13

Bernat Chiva

12/6/2020

```r
library(car)
```

```
## Loading required package: carData
```

```r
library(HH)
```

```
## Loading required package: lattice
## Loading required package: grid
## Loading required package: latticeExtra
## Loading required package: multcomp
## Loading required package: mvtnorm
## Loading required package: survival
## Loading required package: TH.data
## Loading required package: MASS
##
## Attaching package: 'TH.data'
## The following object is masked from 'package:MASS':
##
##     geyser
## Loading required package: gridExtra
##
## Attaching package: 'HH'
## The following objects are masked from 'package:car':
##
##     logit, vif
```

```r
library(tables)
library(RcmdrMisc)
```

```
## Loading required package: sandwich
```

```r
library(doBy)
library(emmeans)
```

```
##
## Attaching package: 'emmeans'

## The following object is masked from 'package:HH':
##
##     as.glht

library(multcompView)
library(readr)
wheat <- read_delim("C:/Users/berna/OneDrive/Desktop/UPC/S1/5. Models
Lineals/datasets/wheat.csv",
    ";", escape_double = FALSE, locale = locale(decimal_mark = ","),
    trim_ws = TRUE)

##
## -- Column specification --------------------------------------------------
------
## cols(
##   VAR = col_character(),
##   W2 = col_double(),
##   PRES = col_character()
## )

wheat$VAR<-as.factor(wheat$VAR)
wheat$PRES<-as.factor(wheat$PRES)

summary(wheat)

##      VAR           W2       PRES
##  ALCOTA:28   Min.   :-8.0   no:207
##  ASTRAL:67   1st Qu.:18.4   si: 34
##  CAJEME: 9   Median :32.6
##  MARIUS:83   Mean   :32.0
##  RINCON: 7   3rd Qu.:44.6
##  SOISSO:31   Max.   :83.6
##  YECORA:16
```

1. Interaction model:

We assume this model: $Y_{ij} = \beta_0 + \beta_1 * F1_i + \beta_2 * F2_j + \beta_3 * F1_i * F2_j$ for our data. With three parameters, one of them the interaction.

```
mod1<-lm(W2~VAR*PRES, wheat)
```

```
summary(mod1)
```

```
##
## Call:
## lm(formula = W2 ~ VAR * PRES, data = wheat)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -39.100 -10.474   0.854  10.320  46.268
##
## Coefficients:
##                   Estimate Std. Error t value Pr(>|t|)
## (Intercept)         31.474      3.481   9.041  < 2e-16 ***
## VARASTRAL            5.858      4.094   1.431 0.153890
## VARCAJEME           36.059     10.248   3.519 0.000524 ***
## VARMARIUS           -2.744      3.973  -0.691 0.490549
## VARRINCON          -20.474      8.238  -2.485 0.013664 *
## VARSOISSO           -9.328      4.779  -1.952 0.052185 .
## VARYECORA           -2.988      5.659  -0.528 0.597997
## PRESsi              -5.274      8.238  -0.640 0.522683
## VARASTRAL:PRESsi     3.171     10.598   0.299 0.765075
## VARCAJEME:PRESsi     3.691     14.395   0.256 0.797892
## VARMARIUS:PRESsi    18.644     10.552   1.767 0.078601 .
## VARRINCON:PRESsi    20.274     16.216   1.250 0.212503
## VARSOISSO:PRESsi     5.508     11.590   0.475 0.635085
## VARYECORA:PRESsi    27.138     15.071   1.801 0.073075 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 16.69 on 227 degrees of freedom
## Multiple R-squared:  0.2429, Adjusted R-squared:  0.1995
## F-statistic: 5.601 on 13 and 227 DF,  p-value: 7.667e-09
```

```
CLD(emmeans(mod1,~VAR),Letters=letters,reversed=T)
```

```
## Warning: 'CLD' will be deprecated. Its use is discouraged.
## See '?cld.emmGrid' for an explanation. Use 'pwpp' or 'multcomp::cld'
instead.

## NOTE: Results may be misleading due to involvement in interactions
```

