Universitätsmedizin GREIFSWALD

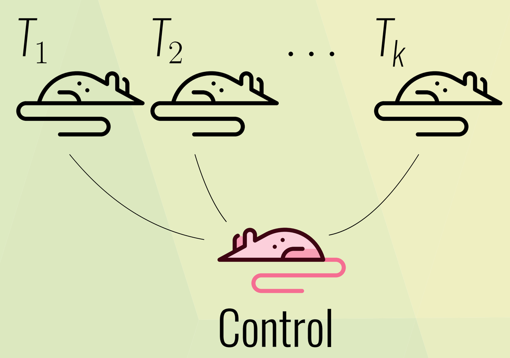
Sample Size Calculation with Unequal Group Sizes for Dunnett's Testing

Marcus Vollmer
Institute of Bioinformatics, University Medicine Greifswald, Germany

A multiple comparison test

Dunnett's Test - What's that?

- Multiple comparison test by Charles Dunnett (1955)[1]
- Post-hoc-Test after ANOVA
- Compare k treatment arms against a control group
  $H_{0i}: \mu_i = \mu_0$
- Similar to performing multiple t-tests
- Designed to hold the family-wise error rate
  FWER=P(number of falsely rejected $H_0 \geq 1) \leq \alpha$



Control

«I don't have any hypothesis about the effect size - it's a pilot study. Why do I need sample size justification?»

Purpose

The three Rs
Principles were developed over 50 years ago as a framework for humane animal research

Replacement	Reduction	Refinement
Methods which avoid or replace the use of animals	Methods which minimise the number of animals used per experiment	Methods which minimise the suffering and improve animal welfare

The importance of sample size estimation
- Which difference is of biological importance?
- Which statistical test should I conduct?
- And how many animals do I need to show a difference?
- Is this feasible (cages, animal keeper, laboratory space, costs)?

Expectations may not always been fulfilled – if so, then you can publish significant results!

Reduction by imbalanced testing

Balanced vs. imbalanced sample sizes

Interestingly, in animal experiments equal sample sizes have been frequently proposed. However, the same statistical power can be achieved by unequally distributed group sizes with a reduction of the total sample size:

n=48: n_0 n_1 n_2 n_3 = 12 12 12 12

n_0 n_1 n_2 n_3 = 18 10 10 10: n=48

$$t = \frac{\bar{X}_0 - \bar{X}_i}{\sigma\sqrt{\frac{1}{n_0}+\frac{1}{n_i}}} = \frac{\Delta\bar{X}}{\sigma}\frac{1}{\sqrt{\frac{2}{12}}} = 2.45\frac{\Delta\bar{X}}{\sigma}$$

$$t = \frac{\bar{X}_0 - \bar{X}_i}{\sigma\sqrt{\frac{1}{18}+\frac{1}{10}}} = 2.54\frac{\Delta\bar{X}}{\sigma}$$

$$\sqrt{\tfrac{1}{12}+\tfrac{1}{12}} = \sqrt{\tfrac{1}{6}} = \sqrt{\tfrac{1}{18}+\tfrac{1}{9}}$$

saved

A general rule for equal effect sizes and equal variances says: $\frac{n_0}{n} \approx \sqrt{k}\frac{\sigma_0}{\sigma}$

Available methods in R

Currently, two packages are available in R, `R:multcomp`[2] and `R:DTK`[3], to perform the special testing problem with unequal group sizes. The computation of the p-values includes the consideration of a multidimensional t-distribution and the adjustment for multiple testing. Unfortunately, a procedure for sample size estimation is missing.
`R:DunnettTests`[4] conducts a sample size calculation, but with identical treatment effect size and pre-specified sample allocation ratio only. In other situations, simulation-based evaluation is suggested, which needs great computational effort.

