CLASS SESSION(S)
Wednesday 8:50 AM - 11:20 AM
Location: Hammer LL205

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 notes.

**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](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:


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). A FREE demo version of MPLUS 7.0 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.

**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 or freely available R software) in a designated computer lab or in class with laptops (depending on class size). Homework will be assigned approximately every week and students may discuss their work with other classmates but are expected to write-up results themselves.

**MAILMAN SCHOOL POLICIES AND EXPECTATIONS**

Students and faculty have a shared commitment to the School’s mission, values and oath. mailman.columbia.edu/about/mission-history

*Academic Integrity*

Students are required to adhere to the Mailman School Honor Code in the student handbook. mailman-handbook.com/2015/node/165

*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. 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), 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 at the weeks progress.

<table>
<thead>
<tr>
<th>Week 1 – Introduction to scope of the course</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Learning Objectives:</strong> Survey 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.</td>
</tr>
<tr>
<td><strong>Reading:</strong> 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.</td>
</tr>
<tr>
<td><strong>Assignment:</strong> Homework I – Read article and give critique</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Week 2 – Fundamental ideas of Measurement</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Learning Objectives:</strong> Learn how to develop conceptual variables. Understand reflective vs. formative latent variables. Understand and know methods for assessing dimensionality, reliability, and validity of scales.</td>
</tr>
<tr>
<td><strong>Required Reading:</strong> Course Notes</td>
</tr>
<tr>
<td><strong>Related Reading:</strong></td>
</tr>
<tr>
<td><strong>Assignment:</strong> Homework 2 – Read article and Data analysis</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Week 3 – Exploratory and Confirmatory Factor Analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Learning Objectives:</strong> Understand, interpret and implement EFA and CFA</td>
</tr>
<tr>
<td><strong>Required Reading:</strong> Course Notes</td>
</tr>
<tr>
<td><strong>Related Reading:</strong></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Week 4 – The Bi-factor Models and Factor models for categorical observed variables</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Learning Objectives:</strong> 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.</td>
</tr>
</tbody>
</table>
Required Reading: Course Notes
Related readings:

Assignment: Homework 3 – Read article and Data analysis

**Week 5 – Item Response Theory Models**

**Learning Objectives:** Understand, interpret and implement IRT models.

**Required Reading:** Course Notes

**Related Reading:**

**Assignment**: Homework 4 – 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:**

**Assignment:** Homework 4 – Read article and Data analysis

**Week 7 – Growth curve modeling**

**Learning Objectives:** Understand what is growth curve modeling and how it can be used for longitudinal data analysis. Be able to describe the models in the hierarchal modeling notation framework and the factor analysis notation framework. Be able to fit a growth curve model and a growth curve mixture model to data using Mplus.

**Required Reading:** Course Notes

**Related Reading:**

**Assignment:** Homework 5 – Read article and Data analysis

---

**Week 8 – Intro to SEM – effect decomposition (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 meditational effects.

**Required Reading:** Course Notes

**Related Reading:**

**Assignment:** Homework 6 – Read article and Data analysis

---

**Week 9 – 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:**

**Assignment:** Homework 6 – Read article and Data analysis

---

**Week 10 – Perspectives on modern causal modeling – Potential outcomes, counterfactuals, propensity scores, instrumental variables**

**Learning Objectives:** Be introduced to theoretical language and ideas of statistical inference for causal modeling and see examples of how propensity scores and instrumental variables can be used to help estimate causal effects.

**Required Reading:** Course Notes

**Related reading:**
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:**
- Mehta PD and Neale M (2005). People are variables too: Multilevel structural equations modeling, *Psychological Methods*

**Assignment:** Homework 7 – 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:**

Week 14 – Group Final Project Presentations

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