Pedagogical

Jordi Valero and Marta Pérez-Casany

12 de noviembre de 2018

```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: RColorBrewer
## 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(car)
library(emmeans)
```

```
##
## Attaching package: 'emmeans'
## The following object is masked from 'package:HH':
##
##     as.glht
## The following object is masked from 'package:multcomp':
##
##     cld
```

```r
library(tables)
```

```
## Loading required package: Hmisc
```

```
## Loading required package: Formula
## Loading required package: ggplot2
##
## Attaching package: 'ggplot2'
## The following object is masked from 'package:latticeExtra':
##
##     layer
##
## Attaching package: 'Hmisc'
## The following objects are masked from 'package:base':
##
##     format.pval, units
```

```
library(RcmdrMisc)
```

```
## Loading required package: sandwich
##
## Attaching package: 'RcmdrMisc'
## The following object is masked from 'package:Hmisc':
##
##     Dotplot
```

ANCOVA

We want to COMPARE TWO PEDAGOGICAL METHODOLOGIES.

Loading the data and printing the top part

```
setwd("G:/PiE2/2018")
pedagogicaldata<-read.csv2("./Dades/compline.csv")
head(pedagogicaldata)
```

```
##   M   C  P PP
## 1 1 110 55 55
## 2 1 121 66 66
## 3 1 108 50 50
## 4 1  95 33 33
## 5 1 107 50 50
## 6 1  89 35 35
```

```
dim(pedagogicaldata)
```

```
## [1] 22  4
```

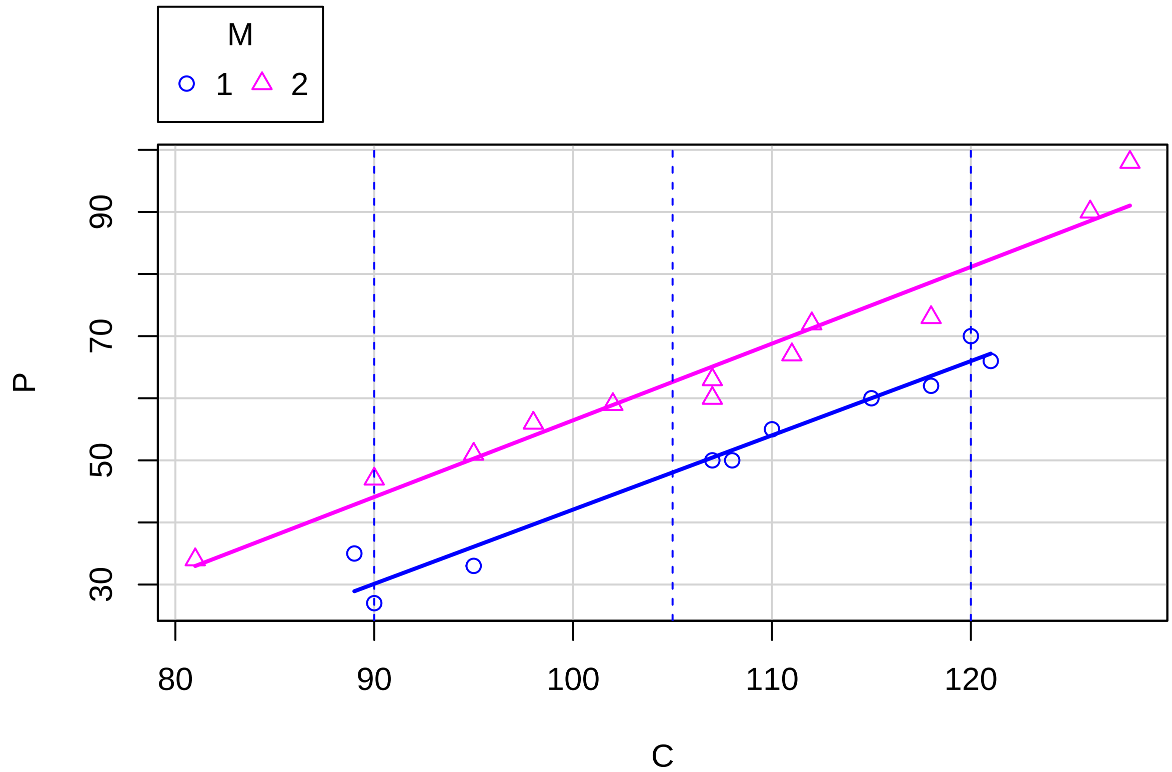
The data set contains 22 rows corresponding to individuals, and four columns. The first column corresponds to the pedagogical methodology followed by the student. The second column contains the coefficient of inteligence of the individual al ther third and fourth columns contain punctuations. we will first work with the column denoted by p.

