# Mètodes R per a Models Lineals

Estimació d'un model lineal (regressió múltiple, ANOVA i ANCOVA)

```
mod <- lm(formula, data, x = TRUE)
```

Model ajustat detallat (Test Omnibus a la darrera línia)

```
summary(mod)
```

Taula de l'ANOVA (per veure els predictors significatius)

```
library(car)
Anova(mod, type=3)
```

Taula de comparació de models niats

```
anova(mod0, mod1) #mod0 model restringit, mod1 model complet
```

VIF (Variance Inflation factors)

```
library(car)
vif(mod)
```

Residus

```
resid(mod)     # Residus crus
rstandard(mod) # Residus Estandarditzats
rstudent(mod)  # Residus Estudentitzats
```

Mesures d'influència:

```
cooks.distance(mod)    # Distàncies de Cook
```

Prediccions

```
ypred<-predict(mod)   #també es pot fer: ypred<-fitted(mod))
```

Plot dels residus versus les prediccions

```
plot(resid(mod)~predict(mod))
abline(h=0)     #Posar una línia horitzontal en el zero (eix d'abscises)
library(car)
scatterplot(resid(mod)~predict(mod),smooth=F,boxplots=F,COL) #alternativa
```

Plot dels residus studentitzats amb línies de referència

```
plot(rstudent(mod))
abline(h=c(-2,0,2)) #Posar una línia horitzontal en el zero (eix d'abscises) i en -2 i 2
```

Intervals de confiança per a les mitjanes

```
library(emmeans)
lsm <- emmeans(mod,formula)    #la formula té només part a la dreta: ~CAT
pairs(lsm)                     #Comparances 2 a 2 (Tukey)
confint(lsm, level=0.95)       #Intervals de confiança
```

Prediccions per una nova variable (Prediction o Confidence)

```
predict(mod, newdata, interval, level)
```

**1. Exemple de Regressió Lineal Múltiple (P,E i H són variables numèriques)**

```
> head(COL)
   E   H    P     C
1 19 174 79.9 189.5
2 15 151 64.5 197.5
3 13 133 52.0 170.5
4 19 173 75.5 180.5
5 17 163 74.0 216.5
6 13 135 54.9 173.5

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

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

> library(car)
> vif(mod)
       P         E         H
 9.489406 20.904776 31.695499

> Anova(mod,type=3)
Anova Table (Type III tests)

Response: C
            Sum Sq Df F value    Pr(>F)
(Intercept) 177887  1 196.219 < 2.2e-16 ***
P           179985  1 198.533 < 2.2e-16 ***
E            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

> (predict(mod, data.frame(P=65,E=15, H=150), interval="prediction", level=.95))
       fit      lwr      upr
1 205.3908 145.3009 265.4807
> (predict(mod, data.frame(P=65,E=15, H=150), interval="confidence", level=.95))
       fit      lwr      upr
1 205.3908 199.1668 211.6148

> scatterplot(fitted(mod),rstandard(mod),smooth=F)  #Plot residus vs.fits

> qqnorm(resid(mod))   #Plot de normalitat dels residus
> qqline(resid(mod))
```

**2. Exemple d'ANOVA (DOSI és una variable categórica amb 5 nivells)**

```
> head(dd)
  DOSI    GMD   IC
1  D00 200.42 2.10
2  D00 190.00 2.19
3  D00 199.33 2.15
4  D00 191.00 2.10
> summary(m1<-lm(GMD~DOSI,dd,x=T))

Call:
lm(formula = GMD ~ DOSI, data = dd, x = T)

Residuals:
    Min      1Q  Median      3Q     Max
-15.152  -6.734  -0.482   8.132  11.518

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  196.350      4.006  49.014  < 2e-16 ***
DOSID08        3.802      5.665   0.671     0.51
DOSID15       27.848      5.665   4.916 8.35e-05 ***
DOSID20       34.714      5.665   6.127 5.48e-06 ***
DOSID30       34.684      5.665   6.122 5.55e-06 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 8.958 on 20 degrees of freedom
Multiple R-squared:  0.7827,   Adjusted R-squared:  0.7392
F-statistic: 18.01 on 4 and 20 DF,  p-value: 2.075e-06

> Anova(m1,type=3)
Anova Table (Type III tests)

Response: GMD
             Sum Sq Df   F value    Pr(>F)
(Intercept) 1172452  1 14611.825 < 2.2e-16 ***
DOSI           5779  4    18.006 2.075e-06 ***
Residuals      1605 20
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

> (lsm<-emmeans(m1,~DOSI))
 DOSI  emmean       SE df lower.CL upper.CL
 D00  196.350 4.005993 20 187.9936 204.7064
 D08  200.152 4.005993 20 191.7956 208.5084
 D15  224.198 4.005993 20 215.8416 232.5544
 D20  231.064 4.005993 20 222.7076 239.4204
 D30  231.034 4.005993 20 222.6776 239.3904

Confidence level used: 0.95
```

### 3. Exemple d'ANCOVA (M és una variable categórica amb 2 nivells i C és numèrica)

```
> head(dades)
  M   C  V VV
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

> summary(mv<-lm(V~M+C+M:C,dades))

Call:
lm(formula = V ~ M + C + M:C, data = dades)

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) -77.49366   10.33232  -7.500 6.07e-07 ***
M2           10.45276   13.05628   0.801    0.434
C             1.19565    0.09575  12.487 2.65e-10 ***
M2:C          0.03924    0.12133   0.323    0.750
---
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

> Anova(mv,ty=3)

Anova Table (Type III tests)

Response: V
            Sum Sq Df  F value    Pr(>F)
(Intercept) 1478.6  1 122.5474 1.828e-09 ***
M              7.7  1   0.6409    0.4338
C           4841.6  1 401.2715 9.372e-14 ***
M:C            1.3  1   0.1046    0.7501
Residuals    217.2 18
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

> (lsm<-emmeans(mv, ~M|C, at=list(C=c(90))))

C =  90:
 M   emmean       SE df lower.CL upper.CL
 1 30.11519 1.987551 18 25.93950 34.29088
 2 44.09963 1.572317 18 40.79631 47.40294
```