Portfolio 1

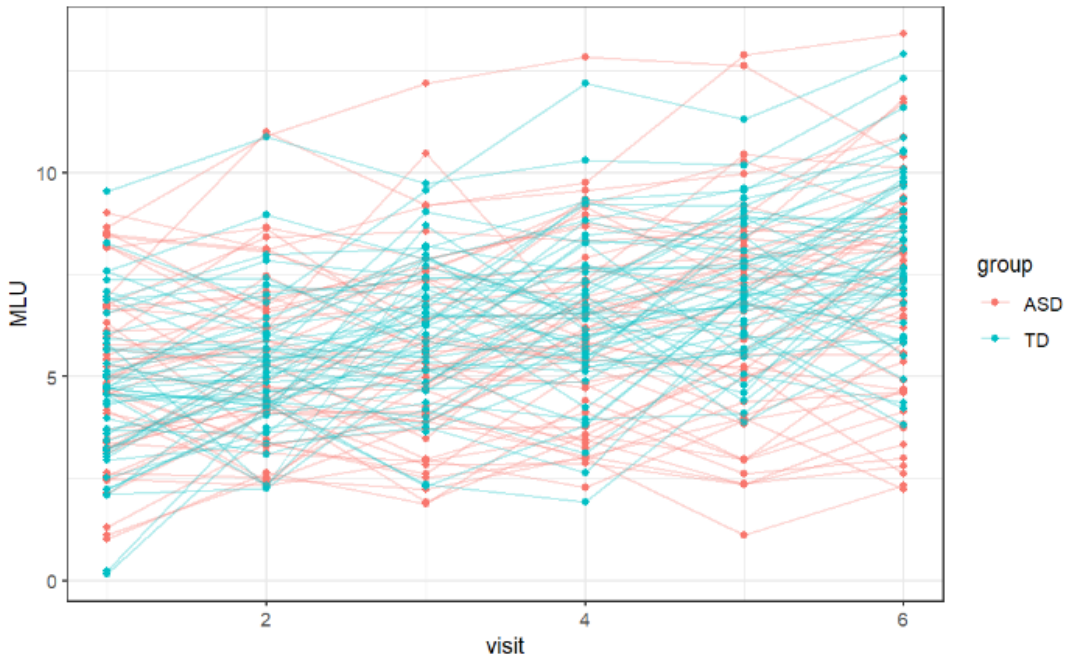
Methods 3, Bachelor of Cognitive Science
By Laura Rahbek, Marie Frederiksen, Ida Dencker & Sofie Mosegaard

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?

We aim to identify if there is a difference between the language development in autistic and typically developing children. Based on previous information of the difference between language development in the two groups, we simulate data to clarify how much data we would have to collect to run a meaningful study. This will give us an idea of how our model should be conducted using Bayesian Analysis in RStudio.

Our simulation data consisted of two populations: Children with an autism Spectrum Disorder (ASD) ($n = 50$, $\mu = 1.5$, $\sigma = 0.5$) and typically developing (TD) children ($n = 50$, $\mu = 1.5$, $\sigma = 0.4$). Each participant had six visits where their naturalistic, verbal interactions were recorded. Each group of participants' means and standard deviation were defined at baselevel (visit 1) as well as their individual changes per visit. This data was simulated to build a model that could estimate how many participants a potential future study requires to investigate the language development in autistic and neurotypical children.

