Exercise: Two-way ANOVA in R

Stephan Huber

2024-01-16

This is an exercise where data management with the dplyr functions pivot\_longer, rename, and bind\_rows is practiced. Moreover, I exemplify how an ANOVA analysis can be executed with R. Doing so I refer to the content of Childs et al. (2021, Chapter 27). All data to this exercise can be found in Huber (2024).

# Read

Read Childs et al. (2021): [27.2 Competition between Calluna and Festuca](https://dzchilds.github.io/stats-for-bio/two-way-anova-in-r.html).

Our goal is to learn how to work with two-way ANOVA models in R, using an example from a plant competition experiment. The work flow is very similar to one-way ANOVA in R. We’ll start with the problem and the data, and then work through model fitting, evaluating assumptions, significance testing, and finally, presenting the results.

# Set up the R session

Download and open the script that you find [here](https://raw.githubusercontent.com/hubchev/ewa/main/rmd_festuca/r_festuca.R). This script contains all the code that is shown below. The solutions to the exercises can be found [here](https://raw.githubusercontent.com/hubchev/ewa/main/rmd_festuca/r_festuca_solutions.R).

setwd("~/Dropbox/hsf/courses/ewa/ewa\_all")  
  
rm(list = ls())  
  
if (!require(pacman)) install.packages("pacman")  
pacman::p\_load(tidyverse, rstatix, ggpubr, agricolae)

# Create a data frame

Plants have an optimal soil pH for growth, and this varies between species. Consequently we would expect that if we grow two plants in competition with each other at different pH values the effect of competition might vary according to the soil pH. In a recent study the growth of the grass Festuca ovina (Sheep’s Fescue) in competition with the heather Calluna vulgaris (Ling) was investigated in soils with different pH. Calluna is well adapted to grow on very acidic soils such as on the Millstone grit and blanket bogs around Sheffield. Festuca grows on soils with a much wider range of pH. We might hypothesise that Calluna will be a better competitor of Festuca in very acid soils than in moderately acid soils. Here are the data: The column (Weight) contains the Festuca dry weights, (pH) contains the codes for the pH treatment (levels: pH3.5, pH5.5), the column (Calluna) contains the codes for the presence or absence of Calluna (levels: Present, Absent).

data\_present <- data.frame(  
 Condition = rep(c("Calluna Present"), each = 5),  
 `pH 3.5` = c(2.76, 2.39, 3.54, 3.71, 2.49),  
 `pH 5.5` = c(3.21, 4.10, 3.04, 4.13, 5.21),  
 check.names = FALSE  
)  
data\_present

## Condition pH 3.5 pH 5.5  
## 1 Calluna Present 2.76 3.21  
## 2 Calluna Present 2.39 4.10  
## 3 Calluna Present 3.54 3.04  
## 4 Calluna Present 3.71 4.13  
## 5 Calluna Present 2.49 5.21

data\_absent <- data.frame(  
 Condition = rep(c("Calluna Absent"), each = 5),  
 `pH 3.5` = c(4.10, 2.72, 2.28, 4.43, 3.31),  
 `pH 5.5` = c(5.92, 7.31, 6.10, 5.25, 7.45),  
 check.names = FALSE  
)  
data\_absent

## Condition pH 3.5 pH 5.5  
## 1 Calluna Absent 4.10 5.92  
## 2 Calluna Absent 2.72 7.31  
## 3 Calluna Absent 2.28 6.10  
## 4 Calluna Absent 4.43 5.25  
## 5 Calluna Absent 3.31 7.45

# Make the data “long”

Read the [R documentation](https://dplyr.tidyverse.org/reference/bind_rows.html) of the function bind\_rows which is part of the dplyr package.

Combine the objects data\_present and data\_absent.

## Condition pH 3.5 pH 5.5  
## 1 Calluna Present 2.76 3.21  
## 2 Calluna Present 2.39 4.10  
## 3 Calluna Present 3.54 3.04  
## 4 Calluna Present 3.71 4.13  
## 5 Calluna Present 2.49 5.21  
## 6 Calluna Absent 4.10 5.92  
## 7 Calluna Absent 2.72 7.31  
## 8 Calluna Absent 2.28 6.10  
## 9 Calluna Absent 4.43 5.25  
## 10 Calluna Absent 3.31 7.45

Read Wickham & Grolemund (2023): [5.3 Lengthening data](https://r4ds.hadley.nz/data-tidy#sec-pivoting).

