Reanalysis of 20-Castagneyrol

Simon Schwab*

18/03/2021

Contents

Reference

Castagneyrol, B., Moreira, X., & Jactel, H. (2018). Drought and plant neighbourhood interactively determine herbivore consumption and performance. Scientific Reports, 8(1), 5930. https://doi.org/10.1038/s41598-018-24299-x

*University of Zurich, simon.schwab@uzh.ch

Notes from reading methods section

- Dependant variable: Relative Consumption Rate (RCR) (log transformed)
- Independent variable:
  - Irrigation (non-irrigated vs irrigated)
  - Composition (B, BG, BP, BQP)
- Covariate: initial weight, w_i
- Design: 2-way 2x4 ANCOVA irrigation x composition (both between) as IV and initial weight as covariate
- N = 96

Reading data

Data is loaded, reshaped if necessary, and factors are specified.

```
PATH = file.path(path.expand("~"), "Data", "ancova") # ancova project folder

d.leaf.lab = read.csv(file.path(PATH, "dataPrimaryStudies", "20-Castagneyrol", "d_leaf_lab.csv"), heade
d.leaf.lab$Composition = factor(d.leaf.lab$Composition, levels = c('B', 'BQ', 'BP', 'BQP'))
dc = read.csv(file.path(PATH, "dataPrimaryStudies", "20-Castagneyrol", "dc.csv"), header = TRUE, sep =
N = 96
```

Preprocessing

Authors preprocessed data to create the outcome variable *Consumption*, and combed the two datasets (code by original author code)

```
## one removed outlier for lava weight
d.leaf.lab$WL.t1[d.leaf.lab$WL.t1 > 0.5] <- NA

d.leaf.lab$Treatment <- as.factor(paste(d.leaf.lab$Irrigation, d.leaf.lab$Composition, sep= '_'))
a <- unlist(lapply(split(d.leaf.lab, d.leaf.lab$Treatment), function(x){coef(lm(WL.t1 ~ SL.t1, x))[[2]]
b <- unlist(lapply(split(d.leaf.lab, d.leaf.lab$Treatment), function(x){coef(lm(WL.t1 ~ SL.t1, x))[[1]]

dl <- summaryBy(d.SL ~ Box.ID + Treatment + Irrigation + Composition, data = d.leaf.lab, FUN = sum, kee

dl$a <- NA
dl$b <- NA

for(i in 1:length(levels(d.leaf.lab$Treatment))){
  dl[dl$Treatment == levels(d.leaf.lab$Treatment)[i], ]$a <- a[i]
  dl[dl$Treatment == levels(d.leaf.lab$Treatment)[i], ]$b <- b[i]
}

dl$Consumption <- with(dl, b + d.SL*a)
dl$Consumption <- dl$Consumption * 1000

dl <- dl[order(dl$Box.ID),]
dc <- dc[order(dc$Box.ID),]
dc$RGR <- with(dc, ((WCf - WCi)/WCi)/8)
# dotchart(sort(dc$RGR))
```

```r
dc$RGR[dc$RGR > 1] <- NA

dcl <- data.frame(
  Box.ID = dl$Box.ID,
  Composition = dl$Composition,
  Irrigation = dl$Irrigation,
  Consumption = dl$Consumption / 8,
  WCi = dc$WCi,
  WCf = dc$WCf,
  RGR = dc$RGR,
  Growth = (dc$WCf - dc$WCi),
  Frass = dc$Frass / 8
)
rm(d.leaf.lab, dl, dc)
```

```r
dcl$Irrigation = as.factor(dcl$Irrigation)
summary(dcl)
```

```
##     Box.ID          Composition     Irrigation  Consumption
##  Length:96          B  :24      Control  :48   Min.   : 2.077
##  Class :character   BQ :24      Irrigated:48   1st Qu.: 5.094
##  Mode  :character   BP :24                     Median : 6.283
##                     BQP:24                     Mean   : 6.864
##                                                3rd Qu.: 8.383
##                                                Max.   :16.777
##
##       WCi                 WCf                 RGR                Growth
##  Min.   :0.00174   Min.   :0.00721   Min.   :-0.09088   Min.   :-0.05809
##  1st Qu.:0.01180   1st Qu.:0.02914   1st Qu.: 0.15852   1st Qu.: 0.01705
##  Median :0.01579   Median :0.04036   Median : 0.21738   Median : 0.02324
##  Mean   :0.01655   Mean   :0.04459   Mean   : 0.22175   Mean   : 0.02804
##  3rd Qu.:0.01978   3rd Qu.:0.05708   3rd Qu.: 0.28699   3rd Qu.: 0.03624
##  Max.   :0.07990   Max.   :0.18420   Max.   : 0.48747   Max.   : 0.17007
##                                      NA's   :2
##      Frass
##  Min.   :0.0005213
##  1st Qu.:0.0026084
##  Median :0.0032900
##  Mean   :0.0038889
##  3rd Qu.:0.0048484
##  Max.   :0.0087312
##
```

