SEM 1: Confirmatory Factor Analysis

Coronavirus measures

Due to the global Coronavirus pandemic, the entire course will be online. All Tuesday lectures are cancelled and all lectures will be made available as video lectures by Tuesday 15:00 the latest. The Thursday practical sessions will start with a general Q&A Zoom meeting on Thursday's at 15:00 and will be followed with online chat Q&A for the remainder of the practical time. Attendance is not mandatory for the Zoom meetings. As we understand that this is a hard time for everyone with a lot of uncertainty, the workload for the course has been reduced. Please let us know if for any reason you cannot keep up with the course (e.g., missing assignments due to illness). While the deadline for assignments is Tuesday at 13:00, we will not post solution files online until the next day (Wednesday), giving you more time to notify us of any potential problems.

We hope you will all stay healthy, and please let us know if you have any suggestions for improving the online format of the course!

General information

Book

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

Note that the book is entirely optional. The course will roughly follow the structure of the book, but not exactly. It is not required to own the book.

Team

Teacher: Sacha Epskamp
Teaching assistants: Myrthe Veenman and Sandra Geiger

Please use Canvas for all communication!

Grading

In this course you will do individual assignments (50%) and a final project (50%). For both, you will get a grade between 0 and 10, which must both be at least 6 to pass the course.

The **assignment grade** will be the average of grades you get for each of the three assignments. In each of the assignments, you can get points per question. The total grade for each assignment is computed with (your points) / (total points) x 10.

The **Final project grade** will be the combination of the following points:

1. Questionnaire items (2 points)
    - 1 point: acceptable (can be used in data collection and some motivation)
    - 2 points: good (well phrased items with strong motivation)
2. Hypotheses (2 points)
    - 1 point: well defined hypotheses, but lack of clarity in writing or portrayal of path diagram
    - 2 points: clear hypotheses and path diagram
3. Poster quality (4 points)
    - 1 point: minimal poster that is not clear
    - 2 points: acceptable poster with minimal required information
    - 3 points: Clear and well formatted poster that could be used at a conference
    - 4 points: High quality poster that could compete for a poster prize
4. Analysis / code quality (2 points)
    - 1 point: ok analyses but with mistakes and/or unclear / not reproducible code
    - 2 points: Clear and reproducible code and no mistakes in analysis

Schedule

Date	Day	Class	Deadlines
March 30	Monday		
March 31	Tuesday	Lecture 1 online	
April 1	Wednesday		
April 2	Thursday	Zoom/chat practical	
April 3	Friday		
April 4	Saturday		
April 5	Sunday		
April 6	Monday		
April 7	Tuesday	Lecture 2 online	15:00 - Deadline assignment 1
April 8	Wednesday		
April 9	Thursday	Zoom/chat practical	15:00 - Deadline items & motivation
April 10	Friday		
April 11	Saturday		
April 12	Sunday		
April 13	Monday		
April 14	Tuesday	Lecture 3 online	15:00 - Deadline assignment 2
April 15	Wednesday		
April 16	Thursday	Zoom/chat practical	15:00 - Deadline hypotheses
April 17	Friday		
April 18	Saturday		
April 19	Sunday		
April 20	Monday		
April 21	Tuesday	Lecture 4 online	15:00 - Deadline assignment 3
April 22	Wednesday		
April 23	Thursday	Zoom/chat practical	
April 24	Friday		15:00 - Deadline poster & code

Note: Data for the final project will be made available in the Zoom/chat practical on Thursday April 16.

Individual Assignments

Each week there will be an assignment. The assignment will be made available 17:00 on Tuesday, and will be due 15:00 the next Tuesday. 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** 13:00. Any assignment uploaded without valid reason after 13:00 but before solutions are posted online will still be graded but with a maximum grade of 6. Assignments uploaded after solutions are posted online will no longer be graded (automatically get a grade of 1).
    - An alternative assignment can be made to redo an assignment that was not graded due to a valid reason (e.g., illness).
- 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

Two important goals in factor analysis are to (1) test if a hypothesized latent variable structure could underly the set of observed indicators (questionnaire items), and (2) if so, study if the latent variable has a different distribution (mean and variance) in different groups. For example, we could be interested in personality differences between psychology and biology students. Then, we could form a questionnaire with different indicators design to measure a set of personality traits. If we find evidence that indeed a set of underlying personality traits may cause response on these indicators (good model fit of the measurement model), and that this measurement model is comparable between groups (measurement invariance), we can then use our model to test if personality traits are on average different in psychology and biology students.

In the final project, we will perform a large study ourselves and study differences between groups on a set of traits during the COVID-19 pandemic. During the first two practical zoom/chat sessions, we will brainstorm on which groups we could compare (e.g., people working at home to people continuing daily routine). You have to think of latent variables you want to measure yourselves, and form items designed to measure these. Next, we will combine these items in a large questionnaire which we will spread through social media. This questionnaire hopefully will lead to a rich dataset we will make publicly available for research and teaching purposes. Finally, you will analyze a part of this dataset yourself and present your findings through an academic poster.

Part 1: questionnaire formation

In the first week of the course, you will think of a construct you wish to measure. You can form groups for this if you wish and collaborate online. Please make active use of the Canvas chat and discussion board functionality to brainstorm about constructs to measure. Once you have an idea, **make a post on the Canvas discussion board** with the construct you wish to measure, or **comment on a post** to join a group measuring the same construct if there is still space in that group. Feel free to ask to join groups even if you don't know the members of that group. To avoid overlap in the questionnaire, multiple groups may not work on the same construct (combine the groups instead and collaborate).

Next, formulate items designed to measure the construct of interest. Ideally, these come from prior literature on the construct. To keep the final questionnaire relatively brief, the maximum number of items is limited to the size of a group:

Group size	Maximum number of items	Maximum word limit
1	3	200
2	4	200
3	6	300
4	8	300
5	10	300

This means that if you wish to use many items to measure a construct (e.g., measure all DSM symptoms for a disorder), you have to form a larger group. The items must be phrased with a maximum of 90 characters and such that they can be scored on a 5-point Likert scale ranging from "strongly disagree" to "strongly agree".

Finally, write a short paragraph with the motivation of items you chose. You can write this separately or in groups (use an online platform to collaborate). This paragraph should be no longer than 200 words in groups of 1 and 2 and 300 words otherwise. The deadline for the items and the motivation is on April 9.

Part 2: formulate hypotheses

Even if you worked in a group in part 1, you will work alone for the remainder of the final project. Before analyzing the data, **form a hypothesis** you wish to test on this data using confirmatory factor analysis. **Your hypothesis must include an analysis of the items you(r group) contributed in part 1 plus something extra.** For example, you could choose to investigate the relationship between your construct and a construct measured with items contributed by another group or you could perform measurement invariance and difference tests on one of the grouping variables. You can also choose to analyze the data in a different way, such as using a software package not discussed in the course (e.g., Mplus, OpenMx). **Each student in a group must perform a different analysis**. Please communicate with each-other to avoid overlap. Hand in your hypothesis, described in at most 50 words, and accompanied with a path diagram on April 16.

Part 3: Analysis and results

After the data is made available, you can work on your planned analysis. Report your findings in an academic poster as if you would present your findings on a conference (target audience: fellow research master students), and hand in the final poster in PDF format as well as your R code on April 24. We will upload all posters to Canvas so you can see what other students did.

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 your findings to your peers

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 set up data collection, perform their own analysis, and to communicate the results through an academic poster.

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)
- Perform your own CFA analysis (final project)
- Communicate your findings through an academic poster (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

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

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

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

Book Chapters 8 and 9

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 – Final Project