This type of analysis is called ANCOVA because as explanatory variables we have categorical as well as continuous variables. In this case we have one of each type. Thus, this situation corresponds to the easiest ANCOVA analysis.

Descriptive statistics

We plot the puctuation as a function of the intelligence coefficient, using a different line for each pedagogical methodology.

```
sp(P~C|M,smooth=F,dat=pedagogicaldata)
abline(v=c(90,105,120),lty=2,col="blue")
```



We clearly see better results for the pedagogical method number 2. The puctuation is always larger independetly of the intelligence coefficient. It can be seen that the lines are quite parallel. It seems that both the intelligence coefficient as well as the method will have a significant influence in the puctuation.

Model with interaction

We start fitting a model with interaction. This means that we allow that the influence of the coefficient of inteligence in the puctuation be different for the two methods. If the interaction term is significant this means that the two slopes are statistically different and that the coefficient of inteligence afects differently in the puctuation for the two methods.

A different intercept will be interpreted as that are the two methods turn out with different results.

We first define the mehtod as a categorical variable

```
M<-as.factor(pedagogicaldata$M)
is.factor(M)
```

```
## [1] TRUE
```

```r
mP<-lm(P~M+C+M:C,pedagogicaldata)

summary(mP)
```

```
##
## Call:
## lm(formula = P ~ M + C + M:C, data = pedagogicaldata)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -5.6767 -1.9789  0.0376  1.3367  6.9744
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) -87.94642   22.15259  -3.970 0.000898 ***
## M            10.45276   13.05628   0.801 0.433808
## C             1.15641    0.20549   5.628 2.44e-05 ***
## M:C           0.03924    0.12133   0.323 0.750112
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.474 on 18 degrees of freedom
## Multiple R-squared:  0.966,  Adjusted R-squared:  0.9603
## F-statistic: 170.4 on 3 and 18 DF,  p-value: 2.112e-13
```

We see from this first analysis that the interaction and the method are not significant. Probably the method is not significant because of the interaction term. doing the Anova type III we will see if the method is significant or not.

```r
Anova(mP,ty=3)
```

```
## Anova Table (Type III tests)
##
## Response: P
##             Sum Sq Df F value    Pr(>F)
## (Intercept) 190.17  1 15.7611 0.0008978 ***
## M             7.73  1  0.6409 0.4338082
## C           382.12  1 31.6705 2.437e-05 ***
## M:C           1.26  1  0.1046 0.7501119
## Residuals   217.18 18
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

It turns out again that the method and the interaction are not significant. Let us fit the model without interaction

Model without interaction (additive model)

```r
mP2<-lm(P~M+C,pedagogicaldata)
summary(mP2)
```

```
##
## Call:
## lm(formula = P ~ M + C, data = pedagogicaldata)
```

```
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -5.503 -2.025 -0.088  1.529  7.296
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) -94.76339    6.65199  -14.25 1.36e-11 ***
## M            14.64776    1.45307   10.08 4.62e-09 ***
## C             1.22009    0.05741   21.25 1.05e-14 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.391 on 19 degrees of freedom
## Multiple R-squared:  0.9658, Adjusted R-squared:  0.9622
## F-statistic: 268.2 on 2 and 19 DF,  p-value: 1.186e-14
```

We see that once the interaction is suppressed, both explanatory variables are significant. Thus, we conclude that the puctuation obtained significatively depend on the coefficient of inteligence but also on the method. the fact that the interaction is not significative means that the way in which the coefficient of inteligence affect the puctuation is the same for the two methods. We will have thus two different lines, one for each method with the same slope but different intercept.

Jointly the two variables explain 96% of the variability in the response variable.