Descriptives

Dependant variable

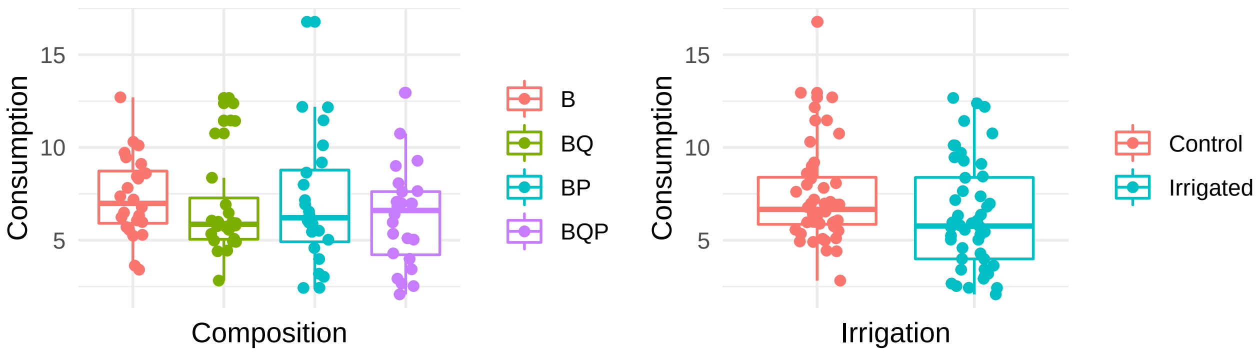
Main effects

```r
p1 = ggplot(dcl, aes(y=Consumption, x=Composition, color=Composition)) +
  geom_boxplot() +
  geom_point(position = position_jitter(width = 0.15, height = 0)) +
```

```
  theme_minimal() +
  theme(axis.text.x = element_blank(), legend.title = element_blank())

p2 = ggplot(dcl, aes(y=Consumption, x=Irrigation, color=Irrigation)) +
  geom_boxplot() +
  geom_point(position = position_jitter(width = 0.15, height = 0)) +
  theme_minimal() +
  theme(axis.text.x = element_blank(), legend.title = element_blank())

plot_grid(p1, p2, nrow=1, ncol=2)
```

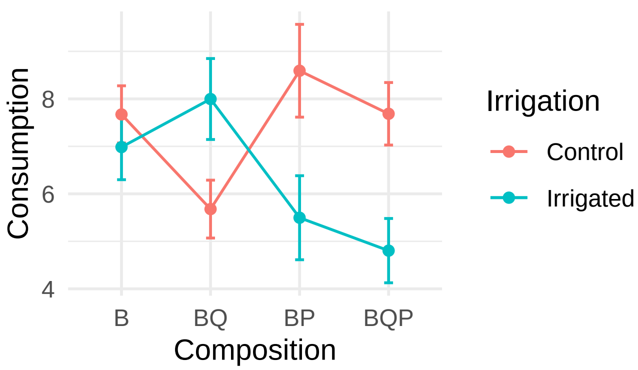


Interaction effects (two-way)

Mean and SE along the 4x2 factorial design, see Figure 1A.

```
d = aggregate(Consumption ~ Composition*Irrigation, data = dcl, FUN = mean)
d$sd = aggregate(Consumption ~ Composition*Irrigation, data = dcl, FUN = sd)[,3]
d$se = d$sd/sqrt(N/4/2)

ggplot(d, aes(y=Consumption, x=Composition, group=Irrigation, color=Irrigation)) +
  geom_errorbar(aes(ymin=Consumption-se, ymax=Consumption+se), width=.1) +
  geom_line() + geom_point() +
  theme_minimal()
```



