Assignment 1, part 2

**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?

The two groups only differ significantly on their language and on all other aspects (verbalIQ, nonverbalIQ, MLU at first visit and gender) there are no significant differences.

The reason why the data had to be non-matched for age is that language development for ASD is usaually slower and therefore matching age and language fluency would have been impossible.

**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)?**

Linguistic development of children MLU is affected significantly by time but not by diagnosis.

We produced a linermixed effects analysis of the relationship between child MLU, time and diagnosis. As fixed effects we entered visit(time) and diagnosis. We also entered random slopes for each subject and random intercepts for visits as random effects.

MLU was not significantly affected by diagnosis (beta = 0.29, SE = 0.15, t = 1.882, p > 0.05).

However, MLU was significantly affected by visit (beta = 0.23, SE = 0.025, t = 9.437, p < 0.05), meaning that for each new visit MLU increases by 0.23.

Our model was not significant compared to a null-model including only the random effects and visit (chisq(1,7) = 2.0177, p > 0.05). THe full model accounted for 81 % of variance whereoff the fixed effects accounted for 22 % of variance (R^2M = 0.2153, R^2C = 0.8063). Thus adding diagnosis as predictor did not add predictive value to the model

The residuals were normally distributed.

We also made a test on an interaction model, where we checked the interaction between visit and diagnosis. This model was significantly better than the model without an interaction effect (chisq(1, 8) = 34.962, p < 0.05).

In the model visit was significant (beta = 0.1044, SE = 0.027, t = 3.686, p < 0.05). Diagnosis was still not significant (beta = -0.217, SE = 0.17, t = -1.260, p > 0.05)

The interaction however was significant (beta = 0.253, SE = 0.038, t = 6.715, p < 0.05). The model with interaction accounted for 82 % of variance in the data, whereoff the fixed effects accounted for 35 % (R^2C = 0.818, R^2M = 0.351)

**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?

MLU of the children's parents is affected significantly by time and by diagnosis.

We produced a linear mixed effects analysis of the relationship between parent MLU, time and diagnosis. As fixed effects we entered visit(time) and diagnosis. We also entered random slopes for each subject and random intercepts for visits as random effects.

MLU of the parents was significantly affected by diagnosis (beta = 0.502, SE = 0.12, t = 6.542, p < 0.05), meaning that parents iwth a typically developing child had a MLU that was 0.502 longer than parents of children with ASD.

MLU of the parents was significantly affected by visit (beta = 0.12, SE = 0.02, t = 6.542, p < 0.05), meaning that for each new visit MLU increases by 0.12.

The model explained 68 percent of variance and the fixed effects explained 23 % of variance (R^2C = 0.6, R^2M = 0.23).

The residuals was normally distributed.

THe model was significantly better than a nullmodel, only including the random effects (chisq(2, 7) = 49.95, p < 0.05)

What is going on is that parents are creating their MLU by time and that parents with children with ASD use a smaller MLU as their children normally use a smaller MLU as well.

**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.

A model including an interaction between ADOS and visit, as well as the variable verbalIQ could explain 82 % of the variance in the data whereoff the fixed effects explained 65 % of the variance (R^2C = 0.82, R^2M = 0.65).

MLU of the child was significantly affected by visit (beta = 0.37, SE = 0.02, t = 15.856, p < 0.05), which means that MLU increased by 0.37 per visit.

MLU of the child was significantly affected by ADOS (beta = 0.03, SE = 0.01, t = 3.607, p < 0.05), which means that children with one point higher ADOS had an MLU that was higher by 0.03 (this is probably due to a slight difference i the means and the interaction will still make the MLU of children with autism lower.)

MLU of the child was significantly affected by verbalIQ (beta = 0.07, sE = 0.01, t = 8.918, p < 0.05) meaning that the childs MLU was 0.07 higher when IQ increased by 1.

MLU of the child was significantly affected by the interaction between visit and ADOS (beta = -0.02, SE = 0.00, t = -8.549, p < 0.05).

**[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?**

The three linguistic measures are MLU, unique words and words said in total.

The developments of these 3 is probably correlated. However, MLU tells more about the complexity of a sentence, unique words are measuring which new words would have been learned and words in total how much the child speaks during a visit.

BUt they probably all express part of the same variance.