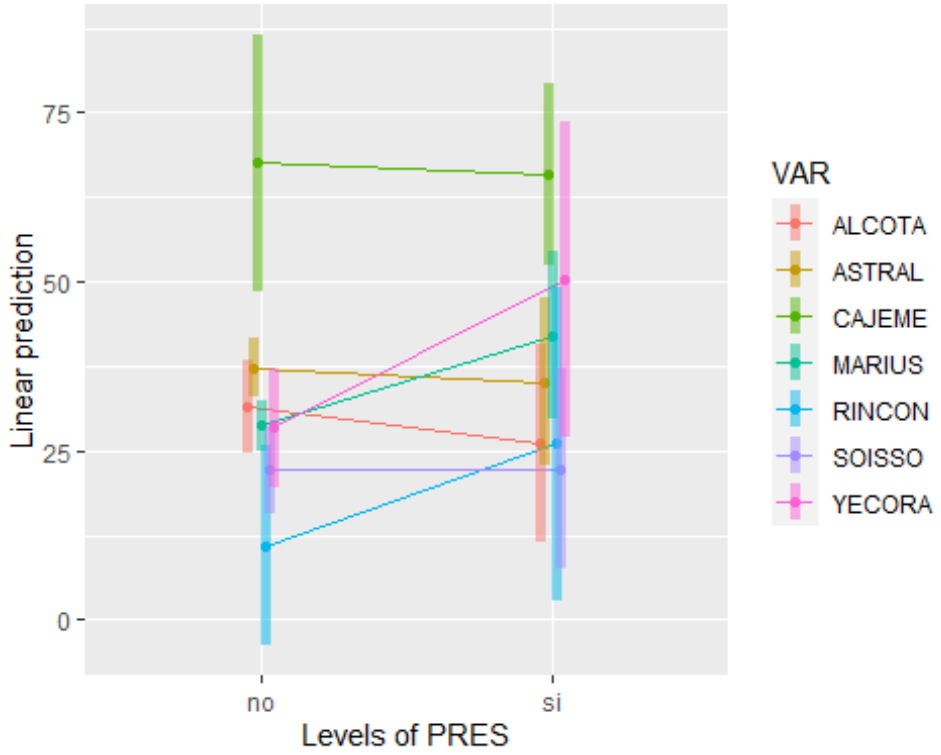
```
##  VAR    emmean   SE  df lower.CL upper.CL .group
##  CAJEME   66.7 5.90 227    55.11     78.4  a
##  YECORA   39.4 6.31 227    26.98     51.9   b
##  ASTRAL   36.3 3.33 227    29.71     42.8   b
##  MARIUS   35.4 3.30 227    28.92     41.9   b
##  ALCOTA   28.8 4.12 227    20.72     37.0   b
##  SOISSO   22.3 4.08 227    14.23     30.3   b
##  RINCON   18.5 6.98 227     4.74     32.3   b
##
## Results are averaged over the levels of: PRES
## Confidence level used: 0.95
## P value adjustment: tukey method for comparing a family of 7 estimates
## significance level used: alpha = 0.05
```

```
CLD(emmeans(mod1,~VAR*PRES),Letters=letters,reversed=T)
```

```
## Warning: 'CLD' will be deprecated. Its use is discouraged.
## See '?cld.emmGrid' for an explanation. Use 'pwpp' or 'multcomp::cld'
instead.
```

```
##  VAR    PRES emmean    SE  df lower.CL upper.CL .group
##  CAJEME no     67.5  9.64 227    48.54     86.5  ab
##  CAJEME si     66.0  6.82 227    52.52     79.4  a
##  YECORA si     50.4 11.81 227    27.09     73.6  abcd
##  MARIUS si     42.1  6.31 227    29.67     54.5  abcd
##  ASTRAL no     37.3  2.16 227    33.08     41.6   bc
##  ASTRAL si     35.2  6.31 227    22.79     47.7  abcd
##  ALCOTA no     31.5  3.48 227    24.61     38.3    cd
##  MARIUS no     28.7  1.92 227    24.96     32.5    cd
##  YECORA no     28.5  4.46 227    19.69     37.3    cd
##  ALCOTA si     26.2  7.47 227    11.49     40.9   bcd
##  RINCON si     26.0 11.81 227     2.74     49.3  abcd
##  SOISSO si     22.4  7.47 227     7.67     37.1    cd
##  SOISSO no     22.1  3.27 227    15.69     28.6     d
##  RINCON no     11.0  7.47 227    -3.71     25.7    cd
##
## Confidence level used: 0.95
## P value adjustment: tukey method for comparing a family of 14 estimates
## significance level used: alpha = 0.05
```

```
(emmip(mod1,VAR~PRES,CIs=TRUE))
```