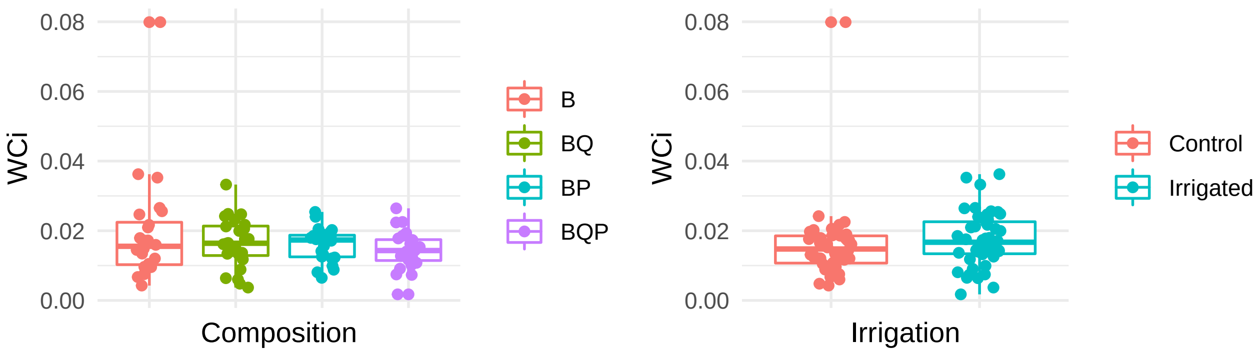
Covariate(s)

```
p1 = ggplot(dcl, aes(y=WCi, x=Composition, color=Composition)) +
  geom_boxplot() +
```

```
  geom_point(position = position_jitter(width = 0.15, height = 0)) +
  theme_minimal() +
  theme(axis.text.x = element_blank(), legend.title = element_blank())

p2 = ggplot(dcl, aes(y=WCi, x=Irrigation, color=Irrigation)) +
  geom_boxplot() +
  geom_point(position = position_jitter(width = 0.15, height = 0)) +
  theme_minimal() +
  theme(axis.text.x = element_blank(), legend.title = element_blank())

plot_grid(p1, p2, nrow=1, ncol=2)
```



Exclusion of outlier

Largest value in the covariate WCi was excluded, but not declared in the paper.

```
dcl.orig = dcl
dcl = subset(dcl.orig, subset = WCi < 0.07)
```

Main analysis ANCOVA

ANCOVA

```
# Orthogonal contrasts
contrasts(dcl$Composition) = contr.helmert(4)
contrasts(dcl$Irrigation)  = contr.helmert(2)

fit.ancova = aov(log(Consumption) ~ WCi + Irrigation * Composition, data = dcl)
# result.ancova = summary(fit.ancova) # Type I
result.ancova = Anova(fit.ancova, type=3) # Type III
print(result.ancova)
```

```
## Anova Table (Type III tests)
##
## Response: log(Consumption)
##                         Sum Sq Df  F value    Pr(>F)
## (Intercept)            19.6766  1 206.0616 < 2.2e-16 ***
## WCi                     4.3487  1  45.5411 1.655e-09 ***
## Irrigation              2.4097  1  25.2354 2.723e-06 ***
## Composition             0.2550  3   0.8900    0.4497
```

```
## Irrigation:Composition  2.5001  3   8.7273 4.066e-05 ***
## Residuals               8.2120 86
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Regression

In the paper, a regression with $lm()$ was used which corresponds to an ANCOVA type I SS. The interaction gives the same result, but main effects and CV are different in type III SS.

```
fit.lm = lm(log(Consumption) ~ WCi + Irrigation * Composition, data = dcl)
result.lm = anova(fit.lm) # Type I SS
# result.lm = Anova(fit.lm, type=3) # Type III SS
print(result.lm)
```

```
## Analysis of Variance Table
##
## Response: log(Consumption)
##                        Df Sum Sq Mean Sq F value    Pr(>F)
## WCi                     1 3.9028  3.9028 40.8719 8.096e-09 ***
## Irrigation              1 2.4378  2.4378 25.5293 2.421e-06 ***
## Composition             3 0.2396  0.0799  0.8364    0.4775
## Irrigation:Composition  3 2.5001  0.8334  8.7273 4.066e-05 ***
## Residuals              86 8.2120  0.0955
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Comparing ANCOVA in original study with reanalysis

