Assignment 1 - Language Development in ASD - part 2

Natalia Makhova, Maria Mareckova, Katarina Michelova, Alex Pasternakova

September 20, 2017

## [1] 11

## Exercise 1) Preliminary Data Exploration

Describe the participant samples in the dataset (e.g. by diagnosis, age, etc.). Do you think the two groups are well balanced? If not, what do you think was the reason?

Participants samples are well balanced according to - Diagnosis - almost the same number of participants, 10 more TD participants - Age - Types\_MOT - spread all over but are slightly different, ASD group peaks at 350 and TD group peaks at 400

not well balanced - MLU\_MOT - mothers of TD children has slightly higher mean than ASD group - CHI\_MLU - TD group says much more words while ASD group says the most from 1-1,5 - types\_CHI - ASD group says much less(up to 50) unique words in comparison to TD group (up to 200) - token\_CHI - TD group said much more words (up to 800) while ASD up to 200

### Exercise 2) Children learning language: the effects of time and ASD

Describe linguistic development in TD and ASD children in terms of Mean Length of Utterance (MLU)?

In our model, we used child’s MLU as an a dependent variable, visit and diagnosis as independent variables. We also included random effects, specifically subject as random intercept and visit as random slope. After analysis, we found the visit as a significant predictor of child’s MLU , β = 0.23,(SE = 0.02), t = 9.516, p < .0001. However, diagnosis did not turn out significant, β = 0.29 ,(SE = 0.15), t = 1.91, p > .05. The result indicates that ASD children’s linguistic performance is not different from non-ASD children. But it also suggests that overall children MLU is changing with each visit. Aditionally, we calcualted r2 for our model, which was following: R2m=0.219 and R2c = 0.803. This indicates that our model with random effects explains significatly more variance but we cannot further generalize it.

### Exercise 3) Child directed speech as a moving target

Describe how parental use of language changes over time in terms of MLU. What do you think is going on?

According to our analysis, mothers’ MLU changes depending on Diagnosis, β = 0.50,(SE = 0.11), t = 4.42, p < .0001. And diagnosis is a significant predictor as well as time (in our case Visit), β = 0.12, (SE = 0.02), t = 6.60, p < .0001. This might suggest that knowing of diagnosis changes mothers’ speech accordingly. Moreover, the R2 for the model: R2m = 0.230 and R2c= 0.676, similarly suggests that including of random effects helps to explain more variance but cannot be generalized.

### Exercise 4) Looking into “individual differences” (demographic, clinical or cognitive profiles)

The dataset contains some additional variables characterizing the kids’ cognitive and clinical profile: ADOS (autism severity), MSEL EL (Expressive Language, that is, verbal IQ, or linguistic skills at first visit as assessed by a psychologist using Mullen Scales of Early Learning), MSEL VR (Visual Reception, used as a proxy for non verbal IQ at first visit), Age, Gender, Ethnicity. Would it make sense to add any of them to your model of linguistic trajectories? Create the best possible model (the one that best explain the data, with MLU as outcome). Next time your model will be tested on new participants, and we will proclaim a winner. Describe your strategy to select the best models (how did you choose the variables to include?) and send the code to Riccardo and Celine.

In order to identify best model, that describes our data, we have chosen as independant variable VISIT to predict Child MLU because we are interested in linguitic trajectory. We belive that we can’t use additional clinical variables (e.g. noverbalIQ or ADOS), because they were collected not every visit, thus we don’t have enough data. Additionally, we see no difference in variable AGE and Visit, so it makes no sence to add Age as IV. In our opinion, ethnicity and gender plays no role. We believe that child MLU is calculated from overall amount of words, which is token\_CHI variable. Therefore it is not reasonable to make token\_CHI as predictor. We did correlation analysis between CHI\_MLU and types\_chi and we discovered strong correlation. Also it safe to assume that types\_chi is a part of token\_chi. In our opinion trying to predict one index of liguistic performance, which is Child MLU, by another index of linguistic performance doesn’tmake much sense, because CHI-MLU was calculated from parts of it,and they also have strong correlation. Also if we do so, our main predictor, which is VISIT becomes not significant. However, our main priority is to see linguistic trajectory over time. Originally the idea was that Diagnosis could be a significant predictor, however, earlier in the code it was clear that diagnosis is not significant.

### [OPTIONAL] Exercise 5) Comment on how the three linguistic variables measure linguistic performance (the so-called “construct validity” of the measures). Do they express the same variance?