Language Development in ASD - part 4

By Dumbledore's Army (Lisa, Astrid, Sebastian, Fredrik, Lena)

Exercise 1

How much power does your study have (if your model estimates are quite right)?

We're testing the power of the model [quadratic: lmer(CHI_MLU ~ (Visits + I(Visits^2)) * Diagnosis + (1+Visits|SUBJ),Data, REML=FALSE)] using simulations.

- Report the power analysis and comment on what you can (or cannot) use its estimates for.
The power for the effect of time-diagnosis interaction is [insert here]. The power of the effect of time interaction is [insert here]. The power of the effect of diagnosis interaction is [insert here].
This is the probability that we find the effect, we are looking for (difference in language acquisition between healthy and ASD children, and the effect of time on language acquisition in children) in our experimental setup, *given* that the effect is actually there.
[the estimates here are missing - our function is having the issue of giving only 0% or 100% the second time we ran through it, we followed the advice you gave in class but it still doesn't work. What could be the issue?]

Exercise 2

How would you perform a more conservative power analysis?
- Identify and justify a minimum effect size for each of your relevant effects

Typical effect sizes in psychological experiments is d=.3 or d=.4. On this level, we will need a minimum of 1600 observations per condition: In that setup this translates into for ASD-condition and for TD-condition, each.
Given our 60-some children, we will need around 25 observations (visits) per child.

In this experiment, however, we realize there is a difference in effect size between the diagnosis-variable and the time-variable. The former has a relatively big effect size (d=1.5), and thus needs less observations in order to reach the 80% power. Time-variable, on the other hand, is so negligibly small d=.1 that the effect can be *very* difficult to track, and would demand many observations.

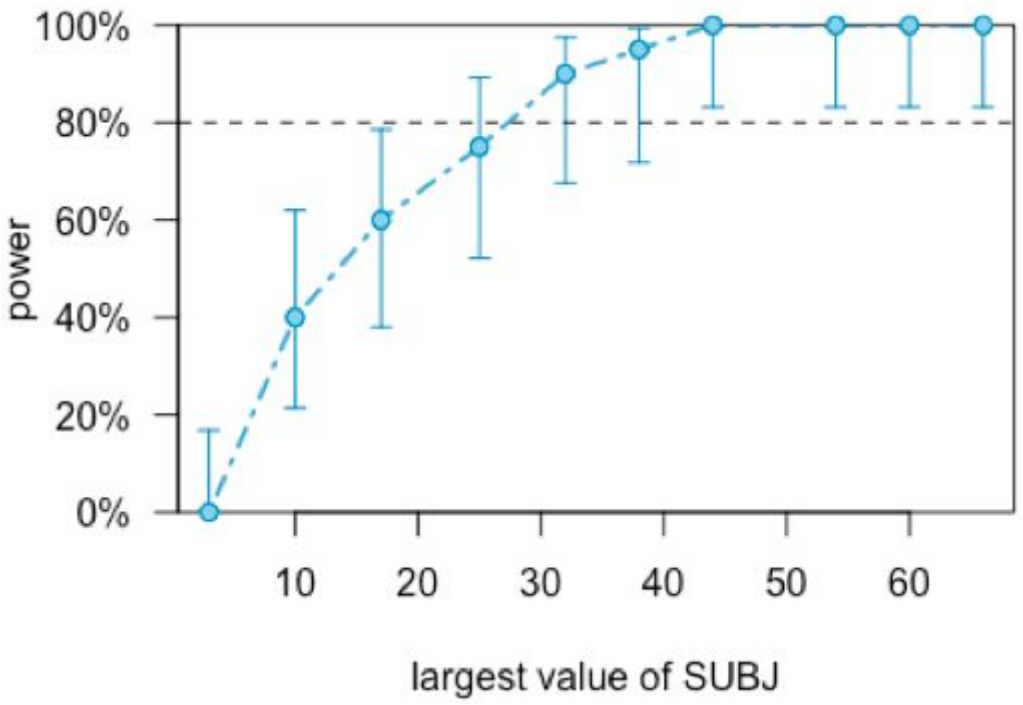
The big effect size of diagnosis, however, might be inflated, as the number of observations are relatively limited. Thus, we shouldn't expect an effect size as high as d=1.5.

