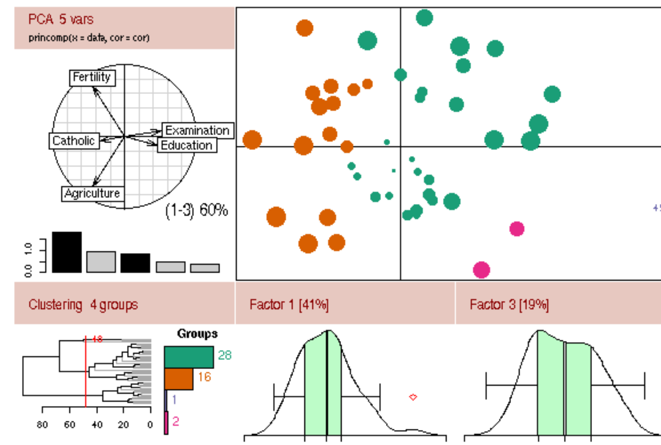
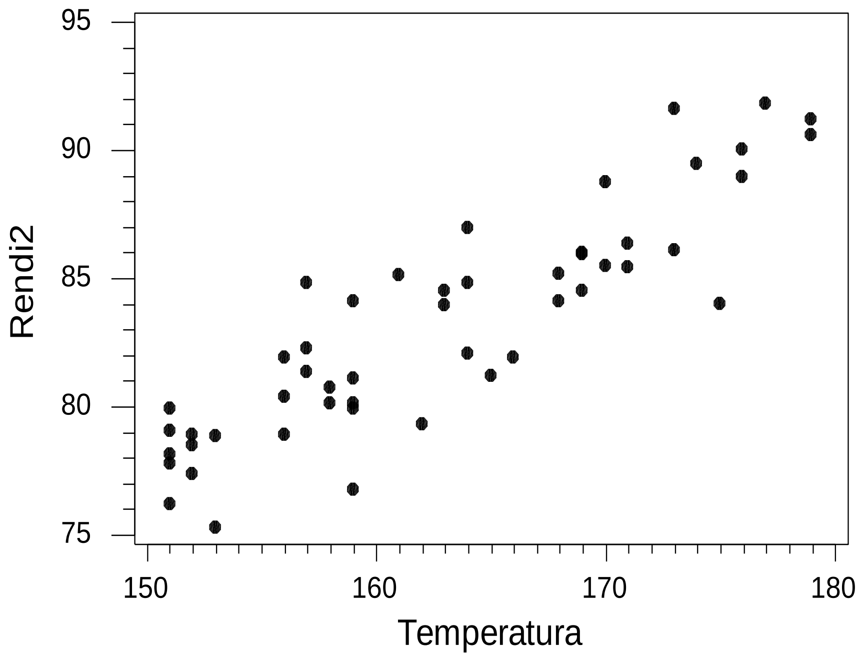
# Repàs de models de regressió lineal

