Contents

Assignments

This chapter contains the details and binding information about the three assignments that comprise the portfolio upon which your course grade is based.

Below, you can find a brief idea of what each assignment will cover.

- For each assignment, you will use R to analyze some real-world data, and you will write up your results in a concise report (not a full research paper).
  - Guidelines for these analyses/reports are delineated in the following three sections.
  - You will submit your reports via Blackboard.
- You will complete Assignments 1 and 2 in your Assignment Groups.
- Ideally, your Assignment 1 Group and Assignment 2 Group will comprise the same people.
  - In the event of group-related issues during A1, you can join a different group for A2.
  - Even if you want to keep the same group members, you still need to enroll in an Assignment 2 Group.
- You will complete the third assignment individually.
- The first two assignments are graded as pass/fail.
  - You must pass both of these assignments to pass the course.
- The third assignment constitutes your course grade.

Assignment 1: Path Analysis

For the first assignment, **you will work in groups** to apply a *path model* that describes how several (observed) variables could be causally related.

The components of the first assignment are described below.

1. Choose a suitable dataset, and describe the data.
   - You can use any of the 8 datasets linked below.
   - Provide some form of citation for the data source
   - Provide descriptive statistics for important demographics (e.g., sex, age, SES) and the substantive variables that you will use to estimate your model.
2. State the research question; define and explicate the theoretical path model.
   - This model must include, at least, three variables.
   - Include a path diagram that represents your theoretical model.
   - Explain the conceptual fit between your theory and your model.
3. Translate your theoretical path model into `lavaan` syntax, and estimate the model.
   - Submit the (clean) R script along with your report.
     - Submit this R script as a separate file, not part of your report.
   - Include the lavaan syntax used to estimate your path models in your report.
     - Place these syntax snippets in an appendix.
4. Report the results
   - Evaluate the model assumptions.

- Provide relevant output in a suitable format.
- Include measures of explained variance for the dependent variables.

5. Discuss the results.
    - Use your results to answer the research question.
    - Consider the strengths and limitations of your analysis.
    - Discuss any important decisions that could have influence your results.

Evaluation

See the Grading section below for more information on how Assignment 1 will be evaluated.

You can access an evaluation matrix for Assignment 1 here.

- This matrix gives an indication of what level of work constitutes *insufficient*, *sufficient*, and *excellent* responses to the five components described above.

Submission

Assignment 1 is due at 23:59 on Wednesday 2 October 2024.

- Submit your report via the Assignment 1 portal on Blackboard.

Assignment 2: Confirmatory Factor Analysis

In the second assignment, **you will work in groups** to run a *CFA* wherein the observed variables are indirect indicators of the unobserved constructs you want to analyze.

Unlike Assignments 1 and 3, you will not submit a research report for Assignment 2. You will report your analysis by filling in a structured worksheet.

- You can download a blank copy of the worksheet here.
- You can download the worksheet template below (in various formats).
    - R Markdown
    - Quarto

Evaluation

See the Grading section below for more information on how Assignment 2 will be evaluated.

Submission

Assignment 2 is due at 23:59 on Wednesday 16 October 2024.

- Submit your report via the Assignment 2 portal on Blackboard.

Assignment 3: Full Structural Equation Model

In the third assignment, **you will work individually** to apply a full *SEM* that describes how several (latent) variables could be causally related.

You will include both your CFA and SEM results in your write-up. The inferences that you draw from an SEM are only valid if the underlying measurement model holds. So, the first step in reporting any SEM (at least any SEM that includes estimated latent constructs) is demonstrating the validity of the measurement model by reporting your CFA results.

The components of the third assignment are described below.

1. Choose a suitable dataset, and describe the data.
    - Ideally, you will work with the same data that you analyzed in Assignments 1 & 2.
    - If you want to switch, you can use any of the 8 datasets linked below.

