relationship; maximum likelihood (ML) estimation; likelihood ratio & Chi-square; pseudo R^2 measures; multinomial-logit models.
5. **Event history analysis**
   Censored observations & times; time-varying covariates; life-tables, proportional hazard models; survival functions, Kaplan-Meyer estimator; Cox regression.

6. Methodology

Each weekly 2-hour session will be divided into two parts: a lecture-type, 1-hour session where the main aspect of the weekly topics will be laid out, and a 1-hour practical session where students will practice their newly acquired knowledge with the help of statistical software

7. Sources of information and didactic resources

- **Basic bibliography**

Kennedy, Peter. 2003. *A Guide to Econometrics, Fifth Edition*. Cambridge, Massachusetts: The MIT Press.

Chatterjee, S., Hadi, A. S., and Price, B. 2000. *Regression Analysis by Example. Third Edition*. New Yor: John Wiley and Sons, Inc.

Norman, G. R. and Streiner, D. L. 2003. *Pretty Darned Quick (PDQ) Statistics. Third Edition*. BD Decker Publishers.

- **Complementary bibliography**

Gujarati, Damodar. 2002. N. Basic Econometrics, Fourth Edition. McGraw-Hill.

Maddala, G.S. 2001. Introduction to Econometrics, Third Edition. Wiley.

Knoke, David; Bohrnstedt, George W., and Mee, Alicia Potter. 2002. *Statistics for Social Data Analysis, Fourth Edition*. Belmont, CA: Wadsworth/Thompson Learning.

Agresti, Alan & Finlay, Barbara. 1997. Statistical Methods for the Social Sciences, International (Third) Edition. New Jersey: Pearson Education – Prentice Hall.

8. Program of activities and schedule of classes and readings

Week	Classroom activity		Readings & Incidences
1	*Lecture*	Introduction to multiple regression: estimation & testing	Kennedy: Chapters 1, 2 & 4
	Lab	Producing & interpreting regression output	
2	*Lecture*	Introduction to multiple regression: assumptions, causality & model building	Kennedy: Chapters 3 & 5
	Lab	Model selection techniques: backward & forward selection	
3	*Lecture*	Sources of bias (1): Specification error	Kennedy: Chapter 6
	Lab	Illustration of omitting/including relevant/irrelevant variables. Non-linear functions & linear transformations	
4	*Lecture*	Sources of bias (2): Selection & measurement error	Kennedy: Chapter 9 & 16.3 First take-home test handed in to students
	Lab	Regression with instrumental variables & split-half regression	
5	*Lecture*	Sources of bias (3): Endogeneity & experimental errors	Kennedy: Chapter 10
	Lab	Regression with reduced equations & 2-stage least squares	
6	*Lecture*	Effects on estimates' accuracy of non-spherical disturbances: heteroskedasticity & autocorrelation	Kennedy: Chapters 7 & 8 First take-home test due at class time
	Lab	Weighted Least Squares & Generalised Linear Models	
7	*Lecture*	Multicollinearity & solving identification problems	Kennedy: Chapters 11 & 12
	Lab	Regression with principal components & ridge regression	
8	*Lecture*	Regression with categorical *independent* variables: ANOVA & ANCOVA	Kennedy: Chapter 14
	Lab	Regression with dummy & continuous variables	
9	*Lecture*	Regression with categorical *dependent* variables: Logistic regression	Kennedy: Chapter 15 Second take-home test handed in to students
	Lab	Logistic & multinomial regression	
10	*Lecture*	Regression with limited dependent variables	Kennedy: Chapter 16.1, 2 & 4
	Lab	Kaplan-Meir and Cox regression	
12	Second take-home test due on Tuesday, April 25 at 6pm in my office		
13	Final test (date & place TBA)		