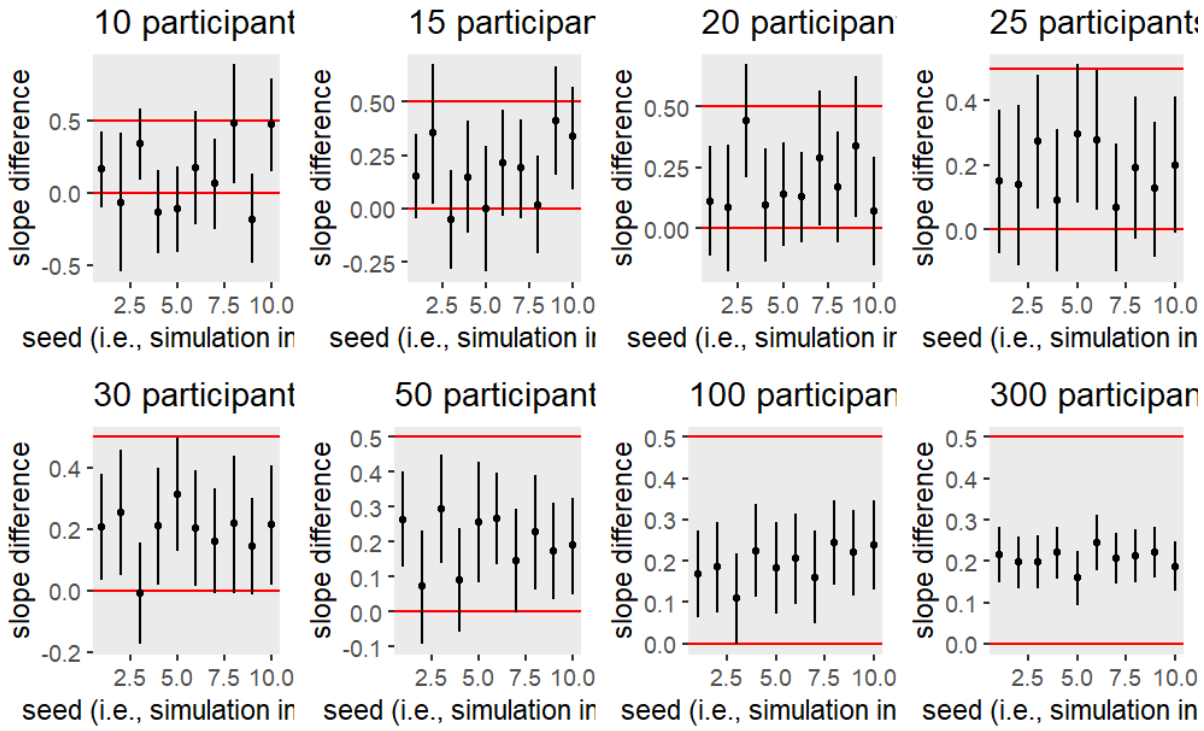
Initially, we proposed a model based on prior assumptions in order to predict and understand patterns. We simulated the data with individual mean length of utterances as the outcome and diagnosis (ASD and TD) and group by visit as fixed predictors. Furthermore, we separately assessed the random effects/predictors for each visit and modelled varying outcomes of participants, i.e., intercepts and slopes, separately for each group and visit. Thereafter, we conducted a model using the libeary brms with weakly informative priors, expecting no effects between diagnosis, to conservatively regularise the model parameters and thereby reducing overfitting and improving predictions. PP-check indicated that there is a difference between our priors and the data, meaning that there is reason to believe that autistic and typically developing children's language acquisition evolve differently. After assessing the model's parameters to ensure that they were satisfactory, we did prior posterior update checks on the model based on our priors and the data, which again supported the theory, that there is a difference in development between the two groups of children.



Plot 1: Mean Length of Utterance (MLU) over the 6 visits between ASD and TD from the simulated dataset.

Plot 1 indicates that ASD children have more divergence in their development of language in our simulated data. As their standard deviation is broader than TD children, the datapoints are placed above and underneath the datapoints of TD children indicating that the groups develop differently.

To conduct a future study it is crucial to be aware of how many participants are needed to reach a proper power and effect size as well as the development of a good predictive model. We found estimations for effect size in different sized simulations of participants (n in each group = 10, 15, 20, 25, 30, 50, 100, 300), where the data was simulated with each n of participants 10 times. The point indicates the difference in slope between the two groups and the vertical lines shows the 95% confidence interval. It is apparent that the effect size is around 0.25, which is a weak effect size. However, only very few simulations with above 30 participants, in each group, include 0 in the confidence intervals, suggesting that a study with 30 participants (again, in each group) would produce a high enough power. The effect size becomes more clear as participant number goes up.



