Quantitative Tutorials - Session 4

Understand scientific articles' statistics and ANOVA II

Dr. Erola Fenollosa

@Erola_Fenollosa

Today's outline

FIRST PART

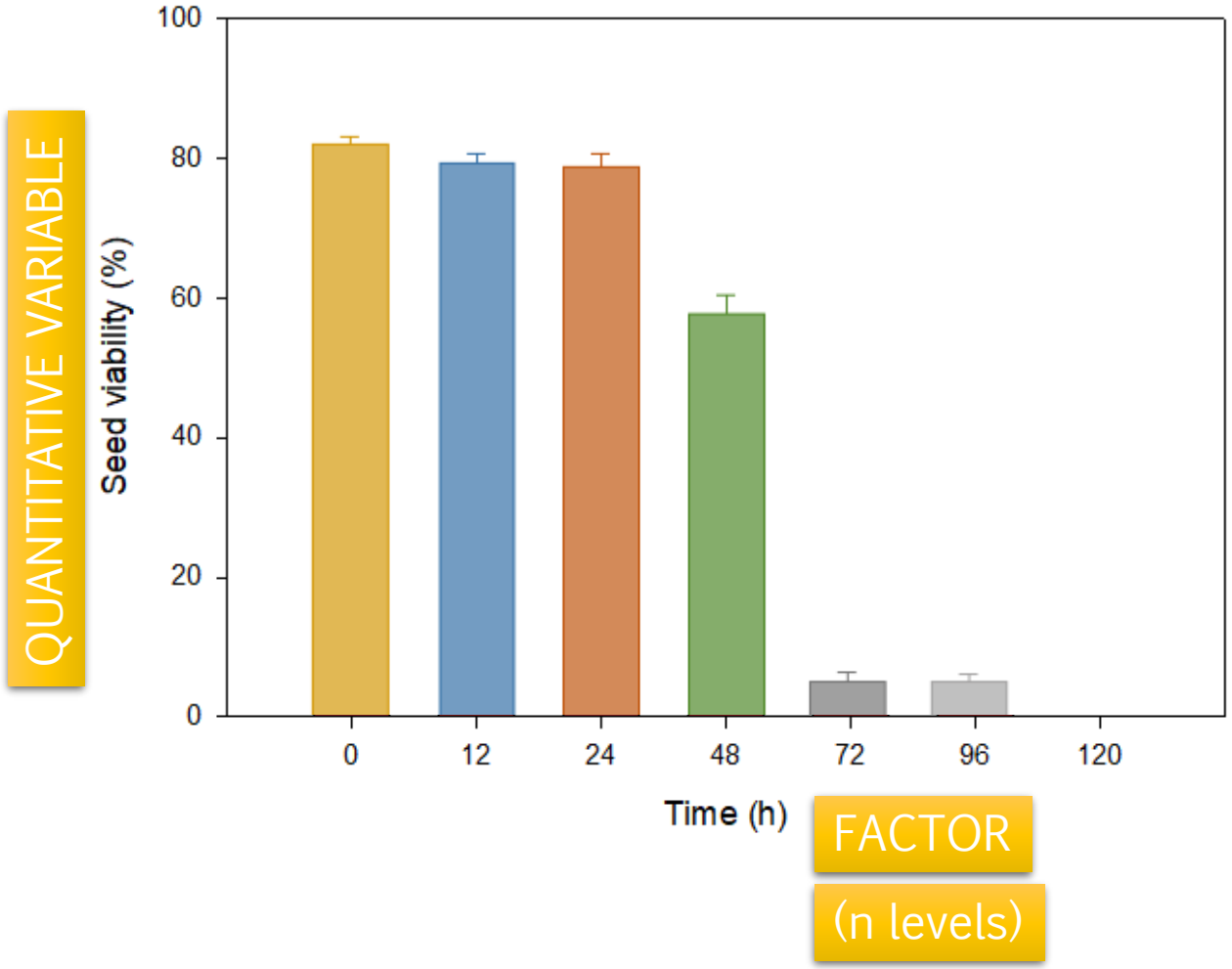
1. Check ANOVA theory from last session
2. Interactive understanding of article's statistics

SECOND PART

3. Check your code of one-way ANOVA
4. Perform a two-way fixed ANOVA following the given Rmd

1. Revise ANOVA theory from last session

Are they different?



Are there significant differences in viability between times?

Hypothesis contrast

H_0 → All are the same

H_1 → At least one is different

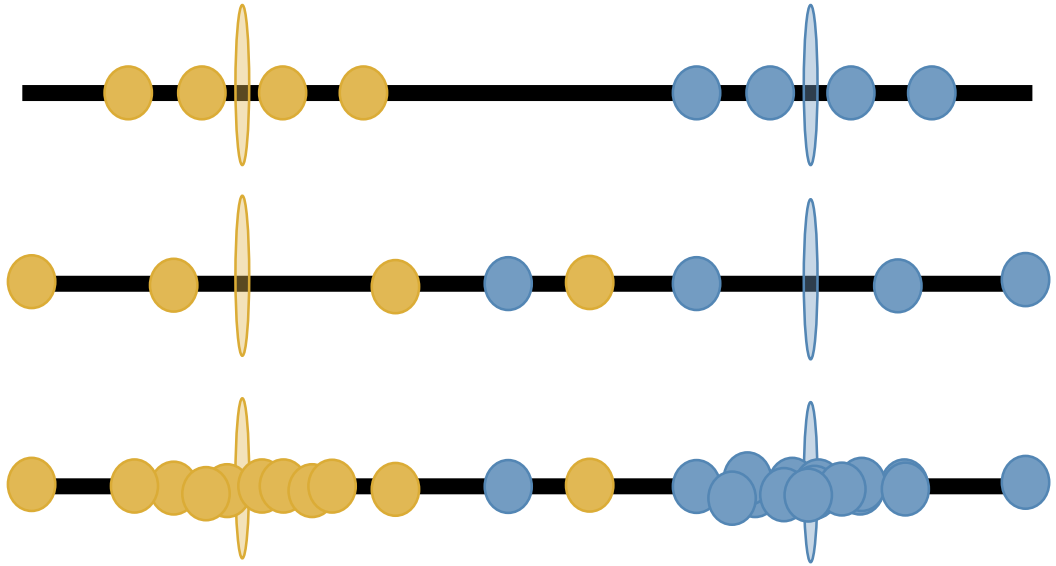
Presumption of inocence

QUICK TASK: Draw in a paper an aproximated plot of your factor and variable

ANOVA = Analysis of the variance

But it's more than that!

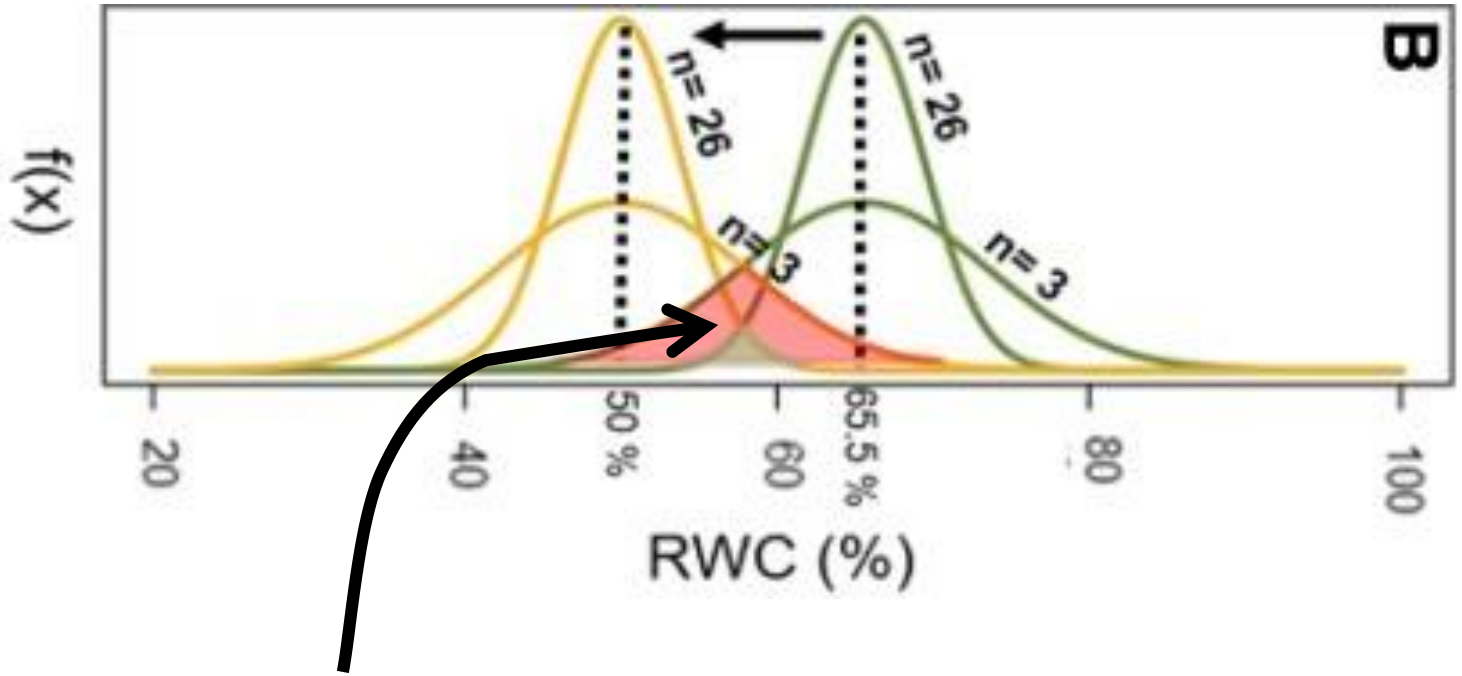
- Difference between means
- Variance
- Sample size



