Victoria Engberg Lowe, Vlada Caraman &
Natasha Becker Bertelsen

BSc in Cognitive Science
Methods 3 E2022
12/10-2022

Methods 3 Assignment 1

Q1: The simulation

1.1 The goals of the simulation

Before we started working with the empirical data, we simulated similar data. The two primary goals for simulating data were:

1. To get a sense of what the empirical data might look like. By simulating data beforehand, we can get a better understanding of the problem before fitting a model to the actual data. For example, the simulated data can inform us about trends in the distributions that might be relevant for us to look after, it can show us visually how the empirical data might look, and it can exemplify the outcomes that we should expect etc.
2. To fit a model to the simulated data, in the hope that the model would turn out to be an appropriate model for the empirical data as well.

1.2 The simulation process

We started out by using the function set.seed at 1000 for generating "random" numbers. We specified the sample size and the number of visits. For this study, we set the sample size at 30 in order to avoid trouble with our model (related to divergence) that a smaller sample size could have caused. The number of visits is 6. Then, we created a new dataset and defined the key parameters for our model and we chose the following values for our parameters:

Sample size	30
Number of visits	6
Average MLU for ASD at Visit 1	log(1.5)
Average individual deviation (population standard deviation)	log(1.5) - log(1.5-0.5)
Average MLU for TD (population mean) at Visit 1	log(1.5)
Average individual deviation (population standard deviation)	log(1.5) - log(1.5-0.3)
Average change in MLU by visit for ASD (population mean)	0.1
Average individual deviation (population standard deviation)	0.03
Average change in MLU by visit for TD (population mean)	0.2
Average individual deviation (population standard deviation)	0.017
Error	0.1

The intercept values and their corresponding standard deviations were determined based on information from the literature. The values were put into a log-scale to ensure that we would only be dealing with positive numbers (as one cannot have a negative MLU). Since these values were in log-scale, we also needed to change the scale for the slope values and their corresponding standard deviations as well as the overall error. These values were determined by plotting histograms with the specified mean and standard deviation for each group by which we could visually ensure that they were scaled in accordance with the log-scaled values.

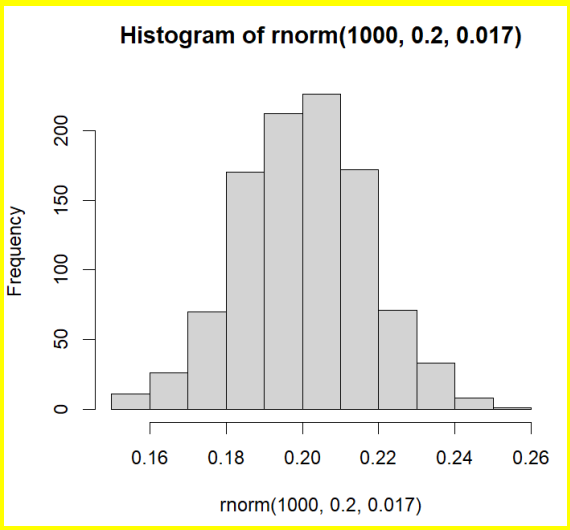


Figure 1: Histogram of average change in MLU by visit for TD with mean 0.2 and standard deviation 0.017.

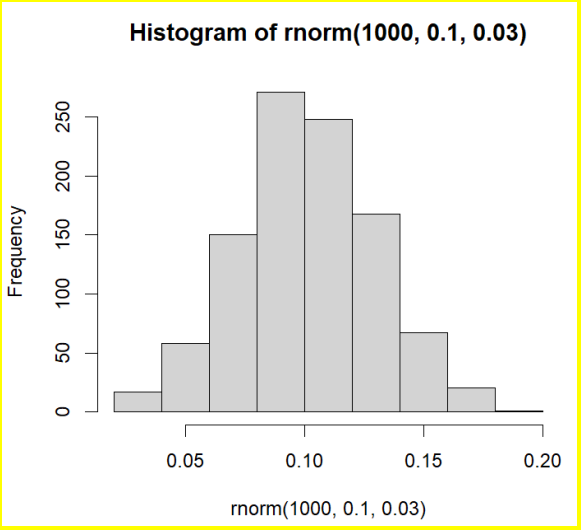


Figure 2: Histogram of average change in MLU by visit for ASD with mean 0.1 and standard deviation 0.03.