Looking into this experiment by itself, we want to ask ourselves: In terms of language production, how big variations between the groups do we want to find, in order to consider it

significant? As the average length of utterance is already quite low (~2 words per utterance), it wouldn't make sense to look for a big, absolute effect. A meaningful difference could be 20%, translating into a variation of ~0.4 words per utterance between the healthy and ASD-children.

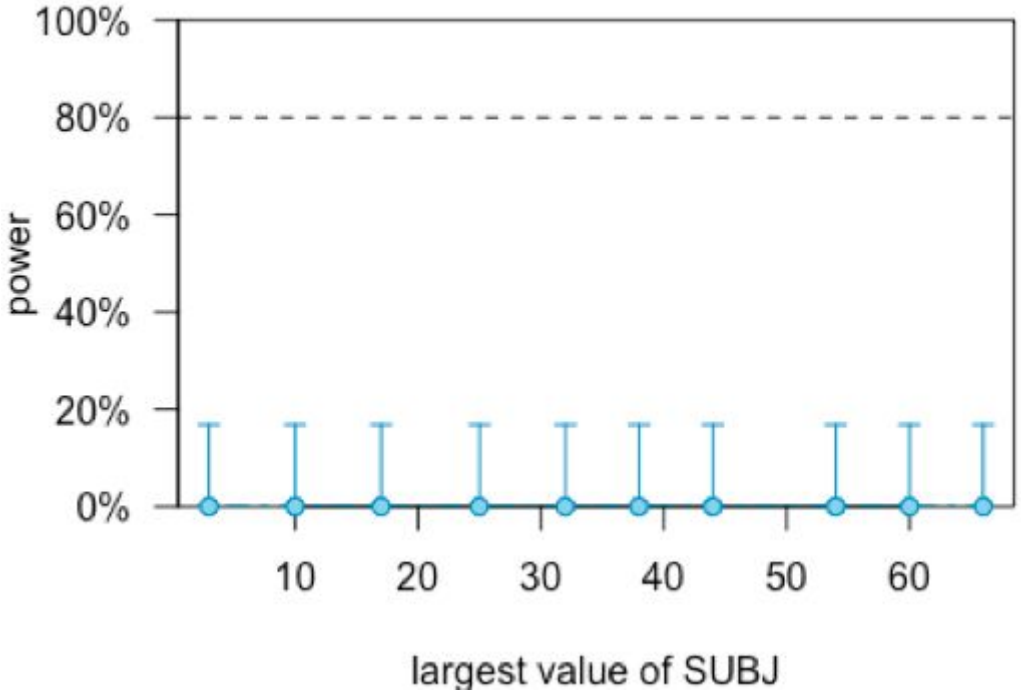
- Take the model from exercise 1 and replace the effects with the minimum effect size that you'd accept. Assess the power curve by Child.ID, identifying an ideal number of participants to estimate each effect

Power curve of interaction between diagnosis and time



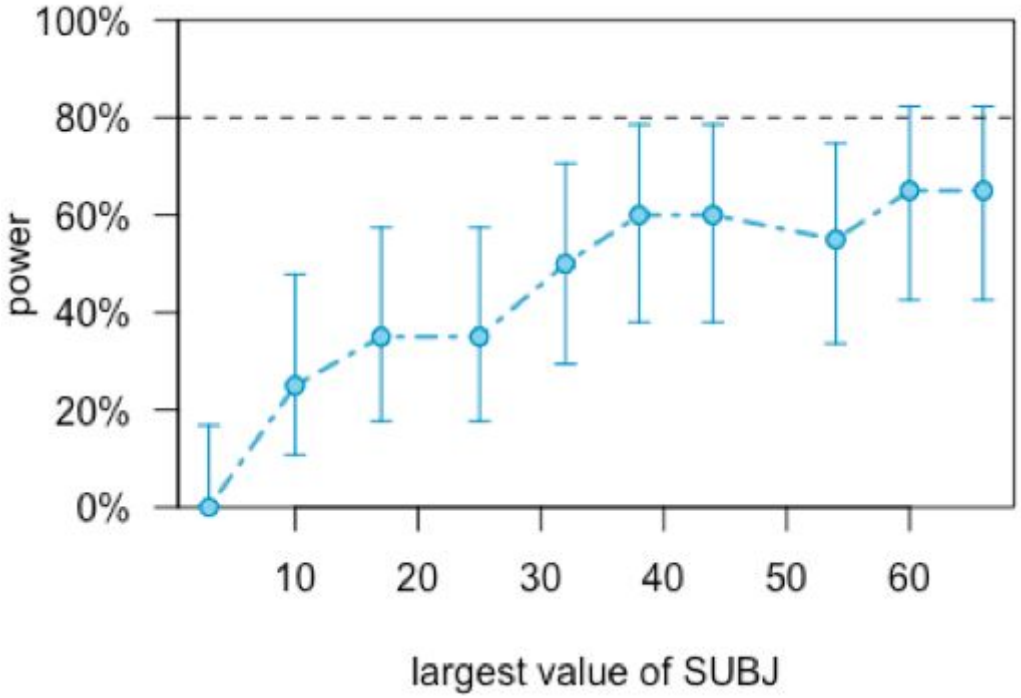
In order to get a power of 80 % of tracking the interaction effect between time and diagnosis, we need just 30 participants in this setup (6 visits/ subject).

