

T1-E04-multi_COL.R

jordi

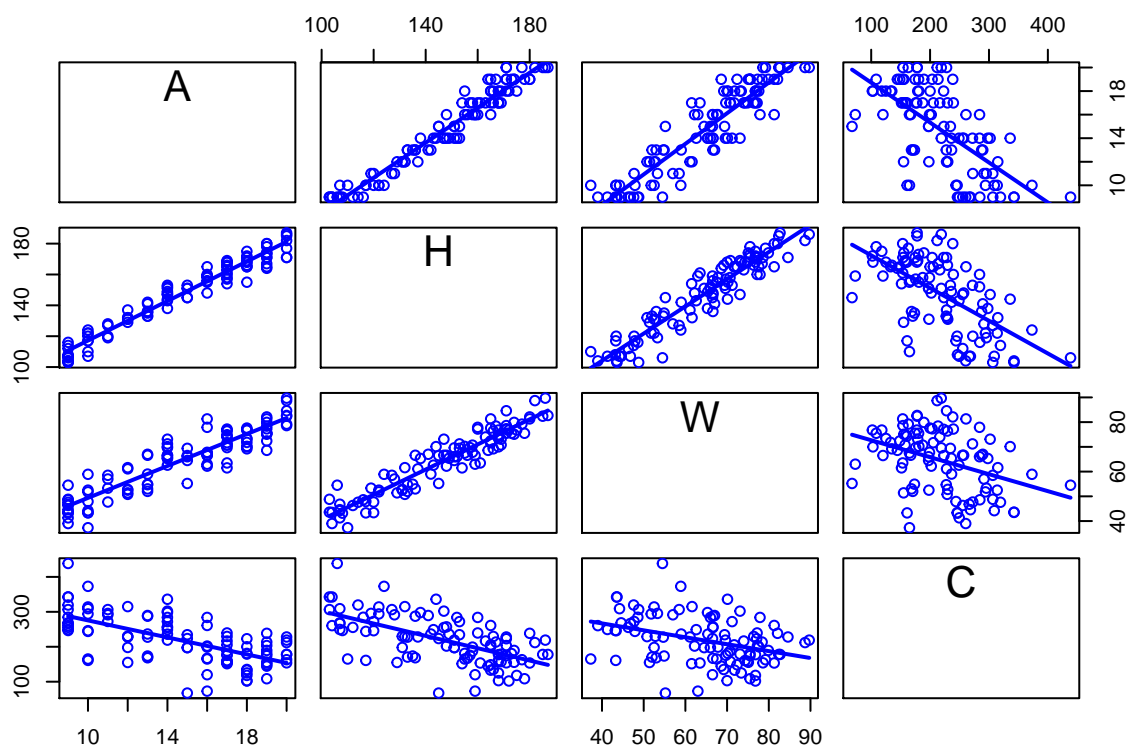
Tue Oct 23 12:21:06 2018

```
setwd("~/Documents/CURS 2018-2019/PIE2")  
#setwd("F:/windows")  
COL <- read.csv2("./Dades/COL.csv")
```

```
p<-4  
n<-dim(COL)[1]  
library(car)
```

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

```
scatterplotMatrix(COL,smooth=F,diagonal=F)
```



```
# MODEL  
mod<-lm(C~W+A+H,COL)  
  
write("-----",",")
```

```
## -----
```

```
write("a), b), c)",",")
```

```
## a), b), c)
```

```
# RESUM DEL MODEL  
summary(mod)
```

```
##
## Call:
## lm(formula = C ~ W + A + H, data = COL)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -74.608 -22.137   1.888  21.156  65.410
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  490.9978    35.0517   14.008 < 2e-16 ***
## W             10.3773     0.7365   14.090 < 2e-16 ***
## A            -13.0195     3.8530   -3.379 0.00105 **
## H             -5.0989     0.7227   -7.055 2.68e-10 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 30.11 on 96 degrees of freedom
## Multiple R-squared:  0.8101, Adjusted R-squared:  0.8041
## F-statistic: 136.5 on 3 and 96 DF,  p-value: < 2.2e-16
#Càlculs opcionals: Intervals de confiança dels paràmetres