Based on the chosen parameter values, we simulated individual intercepts for 30 ASD children and 30 TD children respectively across 6 visits, as well as individual slopes that change by visit. Also, we simulated a MLU value for each child at visit 1. Since the intercept values are defined on a log-scale, we use the exponential function when simulating MLU values to make sure that the outcome values are on the right scale.

We visually inspected the simulated data and could see that the simulation at first hand was generating unrealistically large values (MLU > 15) and so we went back and adjusted the original values accordingly. More specifically, “Average change in MLU by visit for ASD (population mean)” was changed from 0.2 to 0.1, and “Average change in MLU by visit for TD (population mean)” was changed from 0.3 to 0.2. Additionally, “Average

individual deviation (population standard deviation)" and "Average individual deviation (population standard deviation)" were also adjusted from 0.06 and 0.09 to 0.03 and 0.017 respectively. Another visual inspection of the data revealed that we were simulating more appropriate values and therefore we left the values as they were then.

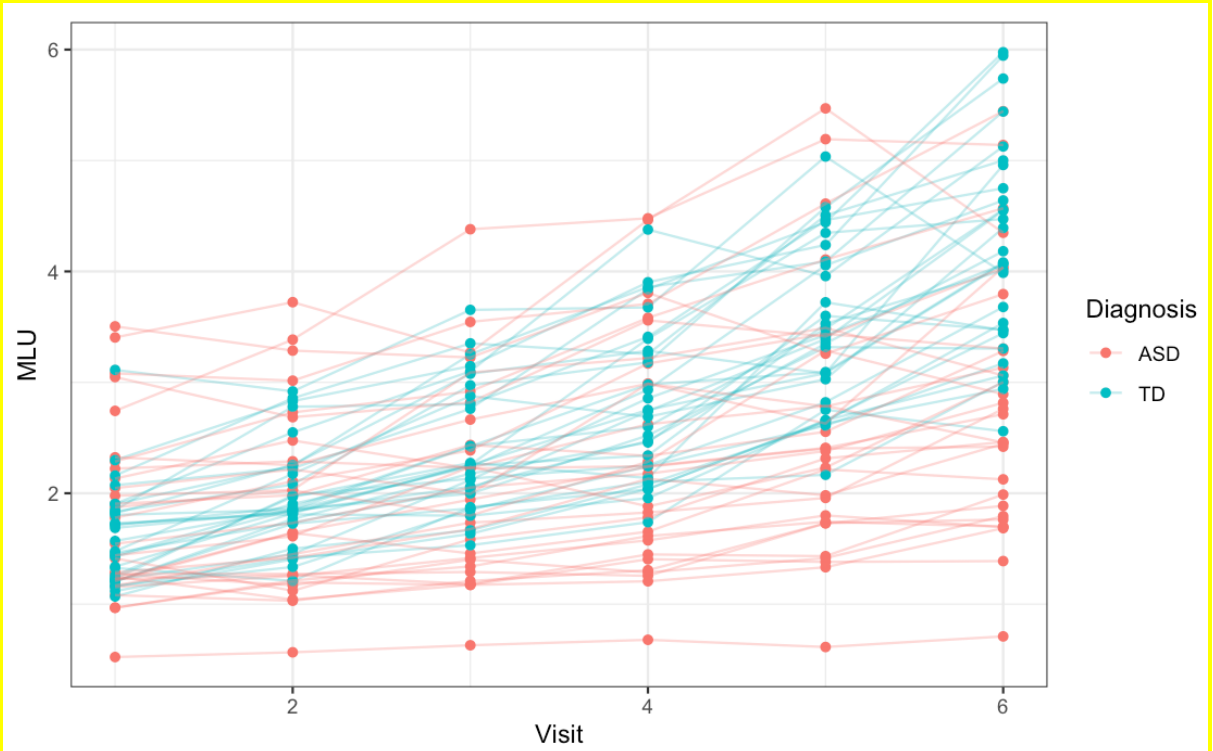


Figure 3: The simulated data.