Plot 2: slope differences with confidence intervals

To quantify our certainty that there is a difference between the two slopes, we can make a bayesian power analysis. The following table (table 1) shows us that our ideal power of 0.8 is reached between 50 and 100 participants, in each group. However, to ensure funding and accessibility we could argue that a participant number between 30 and 50 participants, in each group, would be acceptable.

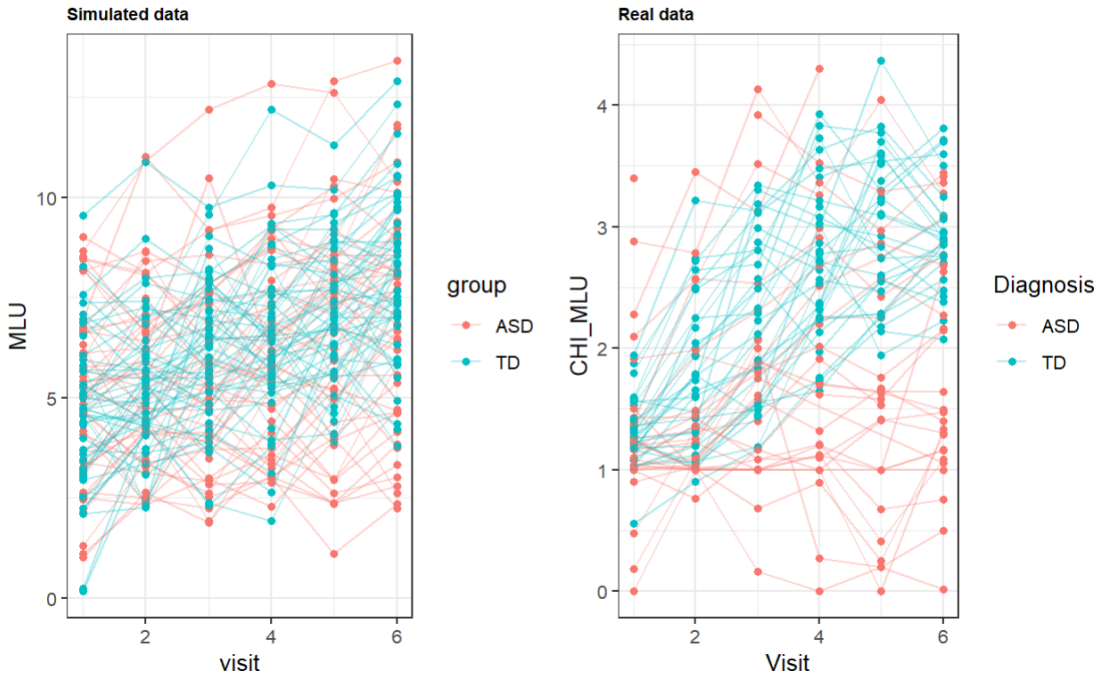
Power	Participants
0.3	10
0.3	15
0.3	20
0.3	25
0.6	30
0.7	50
0.9	100
1.0	300

Table 1: table of ideal power in different simulations

Several interventions could help precision within estimates besides an increased sample size. For example the experiment could be conducted in the private homes of our participants. In experiments involving humans and especially children, there will always be uncertainty in the construct of validity; a child's shyness, tiredness or bad mood could negatively affect the expected, positive, language development over each visit, where a familiar, intimate setting is expected to reduce this tendency.

