Power
one-way ANOVA

Experimental units

CRD

inhomogeneous

homogeneous

Block Designs

RCB

(B)IBD

RCB with factorial treatment structure,

random effects, variance components,

factorial treatment structure (fixed effects),
two-way ANOVA (or more factors),
concept of interaction, 2^k-designs, ...

random effects, mixed effects models, nested factor structure, ...

split-plot, split-plot designs, different models
on whole- and subplots, ...

Similar to Lawson (2015)

One treatment factor

Multiple treatment factors

Multiple treatment factors, varied / randomized on different "scales"
Different Error Rates and Power

- Remember the different error rates of a statistical test:
  - **Type I** error: **Reject** $H_0$ even though it is true.
  - **Type II** error: **Fail** to reject $H_0$ even though $H_A$ holds.

- The probability of a type I error is **controlled** by the **significance level** $\alpha$.

- Type II error: The probability of a type II error is typically denoted by $\beta$ (it is **not** being controlled).

- We need to **assume** a certain setting of the parameters under $H_A$ to be able to calculate $\beta$.

- **Power** of a statistical test is defined as
  \[ P[\text{reject } H_0 \text{ given that a certain setting in the alternative } H_A \text{ holds}] = 1 - \beta. \]

  “probability to get a significant result assuming a certain setting under the alternative $H_A$”
Power

- We are interested in power when **planning** an experiment.
- Why?
  - We believe that \( H_0 \) is **not** true and we want to get a “significant result” out of our experimental data.
  - Power tells us the probability to get a significant result if our \( H_A \) is actually true.
  - Calculating power is like a “thought experiment”, we do **not** need data, but a **precise specification** of the parameter setting under \( H_A \) that we believe in.
Power

- Very low power means: chances are high that you will not get a significant result, even though $H_A$ holds (that is $H_0$ is not true).

- In other words: you wasted time and money with your experiment as it was a priori clear that (with high probability) the result will not be significant.

- E.g., if power is 0.2 it means that only in 1 out of 5 cases (on average) the result will be significant.

- Ideally, power should be of the order of $\geq 80\%$.
Power

- Power depends on
  - design of the experiment (balanced, unbalanced, complete blocks, incomplete blocks)
  - significance level $\alpha$ (typically $0.01, 0.05$)
  - parameter setting under the alternative (incl. error variance $\sigma^2$)
  - sample size $n$

- General rule: the more observations you have, the larger power.

- Typically we choose $n$ such that power is at the appropriate level.

- If $n$ is smaller, the experiment does not satisfy our needs, if $n$ is larger, it is a waste of resources.
Power: Calculations

- For easy situations (like the two-sample \( t \)-test, one-way ANOVA, …) formulas can be derived using new terminology like a so called non-central \( F \)-distribution.

- There are some special functions, like `power.anova.test` (comes with R) or the package `pwr`.

- For more complex designs, **simulations** have to be used.

- This used to be a problem in the old days but is easy nowadays.
Example: One-Way ANOVA (Roth, 2013)

- Say we want to compare $g = 5$ different treatments labelled $A, B, C, D, E$.
- Previous experience tells us that $\sigma^2 = 7.5$ is a realistic value of the error variance.
- We assume that at least two groups differ by 6 units.
- If we use the model
  
  $$Y_{ij} = \mu + \alpha_i + \epsilon_{ij}$$

  with the sum-to-zero constraint it means that we (e.g.) have
  
  $$\alpha_1 = -3, \quad \alpha_2 = 3, \quad \alpha_3 = 0, \quad \alpha_4 = 0, \quad \alpha_5 = 0.$$ 

- See R-File for simulation.
Visualization of 1 Simulated Data Set
Power Analysis: Comments

- Unfortunately, sample size is often determined by your resources (= money).
- Power analysis then gives you an answer whether it is actually worth doing the study or whether it is just a waste of time.
- The difficult part is defining the parameters under the alternative, especially the error variance.
- The nice side-effect of doing a power analysis is that you actually fit your model to (simulated) data and immediately see whether the analysis “works”.