Rearrange the data so that they look like this:

festuca

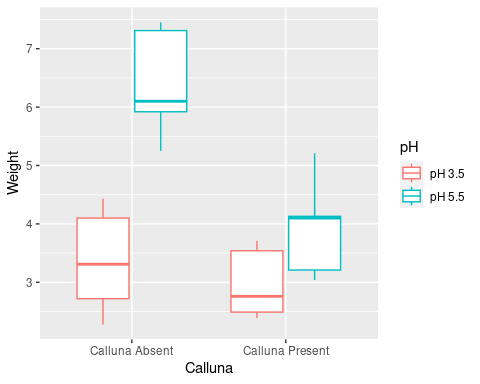
## # A tibble: 20 × 3  
## Calluna pH Weight  
## <chr> <chr> <dbl>  
## 1 Calluna Present pH 3.5 2.76  
## 2 Calluna Present pH 5.5 3.21  
## 3 Calluna Present pH 3.5 2.39  
## 4 Calluna Present pH 5.5 4.1   
## 5 Calluna Present pH 3.5 3.54  
## 6 Calluna Present pH 5.5 3.04  
## 7 Calluna Present pH 3.5 3.71  
## 8 Calluna Present pH 5.5 4.13  
## 9 Calluna Present pH 3.5 2.49  
## 10 Calluna Present pH 5.5 5.21  
## 11 Calluna Absent pH 3.5 4.1   
## 12 Calluna Absent pH 5.5 5.92  
## 13 Calluna Absent pH 3.5 2.72  
## 14 Calluna Absent pH 5.5 7.31  
## 15 Calluna Absent pH 3.5 2.28  
## 16 Calluna Absent pH 5.5 6.1   
## 17 Calluna Absent pH 3.5 4.43  
## 18 Calluna Absent pH 5.5 5.25  
## 19 Calluna Absent pH 3.5 3.31  
## 20 Calluna Absent pH 5.5 7.45

# Descriptive statistics

Calculate the following statistics and graphs:

summary\_stats

## # A tibble: 4 × 4  
## Calluna pH mean var  
## <chr> <chr> <dbl> <dbl>  
## 1 Calluna Absent pH 3.5 3.37 0.818  
## 2 Calluna Absent pH 5.5 6.41 0.893  
## 3 Calluna Present pH 3.5 2.98 0.371  
## 4 Calluna Present pH 5.5 3.94 0.754



# ANOVA

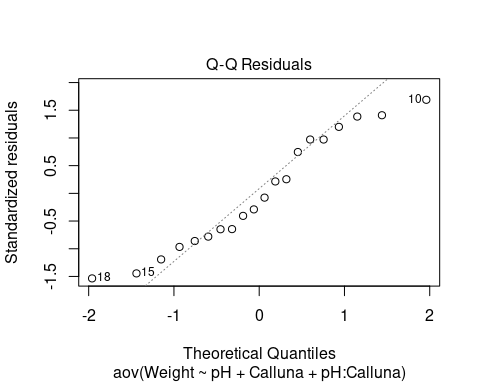
Use this model to calculate the ANOVA: Weight ~ pH + Calluna + pH:Calluna

## Analysis of Variance Table  
##   
## Response: Weight  
## Df Sum Sq Mean Sq F value Pr(>F)   
## pH 1 19.9800 19.9800 28.1792 7.065e-05 \*\*\*  
## Calluna 1 10.2102 10.2102 14.4001 0.00159 \*\*   
## pH:Calluna 1 5.3976 5.3976 7.6126 0.01397 \*   
## Residuals 16 11.3446 0.7090   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

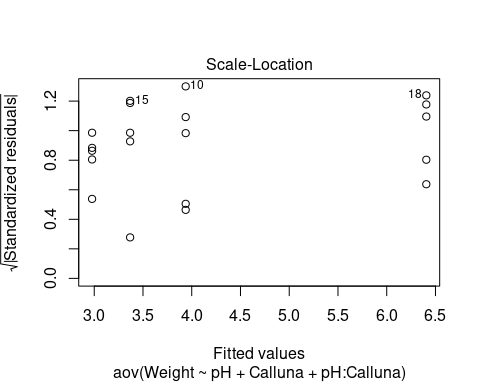
# Diagnostics

Read Childs et al. (2021): [27.5 Diagnostics](https://dzchilds.github.io/stats-for-bio/two-way-anova-in-r.html#diagnostics). Moreover, [this page](https://www.datanovia.com/en/lessons/anova-in-r/) is worth a look. You find some alternatives R functions that you may find helpful for ANOVA diagnostics.