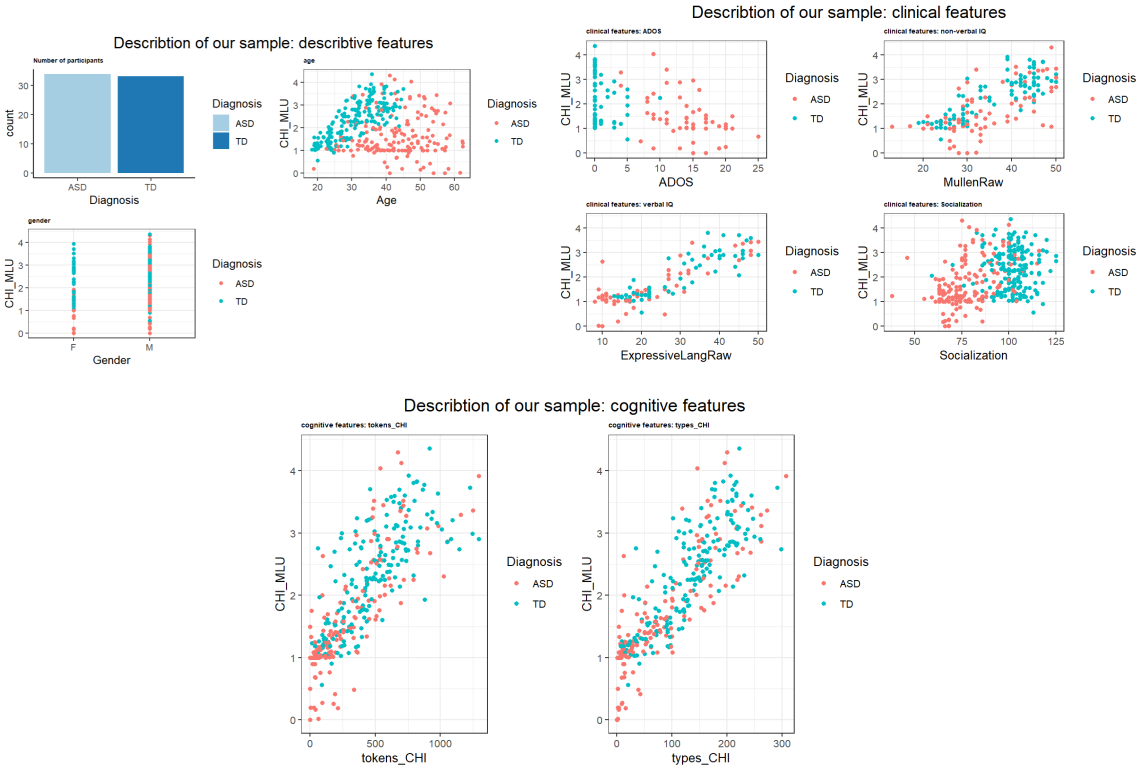
Q2 - Briefly describe the empirical data and how they compare to what you learned from the simulation (what can you learn from them?). Briefly describe your model(s) and model quality. Report the findings: how does development differ between autistic and neurotypical children (N.B. remember to report both population and individual level findings)? which additional factors should be included in the model? Add at least one plot showcasing your findings.

In the second part of this assignment, we aim to analyse the empirical data and interpret the inferential results. First, the simulated data was compared with the real data in plot 3:



Plot 3: simulated vs real data

The plot illustrates the simulated and real data with (MLU) over the 6 visits between ASD and TD. The two plots appear very similar, but the actual data points do have more fluctuations.



Plot 4: the different plots demonstrate the descriptive features of the real data

The empirical data has been plotted with MLU and multiple descriptive features as age, gender, clinical and cognitive features of the two groups. Generally, it can be assessed that the two groups are balanced in their number of participants, gender and non verbal IQ. However, ASD is generally older and scores lower on their clinical features in relation to verbal IQ (ExpressiveLangRaw) and social interaction skills (Socialisation) which can be expected.

Generally speaking, Autistic children have a more scattered development pattern, with some children increasing their MLU rapidly and some with slow development. Typically developed children have a more similar development pattern with modest increasing MLU pr. Visit.

Firstly, we fitted a multilevel model using Diagnosis and diagnosis by visti as predictor, also including random effects..

