

**Applied Statistics and Data Analysis in Management Science**

Professor Marcelo Olivares  
Decision, Risk and Operations  
Columbia Business School

Tuesday 7:10-9:40pm  
Tuesday sessions on 1/27, 2/3, 2/10, 2/24, 3/3.  
Additional class on 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 (35%), a take-home final exam (50%) and class participation (15%). 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:


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

1. Review of Statistical Inference
   - Sampling
   - Limit theorems: Law of Large Numbers and the Central Limit theorem
   - Sampling error and confidence intervals
   - Hypothesis testing
   - Applications: hypothesis testing with survey data, comparing samples, statistical process control.

2. Regression
   - Conditional Expectations, regression and Ordinary Least Squares.
   - Building a regression model.
   - Regression diagnosis.
   - Applications: hedonic pricing.

3. Regression for causal analysis
   - Difference-in-difference and panel data methods
   - Regression with Instrumental Variables
   - Applications: Impact evaluation (e.g. measuring the effect of implementing a buy-online pick-in-store retail channel); Estimating price elasticity from historical data.

4. Probabilistic models for consumer analytics
   - Maximum Likelihood Estimation
   - Analysis of count data
   - Logistic regression
   - Choice modeling
   - Applications: Analyzing website visits; Demand models of discrete choice; Conjoint analysis.

5. Multivariate data analysis
   - Principal components and Exploratory factor analysis
   - Clustering
   - Applications: Interpreting survey data; customer segmentation.