A one-way ANOVA calculates the variance associated with the levels of the factor and compares the mean squares among groups versus within groups

Visual help:

Pérez-Llorca et al., 2018 Plant Physiol



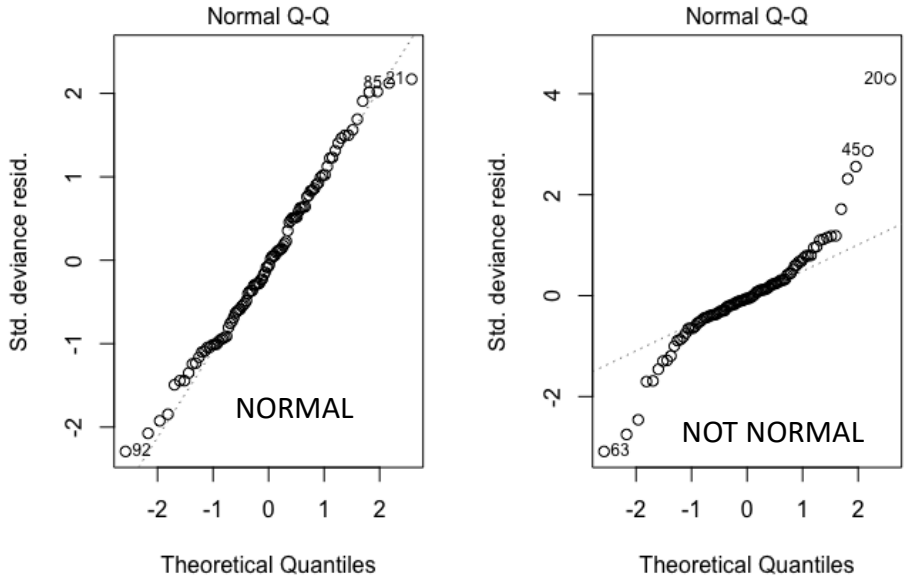
Alpha or confidence level 0.05

Why do we call it one-way?

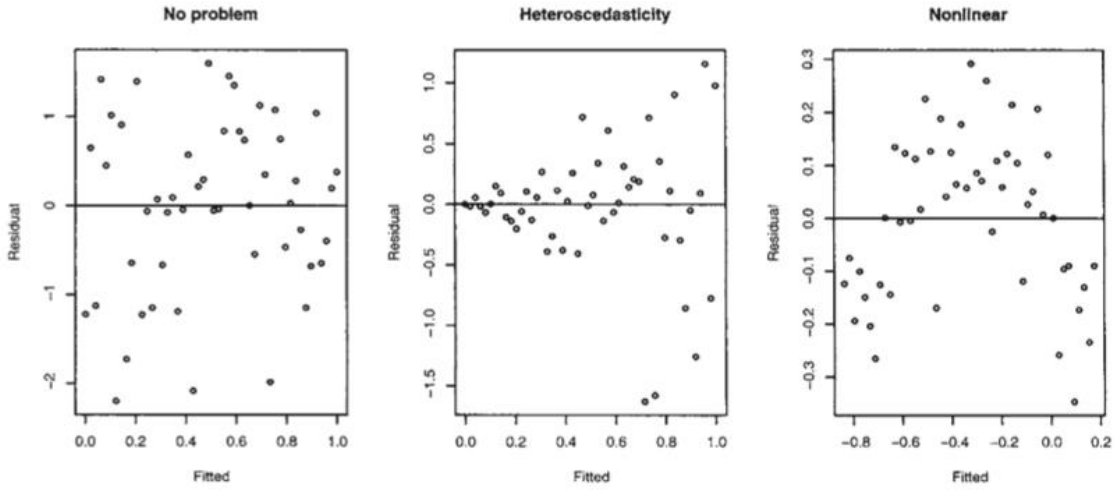
ANOVA requisites ⚠

- Normal distribution of the residuals
  - Kolmogórov-Smirnov
  - Shapiro Wilk
  - Q-Q plot

https://www.seascapemodels.org/rstats/2017/10/06/qqplot-non-normal-glm.html



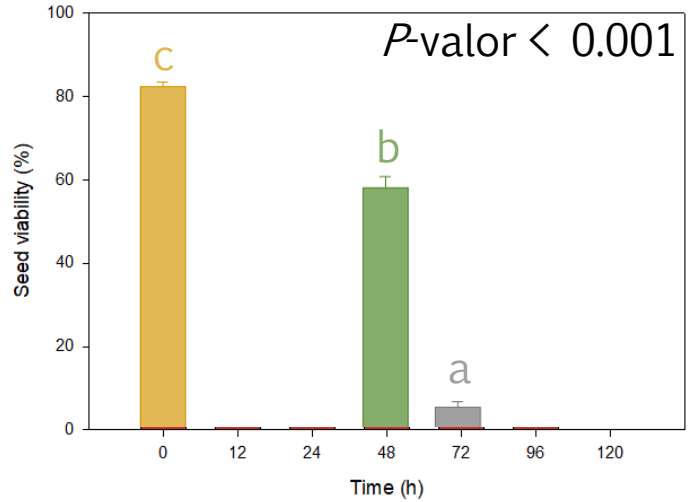
- Homoscedasticity (equal variance of the residuals)
  - Levene
  - Residuals vs. Fitted plot



https://stats.stackexchange.com/questions/76226/interpreting-the-residuals-vs-fitted-values-plot-for-verifying-the-assumptions

The ANOVA workflow

1) Prepare data and examine it first
2) Test ANOVA requisites
    1) Normal distribution (*P*-valor > 0.05)
    2) Homocedasticity (*P*-valor > 0.05)
3) Test ANOVA (*P*-valor < 0.05)
4) Test post-hoc (opcional)



TIME 0	TIME 48	TIME 72
81.8	58.1	10.8
80.5	60.9	5.1
79.3	56.3	11.2
82.0	63.1	8.9
83.4	59.8	9.5

Post-hoc??

Also called: multiple comparisons.
Only when we have more than two levels in a factor, to know actually which ones are different

Too much options...

