APSTA 2094: Factor Analysis and Structural Equation Modeling

Instructor: Peter F. Halpin
Email: peter.halpin@nyu.edu
Phone / Fax: 212-998-5197 / 212-995-4832
Office: Kimball Hall, 246 Greene Street, 204
Office Hours: TBA
Credits: 3

Class Meeting Time / Room: Thursdays 2:00 - 4:45pm.

Course Goals and Orientation: The course provides students with the software skills and theoretical knowledge required to apply structural equation modeling (SEM). First, we review multiple regression and basic concepts from matrix algebra. Next, path analysis and confirmatory factor analysis are developed, leading to more advanced topics, including strategies for translating theory into models, dealing with poor fitting specifications, categorical data, and issues in multigroup and multilevel analysis. The course focuses on cross-sectional applications of SEM, with examples drawn from education and psychology.

The course is conceptually oriented and grounded in examples. While we rely on path diagrams, students will be expected to understand their relation to linear models and covariance structures. There are many software products and associated notations for SEM. The course material will mainly use Mplus (a free demo version is available at https://www.statmodel.com/demo.shtml), but students are also free to utilize other software (e.g., the SEM module in Stata; the lavaan package in R). Lecture and lab are integrated so that a portion of class time is devoted to gaining practical experience with software. The final project requires students to write up an analysis in a manner appropriate for publication in their field of study and present their results to the class. Some class time is devoted to obtaining an appropriate data set for the final project. Students are also invited to use data from their own research, if appropriate, and are encouraged to consult NYU Library Data Services (http://guides.nyu.edu/dataservices)

Prerequisites: Multiple regression at the level of APSTA-GE 2003. APTA-GE 2016 is recommended but not required.

Website: The course uses NYU Classes for lecture notes, handouts, readings, homework assignments, and general information.

Course Text: Students may optionally purchase one or both of the following texts from the NYU Bookstore. Kline is less mathematical and written for applied researchers. Kaplan is written for audiences with a statistics / mathematics background (linear algebra and calculus). Both texts are **optional.**

Kline, R. (2010/2015). Principles and Practice of Structural Equation Modeling. Guilford Press.

Kaplan, D. (2009). Structural Equation Modeling: Foundations and Extensions. Sage Publications.

Grading:
10% Class attendance and participation
40% Three computer-based homework projects
25% Final project presentations (3 presentations: 5%, 5%, 10%; feedback: 5%)
25% Final project write-up

Participation: During the first 11 weeks of class each student receives two participation points per class -- one point for showing up, and another for contributing to the classroom discussion. The final participation grade is based on the average over the classes.

Homework: Three computer-based homework projects are assigned. Students have two weeks to complete assignments, and class time is dedicated to conducting analyses and answering questions. Students can work together, but each person must hand in their own work. Additional homework is assigned

Final Project: The project involves (a) specification of a theoretically motivated model, (b) running the analysis (c) writing-up and interpreting the analysis. Progress on each of these phases is evaluated through a short presentation to the class, given in each of the last three weeks of semester. The final written project is due one week after class ends. Students are encouraged to work together, and may optionally be evaluated in small groups.

Time line: This time line is approximate and may be altered during the semester.

Week	Homework	Topic
1		Overview, review of multiple regression and a “primer” on matrix algebra.
2	HW1 distributed	Path analysis 1: the path diagram; mediation; direct, indirect, and total effects.
3		Path analysis 2: identification; estimation; statistical power; goodness of fit (overlap with week 5) DS Tutorial: using publicly available databases for your research.
4	HW1 due HW2 distributed	Factor analysis 1: confirmatory versus exploratory factor analysis; interpretation of model parameters.
5	HW1 returned	Factor analysis 2: identification; estimation; statistical power; goodness of fit (overlap with week 3)

Week	Homework	Topic
6	HW2 due HW3 distributed	Basic SEM 1: model specification; the role of theory; integrating measurement and structural components.
7	HW2 returned	Basic SEM 2: testing nested models; applications and examples.
8		Spring break
9	HW3 due	Advanced SEM 1: categorical and censored data; relations with item response theory
10	HW3 returned	Advanced SEM 2: multi-group analysis; testing model invariance
11	Prepare Presentation 1	Advanced SEM 3: complex samples (stratification, clustering, and sampling weights); multilevel modeling
12	Prepare Presentation 2	Final project short presentations: theory and model specification
13	Prepare Final Presentation	Final project short presentations: analysis and results
14	Final project write-up due in 1 week	Final project full presentations: the final product

Bibliography (*denotes assigned reading provided by instructor)

Bartholomew, D. J., Knott, M., & Moutsaki, I. (2011). Latent variable models and factor analysis (3rd ed.). London: Arnold. (Covers most of traditional psychometrics from an up-to-date perspective.)

*Bollen, K. A. (1989). Structural equations with latent variables. New York: John Wiley and Sons. (A classic reference for confirmatory factor analysis as well as SEM.)

Fabrigar, L. R., Wegener, D. T., MacCallum, R. C., & Strahan, E. J. (1999). Evaluating the use of exploratory factor analysis in psychological research. Psychological Methods. doi:10.1037/1082-989X.4.3.272 (A review of what is wrong with exploratory factor analysis today.)

