PADP 7120: Data Applications

University of Georgia
Spring 2019
Tu, 3:30-6:15 PM, Candler 115
Professor Anastasopoulos

Office: 280E Baldwin Hall
Email: ljanastas@uga.edu
Office Hours: Tu, Th 12:15-1:15 PM and by appointment.

Course Overview and Objectives

This course is an introduction to statistical reasoning and thinking applied to understanding the policy world. The tools that we will learn in this course will help you not foster a better understanding of policy problems more broadly, but will help clarify the way that you think and reason through problems where data is involved. The topics that we will cover include: research design, measurement, causal inference and understanding how we measure and interpret relationships between events and concepts.

Required Text and Readings

The required texts for the course are:

Textbook: We will use an open source statistics book—you can download a free pdf copy here: *https://www.openintro.org/stat/index.php?stat_book=os*.

Diez, David M., Christopher D. Barr, and Mine Çetinkaya-Ründel. 2016. *OpenIntro Statistics*. 3rd Eds. www.openintro.org. (A hard copy can be purchased on Amazon for $25 or less.). Referred to in the syllabus as *DCM.*

Field, Miles and Field. 2012. *Discovering Statistics Using R*. (this will be used mostly for R exercise reference.). Referred to in the syllabus as *FMF.*

There will also be additional readings available on the course webpage.

Grading

Your course grade will be calculated as follows:

Attendance/Participation 5%
R Labs 5%
Problem Sets 10%
Group project 15%
Midterm exam 30%
Final exam 35%

Letter Grade	Score (% Correct)
A	90-100%
A-	88-89%
B+	84-87%
B	80-83%
B-	78-79%
C+	74-77%
C	67-73%
D	60-66
F	59 and below

Attendance is mandatory and class participation and class discussion is an essential part of this class.

R Labs

Throughout the semester, we will devote about 30 minutes of class time each week to getting acquainted with the statistical package **R.** During these labs, I will introduce a programming topic and then ask you to complete a short assignment in breakout groups of 3 or fewer students.
In preparation for the **R** Labs, please download and install:
1) The latest version of **R**: https://cloud.r-project.org/
2) The latest version of **RStudio**:
   https://www.rstudio.com/products/rstudio/download/.

Problem Sets

Your final grade will be based on 6 problem sets with assignments provided one week before the due date. These assignments will be taken from the required text. They must be turned on on time and late assignments will receive a grade of zero in the absence of a university-approved excuse.

Midterm Exam

There will be one exam for the class: an in-class midterm on **Tuesday, February 26th**. The midterm will be about ½ multiple choice and ½ short answer questions. You will need a simple calculator that can do addition, multiplication and division for both the midterm and the final. The midterm will include pieces of R code that you will be asked to interpret as well.

Group Research Project

The group research project will require you to work in groups of 2-4 on one of a few pre-selected topics that I will provide to you in early March. Groups may also choose their own topic but you must consult with me beforehand if you decide to pursue your own topic.

The group research project will require you to apply the methods that you will learn in the class to a policy problem. There are two main deliverables for the project:

1. A 5-10 page report.

2. A 5-7 minute presentation on your findings.

You will be required to present your findings to the class and submit your reports on **Tuesday April 23rd.** There is more detailed information about the group project here: https://docs.google.com/document/d/1TcznFd3AOgaPEa4W80qVl4_ZgONuEhvn2TVcvfTPe40/edit?usp=sharing.

GROUP PROJECT SIGNUP SHEET: Please sign up to be in a group within the next two weeks using this Google Document: https://docs.google.com/spreadsheets/d/124UWAWzUSG2e3rDVNCWHpj5GtyNGvh6PTYJdfquHF2A/edit?usp=sharing

Group Project Proposal

Shortly after the Spring Break on March 21st, you will be asked to submit a final group project proposal for me to review. The final group project proposal will be a 1-2 paragraph document with the following information:

1. The topic that you are proposing to work on.
2. The position that you plan to take.
3. Data/evidence that you will use to support your argument.

Key Dates

➔ **February 26**: Midterm
➔ **March 21:** Final group project proposal due.
➔ **April 23**: Group research projects and presentations due.

Course Calendar

I reserve the right to change the calendar to ensure that we spend enough time on each topic. If changes become necessary, they will be announced in class.

Week 1 January 15:
Course introduction and data concepts • DCM Ch. 1

R Lab 1: Introduction to R.
- FMF 3.1-3.3

Week 2 January 22:
Sources and nature of data • DCM Ch. 1

R Lab 2: Introduction to R: loading data.
- Read part 1 of this tutorial
  http://web.cs.ucla.edu/~gulzar/rstudio/basic-tutorial.html
- FMF 3.4.

Problem Set 1 Released

Week 3 January 15:
Experiments, observational studies and causality • DCM Ch. 1

R Lab 3: Loading and manipulating data (continued).
- Part 2 of tutorial: http://web.cs.ucla.edu/~gulzar/rstudio/basic-tutorial.html.
- FMF 3.5.

Problem Set 1 Due

Week 3 January 29:
Probability and Distributions • DCM, Chapters 2 & 3

Problem Set 2 Released

Week 4 February 5:
Foundations for inference • DCM, Chapters 4

R Lab 4: Data visualization.
- Part 4 of tutorial: http://web.cs.ucla.edu/~gulzar/rstudio/basic-tutorial.html

Problem Set 2 Due

Week 5 February 12:
Inference for numerical data • DCM, Chapter 5

R Lab 5: Hypothesis testing in R.
- FMF 7.4, 9.4-9.6.

Problem Set 3 Released

Week 6 February 19:
Inference for categorical data • DCM, Chapter 6

R Lab 6: Hypothesis testing in R for categorical data.

Problem Set 3 Due

Week 7 February 26:
Midterm

Week 8 March 5:
Introduction to linear regression I • DCM, Chapter 7

R Lab 7: Introduction to linear regression.

SPRING BREAK NO CLASSES March 12

Week 9 March 19:
Introduction to linear regression II • DCM, Chapter 7

R Lab 8: Introduction to linear regression, continued.

Problem Set 4 Released

Week 10 March 26:
Introduction to multiple regression I • DCM, Chapter 8

R Lab 9: Multiple regression.

Problem Set 4 Due

FINAL GROUP PROJECT PROPOSAL DUE: 3/26

Week 11 April 2:
Introduction to multiple regression II • DCM, Chapter 8

R Lab 10: Multiple regression.

Problem Set 5 Released

Week 12 April 9 :
Introduction to logistic regression • DCM, Chapter 8

R Lab 9 3/28: Logistic regression.

Problem Set 5 Due

Week 13 April 16:
Special topic: Propensity score matching • [Michael Anderson's (Berkeley) Lecture Notes](#)

R Lab 10: Propensity score matching.

Problem Set 6 Released

Week 14 April 23: Group project research presentations.

FINAL GROUP PROJECTS AND PRESENTATIONS DUE: 3/26

Problem Set 6 Due

Week 15 April 30: Group project research presentations.

Statement about Students with Disabilities

Students with special needs that require accommodation should notify me and the Office for Disability Services in the first two weeks of the course so appropriate arrangements can be made. All information and documentation of special needs is confidential.

Statement about Plagiarism and Academic Dishonesty

Students are responsible for maintaining the highest standards of honesty and integrity in every phase of their academic careers. The penalties for academic dishonesty are severe and ignorance of the policy is not an acceptable defense. See also https://ovpi.uga.edu/academic-honesty.