Double Dataset Analysis: Two-Way ANOVA for the Win

Processing and Analysis of Biological Data, BIOS14

Enrico Turato, Final Exam 2022, 2023-JAN-09

Visit Github to have access to the repository containing the code (also inserted in section 3) (and in general all the other files related) used for this small analysis exercise: `https://github.com/EnricoTurato/Final_Exam_BIOS14_2022.git`.

1 Part 1: Blossoms in Greenhouse

1.1 Background and first dataset description

I worked with data from an experiment in which *Dalechampia Plants*, from two populations each of two species ("S" and "L") were exposed to either dry or wet experimental conditions in a greenhouse (treatment). Blossom traits were measured on blossoms in early bisexual condition. The data comprise of many variables among which the important ones for my analysis are the following. sp: species ID; treat: treatment (dry or wet); LBL: lower bract length [mm] and UBL: upper bract length [mm]. The upper and lower bract lengths are averages of the lengths from the bract base to the tip of the three lobes, and their function is similar to petals in flower, acting as advertisement.

1.2 Scientific questions and analysis methods chosen

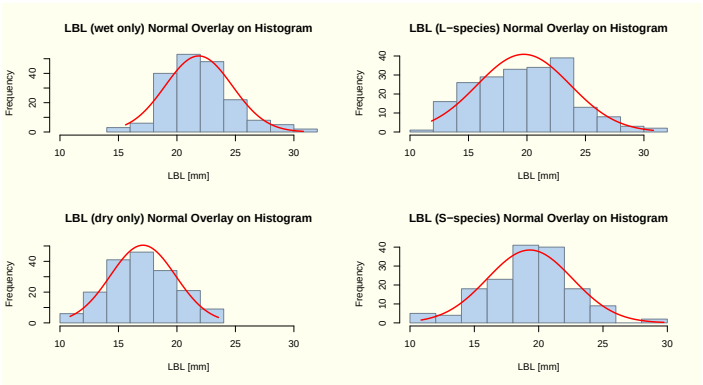
It is interesting to assess the effects of two experimental factors, or categorical variables (in this case the chosen two are "sp" and "treat", meaning the plants species, either S or L and the treatment, either wet or dry), plus their potential interaction, on some chosen response variables that in this case are the lower bract length (LBL) and the upper one (UBL). Following the philosophy of an effect size-centered approach and given the presence of two factors, or categorical variables among all the others, I decided to focus on a two-way ANOVA-type of analysis. So a two-way ANOVA has been performed two times, one for the first mentioned response variable (m = lm(dat$LBL∼species*treatment), anova(m)) and one for the second one (m = lm(dat$UBL∼species*treatment), anova(m)). At the end, estimates as percentage values are given and the main parameters from the ANOVA analysis have been reported.

1.3 Results and Conclusions

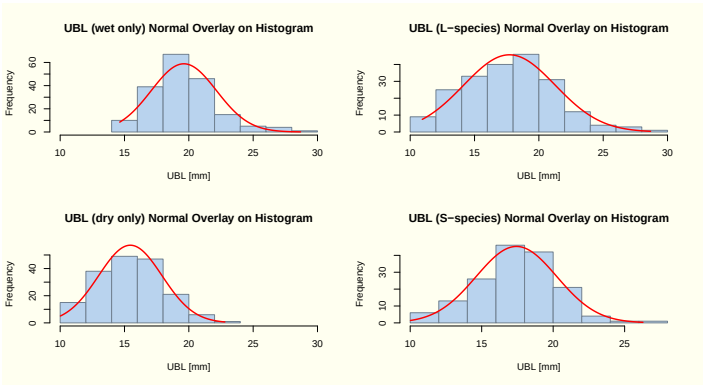
First of all I created four sub-dataset starting from the complete one, just by filtering, in order to group data by the treatments and by species and to observe the distributions of LBL and UBL through histograms to get a rough idea of the values from a qualitative point of view before moving to the ANOVA. The four dataset contained "wet-plants only", "dry-plants only", "L-plants only" and "S-plants only". The results are the following. Figure 1a for LBL and Figure 1b for UBL.

Figure 1: True for both the left image and for the right one: top left, the histogram from the "wet" data; bottom left, the histogram from the "wet" data; top right, the histogram from the "L-species" data; bottom right, the histogram from the "S-species" data. A normal distribution overlay is on all four of them for reference.