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. Structural Equation Modeling: A Multidisciplinary Journal, 6, 1–55. doi:10.1080/10705519909540118 (The controversial current conventions on goodness of fit.)

*Imai, K., Keele, L., & Tingley, D. (2010). A general approach to causal mediation analysis. Psychological Methods, 15(4), 309–34. http://doi.org/10.1037/a0020761

Jak, S., Oort, F. J., & Dolan, C. V. (2013). A Test for Cluster Bias: Detecting Violations of Measurement Invariance Across Clusters in Multilevel Data. Structural Equation Modeling: A Multidisciplinary Journal, 20(2), 265–282. doi: 10.1080/10705511.2013.769392. (Measurement invariance in a multilevel setting)

*Kaplan, D. (1989). Model Modification in Covariance Structure Analysis : Application of the Expected Parameter Change Statistic. Multivariate Behavioral Research, 24, 285–305. http://doi.org/10.1207/s15327906mbr2403 (Explanation of modification indices.)

Li, X., & Beretvas, S. N. (2013). Sample Size Limits for Estimating Upper Level Mediation Models Using Multilevel SEM. Structural Equation Modeling: A Multidisciplinary Journal, 20(2), 241–264. http://doi.org/10.1080/10705511.2013.769391

Lüdtke, O., Marsh, H. W., Robitzsch, A., Trautwein, U., Asparouhov, T., & Muthén, B. O. (2008). The multilevel latent covariate model: a new, more reliable approach to group-level effects in contextual studies. Psychological Methods, 13(3), 203–29. http://doi.org/10.1037/a0012869

MacCallum, R. C., Browne, M. W., & Cai, L. (2006). Testing differences between nested covariance structure models: Power analysis and null hypotheses. Psychological Methods, 11, 19-35.

*MacCallum, R. C., Browne, M. W., & Sugawara, H. M. (1996). Power analysis and determination of sample size for covariance structure modeling. Psychological Methods, 1, 130-149.

*Millsap, R. E. (2001). When trivial constraints are not trivial: The choice of uniqueness constraints in confirmatory factor analysis. Structural Equation Modeling, 8(1), 40–52. http://doi.org/10.1207/S15328007SEM0801 (Overview of identification and problems that can arise.)

Millsap, R. E. (2011). Statistical Approaches to Measurement Invariance. New York: Routeledge. (Multi-group confirmatory factor analysis.)

*Muthén, B. O., & Muthén, L. (2014). Mplus 7.3 [computer software]. Los Angeles, CA: Muthén & Muthén. (The user guide for our software, available at www.statmodel.com.)

*Muthén, B. O., & Kaplan, D. (1985). A comparison of some methodologies for the factor analysis of non-normal Likert variables. British Journal of Mathematical and Statistical Psychology, 38(38), 171–189. http://doi.org/10.1111/j.2044-8317.1992.tb00975.x (Categorical data)

*Muthén, B. O., & Asparouhov, T. (2015). Causal Effects in Mediation Modeling: An Introduction with Applications to Latent Variables. Structural Equation Modeling, 22, 12–23. http://doi.org/10.1080/10705511.2014.935843

*Preacher, K. J., Zyphur, M. J., & Zhang, Z. (2010). A general multilevel SEM framework for assessing multilevel mediation. Psychological Methods, 15(3), 209–33. http://doi.org/10.1037/a0020141

*Preacher, K. J., Zhang, Z., & Zyphur, M. J. (2011). Alternative Methods for Assessing Mediation in Multilevel Data: The Advantages of Multilevel SEM. Structural Equation Modeling: A Multidisciplinary Journal, 18(2), 161–182. http://doi.org/10.1080/10705511.2011.557329

*Pornprasertmanit, S., Lee, J., & Preacher, K. J. (2014). Ignoring Clustering in Confirmatory Factor Analysis : Some Consequences for Model Fit and Standardized Parameter Estimates. Multivariate Behavioral Research, (49), 518–543. http://doi.org/10.1080/00273171.2014.933762

Rabe-hesketh, S., Skrondal, A., & Pickles, A. (2004). Generalized multilevel structural equation modeling. Psychometrika, 69(2), 167–190.

Rabe-hesketh, S., & Skrondal, A. (2008). Classical latent variable models for medical research. Statistical Methods in Medical Research, 17, 5–32. http://doi.org/10.1177/0962280207081236

Rao, C., & Sinharay, S. (2006). Handbook of Statistics: Psychometrics. Amsterdam: Elsevier. (High level of math required; the chapters on factor analysis and SEM are authoritative.)

Rosseel, Y. (2012). lavaan: An R Package for Structural Equation Modeling. Journal of Statistical Software, 48(2), 1–36. (Start here for IRT / factor analysis / SEM with R.)

*Shrout, P. E., & Bolger, N. (2002). Mediation in experimental and nonexperimental studies: new procedures and recommendations. Psychological Methods, 7(4), 422–445. http://doi.org/10.1037/1082-989X.7.4.422

Sörbom, D. (1989). Model modification. Psychometrika, 54(3), 371–384. http://doi.org/10.1007/BF02294623. (Another explanation of modification indices.)

*Wirth, R. J., & Edwards, M. C. (2007). Item factor analysis: current approaches and future directions. Psychological Methods, 12(1), 58–79. doi:10.1037/1082-989X.12.1.58 (A good explanation of why factor analysis and IRT are the same thing.)