



## **CLASS SESSION(S)**

Tuesday 9:30-11:50

Mailman Building 6<sup>th</sup> floor computer lab

## **INSTRUCTOR**

Dr. Melanie M. Wall

Professor

Depts of Biostatistics and Psychiatry

mmwall@columbia.edu

## **COURSE DESCRIPTION**

This course is designed for those students (or any researchers) who want to gain a significant familiarity with a collection of statistical techniques that target the measurement of latent variables (i.e. variables that cannot be measured directly) as well as methods for estimating relationships among variables within causal systems. This course covers exploratory and confirmatory factor analysis, item response theory models, latent class and finite mixture models. The course also covers structural equation modeling including its directed acyclic graphical notation as a means to estimate assumed causal relationships in the presence of confounders, mediators and moderators. Data analysis examples will come from health science applications and practical implementation of all methods will be demonstrated using predominately the Mplus software, but also the R software.

## **PREREQUISITES**

A one year course in applied statistics at the level of 6103-6104 or permission of the instructor.

## **COURSE LEARNING OBJECTIVES**

Students who successfully complete this course will be able to:

- Understand the latent variable measurement problems that can and cannot be addressed using factor analysis and item response theory models
- Be able to implement factor analysis and item response theory models using software on real data and interpret results including checking model assumptions
- Understand the use of latent class and finite mixture models as methods for clustering individuals and be able to fully implement them using software including assessing model fit and interpreting results
- Understand the role of structural equation modeling (SEM) including its directed acyclic graphical notation as a means to estimate assumed causal relationships including the role of confounders, mediators and moderators.
- Identify and be able to estimate using software the total, indirect, and direct effects within a SEM as well as assessing model fit and checking model assumptions.
- Critically read research articles that use the collection of methods learned in the course.

## ASSESSMENT AND GRADING POLICY

Student grades will be derived from a combination of homeworks (50% of grade), one in-class exam (25% of grade) and a group final project (25% of grade including written and oral presentation).

**Homework:** Homework is due in class on date specified. Homework turned in after the time it is due will have 50% deducted. You may only turn in late homework up until the next homework is due, after that there will be no credit given.

**In-class exam:** The in-class exam will be open book, open notes but not open internet.

**Group Final Project:** Depending on the size of the class groups for final projects will be of size 2 or 3 and will be formed by self-selection. For all projects, a written report (6-8 pages not including graphs) outlining the problem, describing the analysis and summarizing the results will be required. The results will be presented to the class orally (15-minute presentation) during the last week of class and possibly during the time slot set aside for finals. Grading of the project will be as follows: 2/3 of the project grade will come from the written report and 1/3 will come from your presentation. Everyone in the group will get the same grade on the project, i.e. no attempt at partitioning credit to one member over the other will be accepted.

The project will entail ONE of the following three options:

- Option 1: A substantive description and detailed statistical analysis of some data set in which the analysis utilizes any of the statistical modeling techniques learned in class.
- Option 2: A detailed description and critique of a published paper in which a latent variable or structural equation modeling technique was employed.
- Option 3: Perform a simulation study or provide some analytical results for a statistical method for latent variable or structural equation modeling.

The grading scale based on the Total percentage is the following and is meant to reflect the Mailman School of Public Health qualitative descriptions at: <http://mailman-handbook.com/2011/node/27>

93-100 A , 90-92 A-, 85-89 B+, 78-84 B, 75-77 B-, 70-74 C+, 65-69 C, 60-64 C-

A grade of A+ will be reserved for highly exceptional achievement and will not necessarily be given every semester. For those taking the course S/N (Pass/Fail), an S will be earned if a grade equivalent to a C- or above is achieved.

For all students, if a grade of at least C- is not achieved, the grade will be F (or N).

## COURSE REQUIREMENTS

Course notes will be made available on a weekly basis through the class website by the instructor and will serve as the **primary source of course material**. Weekly assignments will consist of reading course notes or reading a selected article and performing a data analysis. The course notes for each week will be posted on the CourseWorks site; they range in length from 40 to 60 pages. Assigned articles will also

be available on CourseWorks; these range from 10 to 15 pages. You can anticipate about 4 hours of work outside class for each of the data analyses.

There is no perfect textbook that covers all the topics of this course, but three recommended books are:

- Making Sense of Factor Analysis: *The Use of Factor Analysis for Instrument Development in Health Care Research* (2003) by Marjorie A. Pett, Nancy R. Lackey, and John J. Sullivan, Sage Publishing.
- Item response theory for psychologists. (2000) by Susan E. Embretson and Steven P. Reise, Lawrence Erlbaum Association.
- Kline, R.B. (2011). *Principles and Practice of Structural Equation Modeling*, Third edition, Guilford.