(a) LBL histograms



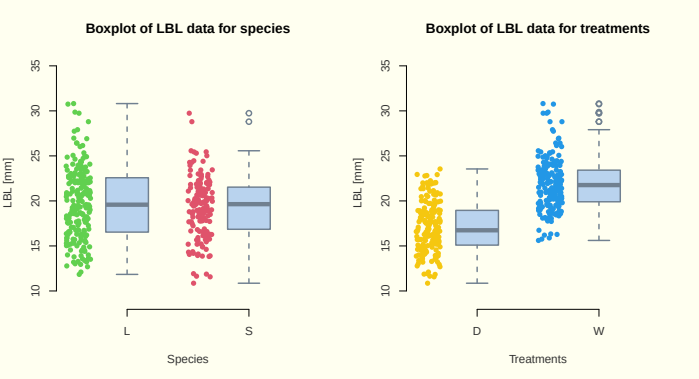
(b) UBL histograms



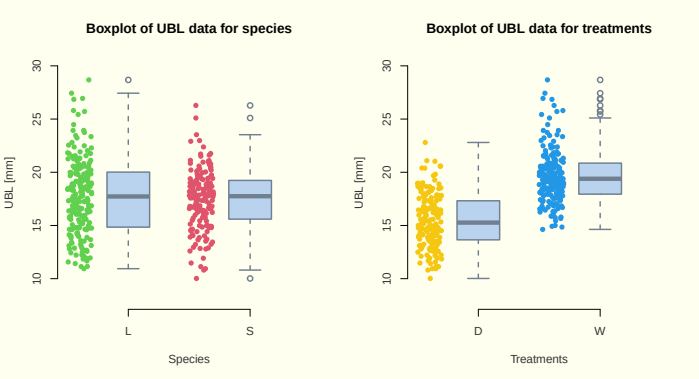
I then checked the distribution of data also with some boxplots, see Figure 2a and Figure 2b.

Figure 2: True for both the left image and for the right one: on the left, boxplots showing the data distribution of the length depending on the species. On the right boxplots showing the data distribution of the length depending on the treatment.

(a) LBL boxplots



(b) UBL boxplots



Next, to get some estimates, I performed the two-way ANOVA analysis. First of all, after computing the means and errors for LBL data and UBL data, see Table 1 left and right respectively, from a graphical point of view I obtained Figure 3.

Table 1: Means [mm] (above) and standard errors [mm] (below) for both LBL (left side) and UBL (right side).

/	Dry	Wet
L-species	17.05	22.55
S-species	17.15	21.09

/	Dry	Wet
L-species	0.27	0.30
S-species	0.34	0.27

/	Dry	Wet
L-species	15.37	20.20
S-species	15.57	18.96

/	Dry	Wet
L-species	0.25	0.27
S-species	0.29	0.23

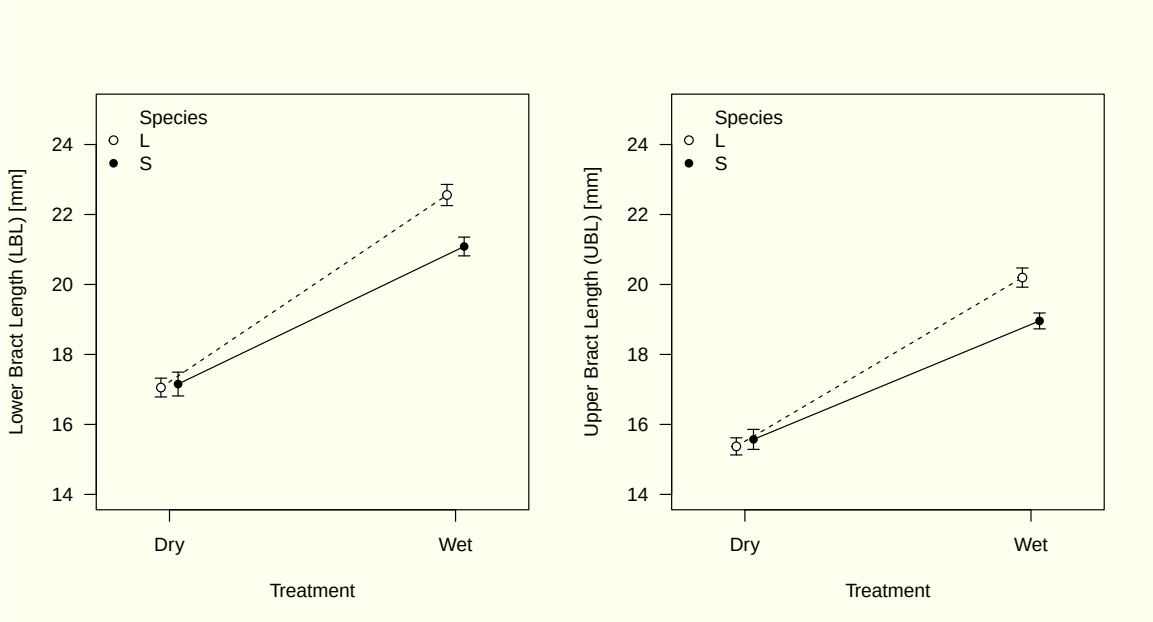


Figure 3: Plots giving a graphical idea of the size effect for the two-way ANOVA analysis performed for both the LBL and UBL when considering species and treatment type as factors.

Talking about the LBL (UBL), to interpret the results, by considering the ANOVA table, one sees that there are detectable effects of mainly treatment compared to the species contribution: more variance is indeed explained by treatment (based on the larger sum of squares: $\approx$ 2101 (1594) and $\approx$ 15 (7), respectively). Furthermore, the effects of species and treatment are not independent, as indicated by enough statistical support (P $\approx$ 0.008 (0.006), $F_{1,360} \approx 7$ (8)). The variance explained by the

