**Before you proceed:**

Remember that your data files and script of the R Markdown should be in the same working directory. Even if you change the working directory in a code chunk of R Markdown, the working directory will be reset to the original working directory as the code chunk gets executed. By defalut the working directory of a R Markdown script is the directory where you have just saved the script.

```
setwd("~/Linear Models")
getwd()
```

```
## [1] "C:/Users/Ankit/Documents/Linear Models"
```

```
COL <- read.csv2("./COL.csv")
p<-4
n<-dim(COL)[1]
library(car)
```

```
## Warning: package 'car' was built under R version 3.4.2
```

Suppose, one wants to perform a linear regression to model the cholesterol level as a function of weight (P) ,height (H) and age (E). So, We will create here the linear model where we are trying to predict the 'Colestrol level (C)' based on the variables Weight(P), Age (E) and Height(H) using the data set COL. The code written below accomplishes this task.

```
mod<-lm(C~P+E+H,COL)
```

We can now see the summary of the model.

```
summary(mod)
```

```
##
## Call:
## lm(formula = C ~ P + E + 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 ***
## P            10.3773     0.7365  14.090  < 2e-16 ***
## E           -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
```

## Optional calculations: Confidence intervals of the parameters

confint(mod,level=0.99)

## Optional calculations: SS1 Test of the parameters with default orderings

anova(mod) anova(lm(C~H+P+E,COL))

## Note: SS3, the tests (F) always coincide with those of the summary (t), F = t ^ 2

Anova(mod,ty=3)

## Diagnosis: TRENDS

plot(predict(mod),resid(mod)) abline(h=0,lty=2)

## Diagnostico: 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)

## Diagnostico: LEVERAGE

plot(hatvalues(mod)) abline(h=c(0,2*mean(hatvalues(mod))),lty=2)

## Diagnosis: INFLUENCE (dffits)

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

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

## Diagnosticos de R

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

## Diagnosticos: Collinearity

vif(mod)

## We remember MODEL SUMMARY

summary(mod)

## Optional calculations: Confidence intervals of the parameters

confint(mod,level=0.99)

## Optional calculations: For some predetermined cases IC of E (Y)

(C0<-data.frame(cbind(P=c(65,75,65),E=c(15,15,12),H=c(150,150,150)), row.names=1:3)) predict(mod, C0, interval="confidence", level=.95, se.fit=T)

## Optional calculations: For some predetermined cases I: Y Prison

predict(mod, C0, interval="prediction", level=.95, se.fit=F)

## Optional calculations: SS1 Test of the parameters with default orderings

anova(mod) #Note: SS3, the tests (F) coincide with those of the summary (t), F = t ^ 2 Anova(mod,ty=3)

## Linear paths in the independent variables:

## center the data, only canvia the independent term

summary(lm(C~I(P-mean(P))+I(E-mean(E))+I(H-mean(H)),COL)) # or, if P = 65, E = 15 and H = 150 are close to the means: summary(lm(C~I(P-65)+I(E-15)+I(H-150),COL))

## Linear paths in the independent variables:

## canvas in some variable, for example, excess weight,

## weight pattern 0.5 * H-10, EP = P- (0.5 * H-10)

summary(mod2<-lm(C~I(P-0.5*H+10)+E+H,COL)) vif(mod2) #Note: Only canvia some parameters and collinearity

## Linear paths in the independent variables:

## remove some independent variable not significant and / or with a lot of colinealidat

## for example H, if excess weight is already used

summary(mod3<-lm(C~I(P-0.5*H+10)+E,COL)) vif(mod3)