Exercises Day 5

PSY8003

Matthias Mittner

spring 2022

Exercise 1: Factor analysis

Use the dataset `psychbfi.dta` which contains data from participant's responses to the NEO-FFI personality questionnaire (Big-Five, https://en.wikipedia.org/wiki/Big_Five_personality_traits). As the name suggests, we expect 5 factors in this dataset:

- openness to experience (inventive/curious vs. consistent/cautious)
- conscientiousness (efficient/organized vs. extravagant/careless)
- extraversion (outgoing/energetic vs. solitary/reserved)
- agreeableness (friendly/compassionate vs. critical/rational)
- neuroticism (sensitive/nervous vs. resilient/confident)

Run an exploratory factor analysis. How many factors are suggested by the data? Does the pattern of factor loadings make sense? How do you interpret the results?

Run, in addition, a confirmatory factor analysis. In this case, we have strong expectations about the factor structure (because the instrument has been used in a lot of earlier research) so this is justified. The names of the variables suggest which factor each variable theoretically belongs to:

- `A`=Agreeableness etc.

Does the CFA confirm the theoretically proposed structure?

Exercise 2: Multi-level regression

Use the `imm10.dta` dataset which contains nested data of students within schools.

These are data containing, at the student level (`stuid`), information about math scores (`math`), socioeconomic status (`ses`), a homework score (`homework`), `sex` and other student characteristics. School level characteristics include mean socioeconomic status (`meanses`), urbanicity, teacher/student ratios (`ratio`), and other characteristics.

- Run a regular regression model (ignoring the multi-level structure in the data) to see whether the `homework`-score is positively related to the student's math performance. Make a scatter plot.
- Is this effect constant across schools? Run an MLM where you allow the slope and the intercept of the regression line for homework to vary across schools. Plot the effect. How do you interpret the results?
- To the last model, add student-level variable `ses` and school-level variable `meanses`. What is the result?