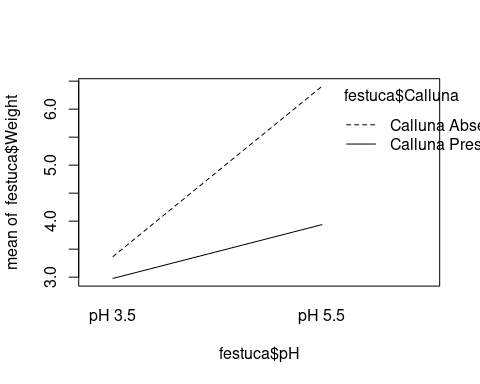
plot(festuca\_model, which = 2, add.smooth = FALSE)



plot(festuca\_model, which = 3, add.smooth = FALSE)



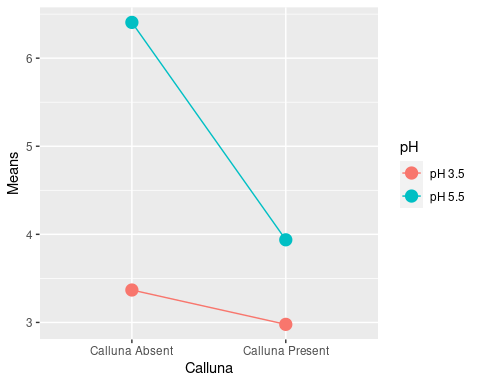
# Interaction diagram

Can you use the function interaction.plot to create the following: 

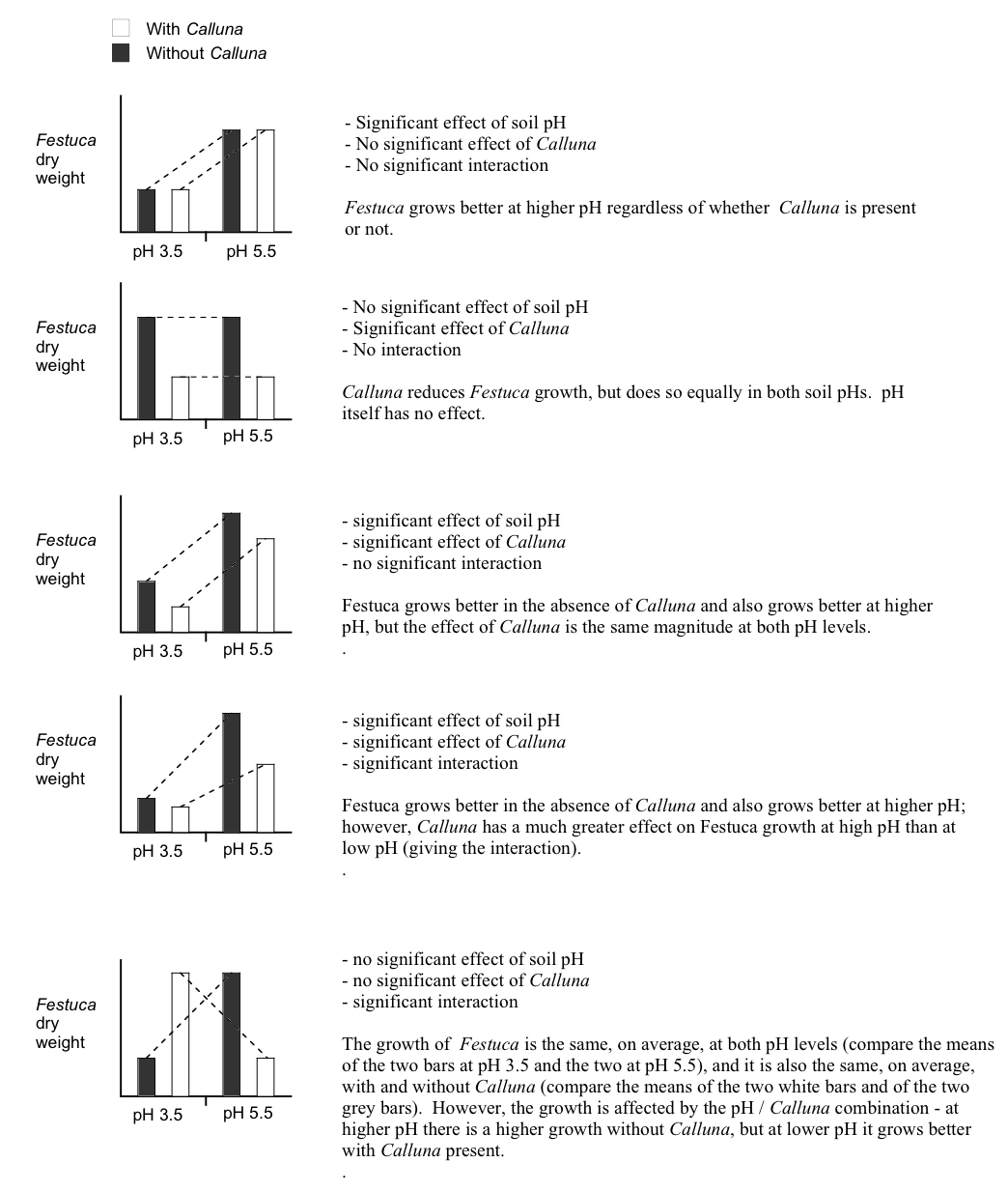
Here is a much nicer and more flexible approach to make interaction plots using tidyverse functions:

