Syllabus

SOC 681: Latent Variable Modeling

Spring 2022

MW 12:30 – 1:45pm
BRNG 1245

Professor Mize
tmize@purdue.edu

Updated: 2022-04-17

Syllabus Contents (click hyperlinks below to jump to a section)

Course Information

- SOC 681: Categorical Data Analysis
  - CRN: 24276
  - MW 12:30 – 1:45pm
  - 3 credit hours
- An updated copy of this syllabus and all other course materials is always available on Brightspace

Instructor Information

- Dr. Trenton Mize, Assistant Professor of Sociology & Core Faculty Member of Advanced Methodologies of the Behavioral, Health, and Social Sciences (AMAP)
  - tmize@purdue.edu
  - Office: Stone Hall (STON) 326B
  - Office hours: Fridays 12:30 – 1:45pm
    - Book an office hours appointment here
      - Specify in-person or Zoom when booking
  - Preferred mode of contact: (1) When possible, attend my office hours to ask questions. If that is not possible, (2) please email me.
  - Preferred pronouns: he/him/his
  - Preferred name: Trent, Professor Mize, or Dr. Mize

Advanced Methodologies in the Behavioral, Health, and Social Sciences at Purdue (AMAP) Graduate Certificate

- AMAP has an interdisciplinary methods certificate
- This course is an approved quantitative course you can apply towards completion of the certificate

Attendance Policy

- You are expected to come to class unless you have a valid reason to miss class.
- You need a valid excuse to miss class. To ask for an excused absence, email me before the class period. I will excuse absences for:
  - Illness (doctors note required if you need to miss more than one class period)
  - Mandated quarantining
  - Religious observances
  - Traveling for academic events (e.g. a conference)
  - *Other valid reasons will be approved on a case-by-case basis

Course Description

Many constructs of fundamental interest to social scientists are not directly observable. A person's political ideology, their level of racial prejudice, someone's personality traits, the propensity to engage in risky behavior, the presence or absence of a mental illness, a person's ability, someone's health lifestyle, and more are all examples of latent variables.

To start, we will learn several different multivariate scaling techniques for modeling latent variables from multiple observed indicators. We will consider scaling strategies that allow continuous, count, ordinal, nominal, and binary observed indicators. We will also consider scaling techniques for both continuous and nominal latent variables. Specific techniques for scaling we cover include exploratory factor analysis, confirmatory factor analysis, latent class analysis, item response theory, and some variants of each of these. We also cover issues of measurement invariance and differential item functioning which applies to all of these methods.

Next, the class covers how to incorporate latent variables into the regression framework. We consider how to estimate and interpret a regression model when either the dependent or independent variable is latent. The course ends by covering methods for simultaneous model estimation. Simultaneously estimating models allows for comparisons of the effects of observed and/or latent variables across multiple models. We will also consider how multiple groups (observed or latent) can be incorporated into a simultaneous equation framework. These types of methods are useful, e.g., when the process being examined may differ across observed groups (e.g. men and women) or across groups that are latent.

Learning Objectives

1. Identify and differentiate the substantive distinctions between observed and latent variables
2. Be able to produce and interpret latent variable statistical models for continuous and categorical observed items
3. Understand how to apply latent variable modeling to substantive research questions and describe and present them to general interest academic audiences

Required Prerequisites

- Course
  - A course in linear regression (or ANOVA) is a required prerequisite
    - E.g. HDFS 590, POL 501, PSY 606, PSY 631, SOC 680, or STAT 512
  - A course on categorical data analysis (or generalized linear models) is a required prerequisite
    - E.g. SOC 681 or STAT 526
- Contact the instructor if you are unsure whether your previous coursework has prepared you for this course

Required Texts, Readings, and Software

Required Readings

- All readings will be provided on Brightspace as PDFs

Lecture Notes

- All lecture notes will be provided on Brightspace as PDFs
  - I prefer students to take notes by hand so I will bring printed versions of the notes for each student to each class. Those who prefer to view the notes electronically on a computer or tablet should inform me during the first week of class.