2. State the research question; define and explicate the theoretical SEM.
    - The structural component of this model must include, at least, three variables.
    - The structural model must include, at least, two latent variables.
    - At least one independent variable in your model must be latent.
    - Use a path diagram to represent your model.
3. Report the results of your CFA.
    - Provide relevant output in a suitable format.
    - Include suitable parameter estimates.
    - Include measures of explained variance for the observed indicators.
    - Include measures of model fit.
    - Evaluate the model assumptions.
4. Translate your theoretical SEM into **lavaan** syntax, and estimate the model.
    - Submit the (clean) R script along with your report.
        - Submit this R script as a separate file, not part of your report.
    - Include the lavaan syntax used to estimate your SEM and CFA models in your report.
        - Place these syntax snippets in an appendix.
5. Report the results of your SEM.
    - Provide relevant output in a suitable format.
    - Include suitable parameter estimates.
    - Include measures of model fit.
    - Include measures of explained variance for the latent dependent variables.
6. Discuss the results.
    - Use your results to answer the research question.
    - Consider the strengths and limitations of your analysis.
    - Discuss any important decisions that could have influence your results.

Evaluation

See the Grading section below for more information on how the component scores represented in the rubric are combined into an overall assignment grade.

You can access an evaluation matrix for Assignment 3 here.

- This matrix gives an indication of what level of work constitutes *insufficient*, *sufficient*, and *excellent* responses to the six components described above.

Submission

Assignment 3 is due at 23:59 on Sunday 3 November 2024.

- Submit your report via the Assignment 3 portal on Blackboard.

Elaboration & Tips

Progression

The three assignments are meant to build upon one another. So, you should plan on analyzing the same dataset for all three assignments and choose a theoretical model that will satisfy the requirements of all three assignments. To facilitate this progression, consider the following points.

- For A2 and A3, your theoretical model must include two or more latent constructs. At least one of the latent constructs must act as in IV in your model.
    - When specifying your model for A1, ensure that you use at least two scale scores constructed from three or more items each.
    - Make sure to include at least one of these scale scores as an IV.
- We will not cover methods for estimating latent variables from categorical items.

  - When specifying the latent constructs/scale scores in your model, do not use nominal items and try to avoid ordinal items with fewer than five levels.
- We will not cover methods for latent variable interactions.
  - When specifying your theoretical model in A1, do not include any interactions involving scale scores (i.e., variables that will be latent constructs in A3).
  - *EXCEPTION:* You may specify an interaction between a scale score/latent construct and a nominal grouping variable (e.g., sex, nationality, ethnicity).
    * You will learn how to test this type of hypothesis via multiple-group SEM in Week 7.

Recycling Prior Work Since each assignment should build on its predecessor, you'll inevitably find yourself in the position of needing to document exactly the same things across assignments. In such circumstances, you'll undoubtedly be wondering if you can reuse some of your previous writing.

Yes. You're free to reuse parts of your reports across all three assignments. There are some minor caveats to this statement, however.

- When reusing prior work, make sure to address any feedback that you received on that work.
  - Uncorrected mistakes in copied passages will incur especially harsh penalties in the grading process.
- Make sure to update any recycled passages to suit the most recent analyses.
  - A path diagram from A1 won't be valid for A2 or A3.
  - A description of your theoretical model from A1 that refers to "scale scores", won't be valid for A3.
- Don't keep irrelevant parts of past assignments.
  - Your grade will be penalized for including extraneous information.

Theoretical Model & Research Question

You need to provide some justification for your model and research question, but only enough to demonstrate that you've actually conceptualized and estimated a theoretically plausible statistical model (as opposed to randomly combining variables until **lavaan** returns a pretty picture).

- You have several ways to show that your model is plausible.
  - Use common-sense arguments.
  - Reference (a small number of) published papers.
  - Replicate an existing model/research question.
- Don't provide a rigorous literature-supported theoretical motivation.
  - You don't have the time to conduct a thorough literature review, and we don't have the time to read such reviews when grading.
  - Literature review is not one of the learning goals for this course, so you cannot get "bonus points" for an extensive literature review.

You are free to test any plausible model that meets the size requirements.