# step 1. calculate means for each treatment combination  
festuca\_means <-   
 festuca %>%   
 group\_by(Calluna, pH) %>% # <- remember to group by \*both\* factors  
 summarise(Means = mean(Weight))

# step 2. plot these as an interaction plot  
ggplot(festuca\_means,   
 aes(x = Calluna, y = Means, colour = pH, group = pH)) +  
 geom\_point(size = 4) + geom\_line()



Read Childs et al. (2021): [27.6.1](https://dzchilds.github.io/stats-for-bio/two-way-anova-in-r.html#understanding-the-model-graphically) and consider the following figure:



Understanding the model graphically

# Multiple comparison tests

TukeyHSD(festuca\_model, which = 'pH:Calluna')

## Tukey multiple comparisons of means  
## 95% family-wise confidence level  
##   
## Fit: aov(formula = Weight ~ pH + Calluna + pH:Calluna, data = festuca)  
##   
## $`pH:Calluna`  
## diff lwr upr p adj  
## pH 5.5:Calluna Absent-pH 3.5:Calluna Absent 3.038 1.5143518 4.5616482 0.0001731  
## pH 3.5:Calluna Present-pH 3.5:Calluna Absent -0.390 -1.9136482 1.1336482 0.8826936  
## pH 5.5:Calluna Present-pH 3.5:Calluna Absent 0.570 -0.9536482 2.0936482 0.7117913  
## pH 3.5:Calluna Present-pH 5.5:Calluna Absent -3.428 -4.9516482 -1.9043518 0.0000443  
## pH 5.5:Calluna Present-pH 5.5:Calluna Absent -2.468 -3.9916482 -0.9443518 0.0014155  
## pH 5.5:Calluna Present-pH 3.5:Calluna Present 0.960 -0.5636482 2.4836482 0.3079685

```r  
HSD.test(festuca\_model, trt = c("pH", "Calluna"), console = TRUE)

##   
## Study: festuca\_model ~ c("pH", "Calluna")  
##   
## HSD Test for Weight   
##   
## Mean Square Error: 0.709035   
##   
## pH:Calluna, means  
##   
## Weight std r se Min Max Q25 Q50 Q75  
## pH 3.5:Calluna Absent 3.368 0.9042511 5 0.3765727 2.28 4.43 2.72 3.31 4.10  
## pH 3.5:Calluna Present 2.978 0.6089089 5 0.3765727 2.39 3.71 2.49 2.76 3.54  
## pH 5.5:Calluna Absent 6.406 0.9451614 5 0.3765727 5.25 7.45 5.92 6.10 7.31  
## pH 5.5:Calluna Present 3.938 0.8685448 5 0.3765727 3.04 5.21 3.21 4.10 4.13  
##   
## Alpha: 0.05 ; DF Error: 16   
## Critical Value of Studentized Range: 4.046093   
##   
## Minimun Significant Difference: 1.523648   
##   
## Treatments with the same letter are not significantly different.  
##   
## Weight groups  
## pH 5.5:Calluna Absent 6.406 a  
## pH 5.5:Calluna Present 3.938 b  
## pH 3.5:Calluna Absent 3.368 b  
## pH 3.5:Calluna Present 2.978 b

# Drawing conclusions and presenting results

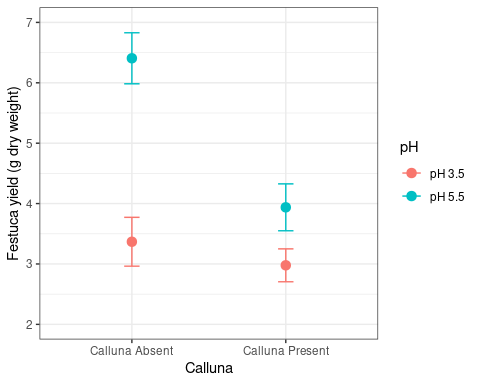
Here are some code examples how the shown graphs above could be made much nicer.