```
Anova(mP2,ty=3)
```

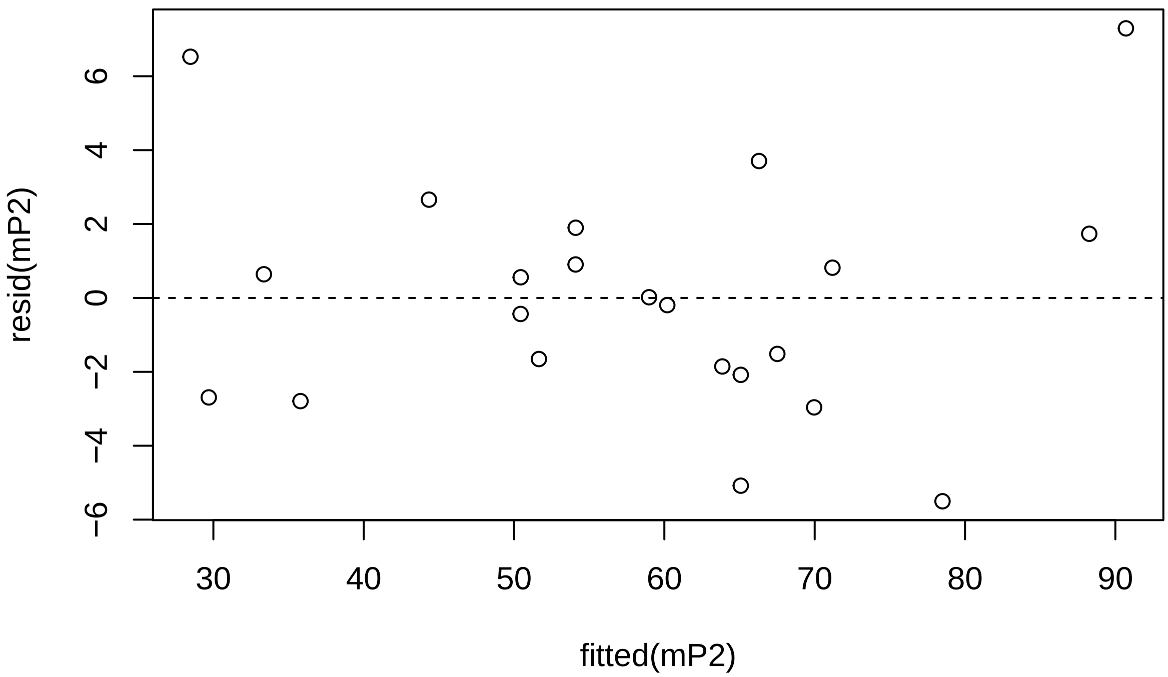
```
## Anova Table (Type III tests)
##
## Response: P
##             Sum Sq Df F value    Pr(>F)
## (Intercept) 2333.3  1  202.94 1.359e-11 ***
## M           1168.3  1  101.62 4.624e-09 ***
## C           5192.8  1  451.67 1.053e-14 ***
## Residuals    218.4 19
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

From the Anova we also see that both explanatory variables have a significant influence in the puctuation.

