«Introduction to Structural Equation Modeling»
Course syllabus

Approved by
Prorgamme Academic Council
Protocol Nr. 01 from 25.06.2018.

Author	Boris Sokolov
Number of credits	4
Contact hours	32
Self-study hours	120
Year of Study	1
Educational format	Without use of online course

I. Aim, Results of Mastering the Discipline and Prerequisites

The course is intended to give an introduction to the foundational concepts and basic computational techniques of structural equation modeling (SEM) and their implementation in a popular statistical software package, R

As a result of this course, students will **be able to**:

- Understand foundational concepts of confirmatory factor analysis (CFA) and structural equation modeling (SEM);
- Understand basic assumptions of CFA and SEM models;
- Understand and apply in practice basic principles of model building, model evaluation and model modification in CFA and SEM
- Build, estimate, assess, compare and modify confirmatory factor an/or structural models using R packages *lavaan* and *semTools*;
- Visualize various types of measurement and structural models using R package *semPlot*;
- Understand the concept of measurement invariance and apply different approaches to measurement invariance testing.
- Conduct multigroup confirmatory factor analysis (MGCFA) and test for measurement invariance using R packages *lavaan* and *semTools*;
- Apply different approaches to theory testing in SEM;
- Understand the concepts of moderation and mediation in SEM;
- Conduct mediation analysis in *lavaan*

To succeed in this course, students are assumed to have basic knowledge of statistics and be familiar with several conventional statistical methods, most importantly regression analysis. In addition, for practical exercises we will use R programming environment, so another major prerequisite is a basic knowledge of R. The course is based on the following previously covered courses:

- Methodology and Research methods in Sociology: Quantitative Research methods
- Research Workshop («Research Design»)
- Basic Statistics and Introduction into „R“

The course is strongly related and complementary to other compulsory and elective courses and provides crucial prerequisites for later courses and research projects as well as skills necessary for the master thesis. The course takes place in the fourth module of the first year of the program, giving students the important skills in designing and conducting their own research as well as assessing the quality of research projects published in a series of working papers and peer-reviewed social science journals.

II. Content of the Course

Lesson 1: Introduction

Course overview. The concept of latent variable. The concept of construct. Reflective vs. Formative measurement. General factor model. Principal component analysis (PCA). Kaiser's rule. Scree plots. Bi-plots. Exploratory factor analysis (EFA). Confirmatory factor analysis (CFA). Differences between CFA and EFA. R implementation of PCA and EFA.

Lesson 2: Confirmatory Factor Analysis – 1: Basics of CFA

Confirmatory factor model: assumptions, key model parameters, notation. Maximum likelihood estimation of CFA-models. Observed vs. predicted VCOV matrices. Model identification. Not identified, just identified and over identified CFA models. Various types of parameter constraints used for model identification. Standardization. Model fit and key fit indices. Estimation of CFA models using the R package *lavaan*. Visualization of CFA models using *semPlot*

Lesson 3: Confirmatory Factor Analysis – 2: Model Correction and Validity Assessment.

Basic principles of CFA model improvement. Modification indices (MI). Expected parameter change (EPC). Standardized residuals. Model comparison: Fit indices and chi-square difference test. Measurement validity. Different types of validity: content validity and construct validity (convergent and discriminant). Reliability: Cronbach's alpha and other measures. Handling missing data: listwise/pairwise deletion and full information maximum likelihood.

Lesson 4: Confirmatory Factor Analysis – 3: Non-normal and categorical data.

Problems with non-normal and categorical data. Which data are (reasonably) non-normal and which are categorical. Robust versions of the maximum likelihood estimator. CFA model for categorical data: key parameters and notation. Weighted least squares estimation. *lavaan* implementation. Interpretation, visualization, assessment, modification, and comparison of categorical CFA models.

Lesson 5: Multi-Group CFA and Measurement Invariance

Multi-group CFA (MGCFA). The concept of measurement invariance (equivalence). Types of invariance: configural, metric and scalar. Invariance testing with *lavaan*. Using *semTools* for automated invariance testing in R. Multi-group categorical CFA (MGCCFA). Required parameter constraints for model identification in MGCCFA. Delta vs. Theta parameterization. MGCCFA implementation in *lavaan*. What to do when measurement invariance is lacking?

Lesson 6: Structural Models

Basic concepts of structural modelling. Measurement model vs. structural model. Path models. Exogenous and endogenous variables. Recursive and non-recursive models. Identification of structural models. Multiple-indicator multiple-cause (MIMIC) models. Structural regression models. *lavaan* implementation. Interpretation, visualization, assessment, modification, and comparison of path models and structural models.

