

# Homework 4

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”

## Dataset for questions 1 and 2 (SEM)

For questions 1 and 2, you are going to use the data (**movie.dat**) and factor model from Homework 3. Just to make sure that we’re all on the same page, we will use the following factor model:

```
model <- 'div =~ div1+div2+div3+div5
         acc =~ acc1+acc2+acc3+acc5
         sat =~ sat1+sat2+sat3+sat4
         diff =~ diff4+diff5'
```

Moreover, we will use the following causal order: Conditions → recommendation diversity → recommendation attractiveness → choice difficulty → choice satisfaction. In this homework we are going to combine the two models to create a Structural Equation Model.

## Question 1. Marginal effects

*Before we test the full model, we are going to check out the marginal effects using a MIMIC model. First, we have to make dummies for the experimental conditions:*

- a. Create a dummy called 'l20' for lists of 20 items (5 items is the baseline).
- b. Create two dummies called 'divmedium' and 'divhigh' for medium and high diversification (low diversification is the baseline).
- c. Create two dummies called 'l20medium' and 'l20high' for the interaction between l20 and divmedium and between l20 and divhigh.

*Next, we run the MIMIC models one by one:*

- d. Run the MIMIC model for recommendation diversity. Make sure you use UVI.
- e. Is the interaction effect significant? (Tip: make sure to use `lavTestWald` to get the omnibus effect!) If not, remove the interaction effect.
- f. Is the main effect of list length significant?
- g. Is the main effect of diversification significant (Tip, make sure to use the `lavTestWald` here as well)
- h. Repeat steps d-g for recommendation attractiveness, choice difficulty, and choice satisfaction.

## Question 2. Full model

*Let's specify, run, and trim our full model:*

- a. Specify and run the saturated model (see the causal order on the first page). Since you did not find any interaction effects in Question 1, you can exclude the l20medium and l20high dummies. Make sure you use UVI.
- b. Trim the model step by step. Keep anything with a p-value < 0.10.
- c. Inspect the modification indices to make sure that you did not miss any effects. Do the indices make any useful suggestions for improving the model? Why (not)?

*Now let's see how well this final model fits:*

- d. Inspect the model chi-square. Does the model show significant misfit? Is this problematic?
- e. Divide the model chi-square by its degrees of freedom. Does this value look good?
- f. Inspect the model CFI and TLI. Do these values look good?
- g. Finally, inspect the RMSEA and its confidence interval. Do these values look good?

*Finally let's report this model:*

- h. Conduct the omnibus tests (use `lavTestWald`) for any remaining effects involving divmedium and divhigh. What is the chi-square value and the p-value for these tests?
- i. Get the r-squares for all the Ys in the model.
- j. Report all your results as on pages 23–25 of the "14-2 SEM part II" slides.

## Dataset for question 3 (Multi-level CFA and SEM)

For the first question, you are going to use **bbc.dat**, a dataset from a recommender system study on BBC's system to watch on-demand TV shows. In this study, we tested the effect of two algorithms against a baseline "algorithm" that just shows the overall most popular TV shows.

### Setup and manipulations

Participants are identified with an ID labeled **user**. They used the system for three weeks, each week with a different algorithm (in random order). **mfe** (matrix factorization with explicit feedback) and **mfi** matrix factorization with implicit feedback) are dummies for the two algorithms.

### Measurements

At the end of each week, participants would be asked to answer the following questions:

- **use1-use6**: 6 five-point scale items measuring the perceived usefulness of the system
- **rec1-rec7**: 7 five-point scale items measuring the perceived quality of the recommendations
- **var1-var5**: 5 five-point scale items measuring the perceived diversity of the list of recommendations

The items are listed in below:

Considered aspects	Label	Items
Choice satisfaction	use1	I would recommend the MyMedia recommender to others.
	use2	The MyMedia recommender has no real benefit for me.
	use3	I can save time using the MyMedia recommender.
	use4	The MyMedia recommender is useful.
	use5	I can find better programmes without the help of the MyMedia recommender.
	use6	The system gave too many recommendations.
Perceived recommendation accuracy	rec1	The recommended items fitted my preference.
	rec2	The recommendations did not include my favorite programmes.
	rec3	The MyMedia recommender is providing good recommendations.
	rec4	The MyMedia recommender is not predicting my ratings accurately.
	rec5	I liked the recommendations provided by the system.
	rec6	I didn't like any of the recommended items.
	rec7	The recommended items were relevant.
Perceived recommendation diversity	var1	The recommendations contained a lot of variety.
	var2	Most programmes were from the same genre.
	var3	The recommendations covered many programme genres.
	var4	All the recommended programmes were similar to each other.
	var5	The MyMedia recommender is recommending interesting content I hadn't previously considered.

### Question 3. Multi-level CFA and SEM

*Let's start with a CFA:*

- a. Specify a CFA model in Mplus. Some tips:
  - The data file has no labels in the top row, but the variables are listed in the order presented in the following order: `user mfe mfi use1 use2 use3 use4 use5 use6 rec1 rec2 rec3 rec4 rec5 rec6 rec7 var1 var2 var3 var4 var5`.
  - Make sure you specify all questionnaire items as categorical.
  - There are some missing questionnaire answers, these are indicated by 9999. You can specify this in Mplus under the "VARIABLE" header via: `missing are all(9999);`
  - Make sure you specify the random intercept for user!
- b. Run the model and report the communalities.
- c. Iteratively remove (one by one) any items with an communalities < 0.4.

*Now let's build an SEM on top of this:*

- d. Specify a saturated structural model on top of this CFA with the following causal order: Conditions → recommendation diversity → recommendation quality → perceived usefulness.
- e. Run the model and report the UVI results (hint: these are listed under STDY).
- f. Interpret the effect of the experimental conditions on perceived recommendation quality. Hints: make sure you mention what you are controlling for, and that you properly explain the comparison you're making.
- g. Is the effect of the experimental conditions on recommendation diversity significant? Hint: make sure you do an omnibus test!
- h. Iteratively trim the model by removing non-significant effects.
- i. Report your final results as on pages 23–25 of the "14-2 SEM part II" slides.