

Applied Statistics and Data Analysis in Management Science B8130-060.

Elective course for the Masters in Management Science and Engineering
IEOR/DRO, Columbia University.

Professor Marcelo Olivares
Decision, Risk and Operations
Columbia Business School
marceloolivar@gmail.com
<http://www.marcelo-olivares.com>

Lectures:

Tuesday 7:10-9:40pm

Tuesday sessions on 2/3, 2/10, 2/24, 3/3. MUDD 1024.

Make-up classes on 2/13 14:15-16:45 (Uris 330) and 2/27 13:00-15:30.

Course summary

“Applied Statistics and Data Analysis” is a hands-on data analysis class where students can learn a broad set of statistics methodologies and econometric models that are useful to analyze management decisions. Through a problem-driven approach, each method is motivated through a real-world example, followed by a rigorous treatment of the theory and mathematical models and the implementation of these methods using software (Stata, R and Excel). The applications include problems related to operations management, marketing and production economics, working hands-on with real data on a series of assignments and in-class exercises.

Grading

The evaluations for the course include computer assignments (40%), a take-home final exam (50%) and class participation (10%). Computer assignments can be done in groups; the other grades are individual.

Course Materials

Most of the course material is self-contained in the class notes provided in class. There is no single textbook covering all the material covered; recommended readings will be given from the following textbooks:

- (1) John A. Rice, *Mathematical Statistics and Data Analysis*, Editorial Thomson Brooks/Cole, Third Edition.
- (2) Jeffrey Wooldridge, *Introductory Econometrics: A Modern Approach*, 2006. Thomson 3rd Edition.
- (3) Joshua Angrist and Jorn-Steffen Pischke, *Mostly Harmless Econometrics*, 2008, Princeton University Press.
- (4) James Lattin-J.Douglas-Paul E.Green, *Analyzing Multivariate Data*, Editorial Brooks/Cole.

We will also cover several cases, which will be distributed in class.

Class topics

Session 1. Introduction to Statistical Inference

Tuesday Feb 3, 2015

Topics

- Random Sampling, Properties of estimators
- Methods con construct Confidence Intervals
- Framework for hypothesis testing
- Testing differences in mean and variance.

Applications: Measuring productivity and quality in services.

Prepare:

- Read the article “Xbox Polling and the Future of Election Prediction”, Harvard Business Review.
- What are the pros and cons of the approach suggested by the article relative to random sampling?

Recommended reading: (1) Chapters 5, 6, 9.1-9.3, 11.

Session 2. Statistical Process Control and ANOVA

Tuesday Feb 10

Topics

- Analysis of Variance (ANOVA)
- Control charts in Statistical Process Control: P-chart and X-bar chart.

Applications:

- Analyzing market research field experiments
- Statistical Process Control in retail stores using video analytics data.

Prepare:

- Read the case “Improving Customer Engagement in Retail” and download the data accompanying the case.

Recommended readings: (1) Chap 12; (4) Chap 11.

Session 3. Linear Regression
Friday Feb 13, 2015

Topics

- Conditional Expectations, Regression and Ordinary Least Squares.
- Building a regression model.
- Regression diagnosis.

Application:

- Hedonic pricing in housing and wine markets.

Prepare:

- Read "Bordeaux Wine Vintage Quality and the Weather".

Assignment due: Problem set 1.

Recommended reading: (3) Chap 3; (2) Chap 2-5.

Session 4. Causal analysis with regression
Tuesday Feb 24, 2015

Topics:

- Difference-in-difference and panel data methods
- Regression with Instrumental Variables

Applications:

- Impact evaluation.
- Measuring the effect of weather on productivity.
- Measuring the effect of inventory levels on sales.
- Estimating price elasticity with market level data.

Prepare:

- Read the paper "Quantifying the Influence of Climate on Human Conflict", Hsiang, Burke and Miguel, Science 2013. (i) What are the problems of estimating the effect of weather using a cross section of countries? ; (ii) What is a "bad control"?
- Read the paper "The Case in Support of Restaurant Hygiene Grade Cards", Jin and Leslie, Choices 2005. What is the main problem of simply using the difference in hospitalization rate before and after the grade card introduction to estimate its effect? (i.e. not using a control group).

Recommended readings: (3) Chap 3, 4.1-4.3, 5.1 and 5.2.

Session 5. Probabilistic Models for Consumer Analysis
Friday Feb 27, 2015

Topics:

- Introduction to parametric probability models
- Maximum Likelihood Estimation
- Modeling count data: Poisson and Negative Binomial Regression
- Duration models in continuous and discrete time.
- Criteria for model evaluation and selection

Applications:

- Analyzing website visits
- Modeling customer retention and churn.

Prepare: TBA

Assignment due: Problem set #2.

Recommended readings: (1) Chapt 8.1-8.5, 9

Session 6. Customer choice modeling
Tuesday March 3, 2015

Topics:

- Logistic regression
- Multinomial logit choice models
- Economic interpretation of Logit and MNL models.
- Extensions of the MNL model.

Applications:

- Analyzing brand choice with scanner panel data.
- Modeling demand for loans.

Prepare: Read the case “Nomis Solutions (B)”

Recommended reading: (2) Chap 17.1; (4) Chapt 13

Problem set #3 due on March 6.