- You can derive your own model/research question or you can replicate a published analysis.

Model Specifications We will not cover methods for modeling categorical outcome variables. So, use only continuous variables as outcomes.

- DVs in path models and the structural parts of SEMs
- Observed indicators of latent factors in CFA/SEM

NOTE: You may treat ordinal items as continuous, for the purposes of these assignments.

We will not cover methods for latent variable interactions.

- Don't specify a theoretical model that requires an interaction involving a latent construct.

There is one exception to the above prohibition. If the moderator is an observed grouping variable, you can estimate the model as a multiple-group model. We'll cover these methods in Week 7.

Assumptions

You need to show that you're thinking about the assumptions and their impact on your results, but you don't need to run thorough model diagnostics. Indeed, the task of checking assumptions isn't nearly as straight forward in path analysis, CFA, and SEM as it is in linear regression modeling. You won't be able to directly apply the methods you have learned for regression diagnostics, for example.

Since all of our models are estimated with normal-theory maximum likelihood, the fundamental assumption of all the models we'll consider in this course boils down to the following.

`All random variables in my model are i.i.d. multivariate normally distributed.`

So, you can get by with basic data screening and checking the observed random variables in your model (i.e., all variables other than fixed predictors) for normality.

- Since checking for multivariate normality is a bit tricky, we'll only ask you to evaluate univariate normality.
- You should do these evaluations via graphical means.

To summarize, we're looking for the following.

Data

- Consider whether the measurement level of your data matches the assumptions of your model.
- Check your variables for univariate outliers.
  - If you find any outliers, either treat them in some way or explain why you are retaining them for the analysis.
- Check for missing data.
  - For the purposes of the assignment, you can use complete case analysis to work around the missing data.
  - If you're up for more of a challenge, feel free to try multiple imputation or full information maximum likelihood.

Model

- Evaluate the univariate normality of any random, observed variables in your model.
  - E.g., DVs in path models, observed IVs modeled as random variables, indicators of latent factors
  - If you fit a multiple-group model for Assignment 3, do this evaluation within groups.
  - Use graphical tools to evaluate the normality assumption.
    - Normal QQ-Plots
    - Histograms

Reporting Standards

What do we mean by reporting your results "in a suitable format"? Basically, put some effort into making your results readable, and don't include a bunch of superfluous information. Part of demonstrating that you understand the analysis is showing that you know which pieces of output convey the important information.

- Tabulate your results; don't directly copy the R output.
- Don't include everything **lavaan** gives you.
- Include only the output needed to understand your results and support your conclusions.

The purpose of statistical data analysis is to provide empirical evidence for or against some theory/hypothesis/model. Therefore, the way you document your analysis and report your results must serve this basic purpose, first and foremost.

- Support any potentially refutable claims with appropriate statistics from your analysis or a suitable citation.
- Not every statement needs supporting evidence.

  - Scientific laws
  - Physical constants
  - Logical (in the technical sense) implications of irrefutable antecedents
  - Your own opinions (assuming you're not trying to pass them off as facts)

The next three subsections cover some specific considerations for reporting common classes of statistical results.

Significance Tests Any claim of a significant effect must be supported by relevant statistical evidence.

- Match your test to your hypothesis
  - Directional hypothesis ⇒ one-tailed test
  - Hypothesis of any non-zero effect ⇒ two-tailed test
- For any test of an estimated parameter (e.g., mean, mean difference, regression coefficient, covariance), report the parameter estimate.
- When using a test statistic (e.g., t, Z, F, χ^2) to conduct the test, report:
  - The estimated test statistic
  - The degrees of freedom for the test statistic
  - The p-value for the test statistic
    - When $p \geq 0.001$, report the computed p-value rounded to three decimal places.
    - Otherwise, report the p-value as $p < 0.001$.
- When using a confidence interval to conduct the test:
  - Clearly indicate the confidence level used to define the interval
  - For directional hypotheses/one-tailed tests, only report the relevant interval bound, and report the other bound as $\pm\infty$.
- When using χ^2 difference tests for significance testing, report
  - The $\Delta\chi^2$ statistic
  - The Δdf
  - The p-value for the $\Delta\chi^2$

Model Fit Judging model fit is always a subjective process. The key is to provide a few pieces of convergent evidence to support the claimed degree of fit.