Following a Bayesian workflow, we subsequently defined our model formula as the following:

```
MLU_f1 <- bf(MLU ~ 0 + Diagnosis + Diagnosis:Visit + (1 + Visit|ID))
```

We specify that the outcome variable that we want to predict is MLU. We tell the model that MLU is predicted by a binary variable (Diagnosis). By adding the number 0 (i.e. 0 + Diagnosis), we inform the model not to take any of the two categories as the intercept, but instead each will have its own estimate. Since we have two separate intercepts, we also inform the model that the slope of visit should be separated by the diagnostic group (Diagnosis:Visit). Lastly, we specify that the model should have random intercepts and slopes, meaning that each ID has its own slope and intercept (1 + Visit|ID).

Then we specified the following priors for our model based on inspiration from already existing literature on the topic as well as by plotting the distributions to see if the range of values produced were reasonable.

```
MLU_p1 <- c(
  prior(normal(0, 0.2), class = b),
  prior(normal(0.4, 0.1), class = b, coef = "DiagnosisASD"),
  prior(normal(0.4, 0.1), class = b, coef = "DiagnosisTD"),
  prior(normal(0, 0.1), class = sd, coef = Intercept, group = ID),
  prior(normal(0, 0.05), class = sd, coef = Visit, group = ID),
  prior(lkj(3), class = "cor")
)
```

We specify that all the betas are normally distributed with a mean of 0 and a standard deviation of 0.2 with the exception of the two intercepts and their corresponding standard deviation as well as the standard deviation for the slope that we specify separately. We do not expect the two groups to have different mean intercepts and slopes because we want the data to be able to tell us whether there is a difference or not. A prior predictive check is used to inspect the choice of prior values.

1.3 What we have learned from the simulation

By looking at our prior-posterior update checks, we can see that, for most cases, our models has learned from the data because the posterior distributions are much more confident (i.e. more pointy) compared to the priors, and the posterior distributions are not being pushed by the priors (except for the variability of intercepts).

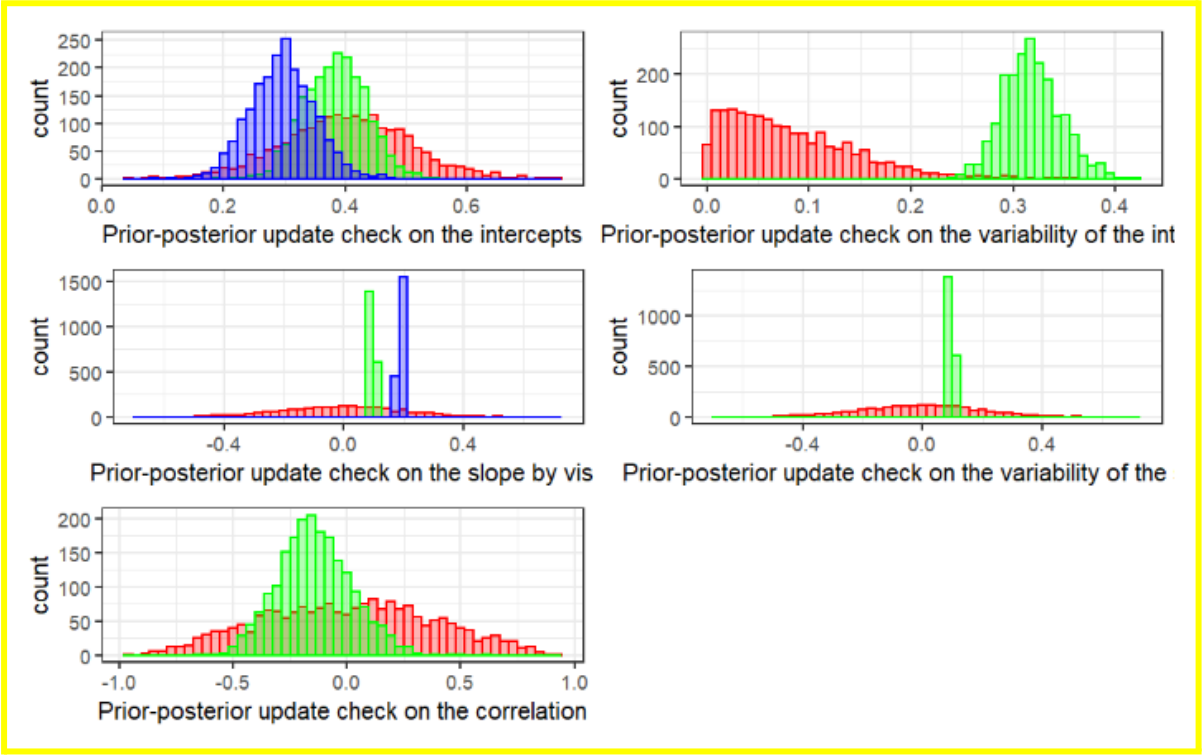


Figure 4: Prior-posterior update checks.

