Portfolio Exam - Part 1 | Methods 3 F2022, CogSci @ AU

Group 3 – Ditlev Kræn Andersen (699440), Bryan Roemelt (670705),

Patrik Molnar (694214), Sára Anna Szabó (690442), Manuela Skov Thomasen (650504)

12/10/2022

## Q1

*- Briefly describe your simulation process, its goals, and what you have learned from the simulation.*

*- Add at least a plot showcasing the results of the simulation. Make a special note on sample size considerations: how much data do you think you will need?*

*- What else could you do to increase the precision of your estimates?*

(BR, MST) Our simulations were designed to provide an informed understanding of the model we need to build for practical purposes, such as predicting the number of participants necessary to run a significant study and to suggest factors with significance in the linguistic development of autistic children.

(BR)Considering that our cleaned experimental data was sourced from the performance of 29 ASD and 32 TD participants, the intention of raising the simulated population sample to 200 was to test the diagnostic precision of the prior data for our estimations and modeling strategy going forward.

(MST, BR) We ran a simulation of 200 participants, 100 classified as typical developed children (TD) and 100 diagnosed with Autistic Spectrum Disorder (ASD), with six independent visits each. The intercept and slope for each participant was sampled from a normal distribution based on the prior data from the literature described in the assignment; these variables were then used to simulate the Mean Length of Utterance (MLU). Plot 1 shows the results of the simulation, where we can see the MLU increases across visits and, as anticipated, it looks like there is a bigger increase in MLU for the TD participants.

*Chart, line chart

Description automatically generated*

(MST) The simulated data was used to test our model and observe how it operates on data. We found that the model learns from the defined priors. To determine an appropriate sample size, a power analysis was run with the objective of finding a sample size that would give us an effect size of 0.8 or above. The result of the power analysis is visualized in plot 2 and 3. As seen in tibble 1 we achieve an effect size of 0.9 at 15 participants.

.Graphical user interface, application, table, Excel

Description automatically generated

**Graphical user interface, application, table, Excel

Description automatically generated**

**Graphical user interface

Description automatically generated**

(MST) The standard error is an indication of the precision of the estimates, so the lower the standard error, the higher the precision. A small standard error means that there is a small variation in the sample statistics across many repeats of the experiment. The standard error depends on the variability in the data and the sample size and is calculated by:

(MST) We have already looked at how an increase in sample size affects the effect size, and the bigger the sample size, the bigger the precision and effect size. Standard deviation describes the variability in the data, which is due to the participants being different, but can also be caused by the methodology or some lurking variable, something that affects the results that has not been considered. Therefore, it is important to investigate what can cause the variability in the data and consider how to avoid it.

## Q2

*- Briefly describe the empirical data and how they compare to what you learned from the simulation (what can you learn from them?).*

(SS) During the data cleaning process, certain data points were removed, so not all participants provided an equal number of data points. This is in accordance with the realistic expectation of children not being equally communicative at all times or having too much noise in the data, etc. (For example, the child with ID 1 only produced one data point that could be included in the final dataset). The data was recorded over the span of 6 visits, where parent-child interactions were recorded and later analyzed.

(SS) As a result of the cleaning, we obtained an R tibble containing 352 observations of 20 variables. Participants were anonymized via numeric IDs. 29 of the children were diagnosed with Autism Spectrum Disorder (ASD) and 32 were typically developing (TD). The variables describe the demographics of each participant, their diagnosis, measures of both verbal and non-verbal IQ, their and their respective mothers' MLU throughout the 6 visits.