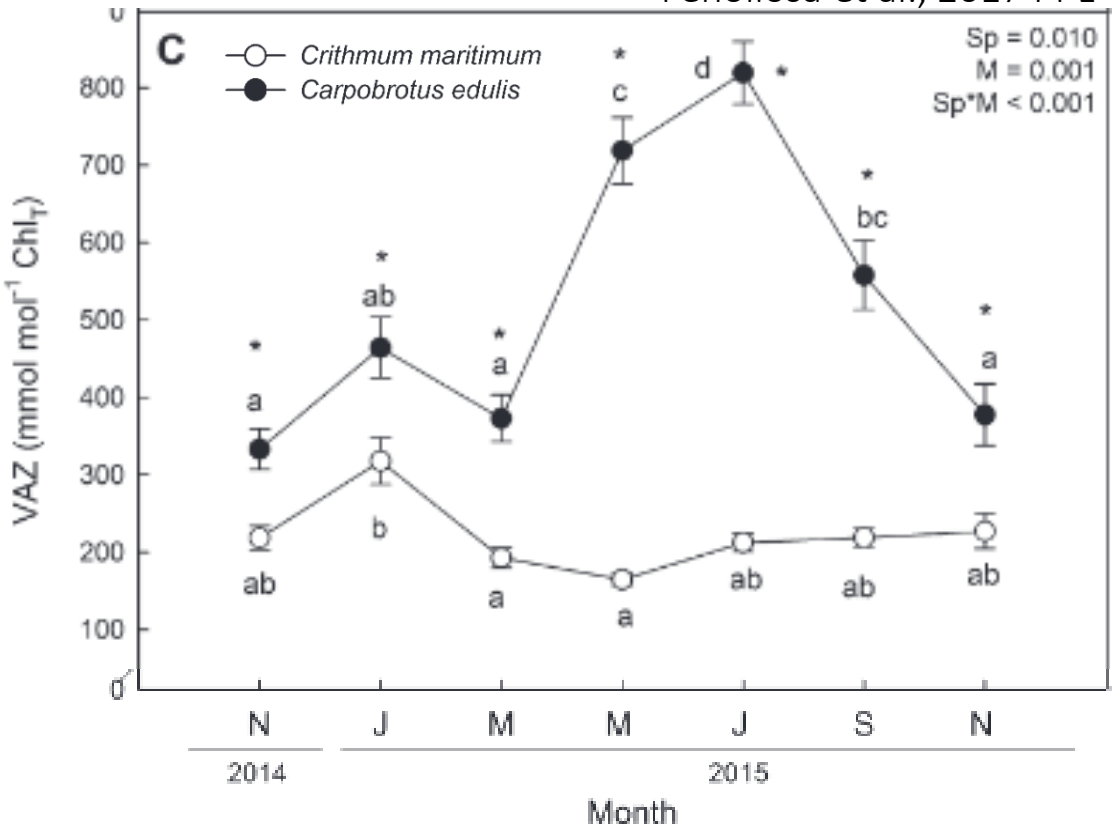
- Bonferroni Procedure
- Duncan's new multiple range test (MRT)
- Dunn's Multiple Comparison Test
- Fisher's Least Significant Difference (LSD)
- Holm-Bonferroni Procedure
- Newman-Keuls
- Rodger's Method
- Scheffé's Method
- Tukey's Test (see also: Studentized Range Distribution)
- Dunnett's correction
- Benjamini-Hochberg (BH) procedure

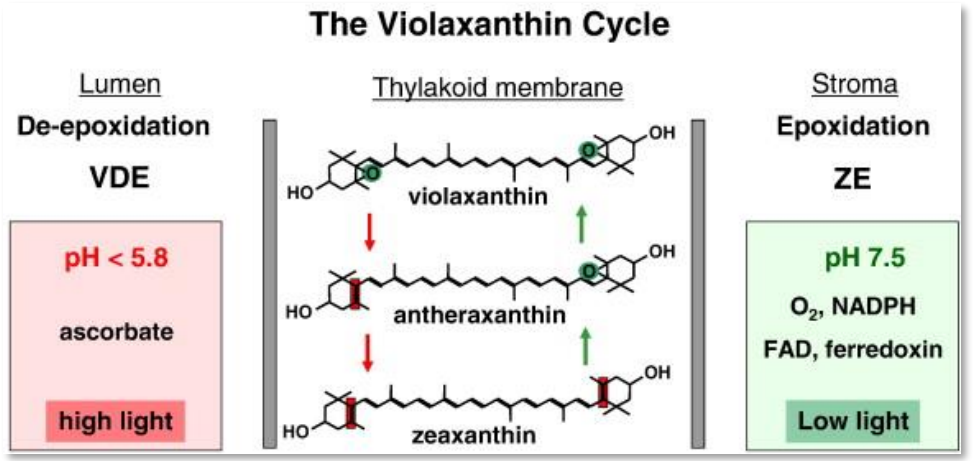
https://www.laverdad.es/sociedad/trabajo-oscuro-20190315105523-ntvo.html

Some are more restrictive than others
So expose always which one you've used.

Explain these results:

Fenollosa et al., 2017 PPL



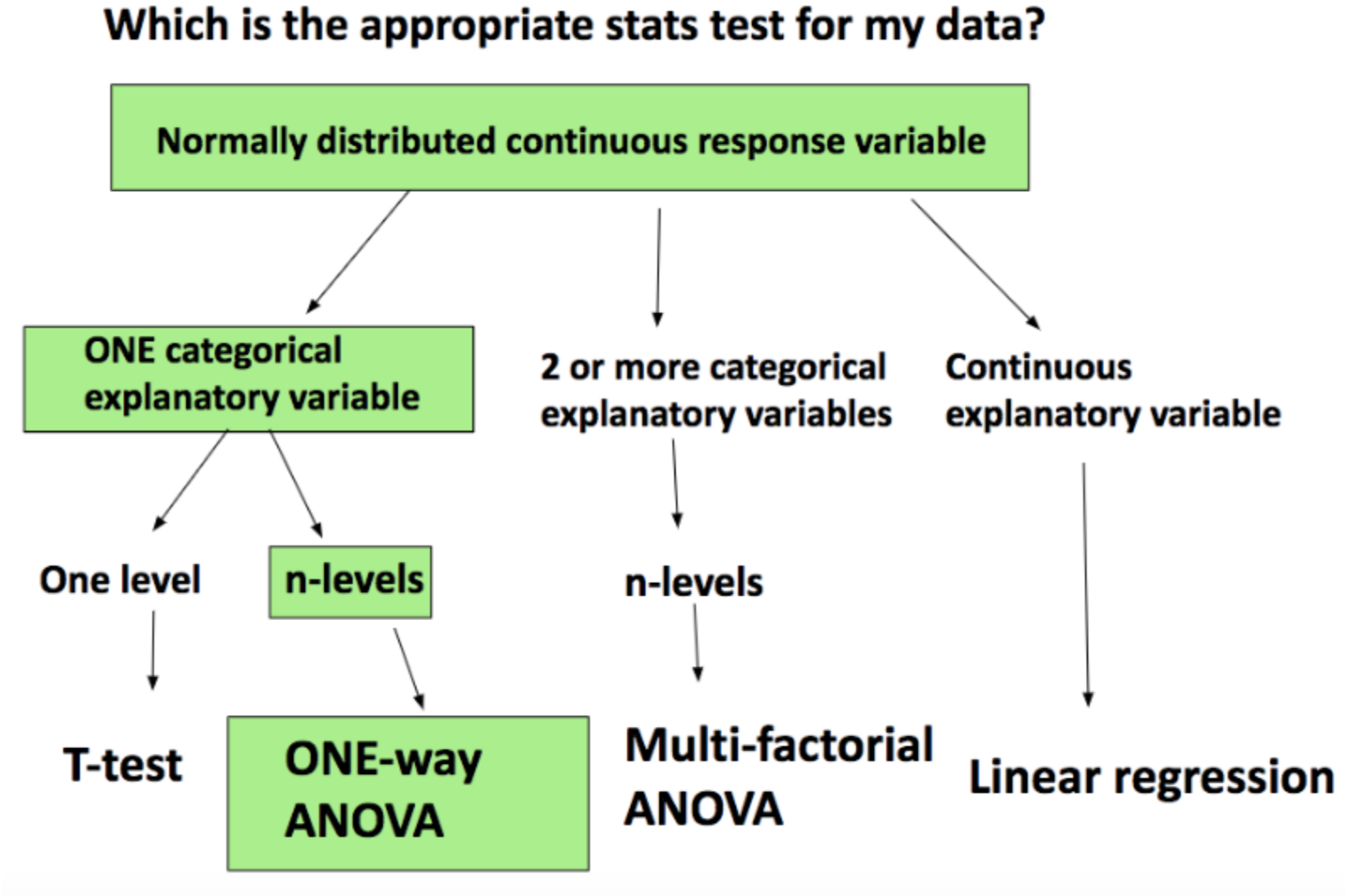


Jahns et al., 2009 BBA Bioenergetics

Crithmum maritimum

Carpobrotus edulis

What may the **asterisks** mean?
What may the **letters** mean?
What may the **top right numbers** mean?



https://ourcodingclub.github.io/tutorials/anova/

What if...

- Repeated measures
- More factors: two-way ANOVA, three-way ANOVA
- Random factors
- Nested factors
- Interaction
- Covariates (ANCOVA)
- Multidimensional (PERMANOVA)

Linear models and ANOVA, are they the same?

We are trying to predict a variable value form the group it belongs. This is very similar to what we do in regression.

