Assignment 3: End of the Course and Hard-Boiled Wonderland

Due date: **11. 2.** 23:59 (**pushed the deadline as much as possible**)

Where to turn in: IS folder "Assignment 3" as PDF

Welcome to yet another assignment of your favourite course on factor analysis!

This time, we will take a closer look on estimation and model fit. Especially model fit is probably the most important practical outcome of this course. I hope we will have covered it sufficiently so that you spread good research practices along with your wings after you finish this course.

The current assignment contains a series of regular questions you might have expected. Remember you can skip many of the tasks and still get a good grade.

Then, there is the advanced topics question – the didactic goal of that is to show you some of the potential avenues one can take after being *introduced* to factor analysis – SEMs, growth models and bifactor models. I believe that after suffering through this course, you would not get lost when studying abroad (*post hoc ergo propter hoc*, anyone?), so the secondary goal is to increase your self-efficacy. The Vectors of Mind point to the skies!

Speaking of being lost, there's also a bonus task milking the n-dimensional mountain analogy, you may recall from the lecture, for a couple of paragraphs. That's the quirky part of this assignment, and I hope you'll enjoy seeing how the stories of our protagonists converge (wink wink). But don't get me wrong, again I had ulterior motives as I wanted to imprint this mental representation into your mind's eye. It is my most fervent wish that this quirky prose does some good.

Have fun!

Mandatory reading

Mulaik Chapter 6

Kline Chapters 12 and 13 (but the mandatory part is only the one concerning model fit and CFA interpretation, other stuff, like relations with IRT, etc. is optional)

Ropovik's article on assessing model fit:
https://www.frontiersin.org/articles/10.3389/fpsyg.2015.01715/full

Quantitude episodes 4x10 (Model identification) and 4x11 (Model fit)
https://open.spotify.com/episode/2uIwUqw6TeZ91pNbh5LAbk?si=12c23fbec4ff4394

https://open.spotify.com/episode/4WibSWWrkE7D8eK79kvyAm?si=185b768b89963485a

I recommend reading the questions before listening. You can replace some of the information by the optional tutorial paper referenced below.

Optional

A tutorial paper on GORIC(A), the best way to do confirmatory analyses:
https://www.tandfonline.com/doi/full/10.1080/10705511.2020.1836967

Anna's and my work in progress tutorial paper on model fit which may help you clarify some stuff from the lecture (see Study Materials)

The supplementary R script located in Study Materials

* * *

Q1: We covered the differences between restricted and unrestricted factor analysis. I highlighted the point that the choice of the FA type does not imply whether you are doing an exploratory or confirmatory analysis. (2)

a) What is your opinion on this matter?

b) Would you do any of your past factor analyses differently given your opinion from a)? If so, how?

c) More specifically, what approach (confirmatory / exploratory) and what technique (restricted / unrestricted FA) would you use to establish a new measure of a reflective latent construct? Why? And how would you proceed if your previous effort failed?

Q2: The greatest magic trick of the factor analysis people was to cast a spell that removed all the unseen parts from the fundamental theorem of factor analysis. What made this miracle possible? How exactly did the spell work? (1)

Q3: Given the population variance-covariance matrix and a perfectly fitting common factor model, what is the determinant of that variance-covariance matrix? Why?
Yes, it's another one of *those* questions! (1)

Q4: Evaluate the following models' fit based on the available information. Explain your answers. It's okay to not give a conclusive judgement. (2.5)

Model 1

$X2(5) = 11.91$, $p = .036$
TLI = .83
RMSEA = .037 [.009 ; .065]
SRMR = .036

Model 2

$X2(16) = 24.43$, $p = .081$
TLI = .90
RMSEA = .132 [0 ; .232]
SRMR = .184

Model 3

$X2(9) = 122.229$, $p < .001$
TLI = .96
RMSEA = .065 [.055 ; .075]
SRMR = .028

Q5: Why do Quantitude say that:

a) Karl Popper stops turning in his grave when you have high *df*
b) Latent variable is like a guest that doesn't bring any food to a party
c) You can scale latent variable to your home address if you want
d) A non-central distribution describes model misspecification

(2):

Q6: A paper describing a simulation study states: *We set the population values for all factor loadings to 0.70 and all error variances to 0.51.*

Why 0.51? What does it mean for the specification of their model, if you assume no important information was omitted? (0.5)

Q7: In this course we didn't have the time to cover various important topics. But since we've already covered A LOT, I leave it up to you to choose what you want to learn more about. Pick at least one of the following topics and answer the corresponding questions. Delving into more topics yields bonus points! And… bonus knowledge. (1+2)

Structural equation modelling – want to model causal paths between different variables? Watch the University of Amsterdam star Sacha Epskamp's course on SEM! Lesson 2-1 through 2-3 should suffice as an introduction. If you want to learn more, Rex Kline's book from the mandatory reading is a great next step.

a) What is the main technical difference between CFA and SEM? In other words, why is CFA a special case of SEM?

b) What are exogenous and endogenous variables? Describe them in your own words but as precisely as possible.

c) What are the residual variances of endogenous variables?

d) What are the implications of specifying an SEM model without a path between two variables? In other words, what does the data-model fit of this SEM tell you?

