***Statistics for Humanities Researchers – day5***

**Read the following instructions with the Day3Exercise1.Rmd file open.**

**The exercises should have 2 outputs:**

* **This document with answers to the questions. Cut and paste any plot you’re asked to produce.**
* **The Rmd file with your code written in and “knitted”**

*Background:* Autism Spectrum Disorder is often related to language impairment. However, this phenomenon has not been empirically traced in detail: i) relying on actual naturalistic language production, ii) over extended periods of time. We therefore videotaped ca 30 kids with ASD and 30 comparison kids (matched by linguistic performance) for ca. 30 minutes of naturalistic interactions with a parent. We repeated the data collection 6 times per kid, with 4 months between each visit. We transcribed the data and counted: i) how many unique words does the kid use in each video? How many does the parent? Ii) how many words per utterance (MLU) does the child use? And the parent? We then want to test the language trajectory of child and parent over time.

*Exercise 1)* Describe the participant samples in the dataset. Do you think it’s well balanced between the two groups?

*Exercise 2)* How does the developmental trajectory of unique words used change over time and by diagnosis? Identify the relevant Independent Variables to add to the model. Which are fixed and which random factors?

Draw a plot and report a statistical test.

Do you think the development follows a linear trend? What about MLU? How would you account for that?

*Exercise 3)* What about the parent? Draw a plot and report a statistical test of parental trajectories in language use. What do you think is going on?

*Exercise 4)* The dataset contains some additional variables characterizing the kids’ cognitive and clinical profile: ADOS (autism severity), Expressive Language (initial linguistic skills at first visit), Mullen1 (non verbal IQ at first visit), Age. Would it make sense to add any of them to your model of linguistic trajectories?

*OPTIONAL Exercise 5):* does the additional variable change the average level of e.g. MLU or its change over time? How would you statistically test the first? How would you statistically test the second?