confint(mod,level=0.99)

##              0.5 %      99.5 %
## (Intercept) 398.881272 583.114304
## W             8.441792 12.312821
## A            -23.145228 -2.893732
## H            -6.998311 -3.199551

write("-----", "")

## -----
write("\b, d)", "")

## b), d)
#Càlculs opcionals: SS1 Test dels paràmetres amb ordenació predeterminada
anova(mod)

## Analysis of Variance Table
##
## Response: C
##      Df Sum Sq Mean Sq F value    Pr(>F)
## W      1  62396   62396   68.826 6.686e-13 ***
## A      1 263670  263670  290.841 < 2.2e-16 ***
## H      1  45123   45123   49.773 2.676e-10 ***
## Residuals 96  87031     907
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
anova(lm(C~H+W+A,COL))

## Analysis of Variance Table
##
## Response: C
##      Df Sum Sq Mean Sq F value    Pr(>F)
```

```
## H          1 171564 171564 189.244 < 2.2e-16 ***
## W          1 189273 189273 208.778 < 2.2e-16 ***
## A          1 10351 10351 11.418 0.001052 **
## Residuals 96 87031 907
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

#Nota: SS3, els tests (F) sempre coincideixen amb els del resum (t),  $F=t^2$ 
Anova(mod)

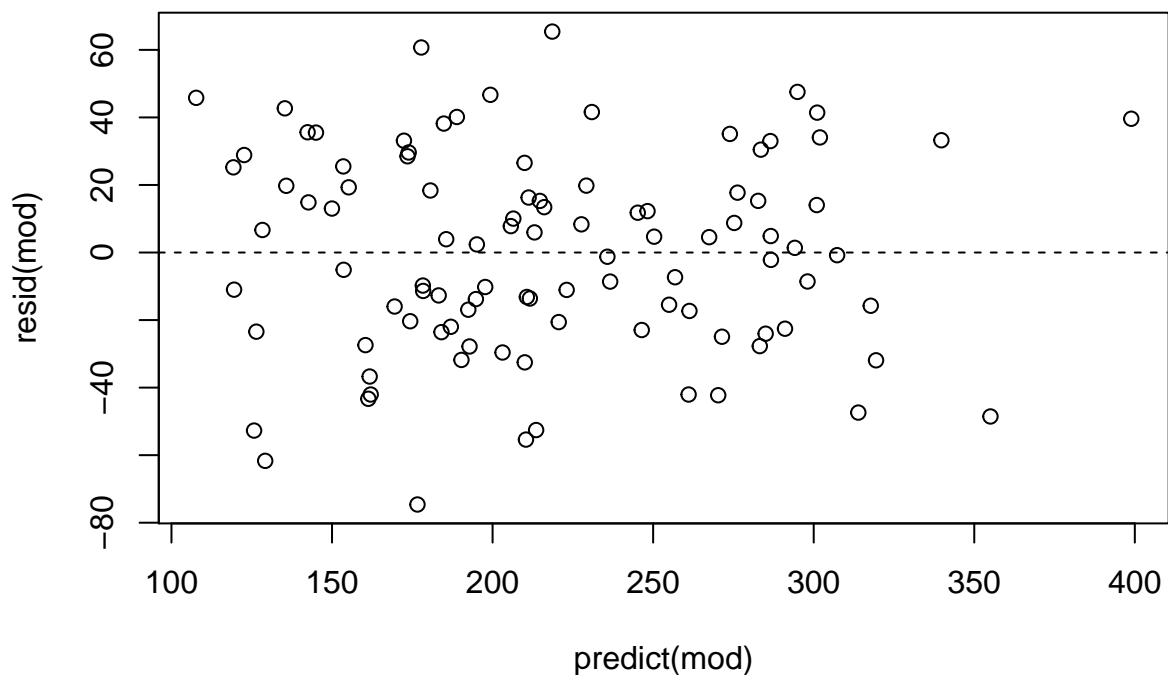
## Anova Table (Type II tests)
##
## Response: C
##          Sum Sq Df F value    Pr(>F)
## W          179985  1 198.533 < 2.2e-16 ***
## A           10351  1  11.418 0.001052 **
## H           45123  1  49.773 2.676e-10 ***
## Residuals  87031 96
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

write("-----", "")

## -----
write("f)", "")

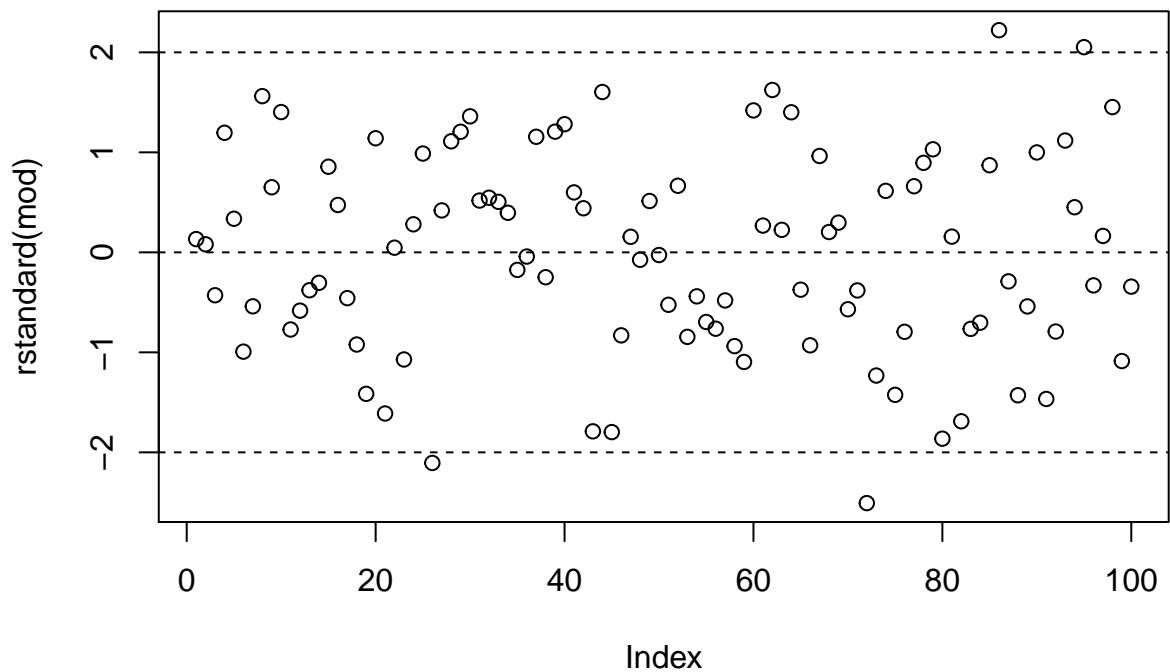
## f)

# Diagnòstic: TENDÈNCIES
plot(predict(mod), resid(mod))
abline(h=0, lty=2)
```



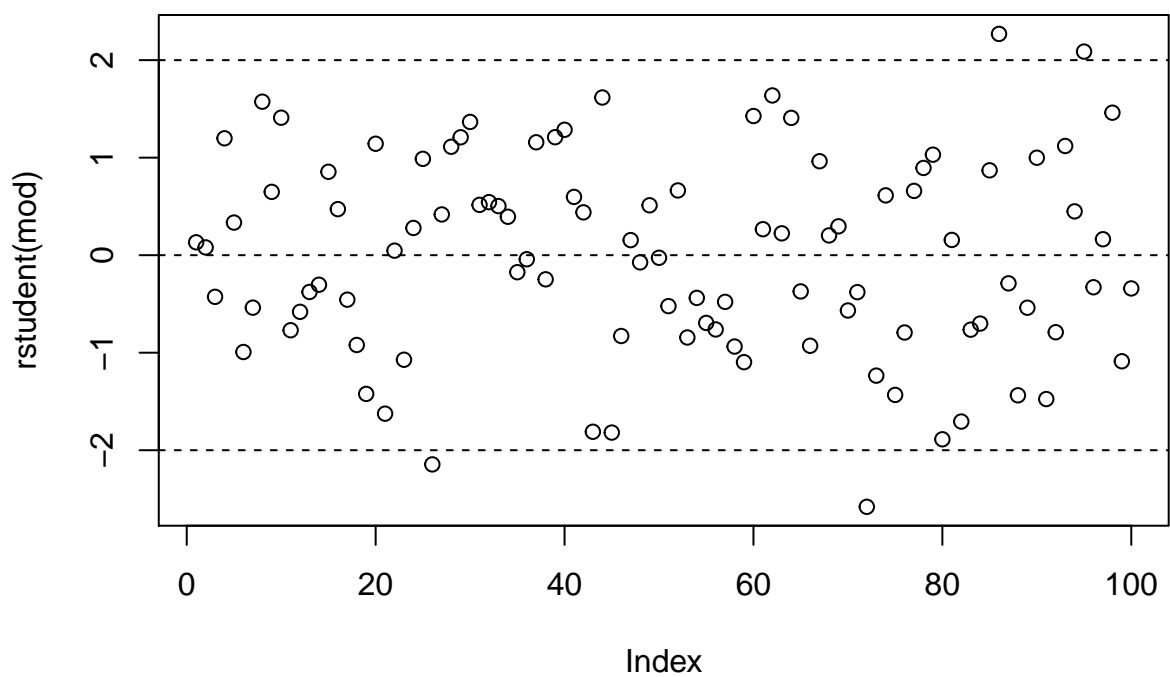
```
# Diagnòstic: OUTLIERS (rstudent)
plot(rstandard(mod))
```

```
abline(h=c(-2,0,2),lty=2)
```

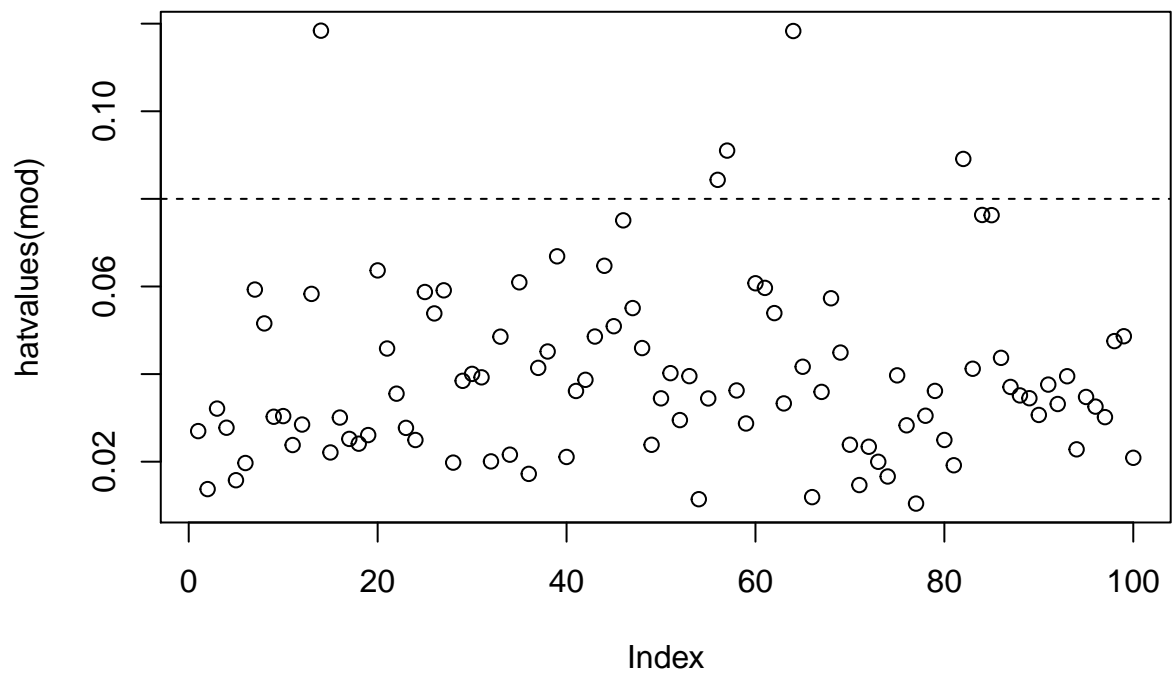


```
plot(rstudent(mod),main="rstudent")
abline(h=c(-2,0,2),lty=2)
```

rstudent



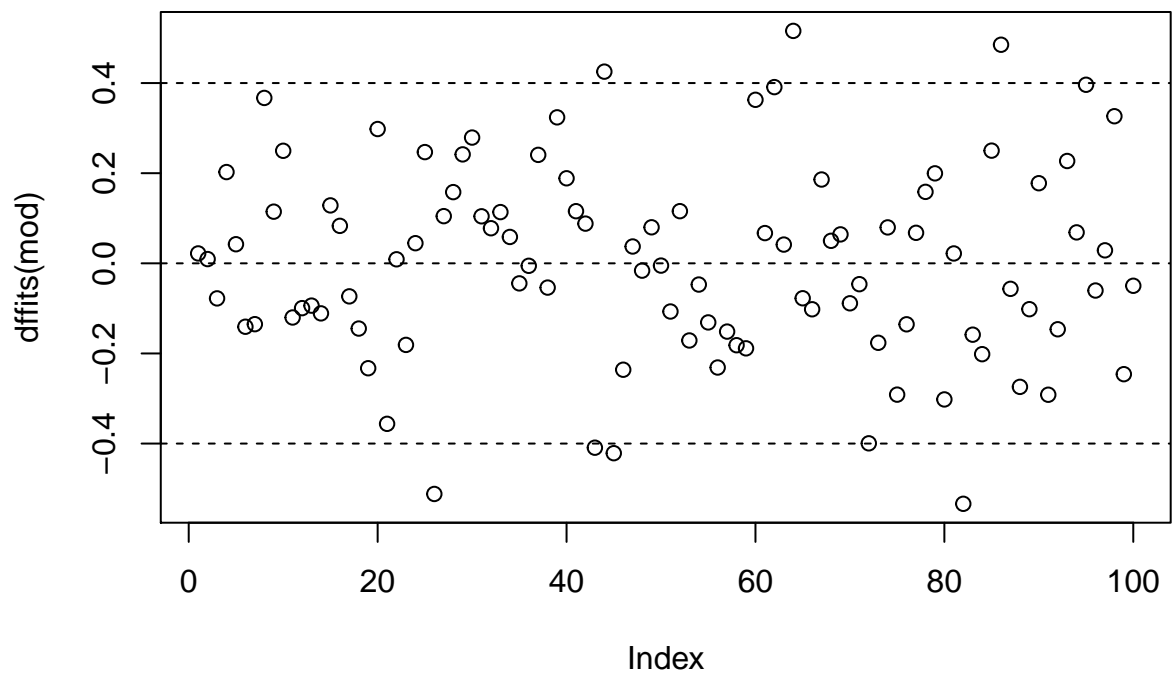
```
# Diagnostic: LEVERAGE
plot(hatvalues(mod))
abline(h=c(0,2*mean(hatvalues(mod))),lty=2)
```



```
# Diagnòstic: INFLUÈNCIA (dffits)

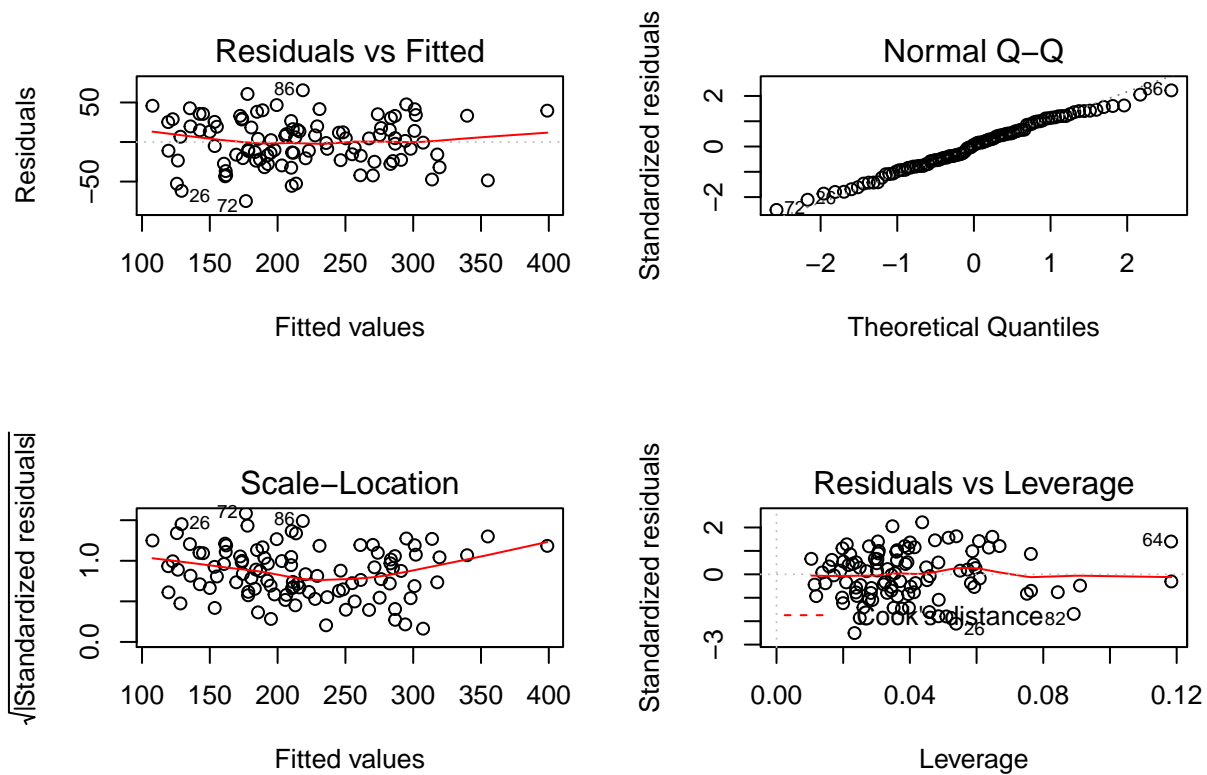
plot(dffits(mod),main="dffits")
abline(h=c(-2*sqrt(p/n),0,2*sqrt(p/n)),lty=2)
```