Independent variables

Main effect irrigation

```
tab.IV = rbind(stats.orig.IV.irrigation, stats.rep.IV.irrigation, stats.rep.IV.irrigation.lm)
rownames(tab.IV) = c("original Study", "reanalysis type 3 SS", "reanalysis type 1 SS")
print(t(tab.IV))
```

```
##         original Study reanalysis type 3 SS reanalysis type 1 SS
## Fvalue  "25.55"        "25.24"              "25.53"
## df1     "1"            "1"                  "1"
## df2     "86"           "86"                 "86"
## pvalue  "< 0.001"      "< 0.0001"           "< 0.0001"
## MD      NA             NA                   NA
## lowerCI NA             NA                   NA
## upperCI NA             NA                   NA
```

Main effect composition

```
tab.IV = rbind(stats.orig.IV.composition, stats.rep.IV.composition, stats.rep.IV.composition.lm)
rownames(tab.IV) = c("original Study", "reanalysis type 3 SS", "reanalysis type 1 SS")
print(t(tab.IV))
```

```
##         original Study reanalysis type 3 SS reanalysis type 1 SS
## Fvalue  "0.83"         "0.89"               "0.84"
```

```
## df1     "3"            "3"                  "3"
## df2     "86"           "86"                 "86"
## pvalue  "0.48"         "0.45"               "0.48"
## MD      NA             NA                   NA
## lowerCI NA             NA                   NA
## upperCI NA             NA                   NA
```

Interaction

```
tab.IV = rbind(stats.orig.IV.interaction, stats.rep.IV.interaction, stats.rep.IV.interaction.lm)
rownames(tab.IV) = c("original Study", "reanalysis type 3 SS", "reanalysis type 1 SS")
print(t(tab.IV))
```

```
##         original Study reanalysis type 3 SS reanalysis type 1 SS
## Fvalue  "8.66"         "8.73"               "8.73"
## df1     "3"            "3"                  "3"
## df2     "86"           "86"                 "86"
## pvalue  "< 0.001"      "< 0.0001"           "< 0.0001"
## MD      NA             NA                   NA
## lowerCI NA             NA                   NA
## upperCI NA             NA                   NA
```

Covariate

```
tab.CV = rbind(stats.orig.CV, stats.rep.CV, stats.rep.CV.lm)
rownames(tab.CV) = c("original Study", "reanalysis type 3 SS", "reanalysis type 1 SS")
print(t(tab.CV))
```

```
##         original Study reanalysis type 3 SS reanalysis type 1 SS
## Fvalue  "40.77"        "45.54"              "40.87"
## df1     "1"            "1"                  "1"
## df2     "86"           "86"                 "86"
## pvalue  "< 0.001"      "< 0.0001"           "< 0.0001"
## MD      NA             NA                   NA
## lowerCI NA             NA                   NA
## upperCI NA             NA                   NA
```

Assumptions

1. Homogeneity of variance

- ANOVA/ANCOVA is fairly robust in terms of the error rate when sample sizes are equal.
- When groups with larger sample sizes have larger variances than the groups with smaller sample sizes, the resulting F-ratio tends to be conservative. That is, it's more likely to produce a non-significant result when a genuine difference does exist in the population.
- Conversely, when the groups with larger sample sizes have smaller variances than the groups with smaller samples sizes, the resulting F-ratio tends to be liberal and can inflate the false positive rate.

```
tapply(dcl$Consumption, dcl$Irrigation, sd)
```

```
##   Control Irrigated
##  2.691209  2.904900
```

```
leveneTest(Consumption ~ Irrigation, data = dcl)
```

```
## Levene's Test for Homogeneity of Variance (center = median)
##       Df F value Pr(>F)
## group  1  0.9261 0.3384
##       93
```

```
tapply(dcl$Consumption, dcl$Composition, sd)
```

```
##        B       BQ       BP      BQP
## 2.253369 2.774892 3.531163 2.698549
```

```
leveneTest(Consumption ~ Composition, data = dcl)
```

```
## Levene's Test for Homogeneity of Variance (center = median)
##       Df F value Pr(>F)
## group  3   0.757 0.5211
##       91
```

2. Independence between covariate and IV

When the covariate and the experimental effect (independent variable) are not independent the treatment effect is obscured, spurious treatment effects can arise and the interpretation of the ANCOVA is seriously compromised.