Growth curve models – a smart way of using factor analysis to model change in time. Watch Curran's course on them. Yes, that Curran!

a) What do latent slope and latent intercept represent? How do you interpret the correlation between the two?

b) What is the basic model specification of a simple growth curve model? How do the FA model matrices look?

c) How do you model non-linear growth in these models?

Bifactor models – I heard you like common factors, so we put a common factor on both sides of your model! This time, it's a bit harder reading on the application of bifactor models in psychopathology. See Bornovalova et al. (2020) in Study Materials.

a) What is the specific factor, substantively speaking?

b) What is its relation to common factors?

c) What are the main arguments for and against using bifactor models?

Bonus Q1: Choose either principal axis factoring (iterative or not), OLS, or ML and describe that estimation technique in a short sales pitch. (0+0.5)

Bonus Q2 (0+2):

Day **I**:

Call me Eigenmael.

It's been God knows how many days since I escaped the sultry summer of 1945. I can still see my pursuers – they keep moving around restlessly but are mere shadows of their past shape. And I mean that quite literally. After following me *here*, their black uniforms with runic insignia and everything underneath turned into volumeless shades. I see them like you see moths crawl behind a lampshade.

Day **I**+12:
When I was of fewer years and dimensions, some time before the war, people used to tell me to visit the Alps. I never heeded that advice only for fate to cast me into a mountain range of my own. Here, the monumental peaks span all the dimensions. You see mountains short and thin yet

piercing through the sixth dimension like bayonets. I really ought to start coming up with names for all the ways mountains can loom large over here.

Day **I**+65:
I met someone new today. And it's not one of the usual faunae you tend to encounter around an n-dimensional hyperspace you get stuck in after escaping from a war-torn empire (and by that I mostly mean shades of all the possible shapes).

This time, it's a full rank person, just like yours truly. I met him meditating by a river, dressed in an Italian suit, arm visibly and uncomfortably broken, distraught for he has lost a map to this place. Stranded here for some time, and being the policeman he claimed to be, he reckoned one should try to make sense of their surroundings. The map, as he explained, marked the spots where paths are the steepest in all possible directions.

His name is Newton Raphson, and it appears we are travelling with one another from now on. Us fullranks need to stick together, *n'est-ce pa*s?

Day **I**+113:
Raphson turned out to be a much more hopeful fellow than I expected from the visage of the overdressed ex-policeman sulking by a river. Meeting another fullrank reinvigorated him. He keeps pushing me to travel on, although none of us know the destination. "In the dark times, should the stars also go out?"

Day **I**+189:
Ever wondered what a couple of hyperplane travelers might be eating all this time? Snakes. It's not the most fanciful of meals but for all the dimensions this place has, its cuisine is depressingly one-dimensional. We've been eating snakes for so long that it felt slightly awkward when one of them started talking. The rhetorically gifted reptile bought his life by telling us the secret of this place.

You need to find the highest peak protruding in all the dimensions, and you will likely return back to normal.

I immediately noticed the burning gaze Raphson directed at me after hearing these hopeful hisses, but I pretended otherwise. Hyperplane stranding or not, I still don't want to needlessly indulge the cop's *Besserwisserei.*

Day **I**+190:
Finding a top spot of a regular mountain is simple. You just keep climbing until you can climb no more. But how do you climb to the top of an n-dimensional mountain range? It's an Escherian nightmare, alright, but the answer is the same. You just keep climbing until you can climb no more.

Day **I**+254:
Still nothing. Even Raphson looks like he's losing hope, but he works hard to cover it up.

Since the last experience with the talkative snake, we always wait a few minutes before we eat them. I keep having dreams about waking up and hearing someone shout the directions to the peak from my stomach only to be silenced by my gastrointestinal tract.

Day **I**+321:
Lately, the climb has been feeling steeper and tougher. I don't want to mention it to avoid Raphson turning into a hopeful golden retriever once again. I like him better a bit depressed.

Day **I**+347:

A. The overly lengthy story above is an analogy of one estimation technique in FA. Which one? Why? **All the following questions are, obviously, supposed to be answered in the context of this technique.**

B. What might be the map Newton Raphson lost?

C. Why is the companion's name Newton Raphson?

D. Why does Eigenmael call Newton and himself “fullranks”?

E. What caused the travelers to be stuck for so long? What would help them progress faster? Why?

F. Given what the snake told the protagonist is true, what might cause the pair to not return to normal after finding the highest peak?