UCLA Anderson

# MGMT 298D: Healthcare Analytics

Spring 2021

## COURSE INFO


Instructor: Professor Elisa Long
elisa.long@anderson.ucla.edu


Office hours: Wednesdays 9:00–11:00 am

Course times: Thursdays 4:10–7:00 pm

Zoom link: https://ucla.zoom.us/j/99978307635?pwd=MjNQTjFPOHdhTjJLOG0vZTJqNUNrQT09

Website: https://ccle.ucla.edu/course/view/21S-MGMT298D-7


Teaching assistants: Robert Quan
robert.quan.2021@anderson.ucla.edu

Apurve Kumar
apurve.kumar.2021@anderson.ucla.edu


## COURSE DESCRIPTION

With US healthcare spending reaching 18% of GDP and the demand for health services continuing to increase, improvements in the quality and efficiency of healthcare delivery are needed. This course explores how analytics tools can be used to describe relationships, predict outcomes, and prescribe solutions to healthcare challenges. Classes will include a combination of lectures, data analysis and software labs, group break-out sessions, and student-led presentations. Although we will do some data analysis in R, this is not a comprehensive programming course, and *no prior experience with R is required*.

## TARGET AUDIENCE

Students interested in consulting, data science, pharma/biotech, medical devices, or non-profits. Although we will be mostly discussing healthcare applications, the general methodologies apply across a range of industries and business sectors. No prior experience in healthcare is needed.

## LEARNING OBJECTIVES

- Identify applications of healthcare data analytics and the trade-offs of various tools.
- Gain practical experience with developing quantitative tools and performing empirical analyses.
- Become familiar with working with data in R, and running linear and logistic regressions.
- Improve public speaking and presentation skills using remote technologies.

## CLASS PREPARATION

To prepare for each class, you should read the relevant sections of Modern Dive and any additional readings posted on CCLE. If you are new to data analysis and programming, I *highly recommend* completing the accompanying exercises in Modern Dive. All due dates will be clearly posted on CCLE.

## COURSE MATERIALS

### Readings

Two HBS cases may be purchased at https://hbsp.harvard.edu/import/817437
Links for all other readings are given below and on CCLE.

### Software

We will conduct in-class exercises using R Studio. Datasets will be posted on CCLE before class.
First, download and install R here: https://cloud.r-project.org/
Second, download and install R Studio here: https://www.rstudio.com/products/rstudio/download/

### R Resources

- "Statistical Inference via Data Science: A Modern Dive into R and the Tidyverse" https://moderndive.com/
  Free online book that includes many tutorials on data visualization and data analysis in R
- UCLA Intro to R seminar https://stats.idre.ucla.edu/r/seminars/intro/
- UCLA Intro to data visualization with `ggplot` https://stats.idre.ucla.edu/r/seminars/ggplot2_intro/
- R Graphics Cookbook https://r-graphics.org/
- The R Graph Gallery https://www.r-graph-gallery.com/index.html

## UCLA POLICIES

The UCLA Anderson Honor Code will apply at all times.
http://www.anderson.ucla.edu/Documents/areas/adm/web/AndersonHonorCode.pdf

All participants in the course are bound by the UCLA Student Conduct Code.
https://www.deanofstudents.ucla.edu/portals/16/documents/uclacodeofconduct_rev030416.pdf

The UCLA Center for Accessible Education (CAE) facilitates academic accommodations for enrolled students with documented permanent and temporary disabilities. Accommodations are designed to promote successful engagement in the UCLA academic experience. Please visit http://www.cae.ucla.edu/ for additional information or email me if you have questions.

## ZOOM ETIQUETTE

- Have a working webcam and microphone.
- Keep your video on, but mute your microphone (unless speaking) to avoid background noise.
- Use the Raise Hand and Chat features in Zoom, and the TAs and I will do our best to monitor these.
- Please keep your attention on the live class discussion, and save any online shopping, Netflix, and laundry for after class.
- Please be professional and respectful of your fellow classmates, the TAs, and instructor. We are all trying our best to make this classroom experience as valuable—and enjoyable—as possible. Your patience is greatly appreciated!
- Some additional tips: https://thewirecutter.com/blog/professional-video-call-from-home/

## FEEDBACK

In these difficult times, remote learning is challenging for everyone involved, and I will try my hardest to make the class a positive experience. In return, I ask for your undivided attention each week and engagement in class discussions, labs, and break-out sessions. If anything is unclear or if you have constructive feedback, please let me know as soon as possible, so that I may revise the course as we go along.

## ASSIGNMENTS AND GRADING

Course grades will be determined as follows:

| | |
|---|---|
| R Labs | 40% |
| Data Visualizations | 10% |
| Final Group Project | 20% |
| Peer Assessment | 10% |
| Class Participation & Discussion Boards | 20% |

### R Labs (40%)

We will have 8 classes with labs in R Studio. I will give a short demonstration on the methodology and then groups will have a break-out session to work together on the analysis in R. Groups should upload their code and solutions to a few questions to CCLE after class.

### Data Visualizations (10%)

There are 2 data visualization assignments, to be completed individually. You should create 1 Powerpoint slide showing a data plot (eg, line graph, bar chart, scatterplot, etc). You are encouraged to try using `ggplot` in R for this, or Excel is fine. I will provide several datasets for you to select from on CCLE.