Linear regression and linear model are complete synonyms

We usually talk about an ANOVA when we're quantifying the effect of a **discrete, or categorical** explanatory variable on a **continuous** response variable. It is also a linear model, but instead of getting a slope we get an estimate of the variable value for each category.

https://ourcodingclub.github.io/tutorials/modelling/

The code

To build the ANOVA table:

- res_aov <- aov (quantitativeVariableName ~ FactorName, data = datasetName)
- summary(res_aov)

To check ANOVA assumptions:

- plot(res_aov)

To do Tukey HSD post-hoc tests:

- library(multcomp)
- post_test <- glht(res_aov, linfct = mcp(FactorName = "Tukey"))
- summary(post_test)

A bit of help coding:

https://ourcodingclub.github.io/tutorials/anova/#model
https://statsandr.com/blog/anova-in-r/#anova

Take home messages

1. What is an ANOVA?
2. Why do we use an ANOVA?
3. What type of variables are we using in an ANOVA? What is the difference between a factor and a level?
4. What is the diference between ANOVA and t-test?
5. Can you describe the H0 and H1 of a one-way ANOVA?
6. What does the “P-value” represents?
7. What requisites do we have to meet to use ANOVA?
8. Is an ANOVA a linear model?
9. What is a posthoc test? Can you name one?
10. What is a one-way ANOVA?

2. Interactive understanding of article’s statistics

Selected articles

Article	Focus Figure	General Topic	Student
Chuang et al 2022	Figure 1	Fish	
Guo et al. 2022	Figure 2	Microbiology	
Nazareth Souto et al 2022	Figure 2	Birds /Pollution	
Dantas Ferreira et al 2022	Figure 1	Rats asthma	
Munné-Bosch et al 2022	Figure 2	Plant Physiology	
Hu et al 2022	Figure 1	Riparian Ecology	
Kharivha et al 2022	Figure 1	Invasive plants	

Chuang et al 2022, Animals

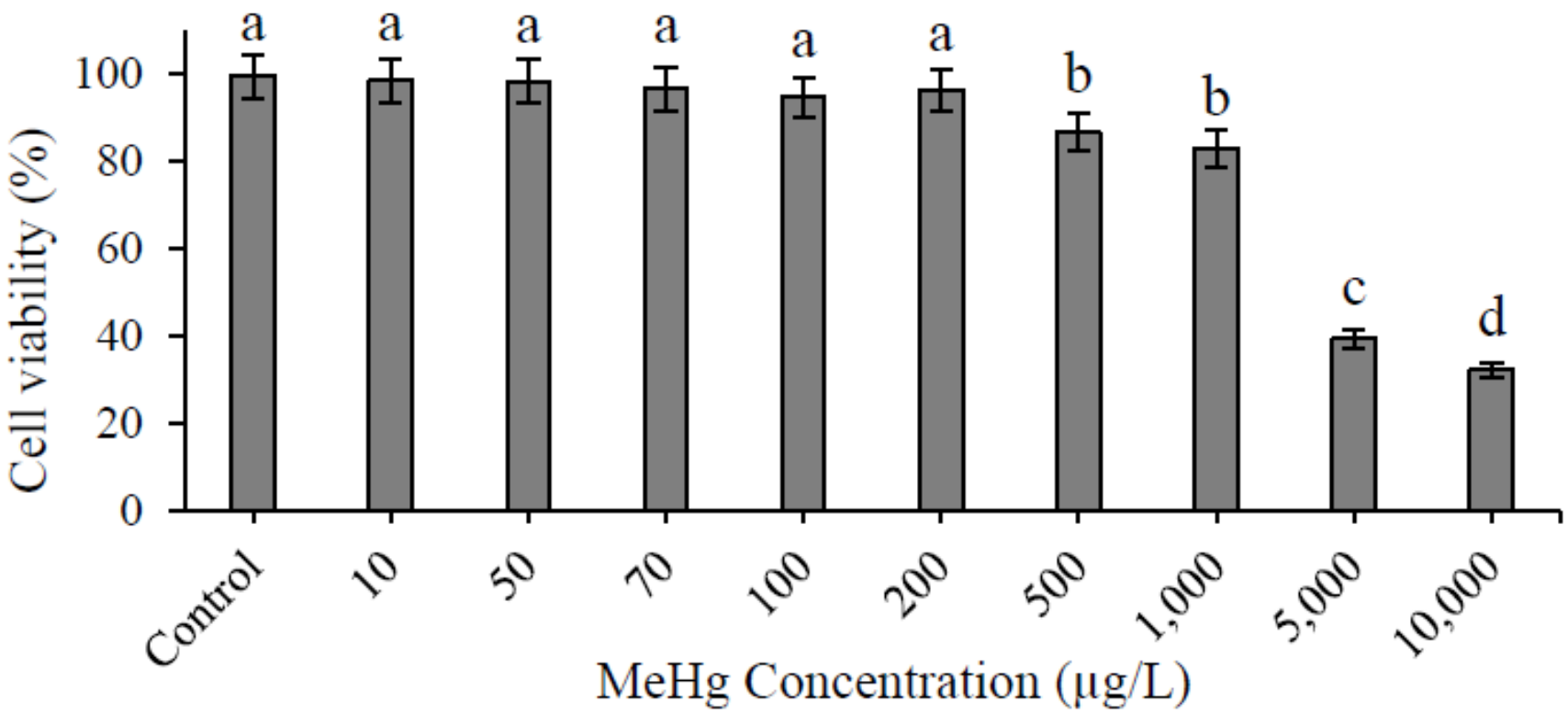
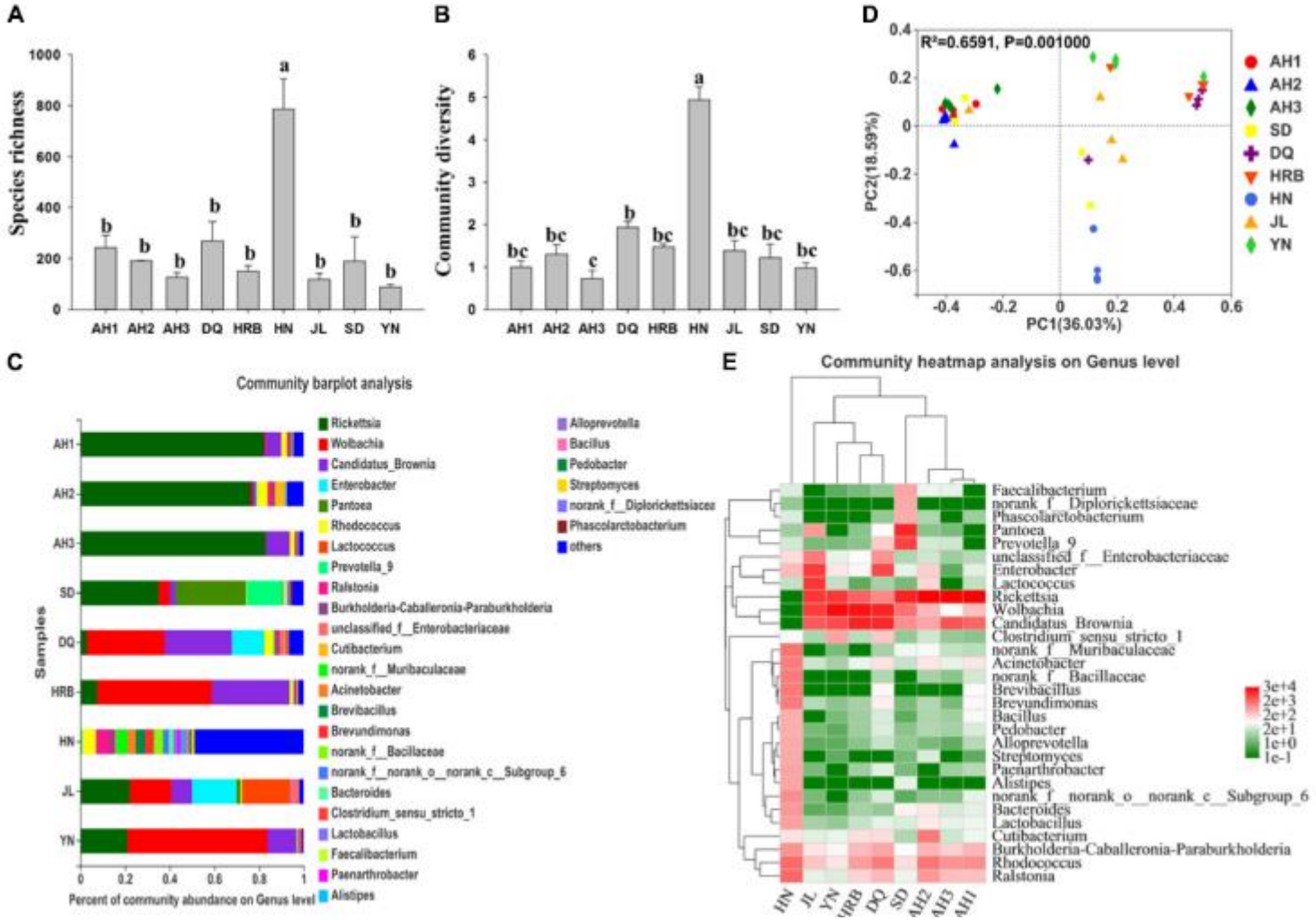


Figure 1. Cell viability of hybrid grouper leukocytes after in vitro incubation with various concentrations of methylmercury for 30 min. One-way ANOVA and Tukey' test were performed to compare the differences between groups. Values are presented as mean ± SD ($n = 3$). Significant differences ($p < 0.05$) between treatment groups are indicated by different letters above the bars.

