

# 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 this homework, you are going to use a dataset from a movie recommender system study. In this study, we tested the effect of the length and diversification of a list of recommended items on the perceived diversity and attractiveness of this list, and how these factors subsequently affected the tradeoff and choice difficulty.

### *Manipulations*

We tested list **length** between subjects, meaning that each participant received lists of either 5, 10, 15, 20, or 25 recommendations (randomly assigned). **Diversification** was manipulated within subjects, meaning that each user received a low-diversity list, a medium-diversity list, and a high-diversity list (in random order). Participants are identified by the variable **userId**.

### *Procedure and measurements*

Participants would first train the recommender by some movies rating movies. The system would then give them three lists of recommendations (low, medium, and high diversity, in random order). Users were asked to inspect the first list, and then answer questions about this list:

- **Perceived\_diversity**: how varied were the recommendations in this list? (measured with five 7-point scale questions)
- **Perceived\_accuracy**: how good were the recommendations in this list? (measured with five 7-point scale questions)
- **Tradeoff\_difficulty**: how easy/difficult was it to compare the recommendations in this list? (measured with one 7-point scale question)
- **Choice\_difficulty**: how easy/difficult would it be to choose a recommendation from this list? (measured with one 7-point scale question)

This procedure was repeated for the other two lists. If you want to learn more about this experiment, you can read Willemsen et al. (2011) "Using Latent Features Diversification to Reduce Choice Difficulty in Recommendation List". Link: <http://ceur-ws.org/Vol-811/paper3.pdf>

## Question 1. Repeated-measures ANOVA with ezANOVA and lme

*We expect that diversification increases the perceived diversity of the recommendations.*

- Turn diversification into a factor with the levels in the order "low", "medium", "high".
- Create a box plot of perceived\_diversity for the different levels of diversification. Do you think there could be an effect? Why (not)?
- Conduct Levene's test to test for heteroscedasticity of perceived\_diversity between different levels of diversification. Interpret the results. Are these results problematic for a repeated measures ANOVA? Why (not)?
- Test for normality of perceived\_diversity between different levels of diversification in terms of skewness, kurtosis, and the Shapiro-Wilk test. Also plot histograms for perceived\_diversity at different levels of diversification. Are these results problematic for a repeated measures ANOVA? Why (not)?

*Let's start with a repeated-measures ANOVA using ezANOVA:*

- Conduct an ezANOVA with perceived\_diversity as the dependent variable and diversification as the within-subjects predictor.
- Interpret the results. Do you have sphericity? What is the effect (F and p-value) of diversification? What is the size of the effect?
- Conduct post-hoc tests using the Bonferroni correction. Which of the differences between diversification conditions are significant?
- Report the results of your ezANOVA and post-hoc test in line with slides 24 and 25 of the slides for "repeated measures".

*Let's conduct a robust version of this test:*

- i. Conduct a robust repeated measures ANOVA with perceived\_diversity as the dependent variable and diversification as the within-subjects predictor. Use 20% trimming, and 2000 bootstrap samples. Also conduct robust post-hoc tests. Do your conclusions change?

*Finally, let's do the same test as a multilevel regression (lme):*

- j. Construct a model "baseModel" with perceived\_diversity as the dependent variable but without any predictors or random intercepts. Use gls, and make sure you choose method="ML".
- k. Construct an lme model "randomModel" with perceived\_diversity as the dependent variable and a random intercept for userId (but no other predictors). Again, choose method="ML".
- l. Add diversification to your model and call it "divModel".
- m. Compare "baseModel", "randomModel", and "divModel" using the anova() function. Interpret the results.
- n. Inspect "divModel" using summary() and interpret the results. What is the effect of medium and high diversification compared to low diversification?

## Question 2. Mixed ANOVA with ezANOVA and lme

*Let's see if list length also increases the perceived diversity:*

- a. Create a new variable: `div$length_factor <- factor(div$length, levels=c(5,10,15,20,25))`
- b. Conduct an ezANOVA with perceived\_diversity as the dependent variable, diversification as the within-subjects predictor, and length\_factor as a between-subjects predictor.
- c. Interpret the results. Do you have sphericity? What is the effect (F and p-value) of diversification? What is the effect of list length? Is there an interaction effect?

*If we run this as an lme, we can use length as a linear variable instead of a factor:*

- d. Add length as a predictor to "divModel" and call it "lengthModel".
- e. Add the interaction of length and diversification to this model and call it "intModel".
- f. Compare "divModel", "lengthModel", and "intModel" using the anova() function. Interpret the results.
- g. Inspect "intModel" using summary(). What is the predicted perceived diversity for a low diversity list with 25 items? What is the predicted perceived diversity for a high diversity list with 10 items?

## Question 3. Mediation analysis

*In questions 1 and 2 we have tested the effect: diversification → perceived diversity. At the beginning of this assignment, I mentioned that we also want to test how these factors subsequently affect tradeoff and choice difficulty. This effect can be described as: diversification*

→ perceived diversity → choice difficulty. In this model, perceived diversity is a **mediator** (it mediates the effect of diversification on choice difficulty). The psychologist David Kenny recommends the following steps to test this mediation model (see <http://davidakenny.net/cm/mediate.htm> for more info):

1. There should be a significant effect of diversification on choice difficulty
2. There should be a significant effect of diversification on perceived diversity (already supported by the tests in questions 1 and 2)
3. Using choice difficulty as the dependent variable and diversification and perceived diversity as predictors, the latter effect should be significant
4. For “full mediation”, the former effect should **not** be significant

Let's perform these tests to see if perceived diversity indeed mediates the effect of diversification on choice difficulty. Let's start with step 1:

- a. Construct a model “baseDiff” with choice\_difficulty as the dependent variable but without any predictors or random intercepts. Use gls, and make sure you choose method=”ML”.
- b. Construct an lme model “randomDiff” with choice\_difficulty as the dependent variable and a random intercept for userID (but no other predictors). Again, choose method=”ML”.
- c. Add diversification to your model and call it “divDiff”.
- d. Compare “baseDiff”, “randomDiff”, and “divDiff” using the anova() function. Interpret the results.
- e. Inspect “divDiff” using summary() and interpret the results. What is the effect of medium and high diversification compared to low diversification?
- f. Is step 1 of the mediation analysis supported by our tests? Why (not)?

Step 2 is already supported, so let's move to step 3:

- g. Add perceived\_diversity as a predictor to “divDiff” and call it “fullDiff”.
- h. Compare “divDiff” and “fullDiff” using the anova() function. Interpret the results.
- i. Inspect “fullDiff” using summary() and interpret the results. What is the effect of perceived\_diversity on choice\_difficulty, controlling for diversification? What is the effect of medium and high diversification compared to low diversification, controlling for perceived\_diversity?
- j. Is step 3 of the mediation analysis supported by our tests? Why (not)?

Finally, let's test step 4:

- k. Add perceived\_diversity to the “randomDiff” model and call it “perceivedDiff”.
- l. Compare “perceivedDiff” and “fullDiff” using the anova() function. Interpret the results.
- m. Is step 4 of the mediation analysis supported by our tests? Why (not)?

If all steps are supported, we can say that perceived\_diversity fully mediates the effect of diversification on choice\_difficulty. In other words, increasing the level of diversification decreases choice difficulty **because** it increases perceived diversity.