dffits



```
#plot(cooks.distance(mod))
#abline(h=c(0,4/n),lty=2)
```

```
#Diagnòstics de R
oldpar <- par( mfrow=c(2,2))
plot(mod,ask=F)
```



```
par(oldpar)

write("-----", "")

## -----
write("g)", "")

## g)
#Diagnòstics: Col·linealitat
vif(mod)

##          W          A          H
## 9.489406 20.904776 31.695499

write("-----", "")

## -----
write("e)", "")

## e)
#Càlculs opcionals: Per alguns casos predeterminats, IC de E(Y)
(CO<-data.frame(cbind(W=c(65,75,65),A=c(15,15,12),H=c(150,150,150)), row.names=1:3))

##      W      A      H
## 1  65  15  150
```

```
## 2 75 15 150
## 3 65 12 150
```

```
predict(mod, C0, interval="confidence", level=.95, se.fit=T)
```

```
## $fit
##      fit      lwr      upr
## 1 205.3908 199.1668 211.6148
## 2 309.1639 294.6188 323.7089
## 3 244.4492 219.8210 269.0774
##
## $se.fit
##      1      2      3
## 3.135539 7.327533 12.407261
##
## $df
## [1] 96
##
## $residual.scale
## [1] 30.1094
```

```
#Càlculs opcionals: Per alguns casos predeterminats, IPredicció de Y
predict(mod, C0, interval="prediction", level=.95, se.fit=F)
```

```
