SEM 1: Confirmatory Factor Analysis

Location

Every Tuesday:

- Lecture
- 15:00 – 17:00 in REC JKB.52

Every Thursday:

- Practical (may include short lecture)
- 15:00 – 17:00 in REC JK3.03B

Grading

- Individual assignments: 50%
- Final project: 50%
    - Pre-data report: 20%
    - Presentation: 10%
    - Post-data report: 20%

Both must have an average grade of at least 5.5 to pass the course.

Book

Brown, T. A. (2015). Confirmatory Factor Analysis for Applied Research. 2nd Ed. London: The Guilford Press. ISBN: 9781462515363.

Learning Goals

An encompassing goal of the course is to prepare the student for academic discourse (PhD). That is, the final learning goals align with the Research Master in general, preparing the student to…

- Analyze data using sophisticated statistical techniques
- Work through the empirical cycle, connecting previous literature to current data analysis
- Communicate both your findings and thoughts through academic writing
- Communicate to an audience of your peers through presenting

To accomplish these general goals, weekly assignments will sometimes feature an **essay question**, which is judged on writing quality, in addition to if the question is answered. Throughout the course students will work on a **final project**, which requires the student to gather data, perform their own analysis, and to communicate the results both in writing and presenting at the end of the course.

Further aims of this course are to prepare the student to…

- Understand the principles of common cause modeling and confirmatory factor analysis (CFA; week 1)
- Fit confirmatory factor models to data and communicate the results (week 2)
- Perform steps of measurement invariance testing (week 3)
- Understand and perform advanced factor analysis topics such as non-normal data, missing data, and sample size selection (weeks 3 & 4)
- Perform an exploratory factor analysis (week 4)
- Fit higher order factor models to data (week 4)
- Critically evaluate factor analysis results (weeks 2-4)
- Formalize a theoretical model and write down your expectations in a pre-registered report (final project)
- Gather and analyze your own data (final project)
- Communicate your findings in scientific writing and presenting (final project)

In order to accomplish these goals a mixture of conceptual knowledge and practical skills is required. To this end, every Tuesday there is a **conceptual lecture** and every Thursday there is a **practical**. Skills will be assessed in a weekly assignment, consisting of both **conceptual questions** and **practical questions**, and a final project.

Specific learning goals to each week follow on the next pages.

Outline

Week 1 – Introduction to common cause modeling

Book: chapter 1 - 3
Tuesday April 3: Lecture
Thursday April 5: Practical (start final project)

Topics:

- Measurement
- The common cause model
- The CFA model
- Maximum likelihood estimation

Learning Goals:

- Understand the principles of common cause modeling and confirmatory factor analysis (CFA)
- Derive the implied variance-covariance matrix from a given CFA model
- Calculate the degrees of freedom from a given CFA model
- Understand identification rules in CFA

Week 2 – Fitting and modifying CFA models

Book: chapter 3 - 5
Tuesday April 10: Lecture
Thursday April 12: Practical

Topics:

- Fitting a CFA model
- Chi-square tests and fit indices
- Model selection
- Modification indices

Learning Goals:

- Fit CFA models to data using R, Jasp & Onyx
- Communicate results from your CFA analysis
- Critically evaluate the fit of CFA models
- Understand the differences between fit indices and how these can be interpreted
- Compare the fit of different model using model comparison tests

Week 3 – Measurement invariance, ordinal data and sample size

Book: Chapters 7, 9 and 10
Tuesday April 17: Lecture
Thursday April 19: Practical

Topics:

- Measurement invariance + mean structure
- Non-normal data
- Power and sample-size

Learning Goals:

- Perform steps of measurement invariance testing
- Know how to handle ordinal data
- Perform a-priori power analyses to determine sample size needed

Week 4 – Missing data, EFA and higher order models

Tuesday April 24: Lecture

Topics:

- Higher order factor models
- Missing data
- Exploratory factor analysis

Learning Goals:

- Perform CFA analysis on datasets with missing responses
- Perform exploratory factor analyses (EFA) and understand the differences between EFA and CFA
- Fit higher order CFA models to data
- Critically evaluate the utility of higher order CFA models