Guo et al 2022 Frontiers in Microbiology



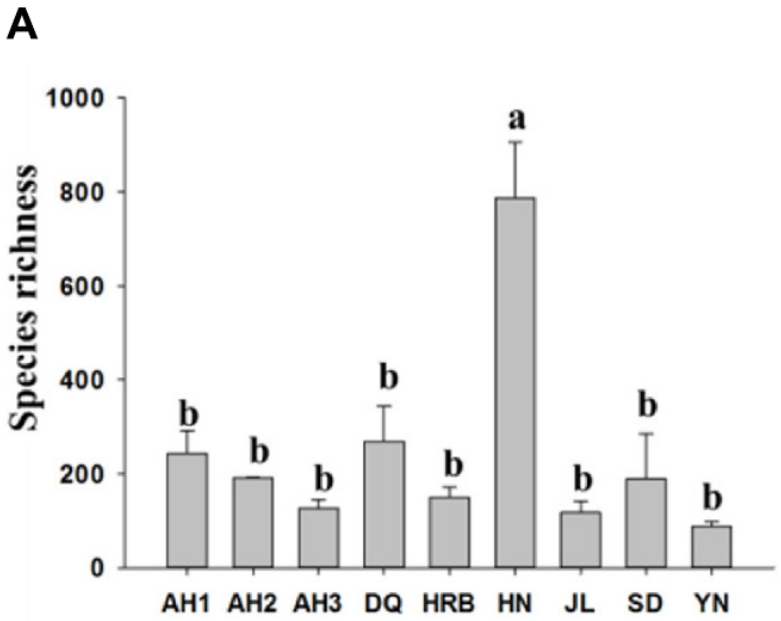


FIGURE 2
Bacterial community composition within and among nine geographic populations of the hemp flea beetle *P. attenuata* in China. **(A)** Histogram of species richness (number of OTUs). **(B)** Histogram of community diversity measured by Shannon index. Letters above each population indicate significant differences (one-way ANOVA, Tukey test, $P < 0.05$, see Table 3) in the mean values. **(C)** Relative abundance of bacterial genera in different populations. **(D)** Principal components analysis (PCoA) plots of bacterial communities based on the Bray-Curtis distance. Each symbol or color represents a sample. The significant differences in beta diversities were analyzed using Adonis analysis with 999 permutations, $P < 0.05$. **(E)** Heatmap showing the relative abundance of top thirty bacterial genera. Hierarchical cluster analysis was based on the Bray-Curtis distance with average method.

Dantas Ferreira et al 2022 Scientific Reports

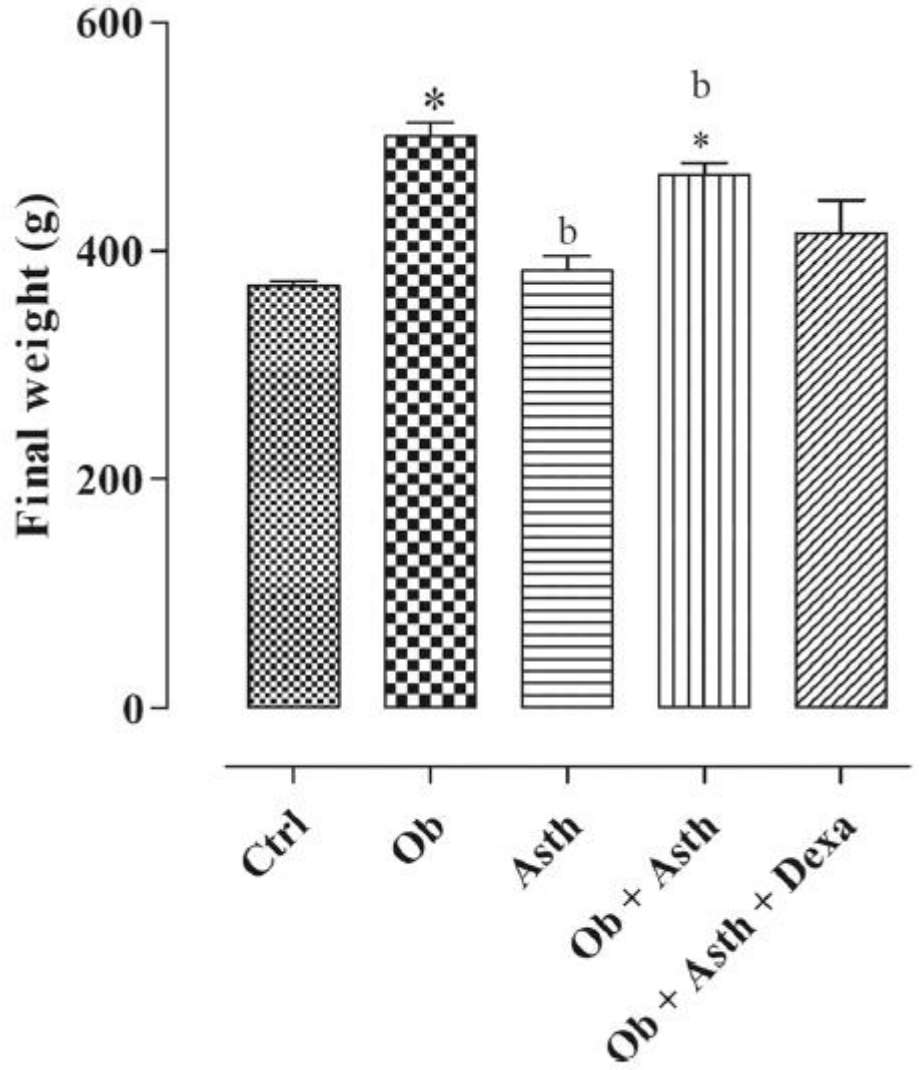


Figure 1. Final body mass values of rats from the Ctrl, Ob, Asth, Ob + Asth and Ob + Asth + Dexa. One-way ANOVA followed by Tukey's posttest (n = 4–6). $^{*}p < 0.05$ (Ctrl vs. Ob, Ob + Asth), $^{b}p < 0.05$ (Asth vs. Ob + Asth).