interaction term is indeed not negligible when compared to the above mentioned ones (sum of squares: $\approx$ 55 (46)).
To quantify the effect size in both cases, I computed the mean LBL and UBL for each treatment and species from Table 1 by working first on columns and then on rows. Based on this, I can finally say the following.
The mean LBL (UBL) (taking the mean between L-species and S-species) was 21.6% (21%) shorter when considering plants with a dry treatment than when having a wet one (mean LBL (UBL) = 17.10 (15.47) mm and 21.82 (19.58) mm, respectively, $F_{1,360} \approx 269$ (261), P « 0.00001, Figure 3). When considering the effect of species instead, the mean LBL (UBL) for S-species (taking the mean between dry treatment and wet treatment) was only 3.5% (2.9%) shorter with respect to L-species (mean LBL (UBL) = 19.12 (17.26) mm and 19.80 (17.78) mm, respectively, $F_{1,360} \approx 2$ (1), $P > 0.16$ (0.28), Figure 3). The difference in LBL (UBL) between treatments was noticeable between the two species as correctly showed by the small above mentioned statistical support for the contribution of the interaction term (24.4% (23.9%) vs. 18.6% (17.9%), respectively L and S species; $P = \approx 0.008$ (0.006)).
All things considered, from the analysis of this particular dataset with the specified methods, it has been shown that treatment is affecting more the bracts length (when wet the bracts are longer in mean) than the existence of two species of plants (S and L) for both LBL and UBL. It has been shown also that there seems to be a non negligible interaction between the treatment and the species even though small.

2 Part 2: Mountain Goats

The following dataset comprises measurements of horn length and body mass of mountain goats. The data included many variables among which the chosen one for the analysis were the following. Sex: M/F; hornL: length of the left horn in mm; hornR: length of the right horn in mm; mass: body mass in kg; density: population density at birth, low / high.

2.1 Scientific questions and analysis methods chosen

Following the philosophy of an effect size-centered approach and given the presence of two factors, or categorical variables among all the others (meaning "sex" and population "density" at birth), it is quite understandable to be curious about the possible relation between such factors and the influence on certain physical response variables from a quantitative point of view such as the horn length and the body mass. I decided to focus on a two-way ANOVA-type of analysis in order to assess the effects of each experimental factor (categorical variables) and their potential interaction on the chosen response variables (horn length and the body mass). Thus a two-way ANOVA has been performed two times, one for the first mentioned response variable (m = lm(dat$horn∼sex*density), anova(m)) and one for the second one (m = lm(dat$mass∼sex*density), anova(m)).

2.2 Results and Conclusions

First of all, given the presence of collected data for each goat of both the left and right horn, to make the analysis easier, I took the difference between the measurement of the length of the left horn and the right one and I created an histogram, see Figure 4. Given the observation that the distribution was a very peaked Gaussian around the zero (mean = -0.15 mm, sigma = 5.69 mm) I have reduced the number of variables by substituting the "hornL" and "hornR" variables with a single one, "mean_horns_lenght", that is exactly the mean between the two sides. I then checked the distribution of this also to be sure is was "non-pathological" (mean = 181.81 mm, sigma = 43.92 mm).

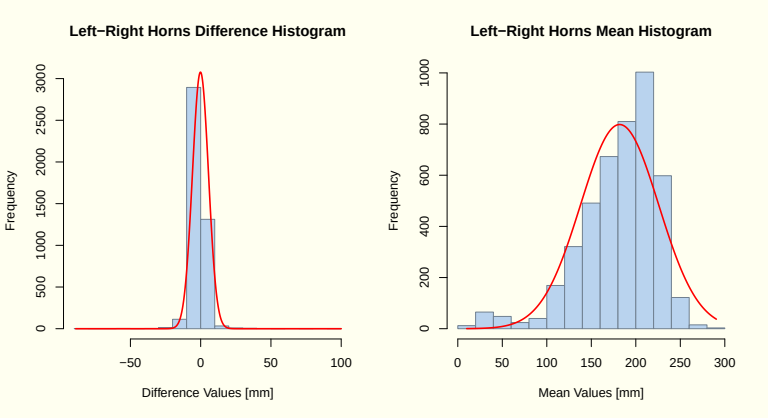


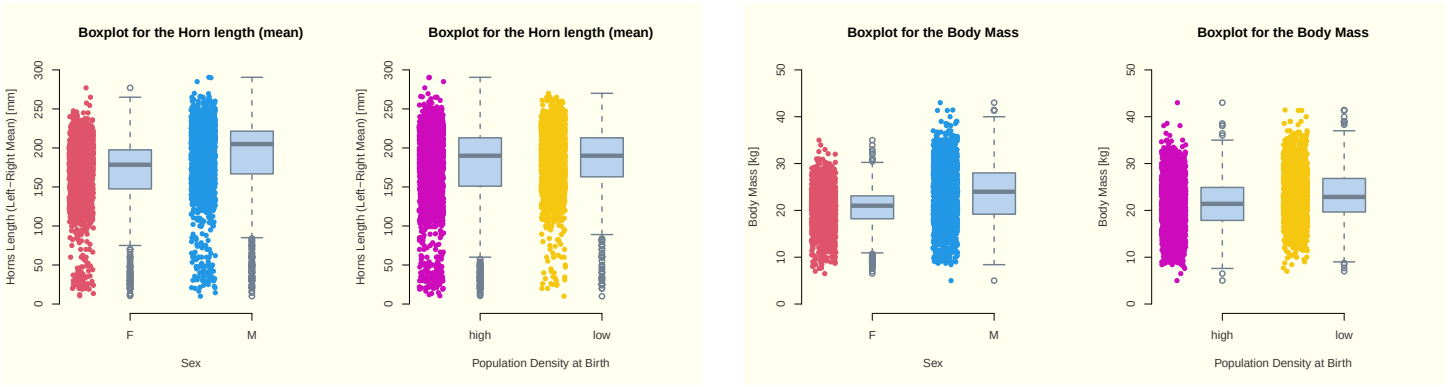
Figure 4: On the left, the histogram containing in the x-axis the difference between the measurement of the length of the left horn and the right one and their frequency on the y-axis with a normal distribution overlay. On the right the values of the the mean between the horns on the two sides on the x-axis [mm] vs their frequency and a normal distribution overlay.