Week 5 - Presentations

Tuesday May 1

- Presentations of group project

Friday May 4 (23:59:99): DEADLINE FINAL PROJECT REPORT

Individual Assignments

Each week there will be an assignment. The assignment will be made available 15:00 on Thursday, and will be due 15:00 the next Thursday. For each assignment, mind the following rules:

- Work on the assignments **alone.**
- Hand in a **PDF** file and an **.R** file (in case R was used). If you use Jasp, hand in the Jasp object as well as a screenshot of the options used.
- Make sure your PDF report is as standalone readable as possible. E.g., if you are asked to report a factor loading matrix, then report it in the PDF and not just say "look at .R file".
- Assignments are due **before** 15:00. I want to discuss assignments in the practicals and will post solutions online. So I cannot accept assignments that are not submitted or late. This includes, for example, uploaded files that were incorrect. It is your responsibility to check this. **If you do not hand in an assignment before 15:00, you will get a 1.**
- If you encounter any problems, or have any feedback, please let me know before the deadline, as then I can take it into account or help you.

Final Project

In this project, you will be asked to go through the entire empirical cycle of performing a confirmatory factor analysis. You will have time during the practicals to work on this. During the first practical, form a group of 5. Then, perform the steps below **in order**. You will be asked to come up with a theoretical CFA model, gather data, evaluate and possibly adjust the model, present your findings and write a report on your research. You can choose division of labor yourself, but in your report, provide a short description of who did what parts of the research. To obtain enough observations, you are all expected to fill in each of the questionnaires constructed. Beside that, you can ask friends / family / etcetera and I will also share the questionnaires on social media.

You will be graded on:

- Pre-data report (20%)
- Post-data report (20%)
- Presentation (10%)

Step 1: Research setup

Think of a set of questions, designed to measure some trait(s) you are interested in. The questions are to be rated on a self-report Likert scale with 5 response options. For example:

1. Strongly disagree
2. Disagree
3. Neutral
4. Agree
5. Strongly agree

Make sure the questionnaire is short (e.g., at most 10 questions), but long enough to provide identifiable models (at least three indicators per latent variable). For example:

Single latent variable: 6-8 indicators
Two latent variables: 4 indicators each
Three latent variables: 3 indicators each

You can measure any trait you like, as long as you can reasonably assume that the latent trait causally underlies the item responses. As such, you can aim to measure traits well described in the literature, or traits originating from wild new ideas. Some examples:

- Personality traits. For example, the Big 5 or the Dark triad.
- Motivation
- Stress
- Nonsense. For example, you could aim to form a questionnaire with zero latent traits underlying the data – an independence model.

- True latent variables. For example, you can aim to measure "length" by asking to rate if someone "is a person that often need to look up".

Creativity will be rewarded!

In a later step, you will be asked to gather data using your questionnaire. Most participants will be your fellow students. So keep this target audience in mind! Other important things to think about when constructing the test:

- Even though testing will be done anonymously, do not include too sensitive questions people might not like answering
- Very important in factor analysis is that observed variables have a lot of variance (many individual differences), so do not ask questions most people would respond to in the same way (e.g., severe psychiatric symptoms)
- Ask questions in a clear and unambiguous way

Step 2: Research hypothesis

Before collecting data, write a short report on the questionnaire you designed, the reasoning behind the questionnaire and what you aim to measure. Include a path diagram of your hypothesized model. You do not have to include numeric estimates of your parameters, but you can include some expectations if you like (such as "I expect this factor loading to be the strongest"). Let all latent variables be correlated. If you expect residual covariances between observed indicators, include them in your path diagram. Include in your report when you plan to start collecting data and when you plan to stop collecting data.

Send the report to me (s.epskamp@uva.nl). If you want to do this properly, make an account on https://osf.io, create a project, upload your report and "register" the project (registration -> new registration -> Open-Ended Registration).

You do not need to do a power analysis beforehand, although if you want to do this (you will get some bonus credit then) you can search online and in the book on how to do this (I will discuss this a bit in week 3. Note that this is **far** from trivial.