Kharivha et al 2022 Plants

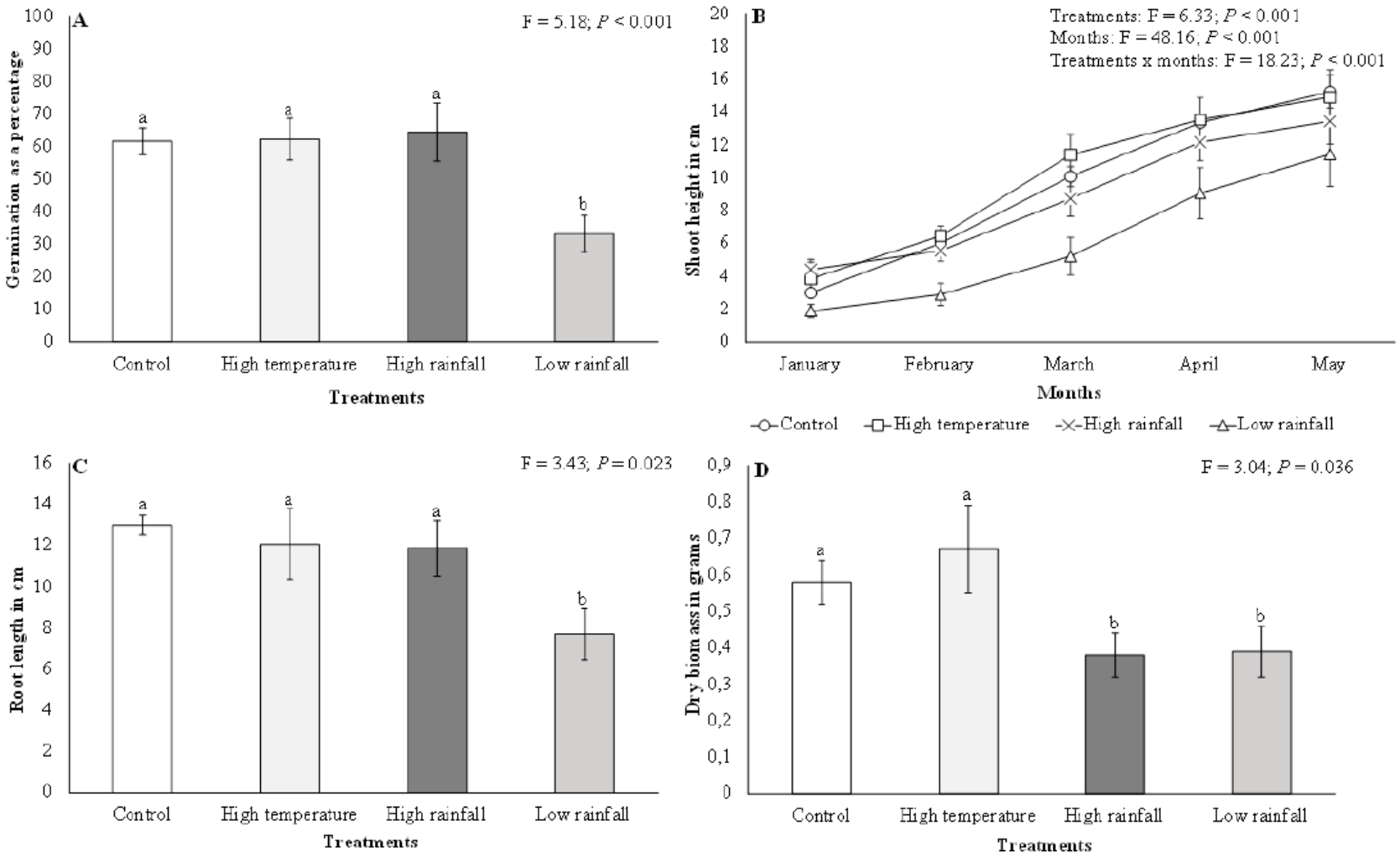


Figure 1. Effects of elevated temperature and high and low rainfall on *Acacia mearnsii* (**A**) germination as a percentage, (**B**) shoot height in cm, (**C**) root length in cm, and (**D**) dry biomass in grams. Bars are means ± SE and ANOVA results are shown. Bars with different letter superscripts are significantly different.

2. Check your code of one-way ANOVA and perform a two-way fixed ANOVA following the given Rmd

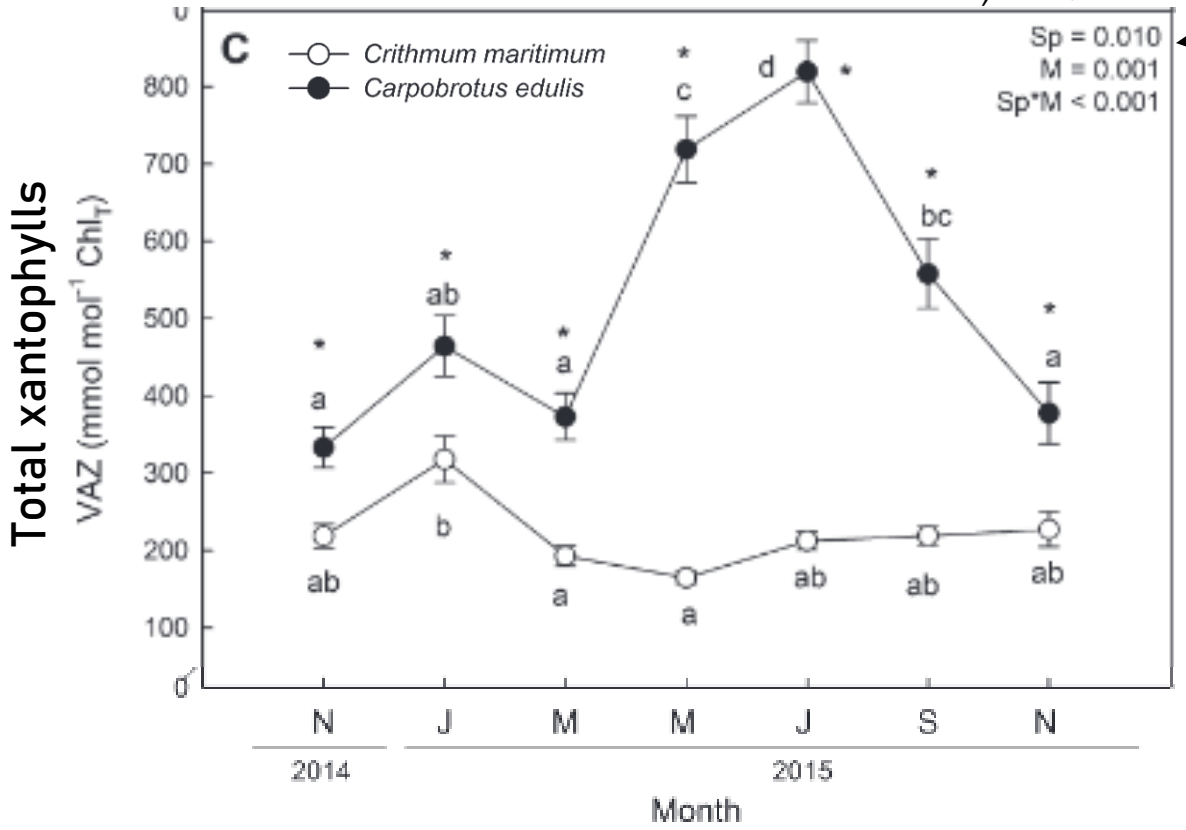
Notes from your previous work:

- Check always sample size before starting, is it balanced? Use *table()*
- Expect something from the visuals
- Discretizing could be useful: *cut()*
- Data transformation to meet ANOVA assumptions: log(), sqrt(), etc.
- The importance of plotting data variability: use jitter or mean +- SE
- Work with RMD and Knit a file, send me an html.

Today's task: Two-way ANOVA

The difference from one-way ANOVA is that we have now two factors. Depending on how those factors interact we can have crossed ANOVA or nested ANOVA. In crossed ANOVA we have a new element: the interaction.

Fenollosa et al., 2017 PPL



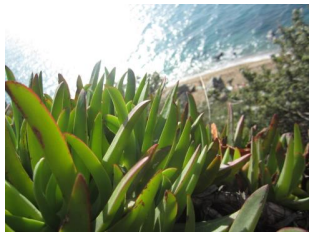
Our ANOVA will now have three p-values

Factor 1

Factor 2

Interaction

Carpobrotus edulis



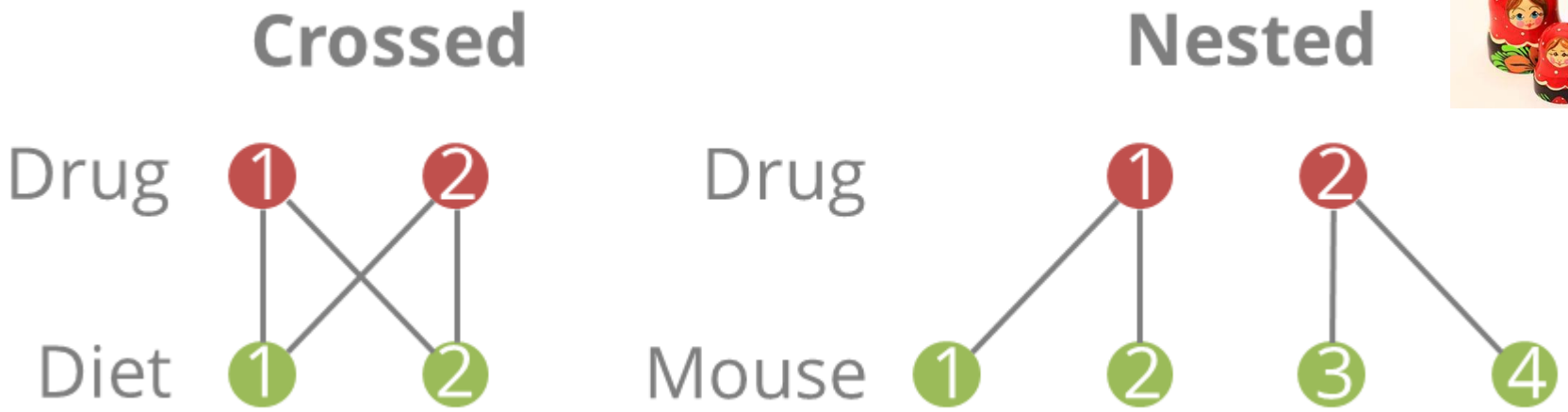
Crithmum maritimum

Nested vs crossed two-way ANOVA

Two factors are crossed when every category (level) of one factor co-occurs in the design with every category of the other factor. In other words, there is at least one observation in every combination of categories for the two factors.

