Improving your data with power test and bootstrap: a summary

ACME
17.11.17

Collecting your data: what can go wrong?

Many things!

- you can lose it
    - start over again
- design can turn out bad
    - change the design, start over again
    - make pros out of your cons
- you can have not enough of it **for your results to be significant**
    - collect some more **how MUCH MORE?**

Power analysis

Estimating a sample size required to detect a preferable effect size.

What you need:

- sample size of your own
- effect size (the bigger the better! but actually its own for different tests)
- sig. level (p-value): 0.05 is usually enough?
- power (1–sig. level)

Power analysis: how?

Pretty simple!

R package **pwr**

- contains different tests (t-test, linear regression, ANOVA...)
- pretty straightforward

Example: **pwr.t.test** (d = 0.8, sig.level = 0.05, power = 0.95)

Result: n = 41

Example: **pwr.t.test** (d = 0.8, sig.level = 0.0005, power = 0.95)

Result: n = 85

Collecting your data: what can go wrong?

Many things!

- you can lose it
  - start over again
- design can turn out bad
  - change the design, start over again
  - make pros out of your cons
- you can have not enough of it **for your results to be significant**
  - collect some more **how MUCH MORE?**
    - I don't really have time
    - improve it with a simple intervention

What's that?

Bootstrapping is resampling your data based on current distribution. Pretty simple:

- you have some data
- you estimate your distribution via a non-parametric test
  - Kernel density estimation
- you generate some more data based on your estimated distribution
- you can now run all of your statistical tests and live happily ever after

Is it OK to use it?

In computational modelling — yes, pretty much.

In experimental research — well…

Examples:

- Proceedings of the Royal Society B (biology): 5 year impact factor = 5,4
- Current Biology (Cell Press): 5 year impact factor = 9,7

Pretty common in biological research (we're pretty close!)

Got it. How do I do it?

Pretty simple.

In R: **boot** package

boot(data, statistic, R)

*meanfun (based on your mean)

For linear regression use the **Boot** function

+   ask Joe for the MATLAB code