

Summer School Questions

Siqi Lyu

September 2020

Script and data for mismatch negativity (mmn)

Files

- Lyu_mmn.R
- smlp2020_SIQI_1_MMN.csv

Question

EEG always comes with a large amount of data points. In this dataset, I have one data point, which is the mean amplitude of the EEG, for each subject's ($n=8$) each event ($n \sim 100$) in each condition (role, $n=2$), in each topographic region (area, $n=2$) at each time interval ($n=10$).

Some of my colleagues first average the EVENT data for each SUBJECT when they preprocess the data in Matlab before importing it into R, and then they run LMM on the averaged dataset, e.g., for main effects of CONDITION and REGION, in different time windows (INTERVALS).

My question is, which way is “correct”/suggested - running LMM on the very raw dataset or the (preliminarily) averaged data?

Script and data for mismatch negativity (spr)

Files

- Lyu_spr.R
- smlp2020_SIQI_2_SPR.csv

Questions

When it comes to a design of more than 3 variables (e.g., $2 \times 2 \times 2 = 8$), I'm lost in model selection and interpretation.

Q1. Model selection

For example, in the current dataset, a full model on region 8 only shows a significant two-way $P \times N$ interaction. Is it the best model? What should I do next? Should I delete the fixed effects in the model one by one, one for each time, e.g., first the N.S. three-way interaction, and then the N.S. $C \times N$ interaction, etc., and compare the model fit after each step of deletion?

If there are four or more variables (and their interactions), there will be a long list of fixed effects from the full model. In which order should I choose the best model? This might be an old question, but I'm always confused when it comes to my own data.

Q2. Model interpretation

If the model output includes not only significant main effects, but also two-way or even three- or four-way interactions, how should I interpret all those significances? What we experimental linguists care about is always the interaction. For example, with the 222 design in the current dataset, I'll be expecting to find the three-way interactions on region 6, and will be interested to see the post-hoc results following that.

My question is, can I ignore the other two-way interactions (if there's any, unfortunately not in the current dataset) while only reporting/focusing on the three-way interaction and post-hoc tests thereof? Are there any necessary relations between the higher-level interactions and the lower-level ones that cannot be neglected/need be reported or taken into account?

I don't have a bigger dataset at hand, but I think the dataset you provided (MRK17_Exp1_extra.csv) is a good example.

- Do I have to discuss all significant effects in the model output?
- Which ones should I keep and which can be left out?
- Any relations between the four-way, three-way, and two-way interactions in the output?
- What's the 'correct'/suggested way to present results in the manuscript?