DACSS 603
Introduction to Quantitative Analysis
Spring 2023

Instructor: Dr. Rosemary Pang
Email: mrpang@umass.edu
Office Location: Bartlett 263 or Zoom

Classroom Assistant: Mekhala Kumar
Email: mjkumar@umass.edu

Course Time and Location:

Tuesday & Thursday 10:00 - 11:15	Multimodal	Machmer W-13 or Zoom
Evening session Thursday 19:00 - 20:15	Online	Zoom

Students do not have to attend the evening sessions if they attend the morning sessions (but those who want are welcome to do so!) and there might be some overlap in material across morning and evening sessions. All class sessions will be recorded and be made available to all students.

Office Hours:

Regular office hours will be held 2:00 PM - 4:00 PM Tu&Th in Bartlett 263 or on Zoom. Please book in advance through Calendly. Make sure you choose the right course and summarize the question you have. If this time does not work, please send me an email for appointment.

Course Description:

This course serves as a rigorous introduction to quantitative empirical research methods, designed for doctoral students in the social sciences and master's degree students with a data analytics or computational social science focus.

Our anchoring framework will be regression (linear and logistic) as tools to explore and describe associations among variables, to predict outcomes of interest, to generalize beyond a sample, and as part of a broader strategy of causal inference. Simulations and data analysis will be conducted in the R statistical environment.

Learning Objectives:

Upon completion, students will be able to understand, evaluate, and criticize the use of common statistical methods for social science research. The course will also serve as a useful basis for more advanced research methods courses.

Prerequisites:

Basic knowledge of algebra, sets, functions, and probability is assumed. More advanced mathematics will be introduced as needed for those who may not have encountered such material previously

(and as review for others), but we will focus on conceptual understanding and enhancing our ability to read mathematical notation in papers. It is best to have had some exposure to R prior to this class, preferably at the level of DACSS 601 Introduction to Data Science, but it is possible to succeed without this background if you are willing to put in extra work towards the beginning of the semester.

Textbook:

The main textbook is a free online book:
[EwR] Hanck, C., Arnold, M., Gerber, A. & Schmelzer, M. (2021) *Introduction to Econometrics with R*.

Other books that we will use excerpts from:
[LSR] Navarro, Danielle. *Learning Statistics with R*.
[ISLR2] James, G., Witten, D., Hastie, T., & Tibshirani, R. (2021). *An Introduction to Statistical Learning*.
[ALR] Weisberg, S. (2014). *Applied Linear Regression*
[SMSS] Agresti, Alan. *Statistical Methods for the Social Sciences*. 5th Edition. Pearson, 2018. (pdf of relevant chapters in Google Classroom)
[RaOS] Gelman, A., Hill, J., & Vehtari, A. *Regression and Other Stories*. Cambridge Univ. Press, 2020. (pdf of relevant chapters in Google Classroom)

Google Classroom and Slack:

All classroom material will be posted in Google Classroom. Class recordings, slides, tutorials, and assignment descriptions will all be found here. Be sure you are logged into your UMass Google account.

The DACSS program has a Slack space that you are encouraged to join and use. You will be invited to join the # dacss603spring23 private channel later this week. If not, please contact the instructor through Slack or email. We encourage you to use that channel to ask questions– chances are that if you are running into an issue or are unsure about something, someone else is too!

You are encouraged to participate in these spaces!

Feedback and Questions:

Students use Google Form to provide feedback and ask questions about course material every week. The instructor will address these questions in the following week.

Course Structure and Grading:
Final grades will be based on:

- **Participation (10%):**
  It is imperative that students actively and regularly participate in class discussion both synchronously (when possible) and asynchronously. Participation does not need to reflect expertise; rather, students should seek to both ask and answer questions regularly and in equal proportion. The main form of asynchronous participation will happen through Slack, where

students are expected to regularly ask and/or answer questions about R, statistics, math, readings and so on.

- **Quizzes (10%):**
  There will be a short take-home quiz most weeks to ensure key concepts from that week are understood. They will serve as feedback for you and for me. They will be auto-graded, mostly multiple choice questions. Quizzes can also include questions from that week's tutorial.

  There will be an R tutorial most weeks to help you with mastering the R functions and concepts necessary to complete the upcoming homework assignment. Tutorials are not themselves graded but questions from the tutorial will appear in the quiz. Completing the tutorials will make it a lot easier for you to complete the homework.

- **Homework (40%):**
  There will be a homework assignment *approximately* every two weeks. The assignments will be made up of exercises to help you better understand concepts and methods covered during class. Collaboration is acceptable, but please write up your own answers; do not hand in identical written responses. The assignments should be completed using R and published in the online course blog using Quarto – a scientific and technical publishing system that allows for easily integrating narrative text, equations, figures, code, and output from code into a single document. Quarto not only renders qmd files, its native format, but is able to render most R Markdown (rmd) and Jupyter notebook (ipynb) files. Authoring your documents in qmd is recommended. Detailed instructions on how to publish to the course blog are provided in Google Classroom. For information on Quarto visit https://quarto.org/.