Books

- **Required book:**
  - Bartholomew et al. 2008. *Analysis of Multivariate Social Science Data*. 2nd Edition. Boca Raton, FL: Chapman and Hall/CRC.
    - e-copy available free through our library
- **Optional/supplemental books on specific topics:**
  - Bollen, Kenneth A. 1989. *Structural Equations with Latent Variables*. New York: Wiley.
  - DeVellis, Robert F. 2017. *Scale Development: Theory and Applications.* 4th Edition. Thousand Oaks, CA: Sage.
  - Collins, Linda M and Stephanie T Lanza. 2010. *Latent Class and Latent Transition Analysis.* Hoboken, NJ: Wiley.
  - Ayala, R.J. de. 2009. *The Theory and Practice of Item Response Theory.* New York, NY: Guilford Press.
  - Roos, J. Micah and Shawn Bauldry. 2022. *Confirmatory Factor Analysis.* Sage.
  - Finch, W. Holmes. 2019. *Exploratory Factor Analysis.* Sage.

Statistical Software

- The focus of the course is on the substance of the models. However, we will also extensively cover estimation and application. The examples in the lecture notes use Stata 17. It is helpful—but not necessary—to be familiar with Stata before the course. I will provide an "Introduction to Stata" guide the first week of class as a crash course for those who are new to Stata.
- **Stata 17**
    - Completing the assignments should be done in Stata as the examples in the lecture notes will use Stata; I also provide template Stata do-files for the assignments.
    - You can download Stata 17 for free through Purdue's software community hub
- **R and Mplus**
    - Most everything we cover in this class can be done in either R or Mplus. However, I am not equipped to support either program for class purposes. That is, I do not have code that accompanies the lecture notes for these programs nor do I have example code for the assignments. Those interested in resources for conducting latent variable analyses in R or Mplus should contact be outside of class time.

Assignments and Required Readings

Applied Data Analysis Assignments

Overview

- The primary source of grades for the course is a series of applied data analysis assignments. These will typically follow the types of models we are covering in class. For example, one assignment will focus on estimating, interpreting, and presenting a confirmatory factor analysis.
    - Assignments will be available on the course website before we begin covering the relevant topic. I encourage you to read over the relevant assignment before we begin covering those topics so that you can ask questions in class about any specific problems you may have as we cover that material.

Working Together

- I encourage you to work with your classmates as you complete the assignments. Seeing the unique findings and issues across different examples will be helpful as you learn the models. In addition, discussing interpretations and presentation strategies with others will improve the clarity of your understanding and writing. However, your work must be original—meaning that the specific models you estimate and interpret in the assignments cannot be identical (or even very close) to those of another student in the class.

Data for Assignments

- The assignments will be applied—meaning you will be using real health/social science data to estimate models and then interpret and present the findings.
- You should use a dataset that you are interested in using for your own work. Any dataset for the assignments must have a sample size of at least 500.
    - If you do not have a dataset you are currently working with come to my office hours. I have several I can recommend and/or give you.

Required Readings and Discussion

- For each section of the class there are one or two required readings. In general, there is no more than one required reading per week.
    - Readings may supplement course material, show a published example of research using the methods we are covering, etc.
- There are two required activities that accompany readings marked as *exemplars* (substantive articles that use the methods we are covering well)
    - A brief two question reflection due on Brightspace
    - In class discussion of the reading; all students are expected to participant in the discussions.
- Collectively, the reading reflections and in-class participation are worth 20% of your final grade

Grading

Course grades refer to the number of points towards your final grade you earn from each assignment or reading discussion. The course has 1,000 total points. Therefore, an assignment worth 100 points is worth 10% of your final grade.

- Exemplar reading reflections on Brightspace and in-class discussion of readings
    - Reflections worth 20 points each (7 exemplar readings * 20 = 140 points)
    - Regular participation in the in-class discussions of the readings is also required and worth 60 points
- Assignments
    - Each assignment is worth 100 points
        - 8 assignments * 100 points = 800 total points

Revising Assignments

Optional Revisions

- All assignments, regardless of your initial grade, can be revised within 10 days from when the assignment is graded. Revisions are optional; you do not have to revise any of your assignments that received a grade on the initial submission.
    - Revised assignments are eligible for a 10-point revision added to the initial grade.
    - If you choose to revise an assignment, you only have to revise your answers that were incorrect; make sure to be attentive to the comments you received on your original assignment. When revising an assignment, please use the Track Changes feature in Word so the changes made are clear. Upload revised assignments to Brightspace.