Lesson 7: Mediation analysis

Substantive importance of mechanism modeling. The concept of mediation. Full and partial mediation. Indirect and total effects and their computation in *lavaan*. Baron-Kenny approach to mediation. Uncertainty estimation: Sobel's test and bootstrap. Interpretation, assessment, and comparison of models with mediation. Other types of third variables: moderation and confounding. Equivalent models.

III. Grading

Your final grade is composed of four components: (a) home assignments (unweighted average), (b) class work, and (c) the final exam. The first two components define your accumulated grade (which comprises 80% of the overall grade) which is then combined with the score for the final exam (20% of the overall grade).

Formally, your cumulative grade for the course is calculated in the following way:

$$\mathbf{G_{cumulative} = G_{home\ assignment} + G_{class\ work}\ (1)}$$

where $G_{home\ assignment}$ is the mean [unweighted] grade for home assignments prepared throughout the course and $G_{class\ work}$ is the binary (i.e. 0 or 1) *post hoc* assessment of your overall participation in class activities. Notice that the latter component is added to your cumulative grade only if it is lower than 9.5.

The final (overall) grade for the course (the one you will have in your course records) is calculated in the following way:

$$\boldsymbol{G_{final} = 0.8 * G_{cumulative} + 0.2 * G_{exam}\ (2)}$$

where $G_{cumulative}$ is the accumulated grade for the course calculated using the formula above and $G_{final\ paper}$ is the grade for the final exam. Both the cumulative grade and the final grade are rounded according to the rules of algebra. All late submissions (that is, submitted after the deadline for a given assignment) are evaluated on a 0-6 scale.

The minimal passing grade for this course is 4.

IV. Grading Tools

Type of Control	Form of Control	Detailed explanation
Weekly	Home assignments	Students are expected to complete written home assignments outside of class time. Some assignments are conceptual, but most of them require using R and empirical data,

		therefore students should plan to spend time in the computer lab outside of regular class hours. All assignments are due at the beginning of the following class time
	Class work	Students are expected to participate in all class activities, such as group discussions, responses to the instructor's questions, and completion of simple conceptual and programming exercises.
Final	Exam	Take-home written assignment, in which you should analyze a real data set using FA/SEM methods. Specifically, you should first use exploratory methods to develop a meaningful, theoretically interpretable factor model. Then you apply the confirmatory approach to assess your model's quality and modify it, if necessary. Finally, you are asked to test whether your latent variable(s) is non-trivially related to a set of external variables. Notice that in the final paper you may either (1) analyze a data set provided by the instructor or (2) analyze your own data. Regardless of your specific data preference, the same grading principles and criteria (see below) will be applied to the assessment of your final submission in both cases. The deadline for submission of your exam paper is noon of June 13th, 2019. Then all papers will be evaluated by the instructor and preliminary grades will be announced. Then those course participants who disagree with their preliminary grades will have an opportunity to defend their project papers (by publicly presenting them in class). Presentations must be given using PowerPoint or LaTeX software.

Grading criteria for final exam (apply to written home assignments as well)

The most important aspects of your final exam paper to be graded are: (a) knowledge of key assumptions behind factor analytic/structural modeling and ability to test them using the data at hand (0-2 points); (b) ability to build a correctly specified factor model for a given data set and theoretical background (0-2 points); (c) ability to interpret the results of your analyses correctly

(0-2 points); (d) appropriate use of FA/SEM terminology (0-1 points); and (e) appropriate use of SEM software tools (specifically, R packages *lavaan*, *semTools*, and *semPlot*) (0-2 points).

Notice that (f) style and formatting issues (e.g. correct formatting of parameter tables and figures) also affect the final grade (0-1 points). I do not expect that your style and grammar will be perfect, but I should be able, at least, to understand from the text of your assignment what exactly you have done,

V. Sources

5.1 Main Literature

Kline, R. B. (2015). *Principles and practice of structural equation modeling*. London, UK: The Guilford Press. URL: http://ebookcentral.proquest.com/lib/hselibrary-ebooks/detail.action?docID=4000663

5.2 Additional Literature

Brown, T.A. (2006) *Confirmatory factor Analysis for Applied Research*. London, UK: The Guilford Press.

5.3 Software

№ п/п	Name	Access conditions
1.	R 3.3	*From the university's internal network (contract)*
2.	Rstudio 1.1.383	*From the university's internal network (contract)*

5.4 Material and technical support

Academic support for the course is provided via a course Dropbox folder, which contains all recommended readings, presentations of lectures, and replication code and data for empirical examples.

Classrooms for lectures on the discipline provide for the use and demonstration of thematic illustrations corresponding to the program of the discipline, consisting of:

- PC with Internet access (operating system, office software, antivirus software);
- multimedia projector with remote control.