Plot 1: Red is the prior for both groups, blue is TD and green ASD.

Plot 2: Red is the prior, green is the posterior.

Plot 3: Red is the prior, blue is TD and green ASD.

Plot 4: Red prior and green sd posterior slope.

Plot 5: Prior red and green posterior.

1.4 Power analysis

Lastly, we did a power analysis to investigate sample size. We conducted the power analysis by 1) creating a function that could simulate the data (this was largely the function we had used to simulate the first set of data). 2) We did a test run of the function to see if the outcome looked realistic. 3) We made a function that could iterate the simulation function and save the outcome in a nested data frame. 4) We ran 100 iterations (i.e. 100 seeds) of the simulations. 5) We modified the nested dataframe so that it only contained the seeds and their respective values for the slope of ASD children and TD children. 6) We looked at the slope posteriors and calculated power, which came out to 1 meaning that there is no overlap between ASD and TD. 8) Lastly we visualized the power analysis (Figure 5).

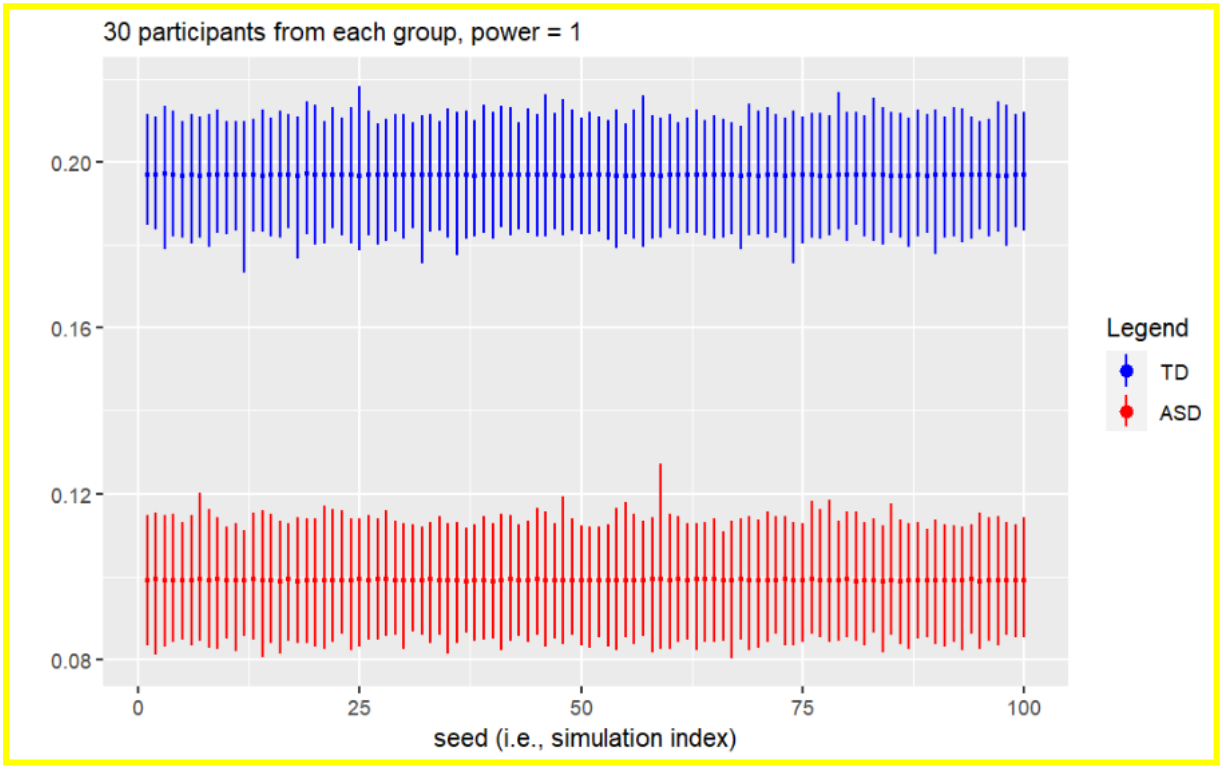


Figure 5: Power analysis plot

Our power analysis and subsequent visualization of it showed that a sample size of 30 participants pr. condition is sufficient (if we assume that the real data is roughly similar to the simulated data) as the power analysis revealed very precise and uniform results. This is also the reason why we opted not to do a precision analysis.