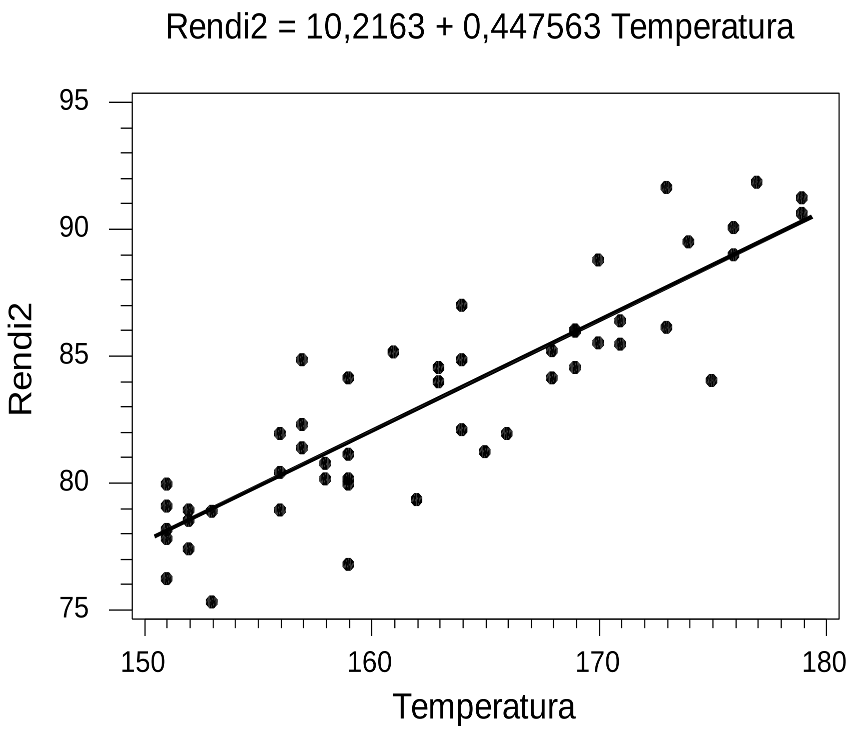
$$y_{ij} = \mu + \alpha_i + e_{ij}$$




UNIVERSITAT DE BARCELONA
UNIVERSITAT POLITÈCNICA DE CATALUNYA BARCELONATECH
Departament d'Estadística

# Models de regressió lineal



# Model de regresión simple

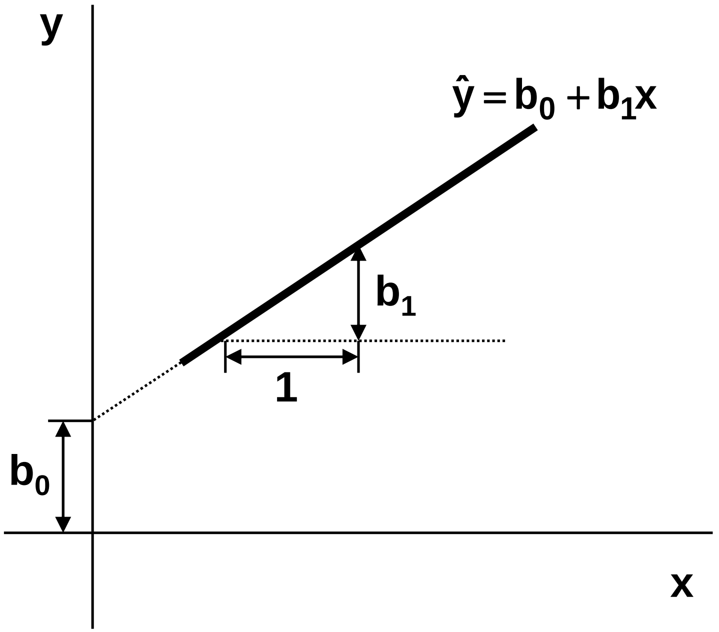
**Model per la població:**

$$y = \beta_0 + \beta_1 x + \varepsilon$$

$$\varepsilon \sim N(0, \sigma)$$

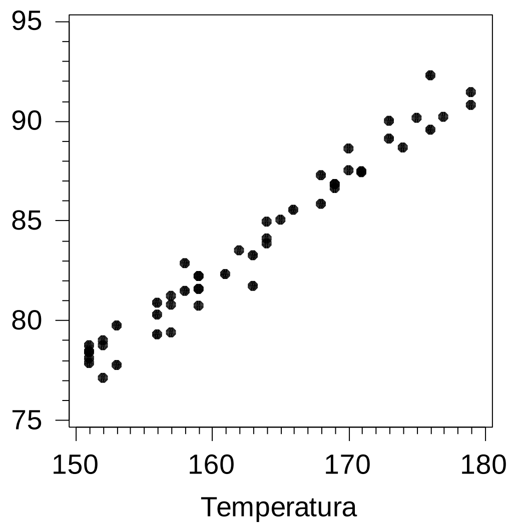
**Recta ajustada:**
**(a partir d'una mostra)**