##      fit      lwr      upr
## 1 205.3908 145.3009 265.4807
## 2 309.1639 247.6528 370.6749
## 3 244.4492 179.8071 309.0914
```

```
write("-----", "")
```

```
## -----
write("h), i)", "")
```

```
## h), i)
```

```
#Canvis lineals en les variables independents:
#canvi en alguna variable, per exemple, excés de pes,
# pes patró 0.5*H-10, WE=W-(0.5*H-10)
```

```
#COL$WE<-COL$W-0.5*COL$H+10
```

```
summary(mod2<-lm(C~I(W-0.5*H+10)+A+H,COL))
```

```
##
## Call:
## lm(formula = C ~ I(W - 0.5 * H + 10) + A + H, data = COL)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -74.608 -22.137   1.888  21.156  65.410
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    387.22473    33.69605   11.492 < 2e-16 ***
## I(W - 0.5 * H + 10)  10.37731     0.73649   14.090 < 2e-16 ***
## A              -13.01948     3.85300   -3.379  0.00105 **
```

```
## H                0.08972    0.58736    0.153    0.87891
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 30.11 on 96 degrees of freedom
## Multiple R-squared:  0.8101, Adjusted R-squared:  0.8041
## F-statistic: 136.5 on 3 and 96 DF,  p-value: < 2.2e-16
```

```
vif(mod2)
```

```