We test whether our groups differ on the CV. If the groups do not significantly differ then is appropriate to use the covariate.

```
fit.cv = aov(WCi ~ Irrigation, data = dcl)
Anova(fit.cv, type=3)
```

```
## Anova Table (Type III tests)
##
## Response: WCi
##                Sum Sq Df  F value  Pr(>F)
## (Intercept) 0.0239163  1 570.2028 < 2e-16 ***
## Irrigation  0.0002358  1   5.6207 0.01981 *
## Residuals   0.0039007 93
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
fit.cv = aov(WCi ~ Composition, data = dcl)
Anova(fit.cv, type=3)
```

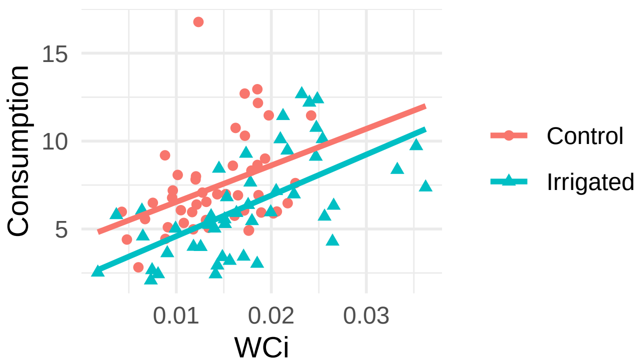
```
## Anova Table (Type III tests)
##
## Response: WCi
##                Sum Sq Df  F value Pr(>F)
## (Intercept) 0.0239896  1 538.0759 <2e-16 ***
## Composition 0.0000793  3   0.5932  0.621
## Residuals   0.0040571 91
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

3. Homogeneity of regression slopes

```
fit.hrs = aov(Consumption ~ WCi*Irrigation, data = dcl)
Anova(fit.hrs, type=3)
```

```
## Anova Table (Type III tests)
##
## Response: Consumption
##                 Sum Sq Df F value    Pr(>F)
## (Intercept)     134.42  1 23.0289 6.223e-06 ***
## WCi             152.76  1 26.1712 1.725e-06 ***
## Irrigation       14.20  1  2.4331    0.1223
## WCi:Irrigation    0.46  1  0.0784    0.7801
## Residuals       531.17 91
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

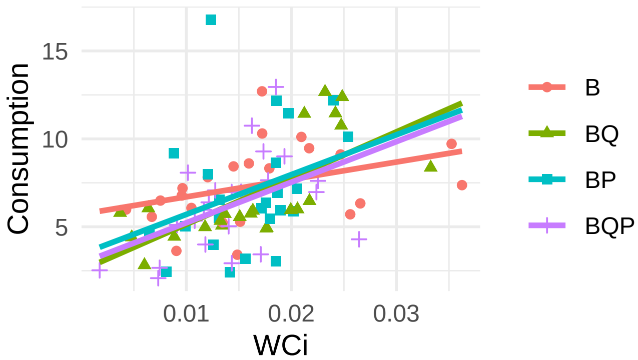
```
ggplot(dcl, aes(y=Consumption, x=WCi, color=Irrigation, shape=Irrigation)) +
  geom_point() +
  geom_smooth(formula = y ~ x,method=lm, se=FALSE, fullrange=TRUE) +
  theme_minimal() +
  theme(legend.title = element_blank())
```




```
fit.hrs = aov(Consumption ~ WCi*Composition, data = dcl)
Anova(fit.hrs, type=3)
```

```
## Anova Table (Type III tests)
##
## Response: Consumption
##                  Sum Sq Df F value    Pr(>F)
## (Intercept)      160.17  1 24.1137 4.215e-06 ***
## WCi              142.24  1 21.4147 1.282e-05 ***
## Composition       25.36  3  1.2729    0.2887
## WCi:Composition   21.54  3  1.0811    0.3614
## Residuals        577.88 87
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
ggplot(dcl, aes(y=Consumption, x=WCi, color=Composition, shape=Composition)) +
  geom_point() +
  geom_smooth(formula = y ~ x,method=lm, se=FALSE, fullrange=TRUE) +
  theme_minimal() +
  theme(legend.title = element_blank())
```



Notes

- We could reproduce the first reported ANCOVA, but had to exclude a single outlier in the covariate which was not clearly declared in the methods section.
- Model was clearly specified and F-values, p-values and Dfs were all reported.
- Type I SS was used, but no large differences compared to type 3 SS.
- Altogether 5 outcome variables and 5 ANCOVAs performed, but not adjusted for multiplicity.
- Assumptions of homogeneity of variances was met.
- Assumption of independence of IV and CV was only met for one IV but not for the other.
- Assumption of homogeneity of regression slopes was met (also thanks to the exclusion of the outlier!).

Data was analyzed according to recommendations by Field, Miles, & Field (2012).