I then checked, with some boxplots, see Figure 5a and Figure 5b, how data were distributed to get an idea of the results I might obtain before moving onto the two-way ANOVA analysis itself.

Figure 5: True both for the left image and for the right one: on the left, boxplots showing the data distribution of the chosen response variable depending on the sex of the goat. On the right boxplots showing the data distribution of the same variable depending on the density of population at the birth.

(a) Horn length boxplots

(b) Body mass boxplots



Now, to get some estimates, I performed the ANOVA analysis. First of all, after summarizing the main values of interest (mean, sd and se for the horns both left and right and for the body mass), Table 2 and Table 3, and next computing the means and errors for the newly created variable "mean_horns_lenght", Table 4, from a graphical point of view I obtained Figure 6.

Table 2: Summary statistics for the hornL and hornR response variables depending on the two factors.

Sex	hornL Mean [mm]	hLSD [mm]	hLSE [mm]	hornR Mean [mm]	hRSD [mm]	hRSE [mm]
Female	169.04	41.32	0.93	169.01	41.19	0.93
Male	191.91	43.47	0.88	192.20	43.49	0.88
Pop. density	hornL Mean [mm]	hLSD [mm]	hLSE [mm]	hornR Mean [mm]	hRSD [mm]	hRSE [mm]
High	177.99	48.57	1.02	178.11	48.64	1.02
Low	185.70	38.23	0.83	185.88	38.14	0.83

Table 3: Summary statistics for the body mass response variable depending on the two factors.

Sex	Mass Mean [kg]	Mass SD [kg]	Mass SE [kg]
Female	20.55	4.05	0.09
Male	23.67	5.90	0.12
Pop. density	Mass Mean [kg]	Mass SD [kg]	Mass SE [kg]
High	21.38	5.33	0.11
Low	23.23	5.29	0.11

Table 4: Means (above) and standard errors (below) after exploiting the new variable "horn_mean_legth" [mm] (on the left side) and body mass [kg] (on the right side).

/	High	Low
Female	165.12	173.30
Male	188.72	195.49

/	High	Low
Female	1.41	1.18
Male	1.39	1.06

/	High	Low
Female	19.88	21.29
Male	22.63	24.74

/	High	Low
Female	0.13	0.12
Male	0.17	0.17

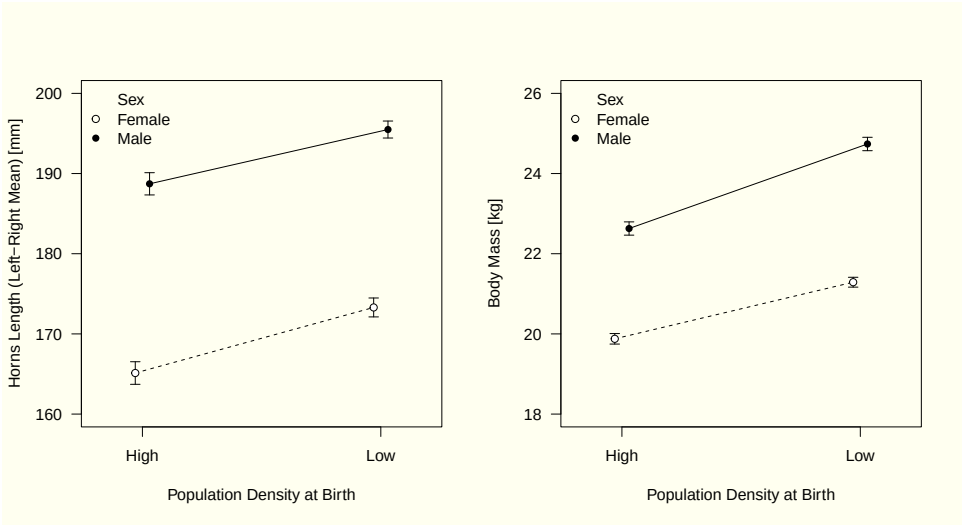


Figure 6: Plots giving a graphical idea of the size effect for the two-way ANOVA analysis performed for both the horn length and the body mass when considering sex and population density at birth as factors.