$$\hat{y} = b_0 + b_1 x$$



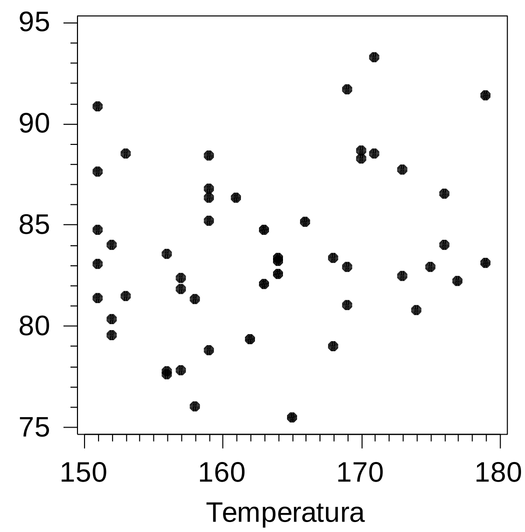
# Regressió simple: comparació de situacions

## Situación 1




```
The regression equation is
Rendi1 = 3,08 + 0,495 Temperatura

Predictor      Coef    SE Coef      T      P
Constant      3,082      2,162   1,43  0,160
Temperat    0,49524    0,01325  37,38  0,000

S = 0,7932   R-Sq = 96,7%   R-Sq(adj) = 96,6%
```

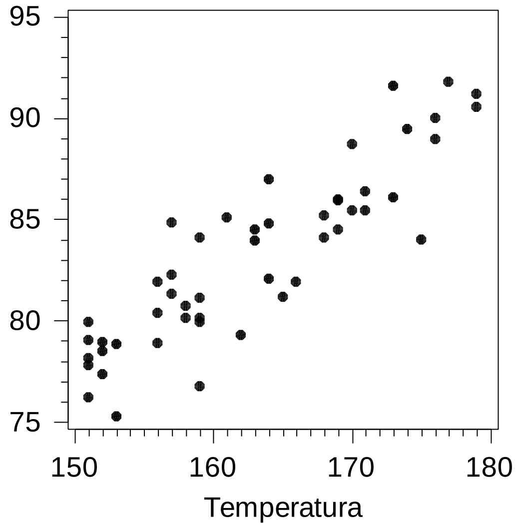
## Situación 3




```
The regression equation is
Rendi3 = 65,7 + 0,110 Temperatura

Predictor      Coef    SE Coef      T      P
Constant      65,72      11,00   5,97  0,000
Temperat    0,11029    0,06743   1,64  0,108

S = 4,037   R-Sq = 5,3%    R-Sq(adj) = 3,3%
```

# Regressió simple: interpretació de resultats

**Situación 2**



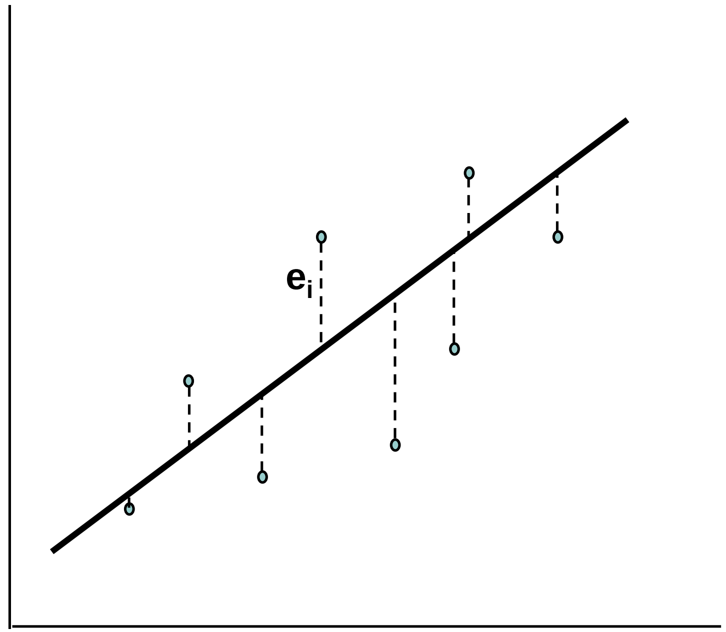

```
The regression equation is
Rendi2 = 10,2 + 0,448 Temperatura

Predictor        Coef     SE Coef        T        P
Constant       10,216       5,497     1,86    0,069
Temperat      0,44756     0,03370    13,28    0,000

S = 2,017      R-Sq = 78,6%      R-Sq(adj) = 78,2%
```

