Final Meetup

DATA 606 - Statistics & Probability for Data Analytics

Jason Bryer, Ph.D.

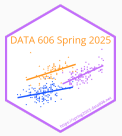
May 7, 2025

Final Exam

- Is now available on Brightspace.
- Due by midnight May 11th.
- You may use your book and course materials.
- We expect you to complete the exam on your own (i.e. do not discuss with classmates, colleagues, significant others, ChatGPT, etc.)
- There are two parts:

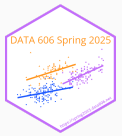
    1. Part one multiple choice questions and short answer questions.
    2. Part two has a small data set to analyze with R, then answer some interpretation questions.

- Put your answers in the Rmarkdown file and submit the PDF file. **Please do not post your answers online!**



Announcements

- You should join the New York Open Statistical Programming Meetup group. https://nyhackr.org
    - They meet monthly, usually at NYU. George Hagstrom has been organizing a group of MSDS students meeting up there each month.
    - Their annual conference will be August 26th and 27th: https://nyhackr.org/events.html
- The Joint Statistical Meeting will be August 2nd to 7th in Nashville: https://ww2.amstat.org/meetings/jsm/2025/
- useR! will be virtual August 1st and in person (Duke University) August 8th to 10th: https://user2025.r-project.org



Propensity Score Analysis

My statistical research interest is in propensity score methods. Propensity score analysis (PSA) is a quasi-experimental design used to estimate causality from observational studies.

Here are some resources for PSA:

- PSA Github repository includes slides slides and Shiny application: https://github.com/jbryer/psa
- Early version of an Intro to PSA book: https://psa.bryer.org
- Recording of a talk given in Fall 2023 for the NYC Meetup group here: https://www.youtube.com/watch?v=JLV4mtFhRMM

`multilevelPSA`

Multilevel PSA

`PSAboot`

Bootstrapping PSA

`TriMatch`

Matching with non-binary treatments

`PSAgraphics`

Graphical analysis of PSA

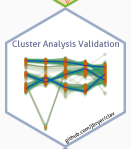
R Packages

Here is list of some other R related projects I have worked on:

`likert` - Analysis and Visualization of Likert Based Items

`ShinyQDA` - R Package and Shiny Application for the Analysis of Qualitative Data

`clav` - Predictive modeling with missing data

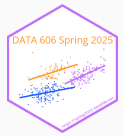


`clav` - Cluster Analysis Validation

`IRRsim` - An R Package for Simulating Inter-Rater Reliability

`mldash` - Machine Learning Dashboard

AmplifyApp, dashboard, and Future Mapping NYC

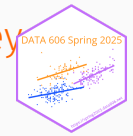


DAACS

The Diagnostic Assessment and Achievement of College Skills (DAACS) is a suite of technological and social supports to optimize student learning. DAACS provides personalized feedback about students' strengths and weaknesses in terms of key academic and self-regulated learning skills, linking them to the resources to help them be successful students. This is currently supported by a five-year $3.8 million grant received in 2021 from the Institute of Education Sciences to test the efficacy at three institutions.

Applications of Data Science:

- We use natural language processing and predictive models to machine score the essays.
  - We had a student this semester work with us to explore whether we can detect AI generated essays specific to the DAACS writing prompt (answer: we)
- We use DAACS data to estimate "risk scores" for students failing so we can target them with resources to help them be successful.
  - Related to this, we have developed a new R package for estimated predictive models with missing data, see medley



Thank You

This has been a great semester. Please don't hesitate to reach out:

Email: jason.bryer@cuny.edu

Github: https://github.com/jbryer

Personal Website: https://bryer.org

LinkedIn: jasonbryer

Mastodon: @jbryer@vis.social

You can download all course materials on Github. Click the clone or download link to download a zip file.