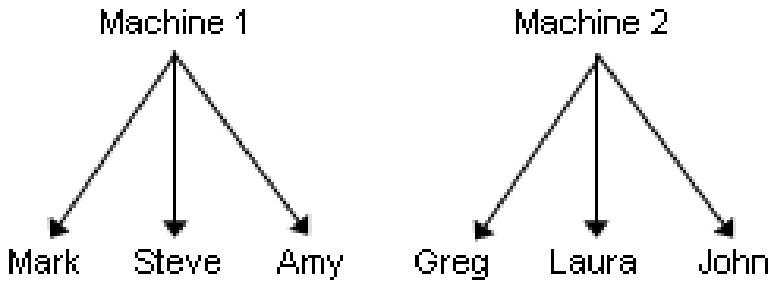
A factor is nested within another factor when each category of the first factor co-occurs with only one category of the other. In other words, an observation has to be within one category of Factor 2 in order to have a specific category of Factor 1. All combinations of categories are not represented.

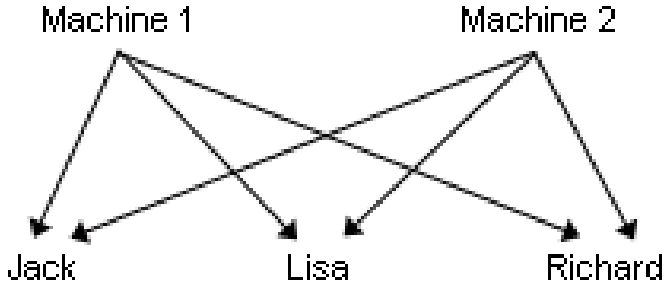
Nested vs crossed two-way ANOVA



https://2d-gel-analysis.com/starters-guides/experimental-design-guide/

Which is a crossed and which is a nested design?





https://support.minitab.com/en-us/minitab/20/help-and-how-to/statistical-modeling/anova/supporting-topics/anova-models/what-are-crossed-and-nested-factors/

Nested vs crossed two-way ANOVA

Nested or crossed?

Two brands are producing three products each. We want to know if the brands have different prices around the world in their respective products (that are different) and also if there is any product that has generally a higher price.

https://statisticsbyjim.com/anova/

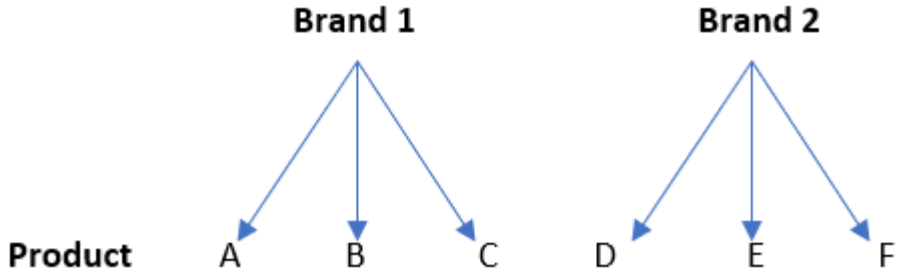
Nested or crossed?

We want to test if the average qualifications a college gets depends on the grad of the students. We've got three schools and three different classes (1rst year, 2nd year, 3rd year)

Nested vs crossed two-way ANOVA

Nested or crossed?

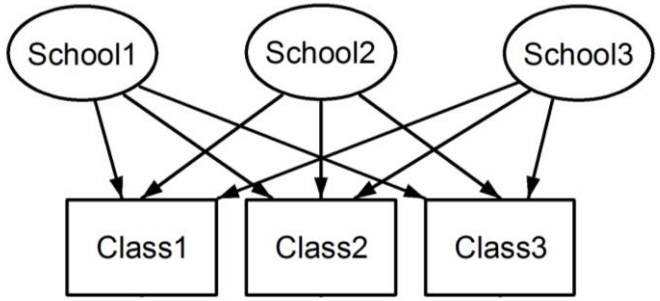
Two brands are producing three products each. We want to know if the brands have different prices around the world in their respective products (that are different) and also if there is any product that has generally a higher price.



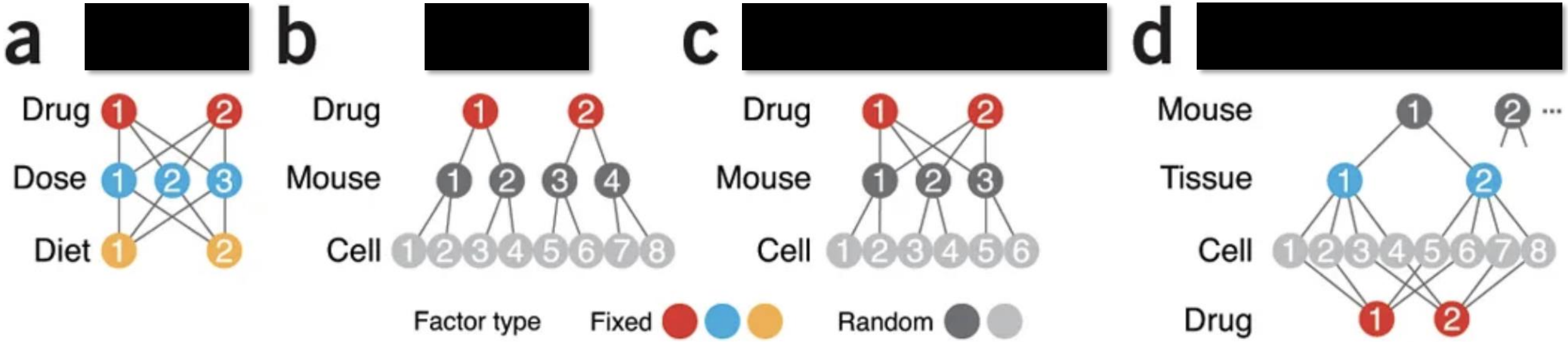
https://statisticsbyjim.com/anova/

Nested or crossed?

We want to test if the average qualifications a college gets depends on the grad of the students. We've got three schools and three different classes (1rst year, 2nd year, 3rd year)



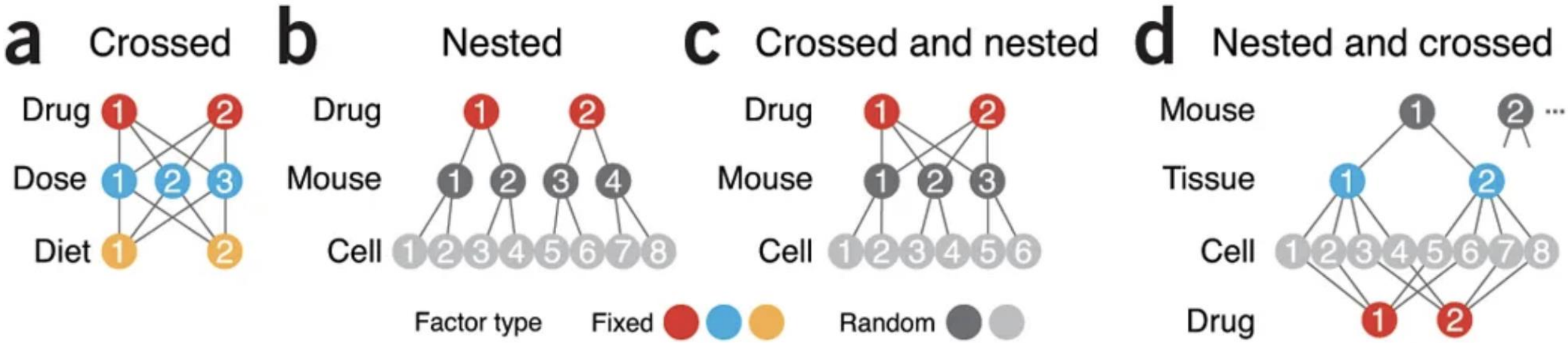
Home practice, guess the relationships between factors:



Note that there are two types of factors, fixed and random. We will focus on fixed factors.