Talking about the horn length (body mass), to interpret the results, by checking first the ANOVA table, one sees that there are detectable effects of both sex and population density, with more variance explained by sex (based on the larger sum of squares: 575479 (10536) and 60026 (3551), respectively). Furthermore, concerning the horn length only, the effects of sex and population density are independent, as indicated by a negligible statistical support ($P > 0.5$, $F_{1,4390} < 0.4$). Talking about the body mass the effects of sex and population density are, instead, not completely independent as showed by a small statistical support in favor of a small interaction ($P \approx 0.02$, $F_{1,4390} \approx 5$). To quantify the effect size in both cases, I computed the mean horn length and body mass for each population density and sex from Table 4 by working first on columns and then on rows. Based on this, I can finally say the following. The mean horn length (taking the mean between male and female, remembering that I already took the mean between left one and right one in each goat at the beginning as already specified) (body mass) was 4% shorter (7.6% smaller) when considering goats born with a high density in population than when having a low population density at birth (mean horn length (body mass) = 176.92 mm (21.25 kg) and 184.39 mm (23.01 kg), respectively, $F_{1,4390} = 33.62$ (137.62), $P \ll 0.00001$, Figure 6). When considering the effect of sex, instead, the mean horn length for females (taking the mean between high density and low density after I took the mean between left one and right one in each goat at the beginning as already specified) (body mass) was 12% shorter (13.1% lower) with respect to males (mean horn length (body mass) = 169.21 mm (20.58 kg) and 192.10 mm (23.68 kg), respectively, $F_{1,4390} = 322.28$ (408.31), $P \ll 0.00001$, Figure 6). The difference in horns length between population densities was very very small between males and females as correctly showed by the negligible above mentioned statistical support for the contribution of the interaction term (3.5% vs. 4.7%, respectively; $P = 0.58$). There was a slight difference in body mass between population densities between males and females as can also be seen by the small but present statistical support in favor of the interaction term in this case (8.5% vs. 6.6%, respectively; $P = 0.02$). All things considered, from the analysis of this particular dataset with the specified methods, it has been shown that sex is affecting more the horn length and the body mass (both higher in males) of mountain goats compared to the low population densities at birth. No strong interaction has been found between sex and the population density at birth even though a non-negligible statistical support in favor of a small interaction has been outlined when it comes to the body mass response variable.

3 Appendix

The R code used for the analysis of the two datasets is inserted here below.