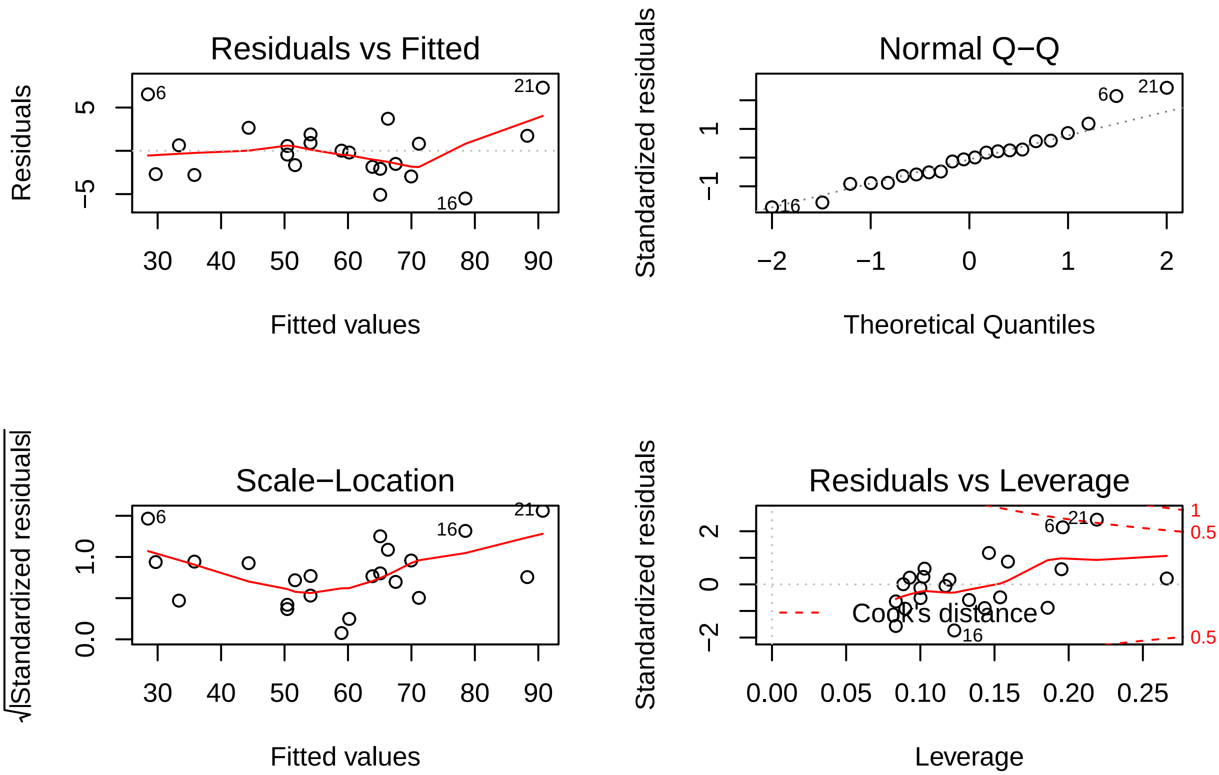
Let us see if the assumptions (hypothesis) of the linear model are satisfied.

Model Diagnostics

```
plot(fitted(mP2),resid(mP2))
abline(h=0,lty=2)
```




```
oldpar <- par( mfrow=c(2,2))
plot(mP2,ask=F)
```




```r
par(oldpar)
```

The Normality, independence and homocedasticity assumptions may be acccepted. Which allow us to accept the additive model as a good model to explain the variability in the puctuation variable.

In what follows we compute the estimated martinal means (emmeans) for the three coefficient of inteligence levels required. Observe that the value of the method appears to be a weighted mean of one and two (the design is not balanced in this case).

```r
(emm2<-emmeans(mP2,~M|C,at=list(C=c(90,105,120))))
```

```
## C =  90:
##         M   emmean        SE df lower.CL upper.CL
##  1.545455 37.68214 1.2019848 19 35.16635 40.19792
##
## C = 105:
##         M   emmean        SE df lower.CL upper.CL
##  1.545455 55.98348 0.7296735 19 54.45626 57.51071
##
## C = 120:
##         M   emmean        SE df lower.CL upper.CL
##  1.545455 74.28483 1.0503334 19 72.08645 76.48320
##
## Confidence level used: 0.95
```

Thus we can predict a puctuation of 37.68 for someone with a coefficient of inteligence equal to $c = 90$ if we do not know which method will atend. The CI for the expected puctuation is equal to $(35.16, 40.19)$. Similarly with the other coefficients.

We leave to the student the analysis of the PP punctuation.