

# practical\_exercise\_3, Methods 3, 2021, autumn semester

[FILL IN YOUR NAME]

[FILL IN THE DATE]

## Exercises and objectives

The objectives of the exercises of this assignment are:

- 1) Download and organise the data and model and plot staircase responses based on fits of logistic functions
- 2) Fit multilevel models for response times
- 3) Fit multilevel models for count data

REMEMBER: In your report, make sure to include code that can reproduce the answers requested in the exercises below (**MAKE A KNITTED VERSION**)

REMEMBER: This assignment will be part of your final portfolio

### Exercise 1

Go to <https://osf.io/ecxsj/files/> and download the files associated with Experiment 2 (there should be 29). The data is associated with Experiment 2 of the article at the following DOI <https://doi.org/10.1016/j.concog.2019.03.007>

- 1) Put the data from all subjects into a single data frame
- 2) Describe the data and construct extra variables from the existing variables
  - i. add a variable to the data frame and call it *correct* (have it be a *logical* variable). Assign a 1 to each row where the subject indicated the correct answer and a 0 to each row where the subject indicated the incorrect answer (**Hint:** the variable *obj.resp* indicates whether the subject answered “even”, *e* or “odd”, *o*, and the variable *target\_type* indicates what was actually presented).
  - ii. describe what the following variables in the data frame contain, *trial.type*, *pas*, *trial*, *target.contrast*, *cue*, *task*, *target\_type*, *rt.subj*, *rt.obj*, *obj.resp*, *subject* and *correct*. (That means you can ignore the rest of the variables in your description). For each of them, indicate and argue for what **class** they should be classified into, e.g. *factor*, *numeric* etc.
  - iii. for the staircasing part **only**, create a plot for each subject where you plot the estimated function (on the *target.contrast* range from 0-1) based on the fitted values of a model (use **glm**) that models *correct* as dependent on *target.contrast*. These plots will be our *no-pooling* model. Comment on the fits - do we have enough data to plot the logistic functions?
  - iv. on top of those plots, add the estimated functions (on the *target.contrast* range from 0-1) for each subject based on partial pooling model (use **glmer** from the package **lme4**) where unique intercepts and slopes for *target.contrast* are modelled for each *subject*
  - v. in your own words, describe how the partial pooling model allows for a better fit for each subject

### Exercise 2

Now we **only** look at the *experiment* trials (*trial.type*)

- 1) Pick four subjects and plot their Quantile-Quantile (Q-Q) plots for the residuals of their objective response times (*rt.obj*) based on a model where only intercept is modelled
  - i. comment on these
  - ii. does a log-transformation of the response time data improve the Q-Q-plots?
- 2) Now do a partial pooling model modelling objective response times as dependent on *task*? (set `REML=FALSE` in your `lmer`-specification)
  - i. which would you include among your random effects and why? (support your choices with relevant measures, taking into account variance explained and number of parameters going into the modelling)
  - ii. explain in your own words what your chosen models says about response times between the different tasks
- 3) Now add *pas* and its interaction with *task* to the fixed effects
  - i. how many types of group intercepts (random effects) can you add without ending up with convergence issues or singular fits?
  - ii. create a model by adding random intercepts (without modelling slopes) that results in a singular fit - then use `print(VarCorr(<your.model>), comp='Variance')` to inspect the variance vector - explain why the fit is singular (Hint: read the first paragraph under details in the help for `isSingular`)
  - iii. in your own words - how could you explain why your model would result in a singular fit?

### Exercise 3

- 1) Initialise a new data frame, `data.count`. *count* should indicate the number of times they categorized their experience as *pas* 1-4 for each *task*. I.e. the data frame would have for subject 1: for task:singles, pas1 was used # times, pas2 was used # times, pas3 was used # times and pas4 was used # times. You would then do the same for task:pairs and task:quadruplet

```
## you can start from this if you want to, but you can also make your own from scratch
data.count <- data.frame(count = numeric(),
                          pas = numeric(), ## remember to make this into a factor afterwards
                          task = numeric(), ## and this too
                          subject = numeric()) ## and this too
```

- 2) Now fit a multilevel model that models a unique “slope” for *pas* for each *subject* with the interaction between *pas* and *task* and their main effects being modelled
  - i. which family should be used?
  - ii. why is a slope for *pas* not really being modelled?
  - iii. if you get a convergence error, try another algorithm (the default is the *Nelder\_Mead*) - try (*bobyqa*) for which the `dfoptim` package is needed. In `glmer`, you can add the following for the `control` argument: `glmerControl(optimizer="bobyqa")` (if you are interested, also have a look at the function `allFit`)
  - iv. when you have a converging fit - fit a model with only the main effects of *pas* and *task*. Compare this with the model that also includes the interaction
  - v. indicate which of the two models, you would choose and why
  - vi. based on your chosen model - write a short report on what this says about the distribution of ratings as dependent on *pas* and *task*

- vii. include a plot that shows the estimated amount of ratings for four subjects of your choosing
- 3) Finally, fit a multilevel model that models *correct* as dependent on *task* with a unique intercept for each *subject*
  - i. does *task* explain performance?
  - ii. add *pas* as a main effect on top of *task* - what are the consequences of that?
  - iii. now fit a multilevel model that models *correct* as dependent on *pas* with a unique intercept for each *subject*
  - iv. finally, fit a model that models the interaction between *task* and *pas* and their main effects
  - v. describe in your words which model is the best in explaining the variance in accuracy