光华管理学院 Guanghua School of Management | UG&PG 本科研究生项目

Advanced Techniques for Management Data

Course No： 02804180
Credit：3
Prerequisite：

Program：Graduate
Instructor：Li Bobai
Semester：2016 Spring

Instructor's resume/brief introduction(Within 500 words):

Bobai Li is a sociologist interested in the fundamental social, cultural, and institutional logic of Chinese business and government organizations. His current research is centered around three themes: 1) the historical and cultural roots of "sponsored mobility" in China and how "sponsored mobility" shapes labor processes and personnel management in contemporary Chinese government bureaucracy and private firms; 2) the dynamics of informal social networks within organizations and how informal structures affect power distribution and labor arrangements; and 3) entrepreneurship and the transmission of control power from founders to successors.

Instructor's contact information：
Room 441, New GSM; Tel: 6275-6267;
Email: libb@gsm.pku.edu.cn

TA's contact information::

Office hour:

Program Learning Goals and Objectives

1 **Learning Goal 1** Graduates will be thoroughly familiar with the specialized knowledge and theories required for the completion of academic research.
- 1.1 Objective 1 Graduates will have a deep understanding of basic knowledge and theories in their specialized area.
- 1.2 Objective 2 Graduates will be familiar with the latest academic findings in their specialized area and will be knowledgeable about related areas.
- 1.3 Objective 3 Graduates will be familiar with research methodologies in their specialized area, and will be able to apply them effectively.

2 **Learning Goal 2** Graduates will be creative scholars, who are able to write and publish high-quality graduation dissertation and research papers.
- 2.1 Objective 1 Graduates will write and publish high-quality graduation dissertation and research papers

2.2 Objective 2 Graduates will be critical thinkers and innovative problems solvers.

3 **Learning Goal 3** Graduates will have a broad vision of globalization and will be able to communicate and cooperate with international scholars
- 3.1 Objective 1 Graduates will have excellent oral and written communication skills
- 3.2 Objective 2 Graduates will be able to conduct efficient academic communication in at least one foreign language

4 **Learning Goal 4** Graduates will be aware of academic ethics and will have a sense of social responsibility.
- 4.1 Objective 1 Graduates will have a sense of social responsibility.
- 4.2 Objective 2 Graduates will be aware of potential ethical issues in their academic career.
- 4.3 Objective 3 Graduates will demonstrate concern for social issues.

Overview

This course covers advanced statistical modeling and data management techniques that are fundamental for quantitative analysis of large management data. It is designed for strategy, macro OB, and marketing modeling IPHD students who intend to use large public data for their research. Statistical topics include (but not limited to) models for categorical (logit/probit, multinomial, and ordered), count (poisson), and limited (censored, truncated, and sample selection) dependent variables, panel analysis, multilevel models, and survival analysis. Statistical methods will be discussed along with processing and handling of public firms data from China (e.g., CSMAR and WIND) and the US (e.g., Compustat & CaptitalIQ) to answer real research questions. Throughout the course, we will highlight the connections among statistics, data, theory, and intuition. Students are required to choose a research topic and a real dataset in the beginning and to go through of the entire process of a research project, from hypothesis development, data handling and transformation, data exploration and modeling, result reporting and interpretation, to manuscript writing. Students completing this course are expected to develop a solid foundation of large data analysis and the ability of seeing through complex data.

Requirements

The discussion of each statistical model will have three components: 1) a mathematical component that cover the statistical foundation of the model; 2) a computing component that works out the mechanisms of the model; and 3) a data component that demonstrates the applications of the model. We will adopt a learning by doing approach with extensive in-class computing exercises to help students better understand statistical procedures and applications. Therefore, this course is computing intensive, and students are required to bring their own laptop computers to class.

To fulfil the learning goals, each student is required to work on a real research project throughout the course. The research project requires students to develop a serious research idea from raw data into a publishable paper. Students should choose a real dataset that they can work on . Students can work on either their own data and ideas or recommended data with suggested research ideas; and they can also choose to work alone or with another fellow student as a team. Data and topics should be decided before Week 3, and students are expect to explore and analyze the data

step by step following the pace of the course. A publishable paper is required for each research project in the end of the course.
Because the course is heavy-loaded, it is recommended only for students who expect to be fully committed. Class attendance is critical, and missing a class would cause serious setbacks in course learning. Students are required to read assigned materials before each class and turn in homework assignments before the deadline; late assignments will not be graded.

Evaluation

Coursework includes homework assignments and a research project.

- Homework assignment (5%_10 =50%): There will be 10 homework assignments, each of which consists of mini problem sets on data management, statistical calculation, estimation and post-estimation, and model interpretation. Students will have a full week to work on each homework assignment.
- Research Project (50%) :) Research topic, data, and collaborator need to be decided by Week 3. The research project consists three stages:
  - data management 10%, from raw data to analysis-ready format, by Week 9;
  - data analysis 10%, model selection and estimation, by Week 15;
  - the final Paper 30%, by the end of the course.

Softwares and Webdrive

- STATA: STATA (www.stata.com) will be our official statistical software. Students should have STATA (Version 10 or later) installed in their computers.
- LATEX: LATEX is the de facto standard of professional publishing, ideal for typesetting technical materials such as mathematics, graphics, and tables. We will use LATEX mainly for formatting STATA results into tables.
  LATEX for Mac OS: http://www.tug.org/mactex/
  LATEX for Windows: http://www.miktex.org
- Mendele: Mendeley (www.mendeley.com) is a free reference manager and academic social network. It is highly recommended.
- WebDrive: The class webdrive hosts all course related materials, e.g., raw data, lecture notes, additional reading materials, sample STATA programs, etc.. It is recommended that students download the desktop version and synchronize webdrive folders regularly.

Course Materials

- 王天夫、李博柏，2008 年，[STATA 实用教程]，中国人民大学出版社。

- STATA Corp, 2011, STATA 12 PDF Documentation.
- Bobai Li. Applied Regression Modeling (Book manuscripts and lecture notes)
- J. Scott Long. and Jeremy Freese, 2006. Regression Models for Categorical Dependent Variables Using Stata. Stata Press.
- StataCorp. 2010. An Introduction to Survival Analysis Using Stata. Stata Press.

Tentative Class Schedule

- Week 1-4: Statistical Computing and Data Management with STATA
- Week 5-7: Models for Categorical Outcomes
- Week 8-9：Count Models
- Week 10-11: Censored and Truncated Regression and Sample Selection Models
- Week 12-14: Panel Models and Multi-level Models
- Week 15-18: Survival and Event History Analysis