Assignment 1 - Language Development in ASD - part 4

Signe Kløve Kjær and Thea Rolskov Sloth

Oct, 2018

Link to code: <https://github.com/YpipY/EXPMETH3/blob/master/power/signe_thea_A1_P4_LangASD_power.Rmd>

# Exercise 1

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

### - [GitHub]Load your dataset, fit your favorite model, assess power for your main effects and interactions of interest.

### - Report the power analysis and comment on what you can (or cannot) use its estimates for.

We tested the following model on the training dataset:

## pow\_model = lmer(CHI\_MLU ~ VISIT \* Diagnosis + (1 + VISIT | SUBJ))

Assessment of power:

By using the powerSim function with the t-test method we calculated the power for each of our fixed effects seperately and as an interaction effect:

The power for predictor 'Diagnosis' is 24.50% with confindence intervals from 21.86: 27.29. Effect size for Diagnosis is -0.21. This is a very low power. However, the effect of most interest is the interaction effect, and therefore the low power for Diagnosis alone is not a major issue, as it contributes to the analysis of interest.

The power for predictor 'VISIT' is 96.60% with confidence intervals from 95.28: 97.63. Effect size for VISIT is 0.10. This is a very high power.

Power for the interaction 'VISIT:Diagnosis' is 100.0%, with confidence intervals from 99.63: 100. Effect size is 0.25. This is obviously the highest power possible. With this level of power we are assured that if there is a true effect, the model would find the effect 100 % of the time.

# Exercise 2

### How would you perform a more conservative power analysis?

### - Identify and justify a minimum effect size for each of your relevant effects

A more conservative power analysis would accept smaller effects of the predictors, and therefore we adjust the effect sizes to slightly smaller values than the effect sizes from the model. We need more participants to find a smaller effect, thus we expect the power to reach a level of 80% at a higher number of participants, than in the previous analysis.

Normally you would make a power curve before doing the actual experiment and the statistical analysis. This is done in order to design your experiment with a sufficient number of participants to ensure a high enough power. When making a power curve you specify the smallest effect size you will accept. This is the smallest difference between levels that you find meaningful. One way to rationalize for a specific minimum effect is to look in the litterature. Another way is to run a pilot study. This can pose a problem as you are likely to find effects which do not generalize to the population, as in a pilot study you would often have a few number of participants to train the model on.

For Diagnosis the effect size is now set to: -0.15

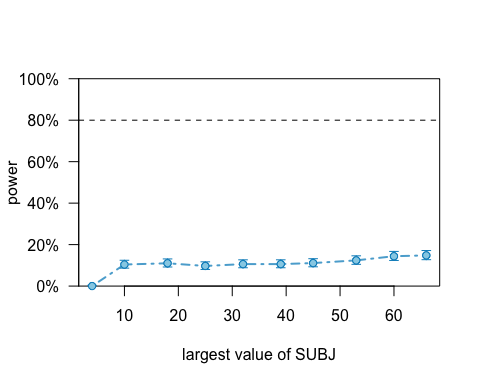
For visit the effect size is now set to: 0.08

For the interaction effect the effect size is now set to: 0.2

# - [GitHub] take the model from exercise 1 and replace the effects with the minimum effect size that you'd accept.

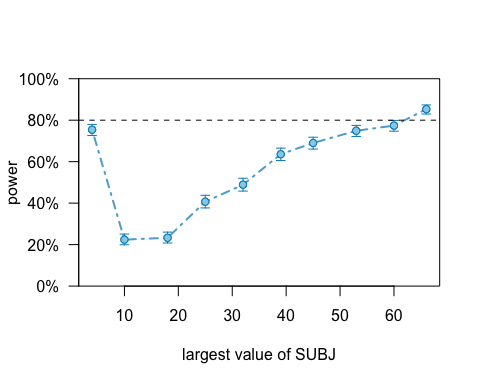
# - [GitHub] assess the power curve by Child.ID, identifying an ideal number of participants to estimate each effect

## Calculating power at 10 sample sizes along SUBJ



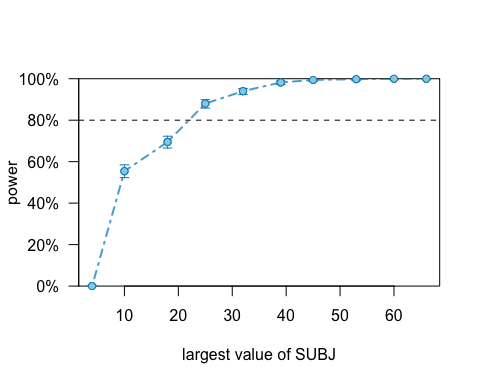
powerCurveDiag

## Power for predictor 'Diagnosis', (95% confidence interval),  
## by largest value of SUBJ:  
## 4: 0.00% ( 0.00, 0.37) - 17 rows  
## 10: 10.40% ( 8.58, 12.46) - 51 rows  
## 18: 11.00% ( 9.13, 13.11) - 92 rows  
## 25: 9.70% ( 7.94, 11.70) - 127 rows  
## 32: 10.60% ( 8.76, 12.68) - 162 rows  
## 39: 10.60% ( 8.76, 12.68) - 204 rows  
## 45: 11.10% ( 9.22, 13.21) - 238 rows  
## 53: 12.40% (10.42, 14.60) - 272 rows  
## 60: 14.40% (12.28, 16.73) - 312 rows  
## 66: 14.80% (12.66, 17.15) - 346 rows