Power curve of time



Effect size of time is so negligible in this sample, that a power of 80 % is out of reach. It's 0 %. Even running more simulations won't help us detect this effect. Effect size of time on language production is seemingly super duper teeny tiny (nonexistent).

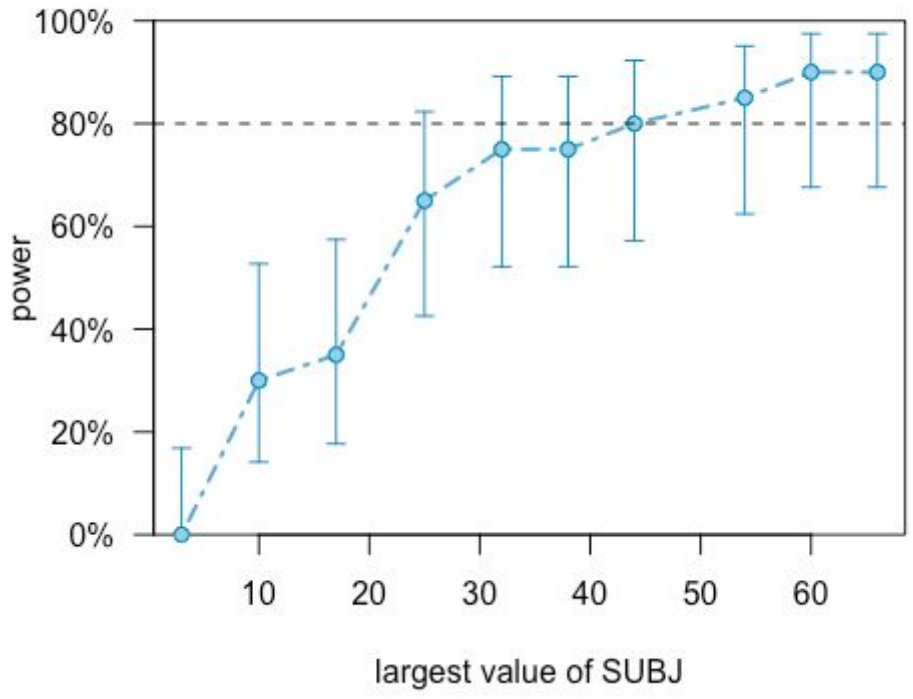
Power curve of diagnosis



Given the 60 participants we have in this experimental setup, we cannot expect reaching a power of 80% for tracking the effect of diagnosis on language production. We must run more simulations to get a better estimate.

- If your power estimates do not reach an acceptable threshold simulate additional participants and repeat the previous analysis

Power curve of diagnosis, vol. II



We used the simulate() function, which generates a new data set based on our model, to double the data.
The 80%-threshold was reached at around 43 participants.

A different approach to reaching the threshold, would be getting more *items*: Rather than getting more participants, we can gather more observations by going on visit 7, visit 8, etc. by the current participants.

- Report the power analysis and comment on what you can (or cannot) use its estimates for. [the estimates here are missing - our function is having the issue of giving only 0% or 100%, we followed the advice you gave in class but it still doesn't work. What could be the issue?]
power <- powerSim(quadratic, simr::fixed (("Diagnosis"), method = "t"), nsim = 200)
[insert est. power here] is the found power for effect of diagnosis. This is the probability that we find the effect, we are looking for (difference in language acquisition between healthy and ASD children) in our experimental setup, *given* that the effect is actually there.