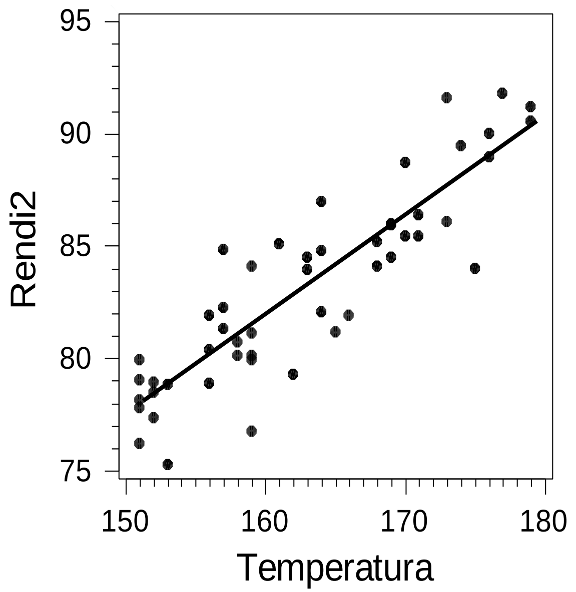
**Desviació tipus dels residus**

**Percentatge de la variabilitat de y explicat per l'equació**

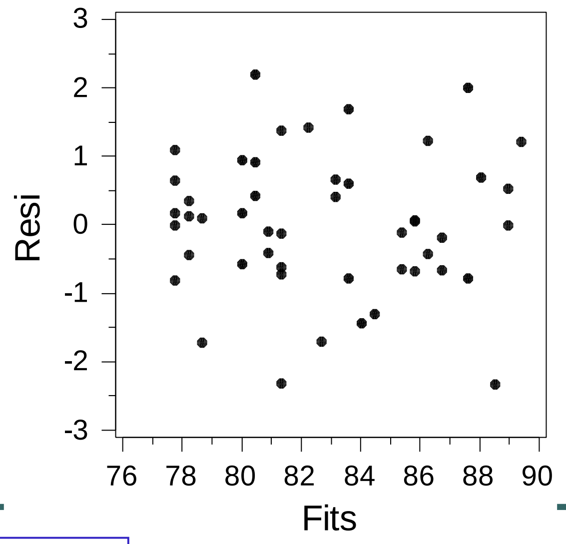
**Proves de significació pels coeficients**