https://yury-zablotski.netlify.app/post/mixed-effects-models-2/

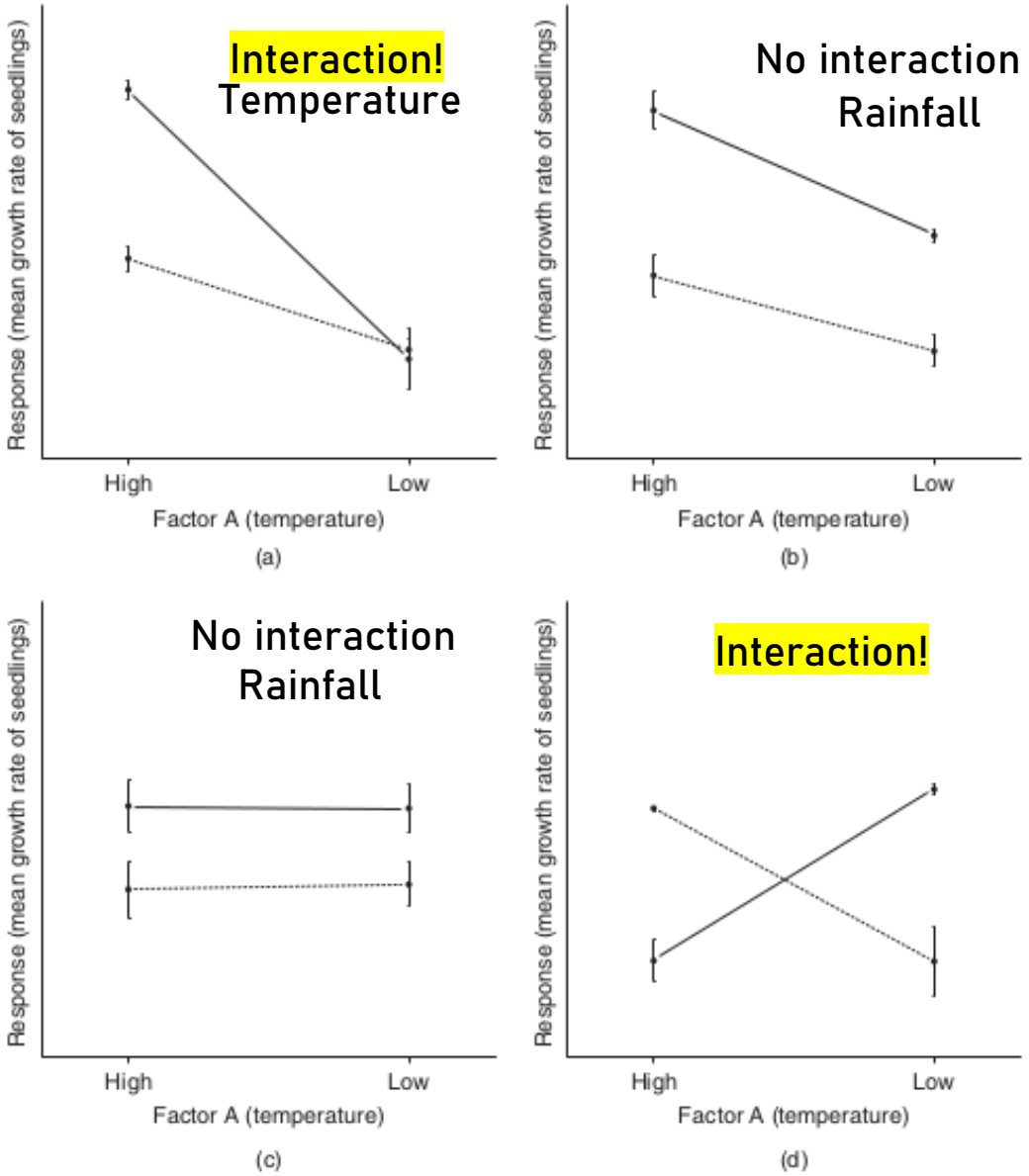
Home practice, guess the relationships between factors:



Note that there are two types of factors, fixed and random. We will focus on fixed factors.

https://yury-zablotski.netlify.app/post/mixed-effects-models-2/

Understanding the interaction



Factor A: Temperature (x-axis)
Factor B: Rainfall (line type)

High rainfall
Low rainfall

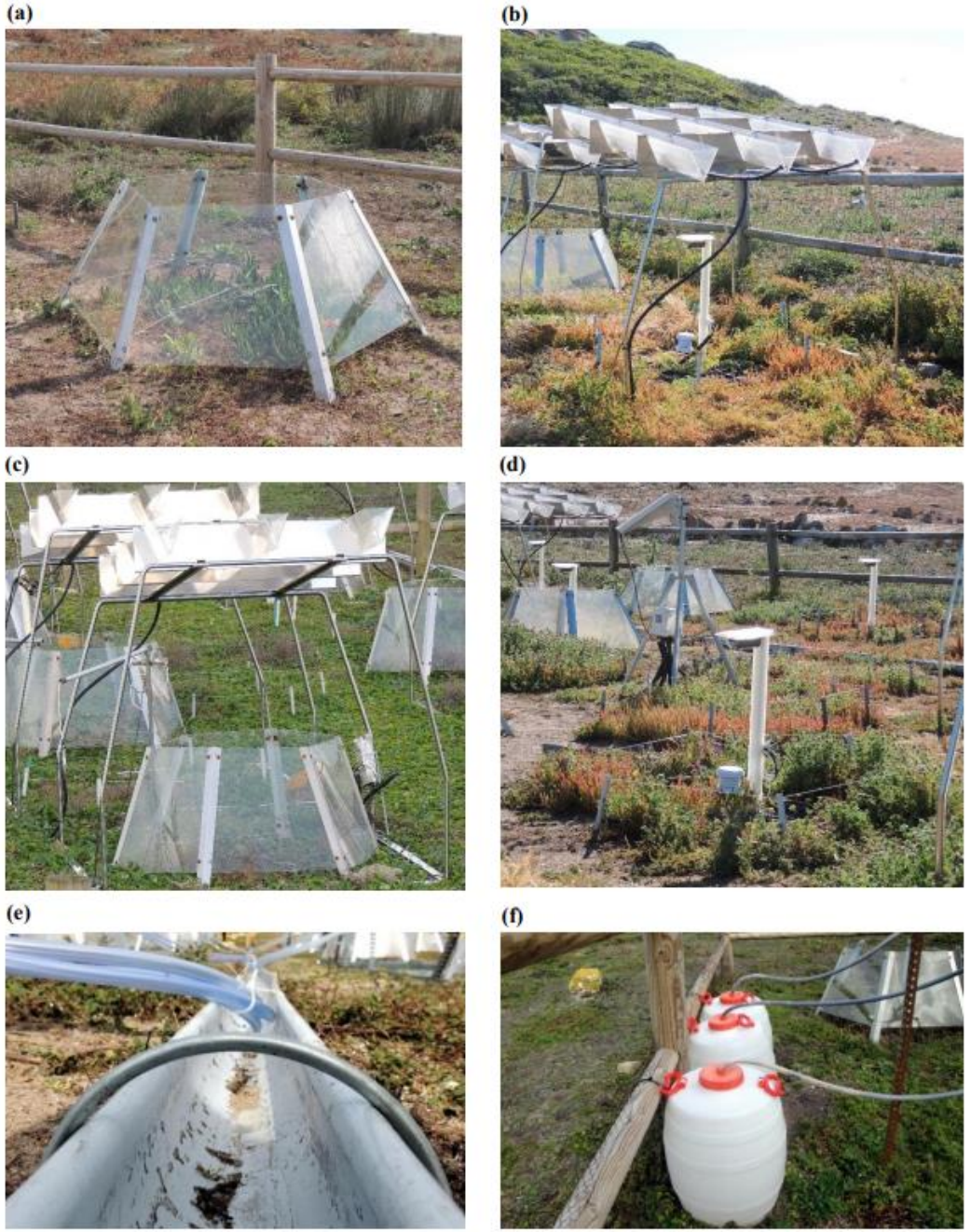
https://www.flutterbys.com.au/stats/tut/tut7.6a.html

Your task

- Open Top Chambers experimental design →
- Two levels of Temperature (Increased/Control)
- Two levels of Rainfall (Natural/Restricted)
- Plant height [cm], Net photosynthetic rate (An) [umolCO_2/m^2·s]
- 6 independent replicates
- 1 year, 1 species

Session4.Rmd

https://www.architecturalplants.com/product/cyperus-papyrus/



Campoy et al 2021 American Journal of Botany