Demo 4

R for statistical analysis

Juulia T. Suvilehto D.Sc.(tech)

Good news!

- We have done most of the hard work
- We do not need learn new concepts related to programming in order to run statistical tests
- This demo will focus on applying what you have already learned to run different types of analyses

Statistical tests

Statistical test syntax

testname(x, y, <other parameters>)

testname(variable1 ~ variable2, <other parameters>)

Tilde (~) is interpreted ‘as a function of’

P-values, a (short) rant

- In general terms, p-values tell us the following: **Assuming null hypothesis is true** (i.e. no difference between the groups), **how likely it is to observe an effect which is as large or larger than the one we observed**
- It does **not** tell us how likely it is that the null hypothesis is true (or how likely it is that alternative hypothesis is false)
- It does **not** tell you how large or important the difference between the groups is
- There is nothing magical about 0.05 – it's a convention

Testing assumptions

Parametric and non-parametric tests

- Many tests have underlying assumptions about the distribution of the data
- Before you run your statistical tests, you need to know these assumptions and make sure your data meets them
- **R will not babysit you**
- Thankfully, testing for assumptions is simple as long as you know it’s a thing & remember to do it
- R has a very good selection of non-parametric tests (tests which do not make these assumptions) which you can use if your data is not suitable for the parametric version

Regression

Running regression analyses in R is easy

- For your typical linear regression, use the function lm() (linear model)
  - Syntax: lm(y ~ x, data = <yourdataframe>)
  - You can include any number of independent variables in the formula, separated by + (for no interactions) or by * (when including interactions)
  - Get information about the model with summary()
  - Compare models with anova()
- For logistic regression, use glm()
  - Syntax: glm(formula = y ~ x, data = <yourdataframe>, family = binomial)
- For multilevel modeling, use lmer() (linear mixed effects in r) from the package lme4

Recap

Data analysis with R (we know how to do all of this!)

- Load the data
- Inspect the data
  - Are the missing values coded appropriately?
  - Are there any outliers that are physiologically impossible (e.g. height >3m, age < 0 years)
  - Are categorical variables coded as factors and continuous variables coded as numeric etc.?
- Are the data organized in a tidy manner
- Modify the data as necessary
- Run analyses/build plots
- Save the outcome

Parting words of wisdom

- You now have a rough idea of how to do all the necessary steps of data analysis in R
- If you're an early career researcher, learning a programming language (like R) is a great asset regardless of your career plans
  - Don't be afraid to try and fail
  - There are so many great tutorials available online
- If you have a system in place already, it can be a really big push to convert all of that to R
  - One approach: try to do one more thing in R in each of your projects
- You know so much more than you did in the beginning of this course, you should be proud of yourself!