A simulation study

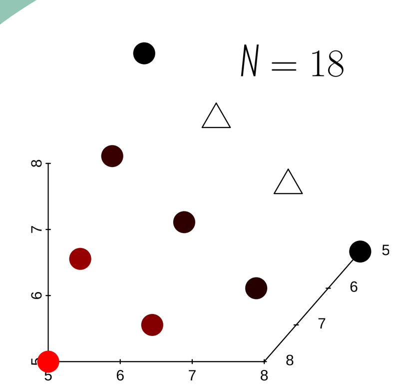
The aim is to determine the sample sizes for multiple treatment groups with different effect sizes (different means and unequal variances). A necessary statistical power of 80% is expected. Ideas for finding the minimal set of group sizes in Monte Carlo experiments:

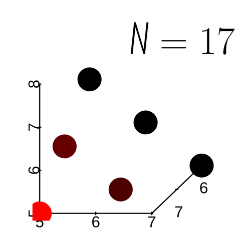
Random search	Modified grid search	Topological concept
Start with initial size. Sample new position based on derived statistical power	Evaluate given parameter sets – start with coarse grid – refine grid – increase accuracy (number of simulations)	Bottom-up or top-down procedure using the topology of different sets of sample sizes

What is the optimal set of sample sizes?

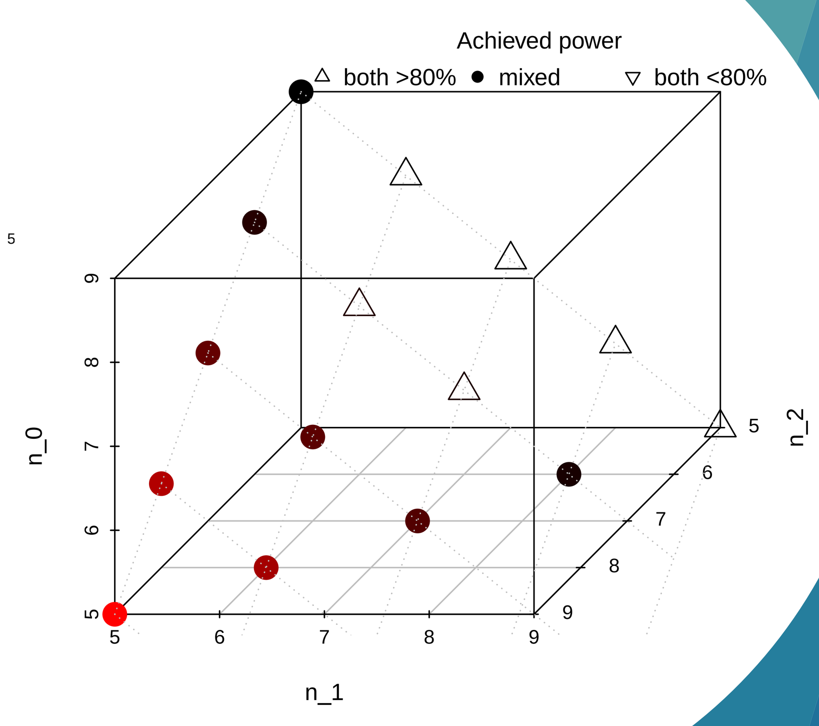
It needs roughly 5e5 samples to estimate the power precisely up to the first decimal place!

An intelligent search at different levels of integer partitions of N can massively reduce the computational size. It is the objective of constructing a random search with the use of topological relations and precision levels.