You will need access to a computer (either your own or in a lab where you can download statistical software). I will be introducing how to implement all the techniques in the course using the Mplus software ([www.statmodel.com](http://www.statmodel.com)). A FREE demo version of MPLUS 6.0 software is available for download from the Mplus website. Some very good resources (including full video lectures by the creator of Mplus, Bengt Muthen) can be found at the UC LA statistics website [http://www.ats.ucla.edu/stat/seminars/muthen\\_08/default.htm](http://www.ats.ucla.edu/stat/seminars/muthen_08/default.htm). In addition Mplus I will introduce some IRT techniques using the freely available R software.

## **COURSE STRUCTURE**

Classroom experience will be a combination of traditional lecture with notes made available electronically before class and hands-on computer lab experience (using the free demo version of Mplus and freely available R software) in a designated computer lab or in class with laptops (depending on class size). Homework will be assigned every other week and students may discuss their work with other class-mates but are expected to write-up results themselves.

## **MAILMAN SCHOOL POLICIES AND EXPECTATIONS**

### *Academic Integrity*

Students are required to adhere to the Mailman School Honor Code, available online at <http://mailman.columbia.edu/honorcode>.

### *Disability Access*

In order to receive disability-related academic accommodations, students must first be registered with the Office of Disability Services (ODS). Students who have, or think they may have a disability are invited to contact ODS for a confidential discussion at 212.854.2388 (V) 212.854.2378 (TTY), or by email at [disability@columbia.edu](mailto:disability@columbia.edu). If you have already registered with ODS, please speak to your instructor to ensure that s/he has been notified of your recommended accommodations by Lillian Morales ([lm31@columbia.edu](mailto:lm31@columbia.edu)), the School's liaison to the Office of Disability Services.

## **COURSE SCHEDULE**

Please see the lecture section of Courseworks to download the readings, exams, and lecture slides.

### **Week 1 – Introduction to scope of the course**

Learning Objectives: Observe a survey of the full range of latent variable and structural equation modeling methods that will be introduced in the course. Learn backgrounds of other students and reason for interest in topic.

Reading: Student is to find a research article on their own that uses any of the methods we will cover in the course and be prepared to describe the main goal of the paper and something in the paper they did not understand and would like to learn more about.

Assignment: Homework I – Read article and Data analysis

### **Week 2 – Fundamental ideas of Measurement**

Learning Objectives: Learn how to develop conceptual variables. Understand reflective vs. formative latent variables. Understand and know methods for assessing dimensionality, reliability, and validity of scales.

Required Reading: Course Notes

Related Reading:

- Scientific Advisory Committee of the Medical Outcomes Trust: Assessing health status and quality of life instruments: attributes and review criteria. *Quality of Life Research* 2002; 11: 193-205
- Diamantopoulos A and Siguaw J (2006) Formative versus reflective indicators in organizational measure development: A comparison and empirical illustration, *British Journal of Management*, 17, 263-282.
- Chapter 1 and 2 of Pett et al and pp 174-205 in Chapter 6 of Pett et al.

### **Week 3 – Exploratory and Confirmatory Factor Analysis**

Learning Objectives: Understand, interpret and implement EFA and CFA

Required Reading: Course Notes

Related Reading:

- Chapter 4, 5, 7 and pp 167-174 in Chapter 6 of Pett et al.
- Kline text pp112-118 and Kline text Chapter 9 on CFA
- Reise SP, Waller NG, and Comrey AL (2000) Factor analysis and scale revision, *Psychological Assessment*, 12(3), 287-297.

Assignment: Homework 2 – Read article and Data analysis

### **Week 4 – The Bi-factor Models and Factor models for categorical observed variables**

Learning Objectives: Be able to Interpret the results from a bi-factor model as a means to assess the existence of subdimensions. Understand the underlying variable factor analysis approach with probit modeling for categorical outcomes.

Required Reading: Course Notes

Related readings:

- Chen FF, West SG, Sousa KH (2006) A Comparison of Bifactor and Second-Order Models of Quality of Life, *Multivariate Behavioral Research* 41(2), 189-225.
- Reise SP, Morizot J, Hays RD (2007). The role of the bifactor model in resolving dimensionality issues in health outcomes measures, *Quality of Life Research*, 16 Suppl: 19-31.

### Week 5 – Item Response Theory Models

Learning Objectives: Understand, interpret and implement IRT models.