Reject & Resubmits: Required Revisions

- Occasionally, I will not give a grade but instead give a "Reject & Resubmit" which is a required revision. Reject & Resubmits require a revision in order to receive a non-zero grade for the assignment.
    - I assign Reject & Resubmits when the quality of the initial submission is well below course standards. This can be due to major problems but can also sometimes be a minor issue (e.g. miscoded dependent variable) that nonetheless populates the rest of the assignment with errors.
    - Your grade will not be penalized on your revision submission—i.e. you can still receive up to a full 100 points on the revision.

Missed or late assignments; incomplete final grade

- Late assignments will be accepted at my discretion and will include a penalty of 10 points per class period late
    - If you expect to have trouble meeting an assignment deadline because of a foreseeable event (e.g. you will be out of town for a conference or you will be observing a religious holiday)—please contact me as early as possible and we can arrange a plan for completing the assignment without penalty.
- A grade of incomplete (I) will be given only in unusual circumstances. To receive an "I" grade, a written request via email must be submitted prior to December 1, and approved by the instructor. The request must describe the circumstances, along with a proposed timeline for completing the course work. Submitting a request does not ensure that an incomplete grade will be granted. If granted, you will be required to fill out and sign an "Incomplete Contract" form that will be turned in with the course grades. Any requests made after the course is completed will not be considered for an incomplete grade.

Grading Scale

Your final grade will be based on the sum of your grade on all assignments and discussion participation which can range from 0 to 1,000 points. Your final point total will be converted to a letter grade using the following category criteria:

- ≥ 970.00 A+
- 930.00 – 960.99 A
- 900.00 – 920.99 A-
- 870.00 – 890.99 B+
- 830.00 – 860.99 B
- 800.00 – 820.99 B-
- 770.00 – 790.99 C+
- 730.00 – 760.99 C
- 700.00 – 720.99 C-
- 670.00 – 690.99 D+
- 630.00 – 660.99 D
- 600.00 – 620.99 D-
- ≤ 590.99 F

Note your final grade percent will not be rounded. I do not change a student's final grade for any reason.

Asking for Help

There are a few basic guidelines to follow when asking for help that will greatly aid my ability to help you with computing problems such as debugging code, recoding a variable, getting a model to estimate correctly, convergence problems, or other common issues. The most important principle is that you should provide me with the files necessary to reproduce the problem you are having. That is, instead of only sending problematic output, also send the data and coding files necessary to reproduce the problem. Keep in mind that these guidelines apply both when you are asking for help over email and when asking for help in person. Specifically:

1. Include basic descriptive statistics of your model variables. If you use Stata, **`desctable`** is an easy way to do this.
2. Send a small version of the dataset you are using that includes all necessary model variables, but excludes extraneous variables in the dataset. E.g. If you are using the GSS, do not send along a dataset with thousands of variables; if you are only using 10 variables for your analysis—only include those in the dataset you send.
3. Send a coding file (e.g. do-file for Stata) that includes only the necessary code to reproduce the problem you are encountering. I.e. This file should load the appropriate dataset and estimate any models and post-estimation necessary *for the specific issue you are asking for help with*. If your problem is with data management, include only the code relevant to recoding/manipulating the specific variable causing the problem. That is, you should not send the files for your entire assignment or project—only send what is necessary to reproduce your issue.
4. Send the relevant output that illustrates the problem and any errors messages you are encountering. E.g. This can be a Stata log-file or a text file. Again, only include output relevant to the problem you are experiencing.
5. To ensure that my email does not reject your message, do not send large files as email attachments. Instead, create a folder in Dropbox, Box, or Google Drive and include the necessary files. All of these services allow you to send a "shareable link" that will allow me to access the files.

Academic Integrity

- Purdue's Honor Pledge states: "As a boilermaker pursuing academic excellence, I pledge to be honest and true in all that I do. Accountable together - we are Purdue."
- I encourage you to study with other students in the class, to have them read over and comment on your notes or assignments, or to provide other feedback.
    - **However, any work you turned in for a grade needs to be completed solely by the student receiving the grade.**
    - For example, while it is acceptable—and encouraged—for you to ask another student to read an assignment or to share code for suggestions and comments, the assignments themselves must be written entirely by the student receiving the grade.
- Incidents of academic misconduct in this course will be addressed by the course instructor and referred to the Office of Student Rights and Responsibilities (OSRR) for review at the university level. Any violation of course policies as it relates to academic integrity will result minimally in a zero grade for that particular assignment. In addition, all incidents of academic misconduct will be forwarded to OSRR, where university penalties, including removal from the university, may be considered.