Q2: The empirical data

2.1 The empirical data and how it compares to the simulated data

The empirical data consists of a sample containing 66 participants, 55 of those were female and 11 were male, however not all participants participated in each visit and therefore the participant count varies from visit to visit, with 55 participants in the 1 visit and only 50 participants in visit 6. The mean age of the participants at visit 1 was 26.4 months old and the mean age of the participants at visit 6 (the last visit) was 47 months. During the study, the participants were tested on their non-verbal-iq, verbal-iq and socialization.

The non-verbal-iq was tested by measuring according to the Mullen Scales of Early Learning (MSEL). The verbal-iq was also measured according to MSEL. Their score of

socialization is an indicator of social interaction skills as well as social responsiveness as measured by Vineland.

We used these measurements to assess the children's clinical and cognitive features, and later we divided the data from the different measurements according to their diagnosis and calculated the mean for each group in the three different measures. We found that the ASD-children scored lower than the TD-children on all three measurements.

Diagnosis <chr>	mean_non_verb <dbl>	mean_verb <dbl>	mean_soc <dbl>
ASD	33.14943	22.61667	77.34286
TD	35.80000	29.28571	101.82990

Table 1: The children's clinical and cognitive features.

One issue for this study is that the sample has an overrepresentation of female participants with a ratio of about 5:1. This is potentially troublesome because in Denmark autism is four times more common (or at least diagnosed) in boys than in girls (Fenger-Grøndahl, Malene, 2017), so the sample seems unrepresentative of its population, this might lead to ungeneralizable results.

Since the study contains 66 participants, and each visit around 50-55 participants, the study has around the same amount of participants as our simulated data had. The power analysis from our simulated data showed that 30 participants pr. condition were enough to get reliable results. However, we have also seen that the empirical data is messier than the simulated data, in addition to this, not all participants participated in all 6 visits and all measurements, which adds to the amount of uncertainty. Because of this and the skewed sample, it might have been better to either get more participants and/or to have implemented a more rigorous exclusion criteria, so that the final data didn't have as many missing data points.

2.2 Model and model quality

The posterior-update check shows us that as expected the real data has more noise in the distribution compared to the simulated data. The plot shows that we have two underlying distributions in the data that are overlapping with each other. One distribution has a peak around the x-value 1.3, and the other distribution has a peak around 2.8. The noise in the data

could be due to the experimental setup of the study. We might get an unrealistic picture of the linguistic ability of the children because they might be affected by the research assistants who bring cameras to videotape the session etc. This is also emphasized by the fact that the data reveals that several children do not produce any words at all or a minimal amount of words during the session.