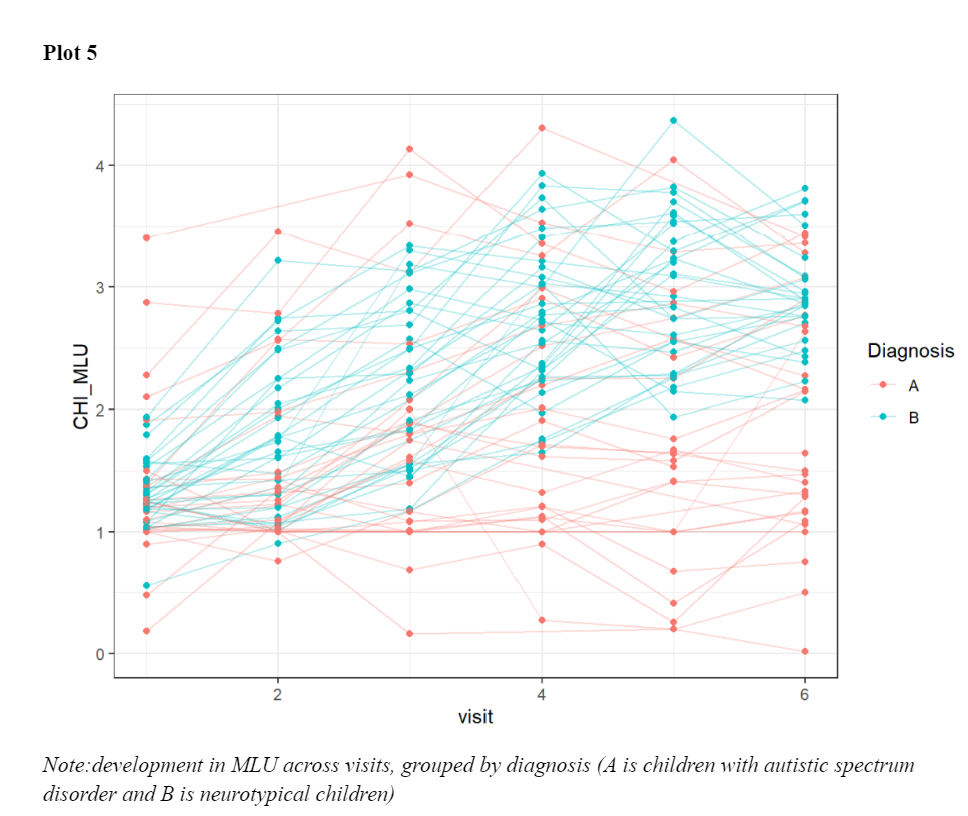
*What did we learn from the simulated data?*

(MST) In contrast to the empirical data that contained incomplete data sets for some child participants, the simulated data contains the full 6 data points per participant. The simulated data showed us that MLU increases across visits and the increase is bigger for TD children. Further we saw that our model became more confident when running on the data and not just priors. Plot 4 is one of the plots made as part of the prior-posterior update check on the model, here we see that the model becomes more confident as the standard deviation decreases for the posterior intercepts. Lastly the simulated data helped us determine a preferable sample size of 15 to get the wished effect size of 0.8 or above.

Chart, histogram

Description automatically generated

(MST) Plot 5 is made on the empirical data and shows the MLU across visits, just like plot 1. With the empirical data we do not see as clear a difference between the ASD and TD group as we did with the simulated data. Plot 5 shows that some children still have an increase in MLU across visits but it is not all children and there is generally more variance in the empirical data.

****

*-Briefly describe your model(s) and model quality. Report the findings: how does development differ between autistic and neurotypical children*

(MST) The model was defined as the following:

Our population-level effects gave an intercept estimate of 0.44 for autistic children with an error of 0.04 and of 0.49 with an error of 0.02 for neurotypical children. The estimate for the difference between visits was -0.01 (error: 0.02) for autistic children and 0.07(error:0.02) for neurotypical children, meaning that the model estimated autistic children to get lower MLU across visits.

The group-level effect estimated a standard deviation of the intercept between participants to be 0.42 (error: 0.05) and of the slope 0.15 (error: 0.02) which tells us that there is a lot of individual variation which was also what we could see in plot 5.

*- which additional factors should be included in the model?*

(PM) We added individual intercepts and slopes to see how it influences the data.

Chart, scatter chart

Description automatically generated

Chart, scatter chart

Description automatically generated