DATA606 – SPRING 2018

Instructor: Jason Bryer, Ph.D.

Class Meetup: Wednesday 8:00pm to 9:00pm

Office Hours: By appointment

Email: jason.bryer@gmail.com

Course Description

This course covers basic techniques in probability and statistics that are important in the field of data analytics. Discrete probability models, sampling from infinite and finite populations, statistical distributions, basic Bayesian statistics, and non-parametric statistical techniques for categorical data are covered in this course. Each of these statistical concepts will be applied in a variety of real-world scenarios through the use of case studies and customized data sets.

Course Learning Outcomes:

By then end of the course, students should be able to:

- Understand the foundations of probability theory and perform basic probability calculations.
- Build basic stochastic models for commonly encountered business problems.
- Model situations involving uncertainty using appropriate probability distributions and conditional techniques.
- Explore and summarize data using descriptive statistics.
- Test hypotheses using classical and modern computational techniques.
- Construct estimators and calculate intervals using classical and modern computational techniques.

- Perform basic Bayesian statistical techniques for estimation and testing hypotheses.

Program Learning Outcomes Addressed By The Course:

- Business Understanding. Learn when probabilistic techniques apply to certain categories of business problems, discuss the sorts of solutions that are possible, and understand the limitations of these techniques.
- Foundational Math Skills. Explore and analyze data, build probabilistic and statistical models, construct estimators, and test hypotheses.
- Predictive Modeling. Learn foundational techniques that underlie predictive modeling algorithms, such as Naïve Bayes.
- Presentation. Complete and submit collaborative assignments using techniques from the course.

How Is This Course Relevant For Data Analytics Professionals?

Probabilistic techniques are the foundation of many data science applications from data exploration and visualization to outlier analysis, stochastic modelling, and data mining algorithms. This course will ensure that students have a strong understanding of these foundations.

Grading

- Homework (16%)
- Labs (40%)
- Data Project (20%)
- Final exam (18%)
- Meetup Presentation (5%)
- Getting Aquainted (1%)

Grade Distribution

Quality of Performance	Letter Grade	Range %	GPA
Excellent - work is of exceptional quality	A	93 - 100	4
Excellent	A-	90 - 92.9	3.7
Good - work is above average	B+	87 - 89.9	3.3
Satisfactory	B	83 - 86.9	3
Below Average	B-	80 - 82.9	2.7
Poor	C+	77 - 79.9	2.3
Poor	C	70 - 76.9	2
Failure	F	< 70	0

How This Course Works:

This course is conducted entirely online. Each week, you will have various resources made available, including weekly readings from the textbooks and occasionally additional readings provided by the instructor. Most weeks will have homework assignments to be submitted. There will also be a presentation required and a forum post introduction required. You are expected to complete all assignments by their due dates.

Meetup presentations will comprise the solution and presentation to the class of one of the suggested problems for study from the weekly materials (not the graded homework problems). Each student must present one problem throughout the semester. Problems are chosen by entering your name and problem in the Google Spreadsheet. Note there is a maximum of three presentations per meetup and presentations should be no more than five minutes. Additionally, prepare your presentation so that the slides or document (I suggest using R Markdown) will be shared on the course website. Problems are assigned first come, first served, so any problem not already chosen by another student is available.

Further details on each of these assignments will be available in Blackboard and/or this Github repository.

Textbooks

Required

Diez, D.M., Barr, C.D., & Çetinkaya-Rundel, M. (2015). *OpenIntro Statistics (3rd Ed).*

> This is an open source textbook and can be downloaded in PDF format here, from the OpenIntro website, or a printed copy can be ordered from Amazon.

Navarro, D. (2015, version 0.5). *Learning Statistics with R*

> This is free textbook that supplements a lot of the material covered in Diez and Barr. We will use the chapter on Bayesian analysis. You can download a PDF version, or buy a print copy from Lulu through the author's website.

Recommended

Wickham, H., & Grolemund, G. (2016) *R for Data Science*. O'Reilly.

> Most of this books is available freely online at r4ds.had.co.nz/ but can be purchased from Amazon.

Kabacoff, R.I. (2011). *R in Action*. Manning Publications.

> You can find a lot of the material in *R in Action* on Kabacoff's website, statmethods.net. You can receive 38% off using the ria38 promo code when ordering from here.

Wickham, H. *Advanced R.* Baca Raton, FL: Taylor & Francis Group.

> Most of this book is available freely online at adv-r.had.co.nz but can be purchased from Amazon.

Kruschke, J.K. (2014). *Doing Bayesian Data Analysis, Second Edition: A Tutorial with R, JAGS, and Stan (2nd Ed)*. London: Academic Press.

> This book can be purchased from Amazon, but also check out the author's webiste (doingbayesiandataanalysis.blogspot.com/) for additional resources.

Other Documents

- Probability Tables
- Inference Guide
- RStudio Cheat Sheets
- Probability Cheat Sheet

Contact

Office Hours (cell phone or using GoToMeeting): By appointment. You're encouraged to schedule an appointment, but you can try to call anytime.

You are encouraged to ask us questions on the "Github Issues" feature on the course repository where other students will be able to benefit from your inquiries. If you wish to ask a question in private, you can email the instructor directly.

For the most part, you can expect me to respond to questions by email within 24 to 48 hours. If you do not hear back from me within 48 hours of sending an email, please resend your message.

I will be checking in on the course regularly, just about every day and likely several times each day. Please do not hesitate to ask if you have questions or concerns.

Accessibility And Accommodations

The CUNY School of Professional Studies is firmly committed to making higher education accessible to students with disabilities by removing architectural barriers and providing programs and support services necessary for them to benefit from the instruction and

resources of the University. Early planning is essential for many of the resources and accommodations provided. Please see:

http://sps.cuny.edu/student_services/disabilityservices.html

Online Etiquette And Anti-Harassment Policy

The University strictly prohibits the use of University online resources or facilities, including Blackboard, for the purpose of harassment of any individual or for the posting of any material that is scandalous, libelous, offensive or otherwise against the University's policies. Please see:

http://media.sps.cuny.edu/filestore/8/4/9_d018dae29d76f89/849_3c7d075b32c268e.pdf

Academic Integrity

Academic dishonesty is unacceptable and will not be tolerated. Cheating, forgery, plagiarism and collusion in dishonest acts undermine the educational mission of the City University of New York and the students' personal and intellectual growth. Please see:

http://media.sps.cuny.edu/filestore/8/3/9_dea303d5822ab91/839_1753cee9c9d90e9.pdf

Student Support Services

If you need any additional help, please visit Student Support Services:

http://sps.cuny.edu/student_resources/