3.1 Blossoms in Greenhouse

```
library(sciplot)
library(magrittr)
library(dplyr)
library(plyr)
library(knitr)
library(rcompanion)
library(MASS)

##############################################################################
##############################################################################
#
# DATASET INTRODUCTION AS USUAL
#
##############################################################################
##############################################################################

dat = read.csv("exam2022_part1.csv")
head(dat)
names(dat)
View(dat)
str(dat)
dat = na.omit(dat)
############################
# creation of sub-datasets
############################

dat_wet = dat[dat$treat=="W",]
dat_dry = dat[dat$treat=="D",]
dat_S = dat[dat$sp=="S",]
dat_L = dat[dat$sp=="L",]

View(dat_wet)
View(dat_dry)
View(dat_S)
View(dat_L)
##############################################################################
# VARIABLES ORGANIZING AND CHECKING PLUS HISTOGRAMS
##############################################################################

populations = as.factor(dat$pop)
species = as.factor(dat$sp)
```

```r
treatment = as.factor(dat$treat)

#####################################################################
# checking LBL and UBL with histograms for each dataset
#####################################################################



#############################################################################
#############################################################################
#
# 4x4 histograms for LBL for the various subdataset
#
#############################################################################
#############################################################################

par(mfrow=c(2,2))
par(bg = "ivory")

plotNormalHistogram(dat_wet$LBL, prob = FALSE, col="slategray2", border="slategray",xlim = c(10,32),
                    main = "LBL (wet only) Normal Overlay on Histogram", xlab = "LBL [mm]",
                    linecol="red", lwd=2 )
plotNormalHistogram(dat_L$LBL, prob = FALSE, col="slategray2", border="slategray",
                    main = "LBL (L-species) Normal Overlay on Histogram", xlab = "LBL [mm]",
                    linecol="red", lwd=2 )
plotNormalHistogram(dat_dry$LBL, prob = FALSE, col="slategray2", border="slategray",xlim = c(10,32),
                    main = "LBL (dry only) Normal Overlay on Histogram", xlab = "LBL [mm]",
                    linecol="red", lwd=2 )
plotNormalHistogram(dat_S$LBL, prob = FALSE, col="slategray2", border="slategray",
                    main = "LBL (S-species) Normal Overlay on Histogram", xlab = "LBL [mm]",
                    linecol="red", lwd=2)

par(oldpar)

#############################################################################
#############################################################################
#
# 4x4 histograms for UBL for the various subdataset
#
#############################################################################
#############################################################################

par(mfrow=c(2,2))
par(bg = "ivory")
plotNormalHistogram(dat_wet$UBL, prob = FALSE, col="slategray2", border="slategray",xlim = c(10,30),
                    main = "UBL (wet only) Normal Overlay on Histogram", xlab = "UBL [mm]",
```

```
                    linecol="red", lwd=2 )
plotNormalHistogram(dat_L$UBL, prob = FALSE, col="slategray2", border="slategray",
                    main = "UBL (L-species) Normal Overlay on Histogram", xlab = "UBL [mm]",
                    linecol="red", lwd=2 )
plotNormalHistogram(dat_dry$UBL, prob = FALSE, col="slategray2", border="slategray",xlim = c(10,30),
                    main = "UBL (dry only) Normal Overlay on Histogram", xlab = "UBL [mm]",
                    linecol="red", lwd=2 )
plotNormalHistogram(dat_S$UBL, prob = FALSE, col="slategray2", border="slategray",
                    main = "UBL (S-species) Normal Overlay on Histogram", xlab = "UBL [mm]",
                    linecol="red", lwd=2)
par(oldpar)

##############################################################################
##############################################################################
#
# Doing boxplots to visualize data of UBL and LBL
#
##############################################################################
##############################################################################

########
# dat
########

########
# first LBL
########

par(mfrow=c(1,2))
par(bg = "ivory")
boxplot(dat$LBL~species, xlab="Species", ylim = c(9, 35),
        ylab="LBL [mm]",boxwex=0.35, main = "Boxplot of LBL data for species", lwd = 2, border=
         ↪  "slategrey", # colour of the box borders
        col = "slategray2", # colour of the inside of the boxes
        col.axis = 'grey20', # colour of the axis numbers
        col.lab = 'grey20', # colour of the axis labels
        frame = F)
stripchart(dat$LBL~species,
           method = "jitter",  main = "method = 'jitter', jitter = 0.2",
           pch = 16, # specify the type of point to use
           cex = 1,
           col = c(3,2),
           vertical = TRUE,
           at = c(0.6,1.6),
           add = TRUE)
```

```r
# at = c(0.75,1.75)

boxplot(dat$LBL~treatment, xlab="Treatments", ylim = c(9, 35),
        ylab="LBL [mm]",boxwex=0.35, main = "Boxplot of LBL data for treatments", lwd = 2, border=