Additional Policies

Welcome statement

In this course, each voice in the classroom has something of value to contribute. Please take care to respect the different experiences, beliefs and values expressed by students and staff involved in this course. I support Purdue's commitment to diversity and welcome individuals of all ages, backgrounds, citizenships, disabilities, sexes, education levels, ethnicities, family statuses, genders, gender identities, geographical locations, languages, military experiences, political views, races, religions, sexual orientations, socioeconomic statuses, and work experiences.

Preferred names and pronouns

- The overall spirit of my policy is that I respect your decision to decide how you would like to be referred to. I expect class members to be similarly respectful.
- Specifically, I will honor the names and pronouns you provide, and your request at any point to address you by your correct name and/or gender pronoun. I also expect class members to honor the names and pronouns peers provide.

Discussion etiquette

We will discuss many topics in this class—some of which may be personal for some students. When discussing topics in class, you are encouraged to comment, question, or critique an idea, but you are not to attack an individual. Our differences, some of which are outlined in the University's nondiscrimination statement above, will add richness to this learning experience. Please consider that sarcasm and humor can be misconstrued (especially in online interactions) and generate unintended disruptions. Working as a community of learners, we can build a polite and respectful course ambience. Some etiquette rules for the course:

- Do not dominate any discussion. Give other students the opportunity to join in the discussion.
- Do not use offensive language. Present ideas appropriately.
- Avoid using vernacular and/or slang language. This could possibly lead to misinterpretation.
- Keep an "open-mind" and be willing to express even your minority opinion.
- Think and edit before you push the "Send" button or make an in-class comment.
- Do not hesitate to ask for feedback from me if you are unsure about something you wish to discuss.

Counseling and psychological services (CAPS)

Purdue University is committed to advancing the mental health and well-being of its students. If you or someone you know is feeling overwhelmed, depressed, and/or in need of support, services are available. For help, such individuals should contact Counseling and Psychological Services (CAPS) at (765)494-6995 and http://www.purdue.edu/caps/ during and after hours, on weekends and holidays, or by going to the CAPS office of the second floor of the Purdue University Student Health Center (PUSH) during business hours.

Disability Resource Center

Purdue University strives to make learning experiences as accessible as possible. If you anticipate or experience physical or academic barriers based on disability, you are welcome to let me know so that we can discuss options. You are also encouraged to contact the Disability Resource Center at: drc@purdue.edu or by phone: 765-494-1247. More details are available on our course Brightspace under Accessibility Information.

Course Schedule

Note that exact topics, readings, dates, and other aspects of the course schedule are subject to change. Always reference the syllabus on Brightspace for the most up to date version of the Course Schedule.

Required readings

- All required readings are to be completed **before the class on the day listed on the syllabus.**
- Readings marked as *exemplars* are substantive research articles that use the methods we cover. You can skim the theory and discussion sections of these articles; you should primarily focus on the Methods and Results sections.

Assignment due dates

- Assignment 1 (100 pts): (Jan 24 @ noon): Latent Variables, Scales, and Datasets of Interest in Your Field
- Assignment 2 (100 pts) (Feb 11 @ 5pm): Exploratory Factor Analysis
- Assignment 3 (100 pts) (Feb 23 @ noon): Confirmatory Factor Analysis
- Assignment 4 (100 pts) (March 4 @ 5pm): Item Response Theory—Binary Items
- Assignment 5 (100 pts) (March 11 @ 5pm): Item Response Theory—Ordinal, Nominal, and/or Count Items
- Assignment 6 (100 pts): (April 1 @ 5pm): Latent Class Analysis
- Assignment 7 (100 pts) (May 2 @ 5pm): Structural equation models

Part 1: Introduction

Week 1 (Jan 10 & 12)

- Mon: Overview of course
- Wed: Introduction to latent variables; scales vs indexes
    - ***Reading due:*** Bollen 2002 *ARPsyc*

Part 2: Summated Scales

Week 2 (Jan 19)