powerCurveVisit

## Power for predictor 'VISIT', (95% confidence interval),  
## by largest value of SUBJ:  
## 4: 75.40% (72.61, 78.04) - 17 rows  
## 10: 22.40% (19.85, 25.11) - 51 rows  
## 18: 23.30% (20.71, 26.05) - 92 rows  
## 25: 40.70% (37.64, 43.82) - 127 rows  
## 32: 48.90% (45.76, 52.05) - 162 rows  
## 39: 63.60% (60.53, 66.59) - 204 rows  
## 45: 69.00% (66.03, 71.86) - 238 rows  
## 53: 74.90% (72.09, 77.56) - 272 rows  
## 60: 77.40% (74.68, 79.96) - 312 rows  
## 66: 85.30% (82.95, 87.44) - 346 rows



## Power for predictor 'VISIT:Diagnosis', (95% confidence interval),  
## by largest value of SUBJ:  
## 4: 0.00% ( 0.00, 0.37) - 17 rows  
## 10: 55.40% (52.26, 58.51) - 51 rows  
## 18: 69.50% (66.54, 72.34) - 92 rows  
## 25: 88.00% (85.82, 89.95) - 127 rows  
## 32: 94.00% (92.34, 95.39) - 162 rows  
## 39: 98.20% (97.17, 98.93) - 204 rows  
## 45: 99.40% (98.70, 99.78) - 238 rows  
## 53: 99.80% (99.28, 99.98) - 272 rows  
## 60: 100.0% (99.63, 100.0) - 312 rows  
## 66: 100.0% (99.63, 100.0) - 346 rows

### - Report the power analysis and comment on what you can (or cannot) use its estimates for.

For the predictor "diagnosis" the curve does not approach the 80% power limit... If we look at the estimate for diagnosis it is -0.21. This number can sound a little weird as it would indicate that when we go from ASD to TD, the value of MLU is 0.21 lower for TD compared to ASD. THis does not make intuitive sense as we would expect the MLU value to be higher for TD compared to ASD. The reason for this is, that we have an interaction effect in the model and it's therefore not really possible to assess the estimates in a meaningful way. It is from these estimates that we rationalized for minimum effects to make a more conservative poweranalysis. It therefore makes sense that the power for this predictor alone is very bad.

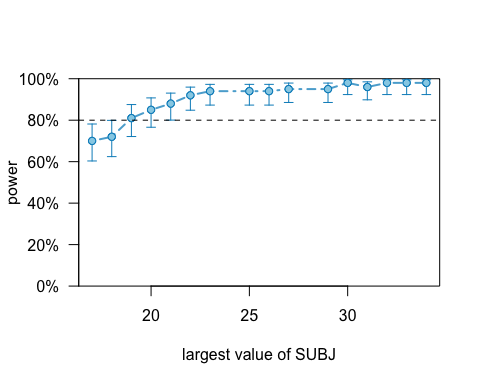
To find an effect size of 0.08 of the predictor "visit" we need around 62 participants to have a power of 80%. Examining the powercurve we can tell that something is wrong as the power with 4 participants is 73% and thereafter drops to to 23% with 10 participants and then rises as would be expected.

To find an effect size of 0.2 of the interaction effect of visit and diagnosis we need around 24 participants to have a power of 80 %.

### 

# 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.



## Power for predictor 'VISIT:Diagnosis', (95% confidence interval),  
## by largest value of SUBJ:  
## 17: 70.00% (60.02, 78.76) - 86 rows  
## 18: 72.00% (62.13, 80.52) - 92 rows  
## 19: 81.00% (71.93, 88.16) - 97 rows  
## 20: 85.00% (76.47, 91.35) - 103 rows  
## 21: 88.00% (79.98, 93.64) - 109 rows  
## 22: 92.00% (84.84, 96.48) - 115 rows  
## 23: 94.00% (87.40, 97.77) - 121 rows  
## 25: 94.00% (87.40, 97.77) - 127 rows  
## 26: 94.00% (87.40, 97.77) - 133 rows  
## 27: 95.00% (88.72, 98.36) - 139 rows  
## 29: 95.00% (88.72, 98.36) - 145 rows  
## 30: 98.00% (92.96, 99.76) - 151 rows  
## 31: 96.00% (90.07, 98.90) - 157 rows  
## 32: 98.00% (92.96, 99.76) - 162 rows  
## 33: 98.00% (92.96, 99.76) - 168 rows  
## 34: 98.00% (92.96, 99.76) - 174 rows

The only relevant effect that we are interested in is the interaction effect of visit and diagnosis.

From looking at the plot and the table we can see that with 30 participants, the power is 95%. It would be worth runnning this study since this level of power is obtained with an even smaller effect size that previously. This means that the power is still high at a rather conservative effect size.