        ↪  "slategrey", # colour of the box borders
        col = "slategray2", # colour of the inside of the boxes
        col.axis = 'grey20', # colour of the axis numbers
        col.lab = 'grey20', # colour of the axis labels
        frame = F)
stripchart(dat$LBL~treatment,
           method = "jitter",  main = "method = 'jitter', jitter = 0.2",
           pch = 16, # specify the type of point to use
           cex = 1,
           col = c(7,4),
           vertical = TRUE,
           at = c(0.6,1.6),
           add = TRUE)

par(oldpar)

########
# now UBL
########



par(mfrow=c(1,2))
par(bg = "ivory")
boxplot(dat$UBL~species, xlab="Species", ylim = c(8, 30),
        ylab="UBL [mm]",boxwex=0.35, main = "Boxplot of UBL data for species", lwd = 2, border=
        ↪  "slategrey", # colour of the box borders
        col = "slategray2", # colour of the inside of the boxes
        col.axis = 'grey20', # colour of the axis numbers
        col.lab = 'grey20', # colour of the axis labels
        frame = F)
stripchart(dat$UBL~species,
           method = "jitter",  main = "method = 'jitter', jitter = 0.2",
           pch = 16, # specify the type of point to use
           cex = 1,
           col = c(3,2),
           vertical = TRUE,
           at = c(0.6,1.6),
           add = TRUE)
# at = c(0.75,1.75)

boxplot(dat$UBL~treatment, xlab="Treatments", ylim = c(8, 30),
```

```
        ylab="UBL [mm]",boxwex=0.35, main = "Boxplot of UBL data for treatments", lwd = 2, border=
        ↪  "slategrey", # colour of the box borders
        col = "slategray2", # colour of the inside of the boxes
        col.axis = 'grey20', # colour of the axis numbers
        col.lab = 'grey20', # colour of the axis labels
        frame = F)
stripchart(dat$UBL~treatment,
           method = "jitter",  main = "method = 'jitter', jitter = 0.2",
           pch = 16, # specify the type of point to use
           cex = 1,
           col = c(7,4),
           vertical = TRUE,
           at = c(0.6,1.6),
           add = TRUE)

par(oldpar)

##############################################################################
##############################################################################
#
# end of boxplots to visualize data
#
##############################################################################
##############################################################################



##############################################################################
# 2-WAY ANOVA FOR HORNS LENGTH For UBL and LBL with factors treatment and species
##############################################################################


##############################################################################
# STARTING THE FIRST 2-WAY ANOVA FOR LBL
##############################################################################

means = tapply(dat$LBL, list(species, treatment), mean)
ses = tapply(dat$LBL,
             list(species, treatment),
             function(x) sd(x)/sqrt(sum(!is.na(x))))
means
ses

m = lm(dat$LBL~species*treatment)
anova(m)
```

```r
summary(m)
colMeans(means)
rowMeans(means)


##############################################################################
# SECOND 2-WAY ANOVA FOR UBL
##############################################################################

means2 = tapply(dat$UBL, list(species, treatment), mean)
ses2 = tapply(dat$UBL,
              list(species, treatment),
              function(x) sd(x)/sqrt(sum(!is.na(x))))
means2
ses2

m3 = lm(dat$UBL~species*treatment)
anova(m3)
summary(m3)
colMeans(means2)
rowMeans(means2)


##############################################################################
# PUTTING PLOTS TOGETHER
##############################################################################

par(mfrow=c(1,2))
par(bg = "ivory")

plot(c(0.97, 1.03), means[,1], ylim=c(14, 25), xlim=c(0.8, 2.2),
     xlab="Treatment",
     ylab="Lower Bract Length (LBL) [mm]",
     xaxt="n", las=1, pch=c(21,16), col="white")
axis(1, 1:2, labels=c("Dry", "Wet"))
arrows(c(0.97,1.03), means[,1]-ses[,1], c(0.97,1.03),
       means[,1]+ses[,1], length=0.05, angle=90, code=3)
arrows(c(1.97,2.03), means[,2]-ses[,2], c(1.97,2.03),
       means[,2]+ses[,2], length=0.05, angle=90, code=3)
segments(0.97, means[1,1], 1.97, means[1,2], lty=2)
segments(1.03, means[2,1], 2.03, means[2,2])
points(c(0.97, 1.03), means[,1], pch=c(21,16), bg="white")
points(c(1.97, 2.03), means[,2], pch=c(21, 16), bg="white")
legend("topleft", c("Species", "L", "S"),
       bty="n", pch=c(NA,21,16))
```

```r
plot(c(0.97, 1.03), means2[,1], ylim=c(14, 25), xlim=c(0.8, 2.2),
     xlab="Treatment",
     ylab="Upper Bract Length (UBL) [mm]",
     xaxt="n", las=1, pch=c(21,16), col="white")
axis(1, 1:2, labels=c("Dry", "Wet"))
arrows(c(0.97,1.03), means2[,1]-ses2[,1], c(0.97,1.03),
       means2[,1]+ses2[,1], length=0.05, angle=90, code=3)
arrows(c(1.97,2.03), means2[,2]-ses2[,2], c(1.97,2.03),
       means2[,2]+ses2[,2], length=0.05, angle=90, code=3)
segments(0.97, means2[1,1], 1.97, means2[1,2], lty=2)
segments(1.03, means2[2,1], 2.03, means2[2,2])
points(c(0.97, 1.03), means2[,1], pch=c(21,16), bg="white")
points(c(1.97, 2.03), means2[,2], pch=c(21, 16), bg="white")
legend("topleft", c("Species", "L", "S"),
       bty="n", pch=c(NA,21,16))


par(oldpar)
```