- **Final Project (40%):**
  The final project will be in the form of a poster presentation of a data analysis project. In-person students will have the opportunity to join a DACSS poster session with students from other classes. Online students who cannot join the in-person poster session should instead record their presentations. The poster should be accompanied by a replication folder that includes the code and data to create the figures and tables in the poster, ideally put together in a .qmd file.

  The final project may be done in groups of maximum two people. Students who want to work as a group should let the instructor know in the first few weeks of the semester by email. Those working as a group should clearly identify each person's contribution at each stage of the project.

  It is important that students make progress on the final project throughout the semester. To ensure that this is the case, students will submit their progress in two check in assignments during the semester and use feedback from them to prepare the final poster

  More details on the final project and the check-in assignments can be found in Google Classroom.

- **Grade Scale:**
  A: 94-100; A-: 90-93; B+: 86-89; B: 81-85; B-: 77-80; C+: 74-76; C: 70-73; FAIL: Below 70

Software:

The course will be taught in the R programming language and the RStudio IDE. R is a free software environment for statistical computing and graphics; it has become the standard in statistics and among increasing numbers of social scientists and policy analysts. Aside from the advantage of being completely free of charge and open-source, R is versatile and powerful, with new packages becoming available all the time for handling tasks central to data analytics and computational social science (e.g. Bayesian inference, social network analysis, text analysis). If you are not familiar RStudio, please find more information in the **Welcome Letter**.

University Policies:

- ACADEMIC HONESTY:
  Since the integrity of the academic enterprise of any institution of higher education requires honesty in scholarship and research, academic honesty is required of all students at the University of Massachusetts Amherst.

  Academic dishonesty is prohibited in all programs of the University. Academic dishonesty includes but is not limited to: cheating, fabrication, plagiarism, and facilitating dishonesty. Appropriate sanctions may be imposed on any student who has committed an act of academic dishonesty. Instructors should take reasonable steps to address academic misconduct. Any person who has reason to believe that a student has committed academic dishonesty should bring such information to the attention of the appropriate course instructor as soon as possible. Instances of academic dishonesty not related to a specific course should be brought to the attention of the appropriate department Head or Chair. The procedures outlined below are intended to provide an efficient and orderly process by which action may be taken if it appears that academic dishonesty has occurred and by which students may appeal such actions.

  Since students are expected to be familiar with this policy and the commonly accepted standards of academic integrity, ignorance of such standards is not normally sufficient evidence of lack of intent.

  For more information about what constitutes academic dishonesty, please see the Dean of Students' website: `https://www.umass.edu/honesty/`.

- STATEMENT ON DISABILITIES:
  The University of Massachusetts Amherst is committed to making reasonable, effective and appropriate accommodations to meet the needs of students with disabilities and help create a barrier free campus.

  If you are in need of accommodation for a documented disability, register with Disability Services to have an accommodation letter sent to your faculty. It is your responsibility to initiate these services and to communicate with faculty ahead of time to manage accommodations in a timely manner. For more information, consult the Disability Services website.

Class Schedule and Readings

The schedule is tentative and subject to change. We may adjust the schedule due to time or interest.

Feb 7 & 9 Introduction and Descriptive Statistics
Syllabus
[LSR] Ch 5

Feb 14 & 16 Probability and Central Limit Theorem
[EwR] Ch 2
[LSR] Ch 9, Ch 10.1, 10.2, 10.3

Feb 21 & 23 Confidence Intervals and Hypothesis Testing
[EwR] Ch 3
[LSR] Ch 10.4, 10.5, 10.6, Ch 11

Feb 28 & Mar 2 T-test, ANOVA, and Chi-square
[LSR] Ch 13.1-13.7, Ch 14.1-14.3, Ch 12.2

Mar 7 & 9 Regression Background and Simple Linear Regression
[EwR] Ch 4, Ch 5.1, 5.2
[LSR] Ch 15.1, 15.2

Mar 14 & 16 No Class: Spring Recess

Final Project Check in 1 due on Mar 21

Mar 21 & 23 Multiple Regression I
[EwR] Ch 6, Ch 5.3

Mar 28 & 30 Multiple Regression II
[EwR] Ch 7

Apr 4 & 6 Transformations
[EwR] Ch 8

Apr 11 & 13 Assumptions, Diagnostics, and Model Evaluation
[RaOS] Ch 11
[SMSS] Ch 14.1
[ISLR2] Ch 2.2

Final Project Check in 2 due on Apr 20

Apr 18 & 20 Review

No class on Apr 18: Patriot's Day

Apr 25 & 27 Regression with Binary Dependent Variable

[EwR] Ch 11

May 2 & 4 Regression with Count Dependent Variable

[ALR] Ch 12

May 9 & 11 Regression with Categorical Dependent Variable

[SMSS] Ch 15.4-6

May 16 & 18 Wrapping Up for Poster Session