[insert est. power here] is the found power for effect of time. This is the probability of finding the effect, we are looking for (the effect of time on language acquisition in children) in our experimental setup, given the effect is there.

[insert est. power here] is the found power for the interaction effect of time and diagnosis. This is the probability of finding the effect, we are looking for (difference in language acquisition between healthy and ASD children, and the development of language through time differing between the two groups) in our experimental setup, given the effect is there.

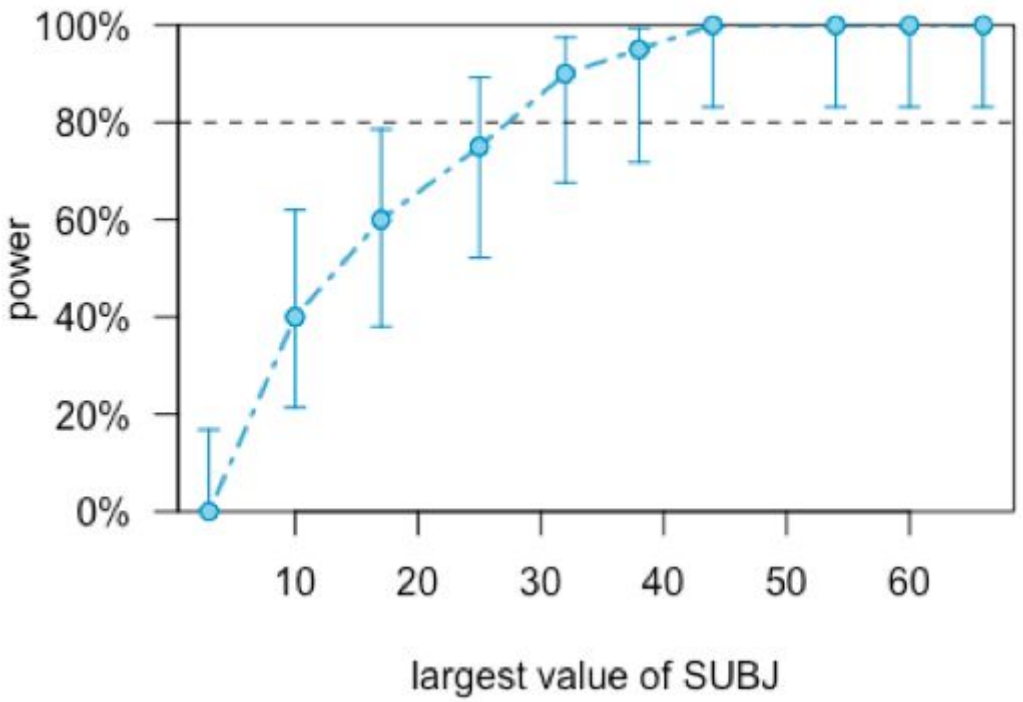
Exercise 3

Assume you have only the resources to collect 30 kids (15 with ASD and 15 TDs). Identify the power for each relevant effect and discuss whether it's worth to run the study and why.
We can read this off of our power curves. Given only 30 participants, we get a power of 40-50% for the interaction effect of visits and diagnosis, given the time-horizon of 6 visits per participant.