3.2 Mountain Goats

```r
library(sciplot)
library(magrittr)
library(dplyr)
library(plyr)
library(knitr)
library(rcompanion)
library(MASS)

################################################################################
################################################################################
#
# DATASET INTRODUCTION AS USUAL
#
################################################################################
################################################################################

dat = read.table("exam2022_part2.txt", header=T)
head(dat)
names(dat)
View(dat)
str(dat)
dat = na.omit(dat)
```

```
##############################################################################
# VARIABLES ORGANIZING AND CHECKING PLUS HISTOGRAMS
##############################################################################

sex = as.factor(dat$sex)
density = as.factor(dat$density)
#plot(dat)
horns_difference = dat$hornL - dat$hornR
mean_horns_lenght = (dat$hornL + dat$hornR)/2


oldpar = par(no.readonly = TRUE)


par(mfrow=c(1,2))
par(bg = "ivory")

plotNormalHistogram(horns_difference, prob = FALSE, col="slategray2", border="slategray",
                    main = "Left-Right Horns Difference Histogram", xlab = "Difference Values [mm]",
                    linecol="red", lwd=2 )


plotNormalHistogram(mean_horns_lenght, prob = FALSE, col="slategray2", border="slategray",
                    main = "Left-Right Horns Mean Histogram", xlab = "Mean Values [mm]",
                    linecol="red", lwd=2 )

par(oldpar)


fitdistr(horns_difference, "normal")


fitdistr(mean_horns_lenght, "normal")

# sex as categorical variable

popstats = ddply(dat, .(sex), summarize,
                  hornLm = mean(hornL, na.rm=T),
                  hornLsd = sd(hornL, na.rm=T),
                  hornLse = se(hornL, na.rm=T),
                  hornRm = mean(hornR, na.rm=T),
                  hornRsd = sd(hornR, na.rm=T),
                  hornRse = se(hornR, na.rm=T),
                  massm = mean(mass, na.rm=T),
                  masssd = sd(mass, na.rm=T),
```

```
                  massse = se(mass, na.rm=T))
popstats[,-1] = round(popstats[,-1], 2)
kable(popstats)

# population density at birth as categorical variable

popstats2 = ddply(dat, .(density), summarize,
                  hornLm = mean(hornL, na.rm=T),
                  hornLsd = sd(hornL, na.rm=T),
                  hornLse = se(hornL, na.rm=T),
                  hornRm = mean(hornR, na.rm=T),
                  hornRsd = sd(hornR, na.rm=T),
                  hornRse = se(hornR, na.rm=T),
                  massm = mean(mass, na.rm=T),
                  masssd = sd(mass, na.rm=T),
                  massse = se(mass, na.rm=T))
popstats2[,-1] = round(popstats2[,-1], 2)
kable(popstats2)
##############################################################################
# TOWARDS 2-WAY ANOVA: LOOKING AT BOXPLOTS
##############################################################################

oldpar = par(no.readonly = TRUE)


par(mfrow=c(1,2))
par(bg = "ivory")
boxplot(mean_horns_lenght~sex, xlab="Sex", ylim = c(0, 300),
        ylab="Horns Length (Left-Right Mean) [mm]",boxwex=0.35, main = "Boxplot for the Horn length
        ↪  (mean)", lwd = 2, border= "slategrey", # colour of the box borders
        col = "slategray2", # colour of the inside of the boxes
        col.axis = 'grey20', # colour of the axis numbers
        col.lab = 'grey20', # colour of the axis labels
        frame = F)
stripchart(mean_horns_lenght~sex,
           method = "jitter",  main = "method = 'jitter', jitter = 0.2",
           pch = 16, # specify the type of point to use
           cex = 1,
           col = c(2,4),
           vertical = TRUE,
           at = c(0.6,1.6),
           add = TRUE)


boxplot(mean_horns_lenght~density, xlab="Population Density at Birth", ylim = c(0, 300),
```

```
        ylab="Horns Length (Left-Right Mean) [mm]",boxwex=0.35, main = "Boxplot for the Horn length
        ↪  (mean)", lwd = 2, border= "slategrey", # colour of the box borders
        col = "slategray2", # colour of the inside of the boxes
        col.axis = 'grey20', # colour of the axis numbers
        col.lab = 'grey20', # colour of the axis labels
        frame = F)
stripchart(mean_horns_lenght~density,
           method = "jitter",  main = "method = 'jitter', jitter = 0.2",
           pch = 16, # specify the type of point to use
           cex = 1,
           col = c(6,7),
           vertical = TRUE,
           at = c(0.6,1.6),
           add = TRUE)


par(oldpar)



par(mfrow=c(1,2))
par(bg = "ivory")


boxplot(dat$mass~sex, xlab="Sex", ylim = c(0, 50),
        ylab="Body Mass [kg]",boxwex=0.35, main = "Boxplot for the Body Mass", lwd = 2, border=
        ↪  "slategrey", # colour of the box borders
        col = "slategray2", # colour of the inside of the boxes
        col.axis = 'grey20', # colour of the axis numbers
        col.lab = 'grey20', # colour of the axis labels
        frame = F)
stripchart(dat$mass~sex,
           method = "jitter",  main = "method = 'jitter', jitter = 0.2",
           pch = 16, # specify the type of point to use
           cex = 1,
           col = c(2,4),
           vertical = TRUE,
           at = c(0.6,1.6),
           add = TRUE)


boxplot(dat$mass~density, xlab="Population Density at Birth", ylim = c(0, 50),
        ylab="Body Mass [kg]",boxwex=0.35, main = "Boxplot for the Body Mass", lwd = 2, border=
        ↪  "slategrey", # colour of the box borders
        col = "slategray2", # colour of the inside of the boxes
```

```
        col.axis = 'grey20', # colour of the axis numbers
        col.lab = 'grey20', # colour of the axis labels
        frame = F)
stripchart(dat$mass~density,
           method = "jitter",  main = "method = 'jitter', jitter = 0.2",
           pch = 16, # specify the type of point to use
           cex = 1,
           col = c(6,7),
           vertical = TRUE,
           at = c(0.6,1.6),
           add = TRUE)
par(oldpar)
##############################################################################
# STARTING THE FIRST 2-WAY ANOVA FOR HORNS LENGTH
##############################################################################

means = tapply(mean_horns_lenght, list(sex, density), mean)
ses = tapply(mean_horns_lenght,
             list(sex, density),
             function(x) sd(x)/sqrt(sum(!is.na(x))))
means
ses


m = lm(mean_horns_lenght~sex*density)
anova(m)
summary(m)
colMeans(means)
rowMeans(means)


##############################################################################
# SECOND 2-WAY ANOVA FOR BODY MASS
##############################################################################

means2 = tapply(dat$mass, list(sex, density), mean)
ses2 = tapply(dat$mass,
             list(sex, density),
             function(x) sd(x)/sqrt(sum(!is.na(x))))
means2
ses2


m3 = lm(dat$mass~sex*density)
anova(m3)
```

```r
summary(m3)
colMeans(means2)
rowMeans(means2)


################################################################################
# PUTTING PLOTS TOGETHER
################################################################################

par(mfrow=c(1,2))
par(bg = "ivory")

plot(c(0.97, 1.03), means[,1], ylim=c(160, 200), xlim=c(0.8, 2.2),
     xlab="Population Density at Birth",
     ylab="Horns Length (Left-Right Mean) [mm]",
     xaxt="n", las=1, pch=c(21,16), col="white")
axis(1, 1:2, labels=c("High", "Low"))
arrows(c(0.97,1.03), means[,1]-ses[,1], c(0.97,1.03),
       means[,1]+ses[,1], length=0.05, angle=90, code=3)
arrows(c(1.97,2.03), means[,2]-ses[,2], c(1.97,2.03),
       means[,2]+ses[,2], length=0.05, angle=90, code=3)
segments(0.97, means[1,1], 1.97, means[1,2], lty=2)
segments(1.03, means[2,1], 2.03, means[2,2])
points(c(0.97, 1.03), means[,1], pch=c(21,16), bg="white")
points(c(1.97, 2.03), means[,2], pch=c(21, 16), bg="white")
legend("topleft", c("Sex", "Female", "Male"),
       bty="n", pch=c(NA,21,16))

plot(c(0.97, 1.03), means2[,1], ylim=c(18, 26), xlim=c(0.8, 2.2),
     xlab="Population Density at Birth",
     ylab="Body Mass [kg]",
     xaxt="n", las=1, pch=c(21,16), col="white")
axis(1, 1:2, labels=c("High", "Low"))
arrows(c(0.97,1.03), means2[,1]-ses2[,1], c(0.97,1.03),
       means2[,1]+ses2[,1], length=0.05, angle=90, code=3)
arrows(c(1.97,2.03), means2[,2]-ses2[,2], c(1.97,2.03),
       means2[,2]+ses2[,2], length=0.05, angle=90, code=3)
segments(0.97, means2[1,1], 1.97, means2[1,2], lty=2)
segments(1.03, means2[2,1], 2.03, means2[2,2])
points(c(0.97, 1.03), means2[,1], pch=c(21,16), bg="white")
points(c(1.97, 2.03), means2[,2], pch=c(21, 16), bg="white")
legend("topleft", c("Sex", "Female", "Male"),
       bty="n", pch=c(NA,21,16))

par(oldpar)
```