Introduction into Applied Statistics and Econometrics

Instructor: Dmitry Arkhangelsky (darkhang@stanford.edu)

Prerequisites: I expect students to have some previous exposure to probability theory (random variables, conditional expectations, multivariate normal distribution, limit laws) and mathematical statistics (decision theory, estimation, testing). Knowledge of basic statistical methods (linear regression, maximal likelihood) will be very helpful. Some programming experience with R will be necessary, but we will organize additional sections for students who don't have this experience.

Short description: There are two main objectives of this course: the first one is to introduce modern statistical methods and their applications to econometrics, the second one is to discuss philosophy, different interpretations and limitations of statistical and econometric methods. In particular, we will discuss model selection and model misspecification.

Goals: At the end of the course I expect students to be able to perform contemporary data analytic tasks in R, understand (on non-technical level) modern econometric methods and discuss methodology of current applied papers.

Details and requirements: Course will consist of 16 lectures that I will read during the first 5 weeks of the Winter semester. There will be lecture notes that you should read **before** the lecture (starting from the second lecture). There will be 4 weekly home assignments (first one due at the beginning of the second week) and one final project. Preliminary devision of weights is the following: 60% on HAs (15% each) and 40% on the final project. Projects should be carried out in groups, optimal (for me) size of the group will be announced once I will have enrollment data. Each group will present its project during the final week of the class.

Literature:*

- Statistics: Theory of point estimation by Lehmann and Casella; Testing statistical hypotheses by Lehmann and Romano; The Elements of Statistical Learning by Hastie, Friedman and Tibshirani
- Econometrics: Econometrics by Bruce Hansen; Causal Inference for Statistics, Social, and Biomedical Sciences by Imbens and Rubin; lecture note by Guido Imbens (MGTECON 640)
- Lecture notes and various papers

*This list will be updated at the beginning of the course

Preliminary Course Outline:

Lecture 1	Why should we care about statistics? Main goals of statistics. Probability theory, statistical models and real life. Basic statistical decision theory: point estimation and testing. Risk, admissibility, minimax and Bayesian approach;
Lectures 2-3	Prediction and classification problems: bias-variance trade-off, loss functions, model selection. Estimation of risk: SURE, Mallow's CP, cross-validation. Variance estimation: bootstrap.
Lecture 4	High-dimensional statistics: large p problems, sparcity, different basises, regularization, relation to Bayesian inference. Why should we care about high-dimensional problems?
Lecture 5	Linear prediction methods: linear regression and ridge regression; splines and kernels.
Lecture 6	Nonlinear methods: decision trees and random forest. LASSO and Gradient Boosting.
Lecture 7	Linear classification methods: logit and LDA. Regularized logit;
Lecture 8	Causal inference and potential outcomes. Structural modeling. Reduced form approach vs. structural approach.
Lecture 9	Model misspecification: what do we estimate in reality? Econometrics vs. Statistics.
Lecture 10	GMM framework. Linear regression and IV examples. Non-linear LS and maximum likelihood.
Lectures 11-12	Semiparametric models and conditional moment restrictions.
Lecture 13	Model misspecification in GMM framework: what happens if some moment conditions aren't satisfied? Sensitivity to model misspecification.
Lecture 14	Program evaluation: unconfoundedness, two-stage LASSO and related methods
Lecture 15	Demand estimation: logit with random coefficients, BLP framework.
Lecture 16	Merger evaluation: Bertrand-Nash equilibrium, cost estimation, welfare.