```r
#type I ss
anova(mod1)
```

```
## Analysis of Variance Table
##
## Response: W2
##            Df Sum Sq Mean Sq F value    Pr(>F)
## VAR         6  17843 2973.81 10.6697 1.932e-10 ***
## PRES        1    449  448.86  1.6104    0.2057
## VAR:PRES    6   2002  333.68  1.1972    0.3086
## Residuals 227  63269  278.72
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
#type III ss
Anova(mod1, type='III')
```

```
## Anova Table (Type III tests)
##
## Response: W2
##             Sum Sq  Df F value    Pr(>F)
## (Intercept)  22784   1 81.7460 < 2.2e-16 ***
## VAR          10926   6  6.5335 2.228e-06 ***
## PRES           114   1  0.4099    0.5227
## VAR:PRES      2002   6  1.1972    0.3086
## Residuals    63269 227
```

```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

a) Assuming that the ANOVA hypothesis are verified, define a factorial model taking into account the factors, variety and presence of insects. Based on the analysis answer the following questions justifying the response:

• Is the flour degradation affected by the wheat type and the presence of insects?

From the anova table we can assure that the wheat type influences the flour degradation, but the presences of insects has no effect, its coefficient has no significance and is similar to 0 value.

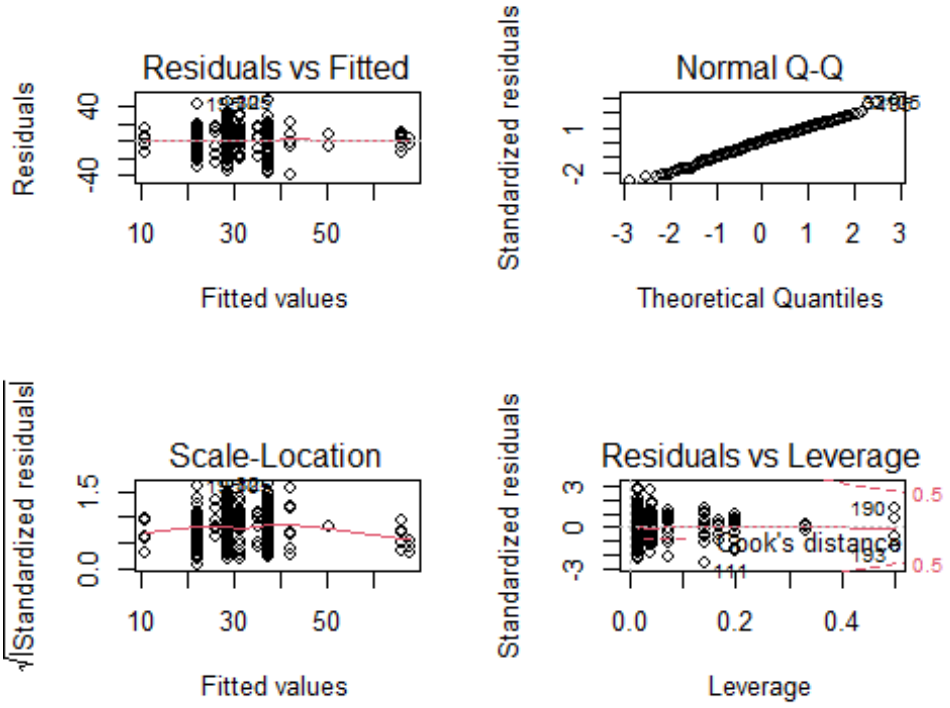
• Compare the results obtained by considering the type I and type III sums of squares. Does the factors order have any influence in the results?

The results from the two types of anova are the same, the only significant variable that affects the flour degradation is the wheat type. The order has no influence on the results in this case.

• In this case, which is the role played by the interaction? Could we avoid the interaction term?

The interaction term has no significance, meaning that is similar to 0. We could suppress it from our model and fit an additive one.