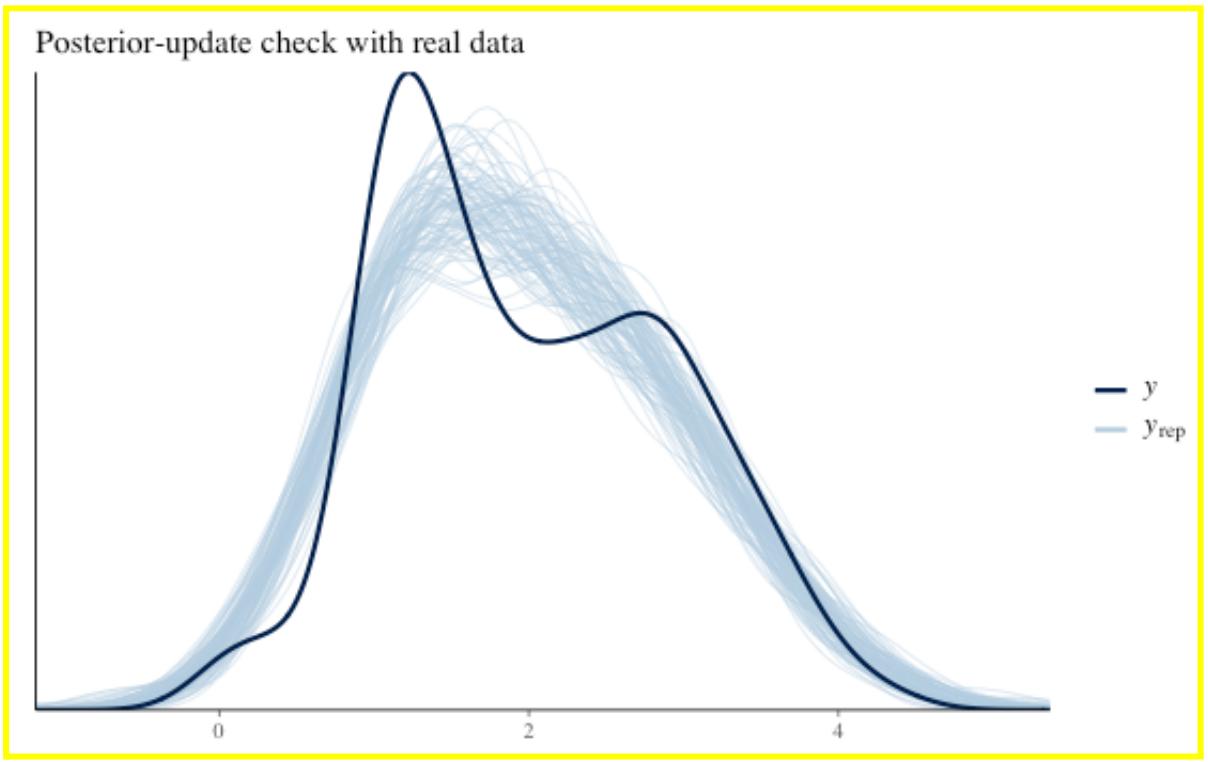
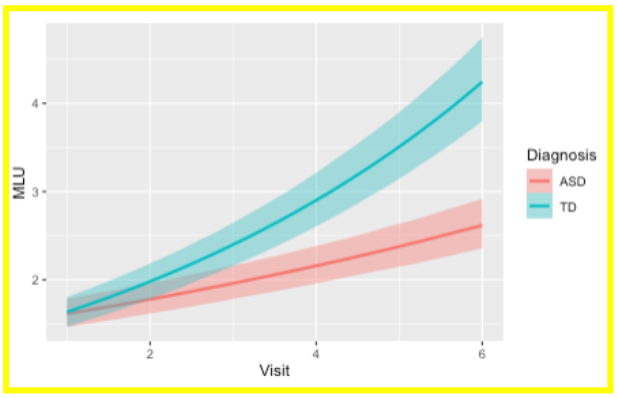


Figure 6: Posterior-update check with empirical data

The following two plots show how the development in linguistic ability differs between autistic and neurotypical children on a population level. The first plot shows the simulated data and the second plot the real data. We see that in general the neurotypical children have a higher linguistic ability compared to the autistic children. Additionally, it is clear that there is a growing difference between the two groups over time. Both the simulated and empirical data show the same general trend between the groups. However, the scale for the MLU values for the empirical data is a bit smaller.



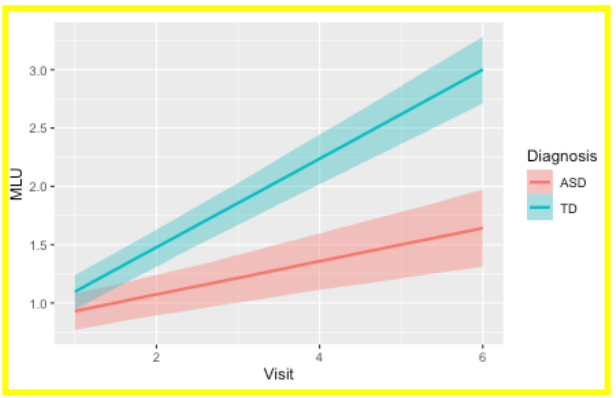


Figure 7: Conditional plot of the difference in linguistic ability between ASD and TD children for simulated data.	Figure 8: Conditional plot of the difference in linguistic ability between ASD and TD children for the empirical data.