### Final Group Project (20%)

Groups should apply one or more analytics tools from class to a real-world dataset (to be posted on CCLE), to test a particular hypothesis. Each group will present their analyses and supporting figures during class 10.

Since we are remote this year, I will assign groups of 4-5 students to try to accommodate time-zone differences. This will be your primary team throughout the quarter. After the last day of class, each student will submit a short peer evaluation of your fellow teammates.

### Peer Assessment (10%)

Given our online presence this year, the course deliverables are more heavily group-based. As such, each team member will assess the contributions of your fellow team members at the end of the quarter.

### Class Participation & Discussion Boards (20%)

Students should aim to attend every class session and be prepared to contribute to class discussions. I will post some advance discussion questions for any assigned readings, and students can contribute to the CCLE online discussion boards.

## COURSE OUTLINE

### Class 1 April 1 COVID analytics

**Aim**: To calculate various summary statistics of the COVID pandemic.
**Lab**: Install R and R Studio, import a dataset, explore dataframes.
**Reading (optional)**: Chen, Chevalier, Long (2021) "Nursing home staff networks and COVID-19", *PNAS*

### Class 2 April 8 Health insurance costs

**Aim**: To examine how age, sex, BMI, and smoking status relate to health costs.
**Lab**: Build a linear regression model in R using insurance claims data.
**Reading**: Moderna HBS Case https://hbsp.harvard.edu/import/817437

### Class 3 April 15 Health tech platforms

**Aim**: To compare physician ratings in online scheduling platform, and describe potential natural experiments in the data.
**Lab**: Use t-tests to compare ratings and explore data visualization with `ggplot`.
**Reading**: Zocdoc https://www.zocdoc.com/about/blog/tech/how-zocdoc-uses-data-to-make-better-products-for-patients/

### Class 4 April 22 Behavioral incentives

**Aim**: To evaluate health and employment outcomes in a randomized trial of a workplace wellness program.
**Lab**: Use t-tests to compare outcomes by treatment group and continue data visualization with `ggplot`.
**Reading**: Jones, Molitor, Reif (2019) "A Reason to be Skeptical of the Workplace Wellness Industry", *Scient Amer.*
**Reading:** Kwoh (2013) "When Your Boss Makes You Pay for Being Fat", *Wall Street Journal*

### Class 5 April 29 Hospital patient outcomes

**Aim**: To introduce causal estimation in healthcare.
**Lab**: Replicate a published study using regression discontinuity based on newborn birth weight and NICU assignment.
**Reading**: Sitka Telehealth https://medium.com/trust-sitka/telehealth-and-senior-living-c9fc471a936c

### Class 6 May 6 Predicting disease

**Aim**: To introduce nonlinear regression models with binary outcome variables.
**Lab**: Build a logistic regression model to predict cardiovascular health outcomes.
**Reading**: Quer et al (2021) "Wearable sensor data and self-reported symptoms for COVID-19 detection", *Nature Medicine*

### Class 7 May 13 Nonprofit analytics

**Aim**: To observe an application of analytics in a nonprofit organization providing suicide prevention for LGBTQ+ youth.
**Lab**: Calculate arrival and service times, by day of week, time of day, and modality (text vs call)
**Reading**: The Trevor Project https://www.thetrevorproject.org/about/innovation-the-trevor-project/
**Reading**: https://www.technologyreview.com/2021/02/26/1020010/trevor-project-ai-suicide-hotline-training

### Class 8 May 20 Race, socioeconomics, and health

**Aim**: To examine how healthcare utilization differs by socioeconomic factors, and identify the potential biases that data analytics in healthcare can introduce.
**Lab**: Merge multiple NHANES datasets and replicate a study on the relationship between income and exercise behavior.
**Reading**: Benjamin (2019) "Assessing risk, automating racism", *Science*
**Reading (optional)**: Obermeyer et al (2019) "Dissecting racial bias in an algorithm used to manage the health of populations", *Science*

### Class 9 May 27 Global health

**Aim**: To illustrate how natural experiments can provide opportunities to evaluate policy interventions.
**Lab**: : Develop a diff-in-diff analysis of country-level HIV outcomes before and after implementing PEPFAR.
**Reading**: Zipline HBS Case https://hbsp.harvard.edu/import/817437

### Class 10 June 3 Final presentations

Each team will give a 10-15 minute presentation on their analysis.

## COURSE SUMMARY

| Class | Date | Healthcare topic | Read before class | R lab in class |
|---|---|---|---|---|
| 1 | April 1 | COVID analytics | | Intro to R |
| 2 | April 8 | Health insurance costs | Moderna (HBS Case) | Linear regression |
| 3 | April 15 | Health tech platforms | Zocdoc (company blog) | A/B testing |
| 4 | April 22 | Behavioral incentives | Workplace wellness programs (Scientific American, WSJ articles) | Data visualization |
| 5 | April 29 | Hospital patient outcomes | Guest Speaker: Sitka Telehealth (Medium article) | Regression discontinuity |
| 6 | May 6 | Predicting disease | Wearable sensor data (Nature Medicine article) | Logistic regression, Classification |
| 7 | May 13 | Nonprofit analytics | Guest Speaker: The Trevor Project (MIT Technology Review) | Operations & queueing |
| 8 | May 20 | Race, socioeconomics, and health | Dissecting racial biases in AI (Science article) | Data merging, regression |
| 9 | May 27 | Global health | Zipline (HBS Case) | Diff-in-diff |
| 10 | June 3 | Final presentations | | |