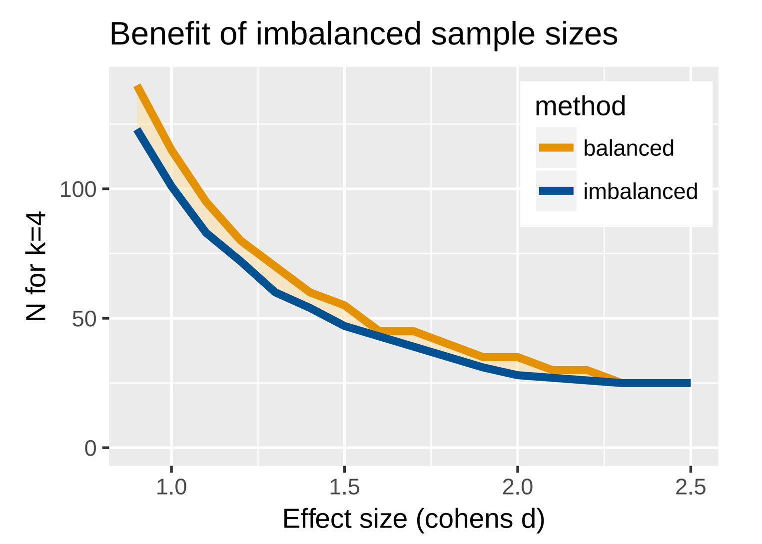


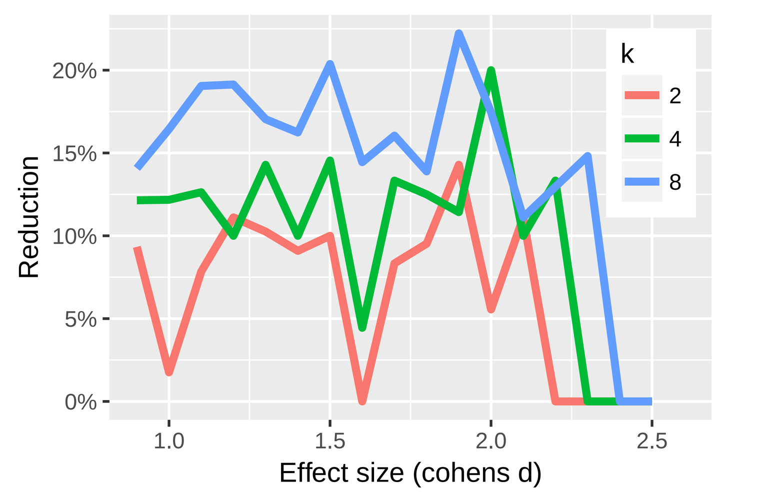
Sets of sample sizes with $N = 19$



Treatment effect sizes: $c_d^1 = 1.85, c_d^2 = 2.25$

Statistical Power — Simulation Study





Passive immunization with glycoforms of IgG

Mice experiment, immunization, Pneumococcal infection
IVIS Spectrum Imaging
- Negative control - Pre-immune IgG: $\mu_0 = 4.58$
- Negative control - Post-immune: $\mu_1 = 5.73$
- 3 Glycoforms: $\mu_{2,3,4} = 3.57$
- equal variance: $\sigma = 0.96$

Grid search results sorted by average power

n0	n1	n2	N	p1	p2	p3	p4	power
34	12	18	100	0.851	0.785	0.780	0.768	0.79600
34	12	19	103	0.868	0.772	0.762	0.783	0.79625
32	11	20	103	0.798	0.809	0.800	0.789	0.79900
33	10	20	103	0.776	0.808	0.807	0.812	0.80075
33	12	19	102	0.843	0.779	0.797	0.785	0.80100
.	.	.	.	.	.	.	.	.
33	11	20	104	0.814	0.804	0.803	0.812	0.80825
32	12	19	101	0.862	0.792	0.800	0.785	0.80975
33	12	20	105	0.864	0.803	0.795	0.806	0.81700
34	11	20	105	0.812	0.807	0.825	0.831	0.81875

[1] C. W. Dunnett, "Pairwise multiple comparisons in the unequal variance case," *Journal of the American Statistical Association*, vol. 75, no. 372, pp. 796–800, 1980.
[2] T. Hothorn, F. Bretz, P. Westfall, R. M. Heiberger, A. Schuetzenmeister, and S. Scheibe, "Package 'multcomp'." Website, 2016. https://cran.r-project.org/package=multcomp.
[3] M. K. Lau, "Package 'DTK'." Website, 2013. https://cran.r-project.org/package=DTK.
[4] F. Xia, "Package 'DunnettTests'." Website, 2013. https://cran.r-project.org/package=DunnettTests.

Stay in contact!
RG Marcus_Vollmer
+49 (0)3834 86-5444
marcus.vollmer@uni-greifswald.de
MarcusVollmer.github.io

07/2017