- Always report the χ^2, its degrees of freedom, and the associated p-value.
- Report at least two additional, non-redundant fit indices (i.e., indices that quantify fit in different ways).
  - If you don't have a particular preference, I'd recommend CFI, RMSEA (and its 90% CI), and SRMR

Parameter Estimates Obviously, you need to report any parameter estimates that directly represent some component of your theory (e.g., regression coefficients that quantify linear associations implied by your theory). You also need to report the significance tests for these parameters.

When evaluating a measurement model, a few key parameter matrices come into play. In addition to showing that your CFA model adequately fits the data, you should report the following parameter estimates:

- Latent variances and covariances
- Factor loadings
- Residual variances

If your CFA includes a mean structure, you should also report any estimated latent means and item intercepts.

Data

Below, you can find links to a few suitable datasets that you can use for the assignments.

- You *must* use one of the following datasets. You may not choose your own data from the wild.

Coping with Covid

- Dataset
- Codebook
- Pre-Registration

Feminist Perspectives Scale

- Dataset
- Article

Hypersensitive Narcissism Scale & Dirty Dozen

- Dataset
- HSNS Article
- DD Article

Kentucky Inventory of Mindfulness Skills

- Dataset
- Article

Depression Anxiety Stress Scale

- Dataset
- DASS Information

Nomophobia

- Dataset

Recycled Water Acceptance

- Dataset
- Article

Procedures

Formatting

Report You must submit your Assignment 1 & 3 reports and your Assignment 2 worksheet in PDF format.

- Each report must include the following information:
  - The names of all assignment authors (i.e., all group members for Assignments 1 & 2, your name for Assignment 3).
  - The Assignment Group number (only for Assignments 1 & 2).

Your Assignment 1 & 3 reports should include appendicies containing the lavaan syntax used to estimate each model include in the respective report.

- Include only the lavaan syntax string and the call to the estimation function used to fit the model (e.g., `cfa()`, `sem()`, `lavaan()`).

For example, the following snippets would be appropriate.

```
cfaMod <- '
f1 =~ a1 + a2 + a3
f2 =~ b1 + b2 + b3
'
```

```
cfaOut <- cfa(cfaMod, data = dat1, std.lv = TRUE)
```

```
semMod <- '
f1 =~ a1 + a2 + a3
f2 =~ b1 + b2 + b3

f1 ~ f2 + x
'

semOut <- sem(semMod, data = dat1, std.lv = TRUE)
```

The snippet below, however, contains additional unecessary commands related to data ingest/processing and summarizing the model.

```
dat0 <- readRDS("../data/dataset.rds")
dat1 <- rename(dat0, x = foo, y = bar, z = baz)

pathMod <- 'foo ~ bar + baz'
pathOut <- sem(pathOut, data = dat1)

summary(pathOut)
```

Syntax File You must also submit an executable script containing the code used for your analyses.

- If you used **KnitR** to embed your analysis code in your report, you may submit the source code that generates your report (i.e., an RMD [R Markdown], QMD [Quarto], or RNW [LaTeX] file).
- Otherwise, submit a standard R script.

Clean the script before submitting.

- Remove extraneous/redundant code and comments.
  - Only include the code for analyses that appear in your report.
  - Include code for data processing/clean/wrangling.
- Order the commands to match your final analysis.
  - E.g., Data ingest → Data processing → Descriptive analysis → Modeling → Processing results → Post-hoc analysis
  - In theory, I should be able to source your script (i.e., run the entire script in "batch mode") to reproduce your results.

Length

You may use as many words as necessary to adequately explain yourself; though, concision and parsimony are encouraged. Note that the assignments are **not** intended to be full-blown papers! The focus should be on the definition of your model, how this model relates to theory (introduction), and what you have learned from your estimated model (discussion).