# step 1. calculate means for each treatment combination  
festuca\_stats <-   
 festuca %>%   
 group\_by(Calluna, pH) %>% # <- remember to group by the two factors  
 summarise(Means = mean(Weight), SEs = sd(Weight)/sqrt(n()))

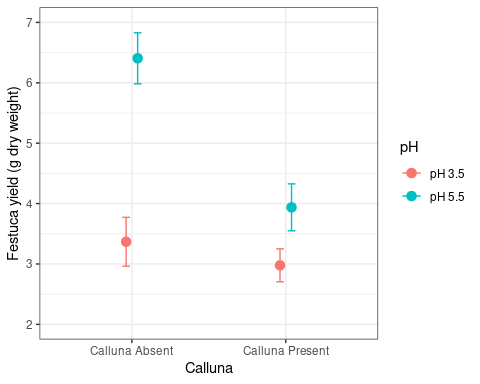
## `summarise()` has grouped output by 'Calluna'. You can override using the `.groups` argument.

# step 1. calculate means for each treatment combination  
festuca\_stats <-   
 festuca %>%   
 group\_by(Calluna, pH) %>% # <- remember to group by the two factors  
 summarise(Means = mean(Weight), SEs = sd(Weight)/sqrt(n()))

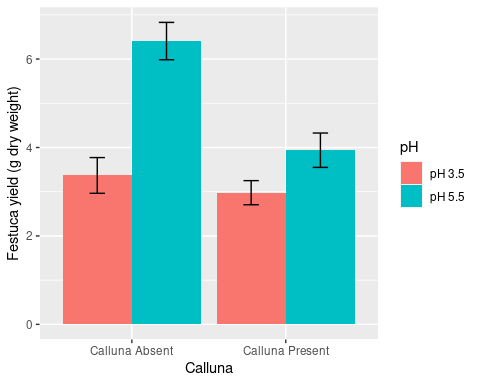
# step 2. plot these as an interaction plot  
ggplot(festuca\_stats,   
 aes(x = Calluna, y = Means, colour = pH,  
 ymin = Means - SEs, ymax = Means + SEs)) +  
 # this adds the mean  
 geom\_point(size = 3) +  
 # this adds the error bars  
 geom\_errorbar(width = 0.1) +  
 # controlling the appearance  
 scale\_y\_continuous(limits = c(2, 7)) +   
 xlab("Calluna") + ylab("Festuca yield (g dry weight)") +   
 # use a more professional theme  
 theme\_bw()



# define a position adjustment   
pos <- position\_dodge(0.15)  
# make the plot  
ggplot(festuca\_stats,   
 aes(x = Calluna, y = Means, colour = pH,  
 ymin = Means - SEs, ymax = Means + SEs)) +  
 # this adds the mean (shift positions with 'position =')  
 geom\_point(size = 3, position = pos) +  
 # this adds the error bars (shift positions with 'position =')  
 geom\_errorbar(width = 0.1, position = pos) +  
 # controlling the appearance  
 scale\_y\_continuous(limits = c(2, 7)) +   
 xlab("Calluna") + ylab("Festuca yield (g dry weight)") +   
 # use a more professional theme  
 theme\_bw()



ggplot(festuca\_stats,   
 aes(x = Calluna, y = Means, fill = pH,  
 ymin = Means - SEs, ymax = Means + SEs)) +  
 # this adds the mean  
 geom\_col(position = position\_dodge()) +  
 # this adds the error bars  
 geom\_errorbar(position = position\_dodge(0.9), width=.2) +  
 # controlling the appearance  
 xlab("Calluna") + ylab("Festuca yield (g dry weight)")



# References

Childs, D. Z., Hindle, B. J., & Warren, P. H. (2021). *APS 240: Data analysis and statistics with r*. online. <https://dzchilds.github.io/stats-for-bio>

Huber, S. (2024). *Empirisch-wissenschaftlich arbeiten (ewa)*. GitHub repository. <https://github.com/hubchev/ewa>

Wickham, H., & Grolemund, G. (2023). *R for data science (2e)*. Accessed January 30, 2023. <https://r4ds.hadley.nz/>