We look at the model's ability to recover the true underlying parameter values by investigating parameters of model output. Starting with the population level: regularly we look at whether the priors we set are in between the CI, however this is not relevant for this case in terms of the intercept, as the relevant thing is that they have the same intercept. We did set a mean of the intercept for both groups of 1.5, in order for data to persuade us of any difference. The CI for intercept of ASD does capture the value of 1.5 which is good. The estimate for the intercept for both groups (TD = 1.09, ,ASD = 1.31) are well under the prior (1.5). However the point of the study was to investigate the development in

language across the six visits, therefore the children were matched at visit one, so their intercepts would be similar. The estimates for the development for each visit (slope) are much lower than the priors (TD = 0.6, ASD = 0.4). But the differences for the two groups are apparent given the estimates (TD = 0.35, ASD = 0.1), with 95% confidence intervals that they do not overlap at all (CI for ASD = 0.05 - 0.10, CI for TD = 0.30 - 0.40). This tells us that there is a definite difference in how fast children with autism develop language compared to typically developing children. The data and model has convinced us that there is indeed a difference. The effective sample size tells us how much of the data from sampling is being kept. We specified 2 chains with each 2000, iterations where 1000 is warmup. ESS from 1437 to 2129 is good, keeping roughly between 35 and 54 % of the sampling. You could improve by increasing iterations, but we see no need for that.

We can also then look at group-level effects (also called random effects). The estimate of 0.56 for the sigma of the children's intercepts indicates the spread. The standard deviation of the children's slopes are estimated to be 0.11. A negative correlation of -0.1 indicates a negative relation between MLU and visit, those with higher intercepts have lower slopes and those with lower intercepts have higher slopes, which the estimates of the population-level support. We can also see from table 2, that our chains have converged sufficiently with a rhat of 1, which is great:

```
 Family: gaussian
  Links: mu = identity; sigma = identity
Formula: CHI_MLU ~ 0 + Diagnosis + Diagnosis:Visit + (1 + Visit | Child.ID)
   Data: df (Number of observations: 352)
  Draws: 2 chains, each with iter = 2000; warmup = 1000; thin = 1;
         total post-warmup draws = 2000

Group-Level Effects:
~Child.ID (Number of levels: 61)
                     Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
sd(Intercept)            0.56      0.08     0.41     0.72 1.00      879     1305
sd(Visit)                0.11      0.02     0.07     0.15 1.00      495      531
cor(Intercept,Visit)    -0.10      0.23    -0.48     0.43 1.00      616      532

Population-Level Effects:
                   Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
DiagnosisASD           1.31      0.13     1.06     1.56 1.00     1437     1463
DiagnosisTD            1.09      0.12     0.86     1.33 1.00     1675     1476
DiagnosisASD:Visit     0.10      0.03     0.05     0.15 1.00     2061     1176
DiagnosisTD:Visit      0.35      0.03     0.30     0.40 1.00     2129     1528

Family Specific Parameters:
      Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
sigma     0.41      0.02     0.37     0.44 1.00      960     1723

Draws were sampled using sampling(NUTS). For each parameter, Bulk_ESS
and Tail_ESS are effective sample size measures, and Rhat is the potential
scale reduction factor on split chains (at convergence, Rhat = 1).
```

Table 2: summary of our model

We then use the hypothesis function to test our hypothesis: slope of TD is higher than ASD. The posterior probability tells that there is 100% probability when sampling 2 kids (one from each) that we would find TD is higher in MLU than ASD. It can be concluded that the distribution of slope of TD and distribution of slope of ASD are significantly different. This could make sense since the IC (for

the difference of the 2 slope distributions) even when accounting for error don't fall below 0, indicating that there is a certain difference.

Finally, we fitted 4 additional models: one also including ADOS, one including socialisation, one including both ADOS and Socialisation and one including verbal predictors of the mother. Note all still had diagnosis as predictor. To test for model comparison, we used the measure LOOIC:

Models	LOOIC
1	462.63
2	187.01
3	452.97
4	188.47
5	418.90

Table 3: the different models and their LOOIC scores

The one with the lowest LOOIC is the better model, being model two with socialisation (*CHI_MLU ~ 0 + Diagnosis + Diagnosis:Visit + Diagnosis:Socialization + (1 + Visit| Child.ID)*). However, it is worth noting that the LOOIC score does not tell us anything about the absolute quality of the model, only the quality in relation to the other models. Nevertheless, model two is the best model for predicting language development in autistic and neurotypical children.