Advanced Statistical Modeling
Psych 2035 – Fall 2024

Lecture Time: Thurs 12:45-2:45pm
Location: WJH B6
Zoom Link (for auditors):
https://harvard.zoom.us/j/93369775662?pwd=iMsFbY5OvJvkgQELiTlkf1T4n7QhUm.1

Lecturer: Dr. Patrick Mair (mair@fas.harvard.edu)
Student Hours: Thurs 2:45pm (WJH 836)

Lab Sections: Fri 12-1:15pm (WJH B6) & Mon 3-4:15pm (WJH B6)
Zoom Links (for auditors):
https://harvard.zoom.us/j/96008459192?pwd=xg5YZiox1TwpiLmtb9VdaVAgHetPlV.1

Teaching Fellow: Dr. Joshua Cetron (jcetron@iq.harvard.edu)
Student Hours: Mon 4:15pm (CGIS Knafel Building, Room K333)

Course Description
Note that the advanced stats classes are in high demand. The maximum class size is 24 students. Should more than 24 students attempt to enroll, I will admit students according to level of priority. If you sign up for this course, I expect you to be committed to complete the class. By dropping the class after a few units you take away a spot for other students. If you can't fully commit to the class, there's always the option for remote auditing.

Course Objectives
The primary objective of this course is to acquaint students with a large set of statistical methodology relevant for their future research in various areas of psychology and related fields. Based on the characteristics of the data and the research questions, they will be able to pick proper methods, carry out the analysis in a leading computational framework, visualize the outputs, and interpret and communicate the results.

Course Website
You can access course website through canvas. We will use the course website extensively, so check it regularly. I will disseminate readings, make weekly announcements, and post links to additional material online.

Course Prerequisites
This course is intended for Harvard doctoral candidates in Psychology. It is required that students completed Psych 1950 and Psych 1952.

Course Requirements/Grading
Exercises
Weekly (optional) exercises will be posted on the website.
Midterm Project
A midterm project where you work on your own data project in R and write a report (teams of 2 students).
Deadline: Fri, Oct 25

Final Project
A final project where you work on your own data project in R and write a report (teams of 2 students).
Deadline: Tues, Dec 3

Do not take this course if you cannot complete either examination due to other (curricular or extracurricular) commitments.

Lecture Slides
I will post annotated lecture slides as soon as possible prior to lecture and advise you to prepare for lecture. That said, these annotated slides are not meant to stand alone: Some of

the material will only become clear after reviewing the readings, attending lecture, consulting online resources, or talking with your peers.

Resources:
Certain assignments in this course will permit the use of generative artificial intelligence (GAI) tools such as ChatGPT. The default is that such use is allowed unless otherwise stated. Any such use must be appropriately acknowledged and cited. It is each student's responsibility to assess the validity and applicability of any GAI output that is submitted; you bear the final responsibility. Violations of this policy will be considered academic misconduct. We draw your attention to the fact that different classes at Harvard could implement different AI policies, and it is the student's responsibility to conform to expectations for each course.

Readings:
Readings and supplemental resources will be announced via Canvas.

Accommodations for students with disabilities
Students needing academic adjustments or accommodations because of a documented disability should present their Faculty Letter from the Accessible Education Office (AEO) and speak with Dr. Mair by the end of the second week of the term. Failure to do so may result in the Course Head's inability to respond in a timely manner. All discussions will remain confidential, although the AEO invites Faculty to discuss appropriate implementation with them.

Week	Day	Date	Event	Unit	Topic
1	Thurs	09/05/24	12:45-2:45	Unit 1 No lab	Course Introduction, Advanced Regression I
2	Thurs	09/12/24	12:45-2:45	Unit 2 Lab 1	Advanced Regression II Advanced Regression
3	Thurs	09/19/24	12:45-2:45	Unit 3 Lab 2	Advanced Regression III Simple Cluster Analysis
4	Thurs	09/26/24	12:45-2:45	Unit 4 Lab 3	Parametric Clustering, Mixture Regression Mixture Regression
5	Thurs	10/03/24	12:45-2:45	Unit 5 Lab 4	Markov and Transition Models Markov Models
6	Thurs	10/10/24	12:45-2:45	Unit 6 Lab 5	Advanced Longitudinal Modeling I Advanced Longitudinal Modeling
7	Thurs	10/17/24		No class No lab	Midterm Project
8	Thurs	10/24/24	12:45-2:45	Unit 7 Lab 6	Advanced Longitudinal Modeling II Advanced Longitudinal Modeling
9	Thurs	10/31/24	12:45-2:45	Unit 8 Lab 7	Regularization Regularization
10	Thurs	11/07/24	12:45-2:45	Unit 9 Lab 8	Model-Based Recursive Partitioning Model-Based Recursive Partitioning
11	Thurs	11/14/24	12:45-2:45	Unit 10 Lab 9	Probabilistic Graphical Models Probabilistic Graphical Models
12	Thurs	11/21/24	12:45-2:45	Unit 11 Lab 10	Big Picture, Q&A Q&A
13	Thurs	11/28/24		No class No lab	Thanksgiving
		12/03/24			Final Project Due

Details:

- Unit 01: Advanced Regression I: influential outliers checks, classical robust regression, quantile regression, t-regression.
- Unit 02: Advanced Regression II: quantile regression, location-shape-scale modeling, GLM link functions, ordinal predictors.
- Unit 03: Advanced Regression III: censored regression, multivariate regression, regression with weights, Bayesian regression using INLA.
- Unit 04: Parametric clustering, mixture regression models: latent class/profile analysis, integrating clustering into regression fits.
- Unit 05: Markov and transition models: (hidden) Markov models, modeling transitions.
- Unit 06: Advanced longitudinal modeling I: basic time series analysis, vector autoregressive models (mlVAR)

- Unit 07: Advanced longitudinal modeling II: fixed effects models, cross-lagged panel models, dynamic process modeling (e.g., Gaussian processes, GIMME, dSEM).
- Unit 08: Regularization techniques: multicollinearity, lasso and ridge regression.
- Unit 09: Model-based recursive partitioning: decision trees for recursive partitioning, integrating partitioning into regression modeling.
- Unit 10: Probabilistic graphical models (Bayesian networks): classical directed acyclic graphs (DAG), cyclic graphs.