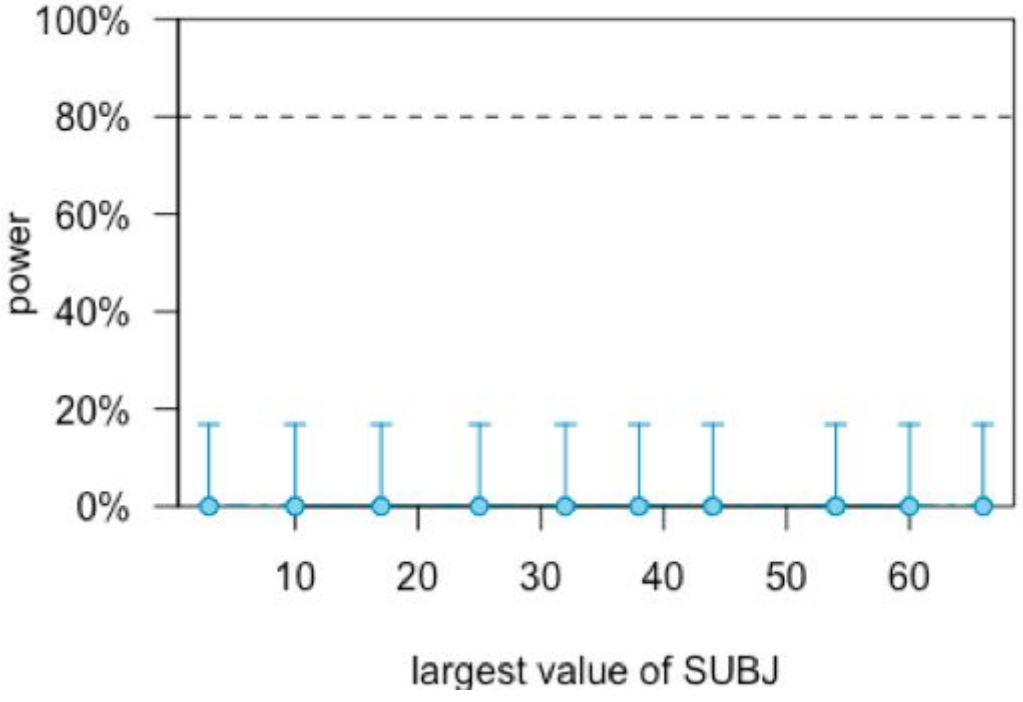
For the the effect of visits only, we are left with a power of [insert % here] given 30 participants.

For the the effect of diagnosis only, the power given 30 participants is [insert % here].

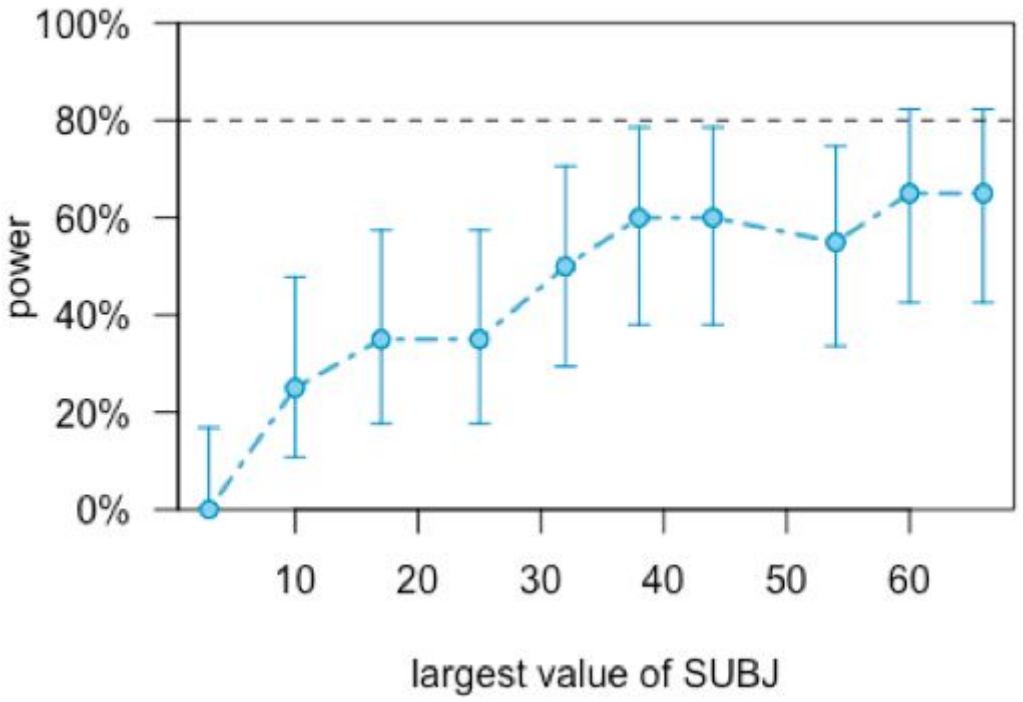
CODE
powerCurveI = powerCurve(quadratic,fixed("Visits:Diagnosis"),along="SUBJ", nsim=20)
plot(powerCurveI)




```
powerCurveV = powerCurve(quadratic,fixed("Visits"),along="SUBJ", nsim=20)
plot(powerCurveV)
```




```
powerCurveD = powerCurve(quadratic,fixed("Diagnosis"),along="SUBJ", nsim=20)
plot(powerCurveD)
```



Link for Gitlab
https://github.com/sebsebar/Alouishes/blob/master/Assignment2_4_LangASD_power.Rmd