Power analysis

04 November 2019

Justification

We plan to analyse our data using Growth Curve Analysis (GCA; Mirman, 2016) fitting a mixed-effects model that predicts the probability of fixation to the target at each time point within a given trial. Following Barr, Levy, Scheepers, & Tily (2013), we will specify a maximal random-effects structure. Our model will include, at least, polynomial transformations of the time domain (first, second, and third order, initially), prime-target phonological relatedness, and prime cognateness, as fixed effects. We will include random intercepts for participants and stimuli, and all combinations of random slopes for fixed effects by participants and items. In case of convergence issues due to model complexity, we will prune terms of the model following a pre-specified order.

There are limitations regarding the calculation of statistical power of mixed-effects models like ours, mainly due to the lack of consensus on how to estimate degrees of freedom. For this reason, we opted for calculating the power for a two-way repeated-measures ANOVA that only includes prime-target relatedness and prime cognateness as within-participant factors. Such analysis would neglect two important sources of variability, making it underspowered compared to the GCA we plan to perform.

The lack of random effect involving stimuli makes the ANOVA neglect two main sources of variability. First, the across-stimuli variability: some prime-target pairs may elicity larger priming effects. Second, this model neglects the temporal unfolding of fixations to the target across trials: the probability of fixation to the target must be summarised for each trial. This means tha radically divergent patterns of looking behaviour within trials can lead to the same average probability of fixation to the target (Barr, 2008). The neglect of these two sources of variation makes this analysis underpowered when compared with the GCA we plan to perform.

For this reason our GCA will achieve a higher power that that resulting from a power analysis of the ANOVA, given identical parameters. Thus, we will perform a power analysis based on an underpowered ANOVA to calculate the under which parameters our GCA will achieve an appropriate power.

Due to contraints related to sample collection, we expect the maximum sample size we will be able to reach to be around 200 toddlers, including those tested in Barcelona and Oxford. Given this fixed sample size, we calculated the smallest effect size of interest we would be able to detect with that sample size ($N = 200$), given a desired 80% statistical power, and a 5% significance criterion.

Power analysis

We used the `pwr` R package (Champely, 2018) to calculate the smallest effect size we would be able to detect with a repeated measures ANOVA the only included prime-target relatedness and prime-cognateness as within-participant factors. This effect size would be an indicator of the sensitivity of our design. We used the following code:

```r
n       <- 100    # sample size (by group)
power   <- 0.80   # desired power
alpha   <- 0.05   # significance criterion
es.eta2 <- 0.06 # effect size (eta-squared)
f       <- sqrt(es.eta2/(1-es.eta2)) # effect size (Cohen's f)

pwr <- pwr::pwr.anova.test(
```
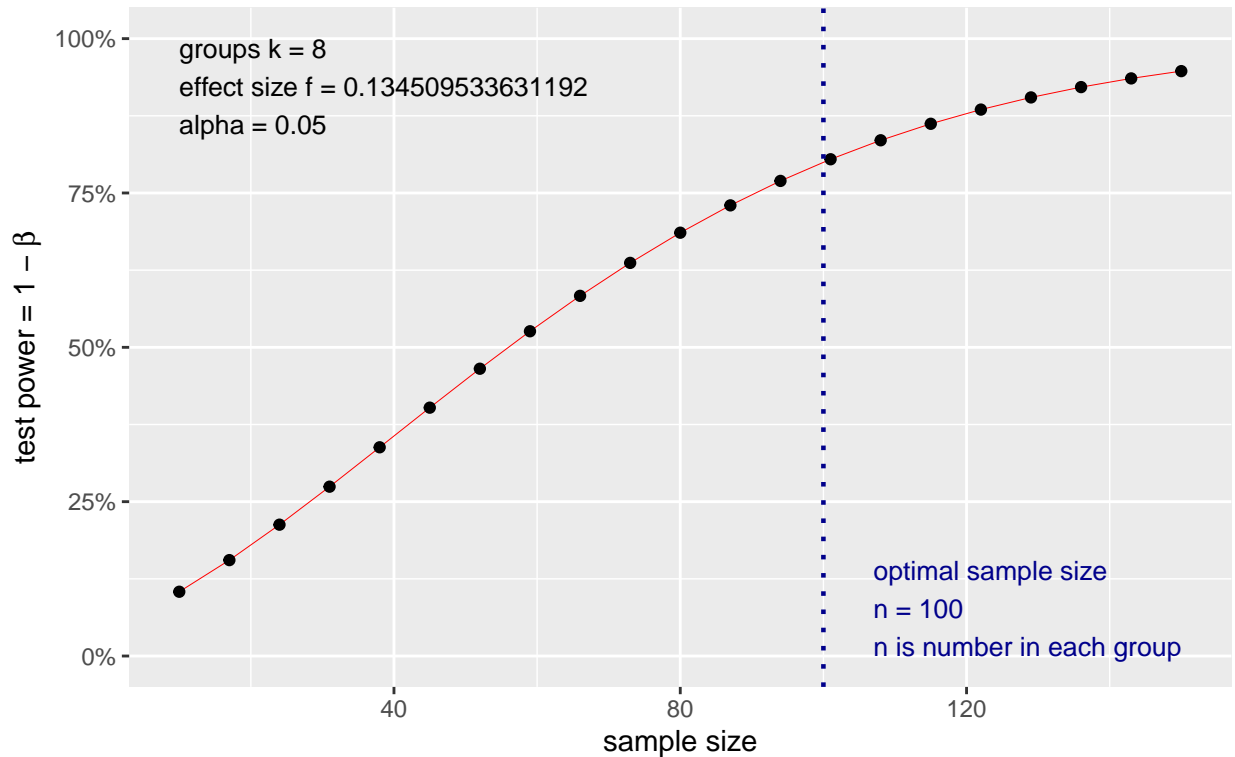
Figure 1: Statistical power as a function of sample size. Given the repeated measures nature of the ANOVA, the number of observations in each condition of the 2x2 design will be equal to the total sample size.

```
k        = 8,     # number of groups (2W x 2W x 2B)
n        = n,     # maximum sample size per condition (1 by participant x 100 participants)
sig.level = alpha, # significance criterion
power    = power
)
```

The smallest effect size our design would detect is a Cohen's $f = 0.135$, small according to Cohen (1969) standards. This effect size is equivalent to $\eta_p^2 = 0.024$. Put in others words our effect would be able to detect an effect that explains 2.415% of the variance. Given that our planned GCA would yield a higher power, our power to detect the same effect size will be higher than 80%. Figure 1 depicts the increase in power associated with each possible sample size, given our parameters.

References

Barr, D. J. (2008). Analyzing "visual world" eyetracking data using multilevel logistic regression. *Journal of Memory and Language*, *59*(4), 457–474. https://doi.org/10.1016/j.jml.2007.09.002

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*(3), 255–278. https://doi.org/10.1016/j.jml.2012.11.001

Champely, S. (2018). *Pwr: Basic functions for power analysis.* Retrieved from https://CRAN.R-project. org/package=pwr

Cohen, J. (1969). *Statistical power analysis for the behavioral sciences.* Academic Press.

Mirman, D. (2016). *Growth Curve Analysis and Visualization Using R.* CRC Press.