## I(W - 0.5 * H + 10)          A          H
##          1.009937          20.904776      20.933520
```

```
#Nota: Només canvia algun paràmetre i la col·linealitat
```

```
#Canvis lineals en les variables independents:
```

```
#eliminar alguna variable independent no significativa i/o amb molta col·linealitat
```

```
#per exemple H, si ja s'utilitza l'exces de pes
```

```
summary(mod3<-lm(C~I(W-0.5*H+10)+A,COL))
```

```
##
## Call:
## lm(formula = C ~ I(W - 0.5 * H + 10) + A, data = COL)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -74.286 -22.638   1.755  20.935  66.244
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    391.9885    12.6975   30.87  <2e-16 ***
## I(W - 0.5 * H + 10)  10.3882     0.7294   14.24  <2e-16 ***
## A              -12.4452     0.8387  -14.84  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 29.96 on 97 degrees of freedom
## Multiple R-squared:  0.81, Adjusted R-squared:  0.8061
## F-statistic: 206.8 on 2 and 97 DF,  p-value: < 2.2e-16
```

```
vif(mod3)
```

```
## I(W - 0.5 * H + 10)          A
##          1.000527          1.000527
```

```
COL$WE<-COL$W-0.5*COL$H+10
```

```
scatterplotMatrix(COL,smooth=F,diagonal=F)
```