# Residus: Què són?

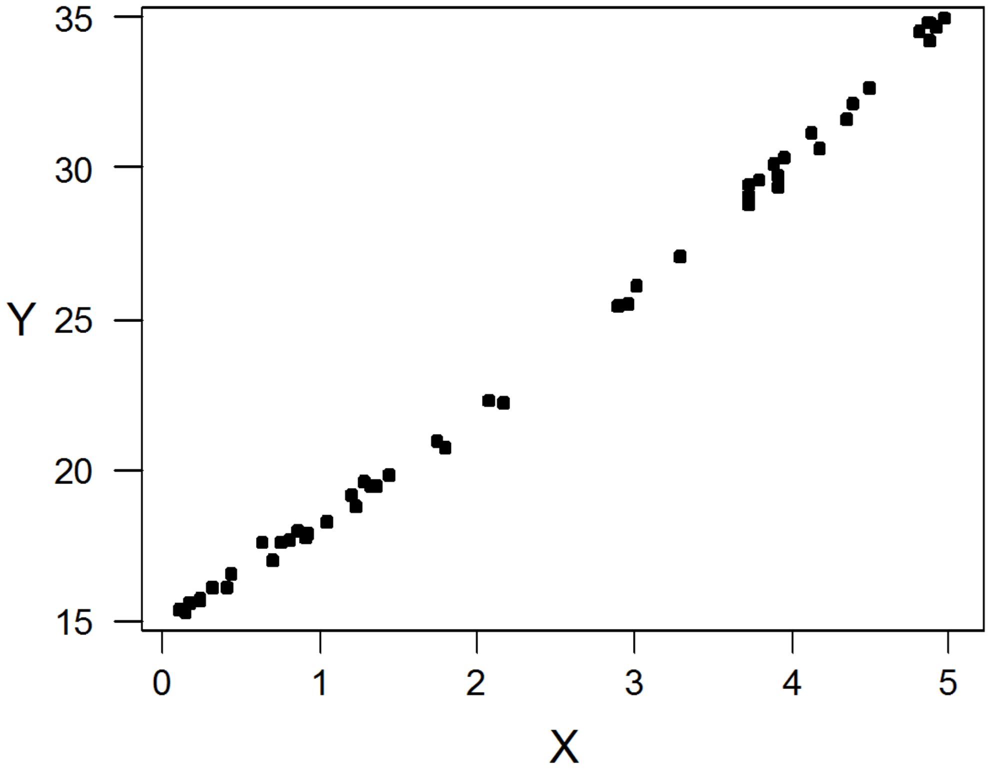



**Part de la variable dependent (y)**
**no explicada pel model**

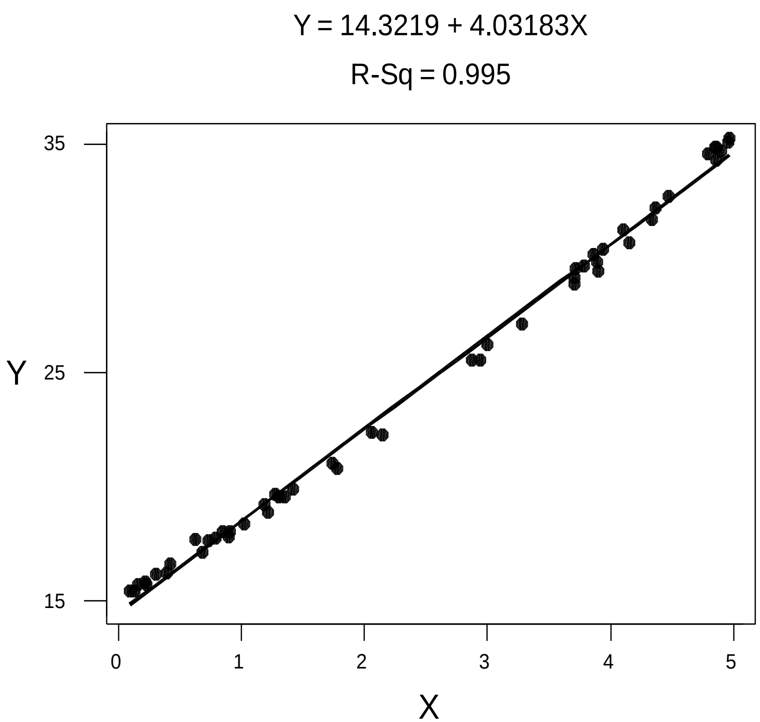


# Anàlisi dels residus

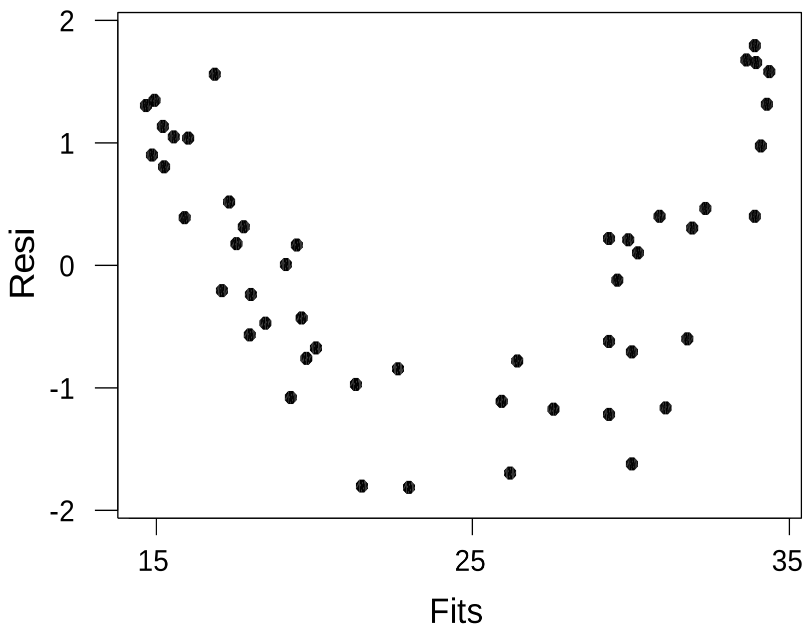


# Análisi dels residus

## Model lineal




## Residus versus