```
#ANOVA assumptions
oldpar <- par( mfrow=c(2,2))
plot(mod1, ask=F)
```




```
par(oldpar)
```

b) Do you think that the ANOVA hypothesis are verified? Justify your answer theoretically and by means of plots.

As we can see in the residuals vs fitted plot, the variance is different among the levels. And although normality can be assume, not homoscedasticity.

2. Additive model:

We assume this model: $Y_{ij} = \beta_0 + \beta_1 * F1_i + \beta_2 * F2_j$. With two parameter and no interaction.

```
#additive model
mod2<-lm(W2~VAR+PRES, wheat)
summary(mod2)
```

```
##
## Call:
## lm(formula = W2 ~ VAR + PRES, data = wheat)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -37.008 -11.516   0.842  10.498  46.921
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  29.7924     3.2166   9.262  < 2e-16 ***
## VARASTRAL     6.8867     3.7742   1.825   0.0693 .
## VARCAJEME    33.9237     6.6092   5.133 6.02e-07 ***
## VARMARIUS    -0.2840     3.6708  -0.077   0.9384
## VARRINCON   -15.6903     7.0814  -2.216   0.0277 *
## VARSOISSO    -8.2767     4.3640  -1.897   0.0591 .
## VARYECORA     0.9085     5.2482   0.173   0.8627
## PRESsi        4.1424     3.2725   1.266   0.2068
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 16.74 on 233 degrees of freedom
## Multiple R-squared:  0.2189, Adjusted R-squared:  0.1954
## F-statistic: 9.328 on 7 and 233 DF,  p-value: 3.504e-10
```

```
anova(mod2)
```

```
## Analysis of Variance Table
##
## Response: W2
##            Df Sum Sq Mean Sq F value    Pr(>F)
## VAR         6  17843 2973.81 10.6158 2.024e-10 ***
## PRES        1    449  448.86  1.6023    0.2068
## Residuals 233  65271  280.13
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
Anova(mod2, type = 'III')
```

```
## Anova Table (Type III tests)
##
## Response: W2
```

```
##             Sum Sq  Df F value    Pr(>F)
## (Intercept)  24032   1 85.7887 < 2.2e-16 ***
## VAR          16243   6  9.6641 1.688e-09 ***
## PRES           449   1  1.6023    0.2068
## Residuals    65271 233
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

a) Assuming that the ANOVA hypothesis are verified, define a factorial model taking into account the factors, variety and presence of insects. Based on the analysis answer the following questions justifying the response:

• Is the flour degradation affected by the wheat type?

Both in the type I and type III anova table, the wheat type affects the flour degradation.

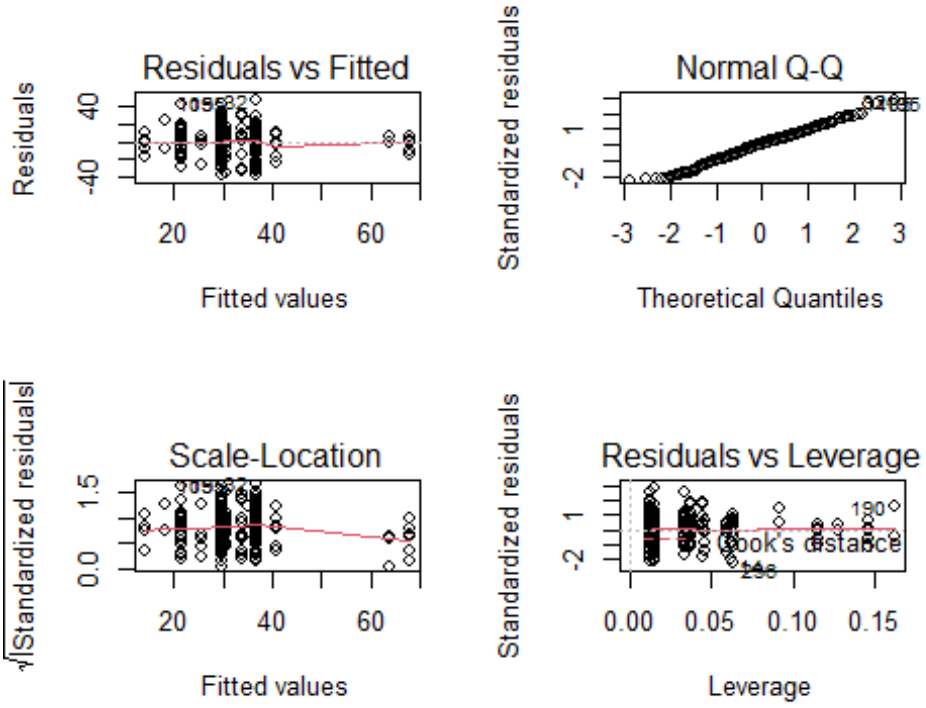
• Is the flour degradation affected by the presence of insects?

Whilst on the type I anova table the presence of insects has no effect on the flour degradation, when we take into account the order, the type III anova, we can see how the presence of insect actually has some effect.

• Compare the results obtained by considering the type I and type III sums of squares. Does the factors order have any influence in the results?

Yes, the results from both anova analysis are different. In the first type only the wheat type is significant, while on the type III anova both variables, wheat type and presence of insects are significant. This means that the order of the factors have an influence in the results, an influence that is masked in the type I anova.

```
#ANOVA assumptions
oldpar <- par( mfrow=c(2,2))
plot(mod2, ask=F)
```




```
par(oldpar)
```

b) Do you think that the ANOVA hypothesis are verified? Justify your answer theoretically and by means of plots.

We can accept the normality and independence properties of the errors, but homoscedasticity seems doubtful. The residuals vs fitted plot varies too much.