QTM 521: Applied II

Contact Hours: 150 minutes of class meeting time, one 50-minute lab, plus 3 hours of regular out-of-class work required as preparation for in-class work per week throughout a full semester
Credit Hours: 4
Prerequisites: QTM 520 (Applied I)

Instructor:	Xxx
Semester	Spring 20xx
Meeting Time and Place:	xxx
Office:	xxx
Office Hours:	xxx
Email/Contact:	xxx
Course Website:	xxx
TA:	xxx

COURSE OBJECTIVES

This course is a continuation of QTM 520: Applied I. It will present the topics of Generalized Linear Methods, Principal component analysis, Multidimensional scaling, Factor analysis, Cluster analysis, Panel data methods, Machine learning techniques.

CLASS REQUIREMENTS

Grades will be based on

- homework assignments (25%)
- midterm exam (25%)
- final exam (30%)
- data analysis project (20%)

HOMEWORK

The homework assignments will consist of analytical problems and data analysis with and without computer work. There will be approximately 6 homework assignments. The homework must be word processed with tables and figures together in the document. If you are using R markdown, make sure to knit it and submit as pdf or html (pdf is preferred) file. Each assignment will have its due date indicated and should be submitted through Canvas. Usually, assignments are due before the class meeting (e.g., if class is Tuesday 11:30 am, the assignment is due by 11:29am on Tuesday morning). Any assignment submitted after the due date/time will earn 0 points. To accommodate unexpected circumstances, your lowest homework grade will be automatically dropped at the end of the semester. Working together on the homework assignments is encouraged, but you must write your own solutions. This includes the computer assignments. It is highly recommended that you make your solo effort on all the problems before consulting others.

EXAMS

The midterm and the final will be in-class exams. **No collaboration is allowed on the exams**.

DATA ANALYSIS PROJECT

The final term project is a group-based work. You will make a team of 4-5 students. 2-3 papers will be provided. Your team must choose one paper and replicate it. The details of the project will be discussed after Spring break.

LAB SESSIONS

Lab sessions will be run by TA. The purpose of the lab sessions is to learn the R programing and data analysis skills useful for regression and other empirical analysis. Students must bring a laptop to these sessions.

HONOR CODE

All students enrolled at Emory are expected to abide by the Emory College Honor Code. Any type of academic misconduct is not allowed which includes 1) receiving or giving information about the content or conduct of an examination knowing that the release of such information is not allowed and 2) plagiarizing, whether intentionally or unintentionally, in any assignment. For the activities that are considered to be academically dishonest, refer to the Honor Code: http://catalog.college.emory.edu/academic/policies-regulations/honor-code.html.

DISABILITY ACCOMMODATIONS

If you are seeking classroom accommodations or academic adjustments under the Americans with Disabilities Act, you are required to register with Office of Accessibility Services (http://accessibility.emory.edu). To receive academic accommodations for this class, please obtain the relevant letter and meet with me at the beginning of the semester. Students are expected to give two weeks notice of the need for accommodations.

REQUIRED TEXTBOOKS

- ~~[GG] Alan Gerber and Donald Green. Field Experiments. W.W. Norton~~
- [MHE] Joshua D. Angrist and Jorn-Steffen Pischke, *Mostly Harmless Econometrics*, Princeton University Press
- [ISL] Gareth James, Daniela Witten, Trevor Hastie, and Robert Tibshirani. Introduction to Statistical Learning. http://faculty.marshall.usc.edu/gareth-james/ISL/
- [CS] Cosma Shalizi. Advanced Data Analysis from an Elementary Point of View
- [IPD] Charles F. Manski, Identification for Prediction and Decisions

TENTATIVE COURSE SCHEDULE

~~Weeks 1 and 2: Experiments~~

- ~~Unbiased estimators with completely randomized design~~
- ~~Consistent estimators with completely randomized design~~
- ~~Unbiased estimators with blocked randomized design~~
- ~~Consistent estimators with blocked randomized design~~
- ~~Standard errors and confidence intervals~~
- ~~Natural experiments~~
- ~~Human subjects research [Ethics component]~~

Reading:
[GG] Chs. 1-4
[MHE] Chs. 1-2

Week 3: Instrumental Variables

Topics:

- Noncompliance in experiments
- "Noncompliance" in natural experiments
- Constant effects models
- Hetergeneous effects models
- Reporting standards for "broken" experiments [Ethics Component]

Reading:
[GG] Chs. 5-6
[MHE] Ch. 4

Weeks 4-6: Causal Inference Using Observational Data

Topics:

- Matching
- Panel Data
- Diff-in-Diff
- Fixed and Random Effects
- Regression Discontinuity
- Reporting standards for observational studies [Ethics Component]
- Reporting standards for subpopulation effects [Ethics Component]

Reading:
[MHE] Chs. 3, 5-6

Week 1-3: Data Acquisition, Software, Control & Management

- Acquiring Distributed Data and Scrapping
- Data Privacy [Ethics Component]
- Proprietary data and Fair Use [Ethics Component]
- Managing Large Datasets
- Version Control and Statistical Packages
- Documentation and Replication

Week 4: Introduction to Machine/Statistical Learning

Topics:

- Regression
- Classification
- Reporting standards [Ethics component]

Reading:
[ISLR] Chs. 1-4

Week 5: Resampling Methods

Topics:

- Cross-Validation
- Jackknife and Bootstrap

Reading:
[ISLR] Ch. 5

Week 6: Predictions versus Causal Inference

Topics:

- Measuring Prediction Quality
- Training and Evaluation
- Optimal Hold-Out Samples

Week 7: Predictions, Bounds and Decision Theory

Topics:

- Predictions for Decision Making
- Prediction for Causal Intervention
- Exploratory Predictions
- Fairness [Ethics Component]
- Algorithmic Bias [Ethics Component]

Reading:
[IPD] Chs. 1, 4, 11

Week 8: Machine Learning and Datafication

Topics:

- Latent Variables
- Measurement
- Classification
- Missing Data

Reading:
[IPD] Chs. 2

Week 9: Sparse Regression

Topics:

- Subset selection
- Shrinkage methods
- Dimensional reduction
- Multiple testing [Ethics component]

Reading:
[ISLR] Ch. 6

Week 10: Non-linearity

Topics:

- Polynomials
- Splines
- Kernel based methods
- Reporting standards [Ethics component]

Reading:
[ISLR] Ch. 7

Week 11: Decision Trees and Boosting

Topics:

- CART
- Random tree methods
- Reporting standards [Ethics component]

Reading:
[ISLR] Ch. 8

Week 12: Introduction to Unsupervised Learning

Topics:

- PCA
- Clustering
- Reporting standards [Ethics component]

Reading:
[ISLR] Ch. 10

Week 13: Factor Analysis

Topics:

- PCA (again)
- From PCA to Factor Analysis
- Factor Analysis
- Reporting standards [Ethics component]

Reading:

CS Chs. 15-16

Week 14: Time Series

Topics:

- Stationary Time Series
- Non-stationary time series
- Inference and forecasting
- Honest forecasting error [Ethics component]

Reading:

CS Chs. 23