Homework 3

Measurement and Evaluation of HCC Systems

How to hand in this homework

- Please email the homework to me as a PDF.
- Late assignments get a penalty of 20% when submitted after the deadline, plus an additional 10% per hour late.
- Make sure you include the R input you used to get to your answer, but do not "dump" the resulting R output on the paper. Copy from the output selectively, and explain it in your own words.
- You may collaborate on this homework, but not copy from others… again, please write your answers in your own words.
- Please include a collaboration statement that says: "I collaborated on this homework with [name]." or "I worked alone on this homework"

Question 1 Dataset

For the first question, you are going to use **movie.dat**, a dataset from a movie recommender system study. In this study, we tested the effect of the length and diversification (in this experiment these were manipulated both between-subjects) of a list of recommended items on the perceived diversity and attractiveness of this list, and how these factors subsequently affected the choice difficulty and choice satisfaction.

Manipulations

We tested list **length** (5 or 20 items) and **diversification** (low, medium, high) between subjects.

Procedure and measurements

Participants would first train the recommender by some movies rating movies. The system would then give them a list of recommendations corresponding to the assigned experimental condition. Users were asked to inspect this list, choose one item from the list, and then answer questions about the list and/or their chosen item:

- **div1-div5**: 5 seven-point scale items measuring the perceived diversity of the list of recommendations

- **acc1-acc5**: 5 seven-point scale items measuring the perceived accuracy of the recommendations
- **diff1-diff5**: 5 seven-point scale items measuring the choice difficulty
- **sat1-sat6:** 6 seven-point scale items measuring the perceived satisfaction with the chosen item

The items are listed in below:

Considered aspects	Label	Items
Perceived recommendation diversity	div1	The list of movies was varied.
	div2	All the movies were similar to each other.
	div3	The list of recommendations contained movies from many different genres.
	div4	Many of the movies in the list differed from other movies in the list.
	div5	The movies differed a lot from each other on different aspects.
Perceived recommendation accuracy	acc1	I would give the recommended movies a high rating.
	acc2	The list of recommended movies showed too many bad items.
	acc3	The list of recommended movies was attractive.
	acc4	I didn't like any of the recommended items.
	acc5	The list of recommendations matched my preferences.
Choice difficulty	diff1	I was in doubt between several movies on the list.
	diff2	I changed my mind several times before making a decision.
	diff3	The task of making a decision was overwhelming.
	diff4	It was easy to select a movie.
	diff5	Comparing the movies took a lot of effort.
Choice satisfaction	sat1	I am satisfied with the movie I chose.
	sat2	My chosen movie could become one of my favorites.
	sat3	I would recommend the chosen movie to others.
	sat4	I think I would enjoy watching the chosen movie.
	sat5	I would rather rent a different movie from the one I chose.
	sat6	I think I chose the best movie from the options.

Question 1. Confirmatory Factor Analysis

Let's create a factor model based on the measured items.

a. Draw the CFA model to be created based on this data. Set up the model using Unit Variance Identification (UVI). Use ovals for factors, and squares/rectangles for items. Don't forget the correlations between the factors and the arrows representing uniqueness. Tip: You can use the drawing tools in Word for this, but PowerPoint works better because you can "attach" the arrows to the boxes. You can also use drawings.google.com.

b. Specify the model in R.

c. Run the model. Make sure you use UVI, and make sure to treat the items as ordered categorical.

d. Get the model output. Make sure you ask R to display the R-squares.

Let's see if we can iteratively improve our model.

e. Are there any items with a really high uniqueness (Say, > 0.7)? Which one(s)? Tip: Remember that uniqueness = 1 - r-squared!
f. Can you think of a reason why it/they don't fit? (look at the item text in the table)
g. Remove the item with the highest uniqueness from the model, run it again, and inspect the output again. Are there any items with uniqueness > 0.7 left?
h. Get the modification indices of this model. Which item is the biggest problem based on the modification indices? Can you think of a reason why this item is problematic?
i. Remove the most problematic item from the model, run it again, and inspect the output again.
j. Are there any items with a really high uniqueness in this model? Which one? Remove this item as well.
k. Continue removing items with high uniqueness (> 0.7) or high modification indices (> 20) from the model. Make sure to remove them one by one. Choice difficulty (one of the hardest things to measure) will end up with only 2 items, but that's okay.

Let's inspect the factor and model fit.

l. In your final model, calculate the AVE for each factor. Do you have convergent validity for all factors?
m. Compare √(AVE) with the correlations between factors. Do you have discriminant validity for all factors?
n. Inspect the model chi-square. Does the model show significant misfit? Is this problematic?
o. Divide the model chi-square by its degrees of freedom. Does this value look good?
p. Inspect the model CFI and TLI. Do these values good?
q. Finally, inspect the RMSEA and its confidence interval. Do these values look good?

Finally, some reporting.

r. Report the results like on slides 40-43 of the "CFA practice" slides. The only thing you currently don't have is Cronbach's Alpha. We are going to calculate that in Question 2.

Question 2 Dataset

For the section question, you are going to use **privbeh.dat**, a dataset of users' tendency to use various Facebook privacy features. The use of each feature is rated on a 6- or 7-point scale.

Label	Items
pb1	Changed friend subscription
pb2	Reported a story or marked as spam
pb3	Unsubscribed to a friend
pb4	Unsubscribed to status updates
pb5	Deleted content from Timeline/ Wall
pb6	Reported/marked content as spam
pb7	Hid a story from Timeline/Wall
pb8	Withheld/restricted "Interested In"
pb9	Withheld/restricted religion
pb10	Withheld/restricted political views
pb11	Withheld/restricted birthday
pb12	Withheld/restricted relationship status
pb13	Withheld/restricted phone number
pb14	Withheld/restricted email address
pb15	Withheld/restricted IM screen name
pb16	Withheld/restricted street address

Question 2. Exploratory Factor Analysis

a. Create a scree plot for the dataset. How many factors do you expect to find based on this plot?
b. Use the "fa" function to run a factor analysis with 3 factors using the "wls" extraction method and the "geominQ" rotation.
c. Repeat this for 4 factors and 5 factors.
d. Inspect the models. Which of the factor models is the most parsimonious? Why?

Let's see if we can improve the fit of the model

e. Rerun the 4- and 5-factor solution, but remove the two items with the lowest communality as well as item bp2.
f. Report and interpret the chi-square fit (statistic, df, p-value), TLI, and RMSEA of the 4-factor solution.
g. Report and interpret the chi-square fit (statistic, df, p-value), TLI, and RMSEA of the 5-factor solution.
h. Conduct a model comparison between the 4- and 5-factor solution. Report and interpret the model comparison chi-square test (statistic, df, p-value).
i. Is the 4-factor solution significantly worse than the 5-factor solution?
j. Given your answers, which model would you use for further research? Why?