Step 3: Data collection

Make your questionnaire available at an online platform and share it to start collecting data. Also include a few grouping variables (such as gender) so you could perform exploratory measurement invariance tests if you want. Send me the link, so I can share it with your fellow students and via social media. You are allowed to fill in the questionnaire yourself as well (although only once of course per person). Aim to gather as much data as possible. At least $N = 50$, but preferably around $N = 100$ should be obtainable I hope. One tool you can use is Qualtrix, please mail lcto-psy@uva.nl for an account.

Step 4: Data analysis

Check assumptions and fit your expected CFA model to the data you obtained. You might need to specify your data as being ordinal. Does the model fit? If not, perform exploratory model modifications to your model. You are also allowed to do other exploratory analysis (such as network analysis).

Step 5: Present your findings

Present your findings in both a **report** (max 5000 words) and a **presentation** (max 15 minutes including discussion). Your presentation should include the reasoning behind your questionnaire, the expected model, the eventually fitted model and discussion. Your report should be written as a scientific paper and follow the recommendations of Brown as much as possible (if discussed in the course). See the tables on the next pages. Critically discuss your findings. If possible, link your findings to other research on these traits as well. Also discuss limitations and recommendations for future researchers aiming to measure the same traits (which items should be included, which items could be changed, etcetera).

TABLE 4.6. Information to Report in a CFA Study

Model Specification

- Conceptual/empirical justification for the hypothesized model
- Complete description of the parameter specification of the model
  - —List the indicators for each factor
  - —Indicate how the metric of the factors was defined (e.g, specify which observed variables were used as marker indicators)
  - —Describe all freely estimated, fixed, and constrained parameters (e.g., factor loadings and cross-loadings, random and correlated indicator errors, factor correlations, intercepts and factor means[a])
- Demonstrate that the model is identified (e.g., positive model *df*, scaling of latent variables, absence of empirical underidentification)

Input Data

- Description of sample characteristics, sample size, and sampling method
- Description of the type of data used (e.g., nominal, interval; scale range of indicators)
- Tests of estimator assumptions (e.g., multivariate normality of input indicators)
- Extent and nature of missing data, and the method of missing data management (e.g., direct ML, multiple imputation[b])
- Provide sample correlation matrix and indicator *SD*s (and means, if applicable[a]), or make such data available on request

Model Estimation

- Indicate the software and version used (e.g., LISREL 8.72)
- Indicate the type of data/matrices analyzed (e.g., variance–covariance, tetrachoric correlations/asymptotic covariances[b])
- Indicate the estimator used (e.g., ML, weighted least squares[b]; as justified by properties of the input data)

Model Evaluation

- Overall goodness-of-fit
  - —Report model χ^2 along with its *df* and *p* value
  - —Report multiple fit indices (e.g., SRMR, RMSEA, CFI) and indicate cutoffs used (e.g., RMSEA ≤ .06); provide confidence intervals, if applicable (e.g., RMSEA)
- Localized areas of ill fit
  - —Report strategies used to assess for focal strains in the solution (e.g., modification indices/Lagrange multipliers, standardized residuals, Wald tests, EPC values)
  - —Report absence of areas of ill fit (e.g., largest modification index) or indicate the areas of strain in the model (e.g., modification index, EPC value)
- If model is respecified, provide a compelling substantive rationale for the added or removed parameters and clearly document (improvement in) fit of the modified models

(cont.)

TABLE 4.6. *(cont.)*

- Parameter estimates
  - —Provide all parameter estimates (e.g., factor loadings, error variances, factor variances), including any nonsignificant estimates
  - —Consider the clinical as well as the statistical significance of the parameter estimates (e.g., are all indicators meaningfully related to the factors?)
  - —Ideally, include the standard errors or confidence intervals of the parameter estimates
- If necessary (e.g., suitability of *N* could be questioned), report steps taken to verify the power and precision of the model estimates (e.g., Monte Carlo evaluation using the model estimates as population values[c])

Substantive Conclusions

- Discuss CFA results in regard to their substantive implications, directions for future research, and so on.
- Interpret the findings in context of study limitations (e.g., range and properties of the indicators and sample) and other important considerations (e.g., equivalent CFA models[d])

[a]see Chapter 7; [b]see Chapter 9; [c]see Chapter 10; [d]see Chapter 5.