For each of the assignments, you should be able to get the job done in fewer than 10 pages of text (excluding title page, figures, appendices, and references).

Submission

You will submit your reports through Blackboard.

- Each assignment has a corresponding item in the "Assignments" section of the BB page through which you will submit your reports.
- For Assignments 1 & 2, you may only submit one report per group.
  - Designate one group member to submit the report.

  - The grade for this submission will apply to all group members.
- If something goes wrong with the submission, or you notice a mistake (before the deadline) that you want to correct, you may upload a new version of your report.
  - We will grade the final submitted version.
- The submissions will be screened with Ouriginal.

Grading

Group Assignments

Assignments 1 & 2 are simply graded as pass/fail. To pass, you must:

1. Make a reasonable effort to complete all required work
   - Include all required components listed above in your A1 report
   - Complete every section of the A2 worksheet
2. Demonstrate a satisfactory ability to apply path analysis (A1) and CFA (A2)
   - Does your work demonstrate that you understand what you're doing?
     - Mindlessly implementing a sequence of memorized actions is not sufficient.
   - Does your report show that you understand what you've done, why you've done it, and the implications of your results?
   - Can we understand your analysis and results after reading your report?
3. Submit your assignment before the deadline

Otherwise, you will fail the assignment.

Individual Assignment

Assignment 3 will be fully graded on the usual 10-point scale. We will follow the procedure described below to compute your A3 grade.

- Each report starts with a baseline grade of 7.
- We will evaluate the quality of your work to adjust your grade above or below this baseline.
  - As we read through the report, we add credit (e.g., 0.1 points, 0.25 points, 0.5 points) for each "good" aspect that demonstrates above-average understanding or effort.
  - Likewise, we deduct credit for each "bad" aspect that indicates below-average understanding or effort.

So, if your report aligns with a satisfactory (but not exceptional) level of understanding and effort, your grade will be 7. Your mark will be higher, however, to the extent that you can demonstrate more-than-satisfactory levels of mastery and/or effort. Likewise, your grade will be lower if you do not meet expectations. Importantly, the added and deducted credit will be assigned independently, so the additional points you get for good aspects can counteract any points you may lose for negative aspects.

- The adjustment points will be allocated according to the extent to which your submission addresses the required components listed above.
- The evaluation matrix gives an indication of how these points will be apportioned.

Assuming your group passes the first two assignments, your final course grade will simply be your Assignment 3 grade.

Resits

You must get a "pass" for Assignments 1 & 2 and score at least 5.5 on Assignment 3 to pass the course. If you fail any of the assignments, you will have the opportunity to resit the failed assignment(s).

Procedure The resit procedure is very simple. You simply revise the failed assignment, and your score on the revised submission will replace your failing grade (assuming a better grade for the revision).

Take note of the following points.

- The same submission deadline applies to all three assignments. You are, of course, free to submit your resits early.
- You must resit each failed assignment. If, for example, you failed A2 and A3, you must revise both A2 and A3 and submit both revisions before the deadline.
- You must complete the A1 and A2 resits in your original groups. If some group members have dropped the course, complete the resits with the remaining group members.
- You must complete the A3 resit individually.
- Your revised grade for A3 cannot be higher than 6.
- Submit your resit assignments through the *Assignments* section on Blackboard.

Grading Standards Your resit score will not be a simple additive transformation of your original score. We will not grade the resits by comparing your original report to your resit and checking if you corrected "enough" issues to have produced a passing score in the context of the original report. Such an approach would be the educational testing equivalent of HARKing. So, if you score a 5 on your original submission, for example, and you correct two big issues (the absence of which would have implied a passing score in the original report), you can still fail the resit. The resits are your second chance to demonstrate mastery of the relevant concepts (i.e., Path analysis, CFA, SEM). If your corrections bring the report up to a level that seems to suggest that you "know what you are doing", you will pass. If your report still suggests fundamental misunderstandings (or is otherwise substantially deficient), you will not pass. A passing grade in this course is supposed to show that you can use lavaan to conduct a sensible path analysis, CFA, and SEM, not that you can hack some haggard instructor's goofy evaluation scheme.