Required Reading: Course Notes

Related Reading:

- Edwards MC (2009). An introduction to item response theory using the need for cognition scale, *Social and Personality Psychology Compass*, 3/4, 507-529.
- Woods CM (2009) Evaluation of MIMIC-Model Methods for DIF testing with comparison to two group analysis, *Multivariate Behavioral Research*, 44:1-27.
- Edelen Orlando M, Thissen D, Teresi JA, et al. Identification of differential item functioning using item response theory and the likelihood-based model comparison approach: application to the Mini-Mental State Examination. *Med Care* 2006;44(Suppl 3):S134–S142

Assignment: Homework 3 – Read article and Data analysis

### Week 6 – Latent Class Analysis

Learning Objectives: Understanding, interpret and implement LCA for 3 main applications: measuring categorical latent variables, diagnostic testing, and longitudinal trajectory clustering

Required Reading: Course Notes

Related Reading:

- Rindskopf D and Rindskopf W (1986). The value of LCA in medical diagnosis. *Statistics in Medicine*, 5, 21-27. Used LCA for assessing Myocardial Infarction with 4 clinical measurements.
- Jung T and Wickrama AS (2008) An introduction to latent class growth analysis and growth mixture modeling. *Social and Personality Psychology Compass* 2/1, 302-317.

### Week 7 – Factor Mixture Models

Learning Objectives: Understanding the structure of factor mixture models as a hybrid between latent class and factor analysis models. Be exposed to some applications where they are being used.

Required Reading: Course Notes

Related Reading:

- Muthen B and Asparouhov T (2006) Item response mixture modeling: Application to dependence criteria. *Addictive Behaviors*, 31, 1050-1066.
- Muthen B (2006) Should substance use disorders be considered as categorical or dimensional? *Addiction* 101 Supple 1 6-16.

Assignment: Homework 4 – Read article and Data analysis

### **Week 8 – Intro to Causality and what Structural equation modeling can and cannot do**

Learning Objectives: Learn brief history of structural equation modeling and be exposed to general theoretical frameworks and how they are translated into graphical notation.

Required Reading: Course Notes

Related Reading:

- Chapter 5 and 10 in Kline text.
- Meehl P and Waller M (2002) The Path Analysis Controversy: A New Statistical Approach to Strong Appraisal of Verisimilitude, *Psychological Methods*, Vol. 7, No. 3, 283-300

### **Week 9 – Effect decomposition including mediation of bivariate relationships**

Learning Objectives: Identify and be able to estimate using software the total, indirect, and direct effects within a SEM. Be able to estimate and test for mediational effects.

Required Reading: Course Notes

Related Reading:

- James LR (2008) On the path to mediation, *Organizational Research Methods*, 11:359.
- Gelfand LA, Mensinger JL, Tenhave T (2009) Mediation Analysis: A retrospective snapshot of practice and more recent directions, *Journal of General Psychology*, 136(2): 153-176.

Assignment: Homework 5 – Read article and Data analysis

### **Week 10 – Examining Moderators in SEM**

Learning Objectives: Understand the meaning of effect modification and be able to test interactions within a SEM.

Required Reading: Course Notes

Related reading:

- Kline text, pp. 327-340
- Marsh, H. W., Wen, Z., & Hau, K. T. (2004). Structural equation models of latent interactions: Evaluation of alternative estimation strategies and indicator construction. *Psychological Methods*, 9, 275-300.
- Frazier PA, Tix AP, Barron KE (2004). Testing Moderator and Mediator effects in Counseling Psychology Research, *Journal of Counseling Psychology*, 51(1) 115-134.

### **Week 11 – In class Exam**

Learning Objectives: Assimilation of all topics covered during Weeks 1-10. Review.

**IN-CLASS EXAM**

### **Week 12 – Multilevel SEM**

Learning Objectives: Understand the usefulness and application of multilevel SEM

Required Reading: Course Notes

Related Reading:

- Kline text pp. 345-354
- Kaplan and Elliott (1997). A didactic example of multilevel SEM applicable to the study of organizations. *Structural equation modeling* 4(1), 1-24
- Mehta PD and Neale M (2005). People are variables too: Multilevel structural equations modeling, *Psychological Methods*

Assignment: Homework 6 – Write one page proposal for final project

### **Week 13 – Special topics – Spatial structural equation modeling**

Learning Objectives: Be exposed to new extensions to the use of structural equation modeling including spatially (geographic) referenced data and any other topics suggested by students.

Related Reading:

- Wall MM and Liu X (2009) "Spatial Latent Class Analysis Model for Spatially Distributed Multivariate Binary Data", *Computational Statistics and Data Analysis*, 53, 3057-3069.
- Wall MM (2011). Spatial structural equation modeling with an application to U.S. behavioral risk factor surveillance survey data, to appear In the *Handbook of Structural Equation Modeling*, ed Rick Hoyle, Guilford Press.

### **Week 14 – Group Final Project Presentations**

Learning Objectives: Enhance presentation skills and critical listening skills by participating in group presentation and discussion.