- Mon: **No class—MLK Day**
- Wed: Summated scales; test theory; measurement and assumptions
    - ***Reading due:*** Introduction to Stata

Week 3a (Jan 24)

- Mon: Scale validation; validity and precision; reliability and internal consistency
    - ***Assignment 1 due at noon***
    - ***Reading & reflection due:*** Kervyn et al. 2013 *EJSP* (exemplar)

Part 3: Principal Components Analysis & Exploratory Factor Analysis

- *Supplemental reading:* Bartholomew et al. 2008 Chs. 5 & 7

Week 3b (Jan 26)

- Wed: Principal components analysis

Week 4 (Jan 31 & Feb 2)

- Mon: Exploratory factor analysis; choosing the number of factors
- Wed: Multiple latent dimensions; orthogonal and correlated factors

Week 5 (Feb 7, 9, & 11)

- Mon: **Class cancelled (Prof Mize out of town)**
- Wed: Rotation in exploratory factor analysis
    - ***Reading & reflection due:*** Hahn and Belt 2004 *JHSB* (exemplar)
- Fri: ***Assignment 2 due at 5pm***

Week 6a (Feb 14)

- Mon: Factor scores

Part 4: Confirmatory Factor Analysis

- *Supplemental reading:* Bartholomew et al. 2008 Ch. 11 (first half)

Week 6b (Feb 16)

- Wed: Confirmatory factor analysis (CFA); model identification

Week 7 (Feb 21, 23, 25)

- Mon: Absolute model fit and diagnostics in CFA
- Wed: Modification indices; comparative model fit

Week 8 (Feb 28 & March 2, 4)

- Mon: Factor scores and scale reliability
  - ***Reading & reflection due:*** Agarwala and Lynch 2006 *SF* (exemplar)
- Wed: Higher-order factor analysis; multitrait-multimethod models
- Fri: ***Assignment 3 due at 5pm***

Part 5: Nonlinear Factor Analysis and Item Response Theory

- *Supplemental reading:* Bartholomew et al. 2008 Ch. 8 & 9

Week 9 (March 7, 9, 11)

- Mon: Nonlinear factor analysis (i.e. factor analysis with categorical indicators); item response theory (IRT)
- Wed: Tetrachoric and polychoric correlations

March 14 – 18

- **No class; Spring break**

Week 10 (March 21, 23, 25)

- Mon: Binary indicators in IRT; Rasch (one-parameter logistic) model
- Wed: Item characteristics curves; category characteristic curves

Week 11 (March 28 & 30, April 1)

- Mon: Two-parameter logistic model; marginal effects interpretations
  - ***Reading & reflection due:*** MacIntosh 1998 *ASR* (exemplar)
- Wed: Three-parameter logistic model
  - ***Assignment 4 due at 5pm***

Week 12 (April 4, 6, 8)

- Mon: Nominal, ordinal, and count indicators in IRT
- Wed: *No class; class moved to Friday*
- Fri: Hybrid IRT (class via Zoom)
  - ***Assignment 5 due at 5pm***

Part 6: Latent Class Analysis

- *Supplemental reading:* Bartholomew et al. 2008 Ch. 10

Week 13 (April 11, 13, 15)

- Mon: Choosing the number of latent classes in LCA; Naming and interpreting latent classes in LCA
- Wed: Model fit and diagnostics in LCA
    - ***Reading & reflection due:*** Scarborough et al. 2021 *ASR* (exemplar)
- Fri: ***Assignment 6 due at 5pm***

Part 7: Structural Equation Modeling

- *Supplemental reading:* Bartholomew et al. 2008 Ch. 11 (second half)

Week 14 (April 18, 20, 22)

- Mon: Latent variables as independent and dependent variables in regression analysis
- Wed: Path modeling; multiple groups analysis
    - ***Reading due:*** Davis 1985 Ch. 1
    - ***Reading & reflection due:*** Simons et al. 2011 *ASR* (exemplar)

Part 8: Measurement Invariance & Differential Item Functioning

Week 15 (April 25 & 27)

- Mon: Measurement invariance in CFA; Differential item functioning (DIF) in IRT
- Wed: Measurement invariance in LCA
    - ***Reading & reflection due:*** Perreira et al. 2005 *SF* (exemplar)

Week 16 (May 2)

- Mon: ***Assignment 7 due @5pm***