

Statistics part: what we will cover each day

Ramon Diaz-Uriarte

(None) (Release: Rev: (None))

How to read this

“Day 2” refers to the second session or lesson of the classes devoted to stats. For example, that will most likely be 13-November-2024. (This notation allows us to reuse this file, even if we change the exact date of a lesson).

“Section 5”: refers to the section in the main PDF, “R-basic-stats.pdf”, titled “Some basic statistics with R”. To be redundant, I often give in parentheses a short version of the section (e.g., “section 5 (plots)”, where “plots” stands for “Looking at the data: plots” that is the full, long, section title of section number 5).

Speed

The plan below is not the ideal plan. The ideal plan includes the “omics” slides and the categorical data analysis PDF, as well as a programming example of writing the code for a permutation test, which would be included at the end. And we want to go over what is an R package too. Thus, I will try to go slightly faster than what is shown below. This will be possible if you read the notes before coming to class.

Day 1

- Sections: 5 (plots), 6 (two-sample t; including supplementary PDF about confidence intervals referred to in section 6.2), 7 (one and two-tailed), 8 (power)
- For those interested, we will discuss **at the end of class (i.e., expect to stay after 19:00 on 11-November-2024)**, any doubts about the projects for the practical programming exercise.

Day 2

- Whatever remains from Day 1. Sections: 9 (equiv. testing), 10 (bayesian), 11 (conf. int: see longer slides).

Day 3

- Paired stuff: 12.1, 12.2.1, 12.2.2, 12.2.3, 12.2.8, 12.3, 12.4, 12.5 (plots for paired), 12.6, 13 (one-sample), 14 (non-par), 15 (non-indep data), 16 (symmetry and paired t). 12.7 (a first taste

of lin.mods. —only if time). Start section 18. **Read the external files on your own BEFORE class, I will answer questions about them, but I won't go over them in class..**

Day 4

- Whatever remains from Day 3: Section 18, including the two external files (anova basic theory and anova theory even simpler); section 19 (FWER and FDR), some 20 (two-way anova). **Look at the notes for section 20 BEFORE class.**

Unless you look at the notes **before you come to class** this will not make any sense. The two-way anova section is long (more than 50 pages), possibly completely new, and possibly not intuitive nor easy.

Day 5

- sections: 20 (two-way anova), 21 (regression).

Do not expect to understand this starting from 0 in just two hours of class. Again, come to class having looked at the material, **even if during the on your own, before-class, reading it seems confusing.**

Day 6

- Sections 22 (multiple regression), 23 (ancova), 24 (interactions, summary)

Day 7

- Sections 25 (diagnostics), 26 (variable/model selection), 28 (experimental design)

Day 8

- Section 29 (causal inference: see additional PDF — **really read the additional PDF on your own before coming to class**; otherwise, this will make no sense in just the time we have in class).
- If time left, R packages and miscell programming stuff (e.g., permutation test); 30 (ratios).

Day 9: the hour after the statistics exam

- R packages and miscell programming stuff; omics slides; categorical data analysis; 30 (ratios). (Obviously, we cannot do all of this in one hour). **The exam will last one hour, and we will use the remining hour to go over these topics.**