Example for reproducible sample size calculation

Below you can find an example of a transparent and reproducible write-up of the sample size estimation for the example that we used in class. In this example we calculated the necessary sample size for replicating the study by Liu and colleagues, and we wanted to power our study to be able to detect an effect of the mindfulness intervention vs. control if this effect exists (Liu et al., 2013).

The main hypothesis is that the mindfulness intervention can increase pain tolerance in the cold pressor task, and this hypothesis is tested in a within-between mixed study design. Participants are randomly allocated into two groups: an intervention group and a control group. Both groups start first by undergoing the cold pressor task (a standard pain induction procedure), and baseline pain tolerance is recorded. After the baseline measurement, the intervention group gets mindfulness-based instructions for pain reduction, while the control group only gets a very simple instruction which is not related to mindfulness. Eventually in a few minutes the cold pressor task is repeated.

The primary hypothesis is that pain tolerance will increase more in the intervention group from pretest to posttest than in the control group. This hypothesis will be tested using a repeated measures ANOVA (or equivalent, like a mixed linear regression model), by looking at the time * group interaction effect.

Sample size estimation

I conducted an a-priori sample size estimation based on data from the literature used G*power (Faul et al., 2009). The study was powered to be able to test the main research hypothesis with 90% power. We will test this hypothesis using a repeated measures ANOVA where we will look at the significance of the time * group interaction, so I chose the ANOVA: repeated measures, within-between interaction option in g*power. The effect size was estimated from information reported in three research studies using similar research designs as the one planned in this study (Forsyth & Hayes, 2014; Liu et al., 2013; Swain & Trevena, 2014). Since the papers did not report standardized effect sizes, I used the mean gain scores and pre and post test standard deviations reported in the papers to calculate Cohen's d effect size using the effect size calculator web-page of the Practical meta-analysis book (Mark W. Lipsey & Wilson, 2001, 2001). Pre- to posttest correlation coefficient was not reported in any of the three papers. Based on my experience on the field working with experimental pain scores that are recorded a few minutes after each other an $r = 0.7$ is a reasonable estimate, so I used this value for effect size calculation and later entered this value in the sample size calculation as well. The effect sizes and confidence intervals extracted from these papers were 0.89 (0.37-1.4), 0.84 (0.33-1.36), and 1.01 (0.66-1.36) respectively (For the exact values extracted from the studies, see the online supplement.). I wanted to be able to detect an effect even if it is closer to the lower bound of the confidence intervals, so I chose cohen's $d = 0.33$ ($f = 0.165$) as a minimal effect size that the study should be able to detect. Furthermore, the desired alpha error rate was set to 0.05, and I assumed that the assumption of sphericity will hold true in the analysis. Entering these values into G*power, I estimated that 60 participants will be needed in the study to achieve the desired 90% power.

In this research study it is possible that some participants will find the cold pressor task (the pain induction procedure) so uncomfortable at pretest, that they would decline participation in the second administration of the pain stimulus. Participants with missing data

will be omitted from the analysis, so I estimate that I will need to accrue 66 participants, accounting for 10% dropout rate from pretest to posttest.

References

Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods*, *41*(4), 1149–1160. https://doi.org/10.3758/BRM.41.4.1149

Forsyth, L., & Hayes, L. L. (2014). The Effects of Acceptance of Thoughts, Mindful Awareness of Breathing, and Spontaneous Coping on an Experimentally Induced Pain Task. *The Psychological Record*, *64*(3), 447–455. https://doi.org/10.1007/s40732-014-0010-6

Lipsey, Mark W., & Wilson, D. B. (2001). *Practical meta-analysis*. Sage Publications, Inc.

Lipsey, Mark W., & Wilson, David. B. (2001). *Effect Size Calculator*. https://campbellcollaboration.org/escalc/html/EffectSizeCalculator-Home.php

Liu, X., Wang, S., Chang, S., Chen, W., & Si, M. (2013). Effect of brief mindfulness intervention on tolerance and distress of pain induced by cold-pressor task. *Stress and Health: Journal of the International Society for the Investigation of Stress*, *29*(3), 199–204. https://doi.org/10.1002/smi.2446

Swain, N. R., & Trevena, J. (2014). A Comparison of Therapist-Present or Therapist-Free Delivery of Very Brief Mindfulness and Hypnosis for Acute Experimental Pain. *New Zealand Journal of Psychology*, *43*(3), 22.