I don't want any of you to fail this course. I derive no pleasure from your suffering. Anyway, from a purely selfish perspective, failing grades make much more work for the instructors than passing grades do. That being said, I do want the degree you earn from UU to be a meaningful marker of the skills and knowledge that your study program is meant to teach. So, if you did not achieve the learning goals for this course, you should not pass the course, even though that's an unpleasant outcome for all parties.

Example Assignment

You can find an example of a good submission (for an older version of Assignment 2) here.

This example is not perfect (no paper ever is), and several points could be improved. That being said, this submission exemplifies what we're looking for in your project reports. So, following the spirit of this example would earn you a high grade.

Rules

Resources

For all three assignments, you may use any reference materials you like, including:

- All course materials
- The course GitBook
- Additional books and papers
- The internet

Collaboration

You will complete the first two assignments in groups.

- Although you will work in groups, your group may not work together with other groups.

You will complete the final assignment individually.

- For this assignment, you may not work with anyone else.

For all three assignments, you are obligated to submit original work (i.e., work conducted for this course by you or your group).

- Submitting an assignment that violates this condition constitutes fraud.
- Such cases of fraud will be addressed according to the University's standard policy.

Academic integrity

Hopefully, you also feel a moral obligation to obey the rules. For this course, we have implemented an examination that allows you to showcase what you have learned in a more realistic way than a written exam would allow.

- This assessment format spares you the stress of long exams (the two exams for this course used to be 4 hours each) and the attendant studying/cramming.
- The assignments will also help you assess your ability to independently analyse data, which is important to know for your future courses and/or career.

However, this format also assumes that you complete the assignments in good faith. So, I simply ask that you hold up your end of the bargain, and submit your original work to show us what you've learned.

Using AI tools

You are free to use AI tools like OpenAI's ChatGPT or GitHub CoPilot, but you must cite the tools you used in the same way that you would cite any other software. As with citing any other source, properly citing AI tools both attributes the appropriate credit to the tool's developers and helps you clearly and accurately describe your own work. In addition to providing a valid citation, you must also explain how you used the AI tool. Specifically, did you use AI to:

1. Create new content (e.g., use ChatGPT to draft some parts of the text)
2. Modify content that you created (e.g., use CoPilot to debug your R code)

A convenient way to satisfy these requirements would be to provide the necessary information in a short section at the end of your report. You can think of this section as being analogous to the Conflict of Interest statements you often see in journal articles.

AI tools & academic integrity Generative AI may be shiny new tech, but that just means AI is like the newest, shiniest gizmo in a toolbox full of time-tested options, not some paradigm shift in carpentry that deprecates the hammer. The normal ethics of scientific communication still apply:

1. Transparently describe your work
2. Honestly attribute credit/blame

Of course, the natural corollary of the above is that all the standard penalties for bad behavior also apply. Obfuscating your use of AI tools or attempting to claim AI-generated content as your own work is a form of academic fraud that is more-or-less equivalent to plagiarism. We will treat any such attempted subterfuge as academic dishonesty.

Finally, don't trust The Robots too much! You are ultimately responsible for the work you submit. So, you will have to own any "mistakes" that wily AI models induce in your work.

Strict stuff

By submitting your assignments (both group and individual), you confirm the following:

1. You have completed the assignment yourself (or with your group)
2. You are submitting work that you have written yourself (or with your group)
3. You are using your own UU credentials to submit the assignment
4. You have not had outside help that violates the conditions delineated above while completing the assignment

All assignments will be submitted via Ouriginal in Blackboard and, thereby, checked for plagiarism. If fraud or plagiarism is detected or suspected, we will inform the Board of Examiners in the usual manner. In the

event of demonstrable fraud, the sanctions delineated in Article 5.15 of the Education and Examination Regulations (EER) will apply.