The following plot illustrates the same population trend that the typically developing group in general have a higher MLU value compared to the autistic group. Also, the plot shows individual level findings. For example, we can see that the data for the autistic children have a lot more variability, and therefore there are autistic children who perform better than typically developing children at some visits, and there are other autistic children who have a very low MLU.

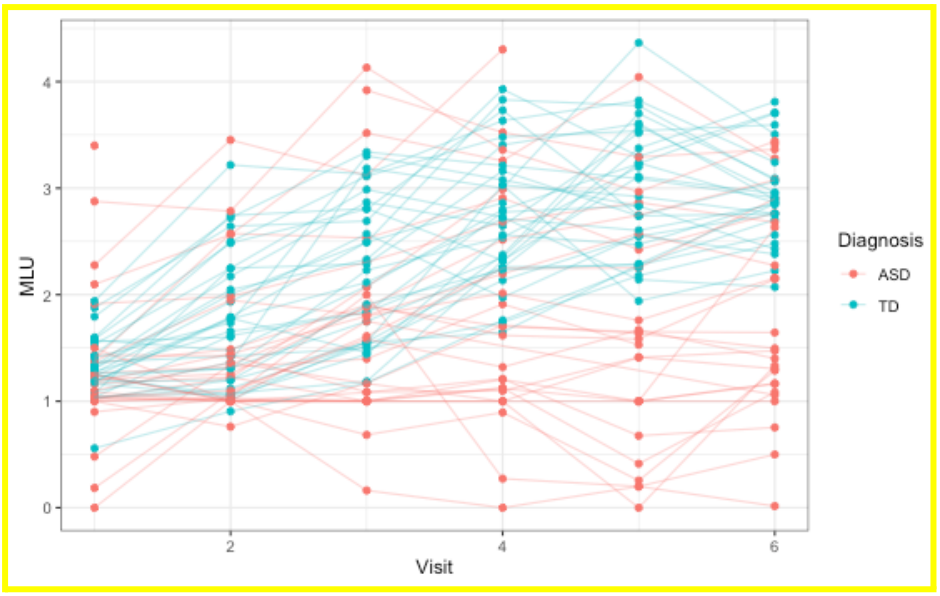


Figure 9: A plot of the empirical data with the visit number on the x-axis and the MLU value on the y-axis and data points grouped by diagnosis (red = ASD, blue = TD).

The model output on the empirical data tells us that the intercept for ASD children is 0.79 with an error of 0.08, and the slope is 0.14 with an error of 0.03. The intercept for TD children is 0.71 with an error of 0.08, and the slope for TD is 0.14 with an error of 0.03. The individual variability for the intercept is 0.48, and the individual variability for the slope is 0.09. The correlation is 0.12.

2.3 Additional factors in the model

We have added additional factors to our model to explore if the development of MLU could be affected by other factors that we have not yet explored. Therefore, we have specified the following models that include the MLU of the mother, the verbal IQ and the socialization variable:

```
MLU_f2 <- bf(MLU ~ 0 + Diagnosis + Diagnosis:Visit + Diagnosis:MOT_MLU + (1 + Visit|ID))
MLU_f3 <- bf(MLU ~ 0 + Diagnosis + Diagnosis:Visit + Diagnosis:VerbalIQ + (1 + Visit|ID))
MLU_f4 <- bf(MLU ~ 0 + Diagnosis + Diagnosis:Visit + Diagnosis:Socialization (1 + Visit|ID))
```

We have chosen to include the mothers' MLU in the model due to the fact that previous studies have found that the development of the mother's MLU increases in parallel with the child's MLU and therefore it would be interesting to investigate to which degree it might affect the results in this study. Additionally, we made a new model which also takes the child's verbal IQ into account, as that should also be indicative of how many words are available to the child. In one new model we have also included socialization, which indicates the children' social interaction skills and social responsiveness. We think that socialization could be an interesting predictor for the kids' language development.

Prospectively, for the model comparison we will go back and standardize all measures. Since the new variables are also included in the prior for the overall beta, we can use the same priors again. We will do prior and posterior update checks to evaluate the models and do a looic model comparison which will tell us which is the best model to use.

References:

Fenger-Grøndahl, Malene. (2017, April). *Autisme*. Faktalink. https://faktalink.dk/titelliste/auti