valors previstos




**Els residus contenen informació**

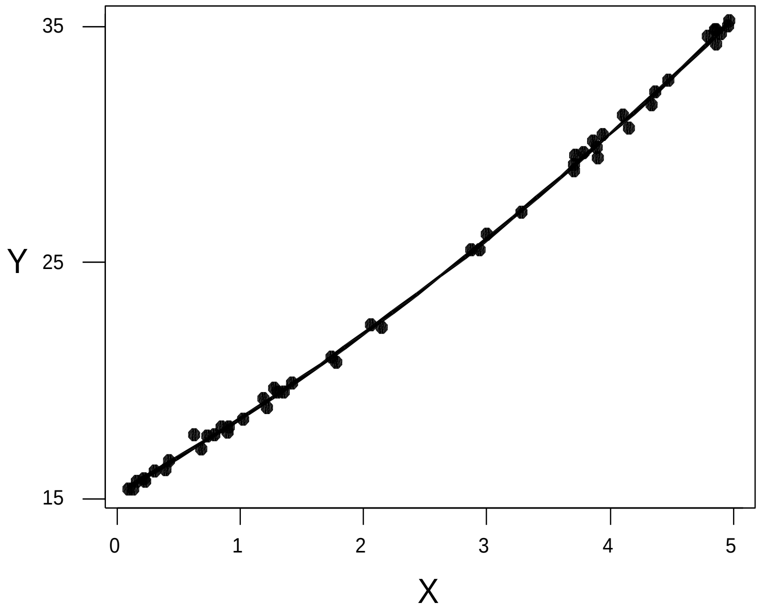
**Model no adequat**

# Anàlisi dels residus

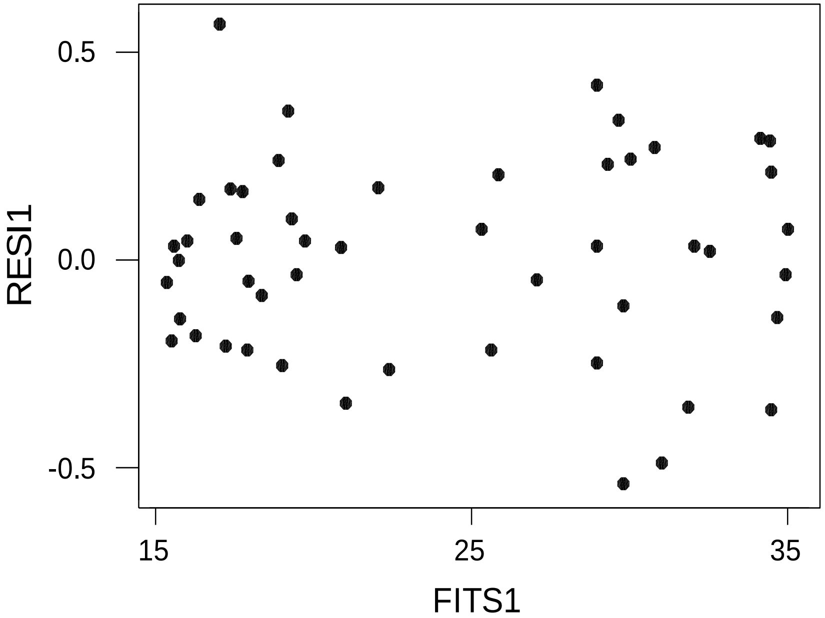
## Model quadràtic

Y = 15.0867 + 2.94152X + 0.214238X**2

R-Sq = 0.999




## Residus versus valors previstos

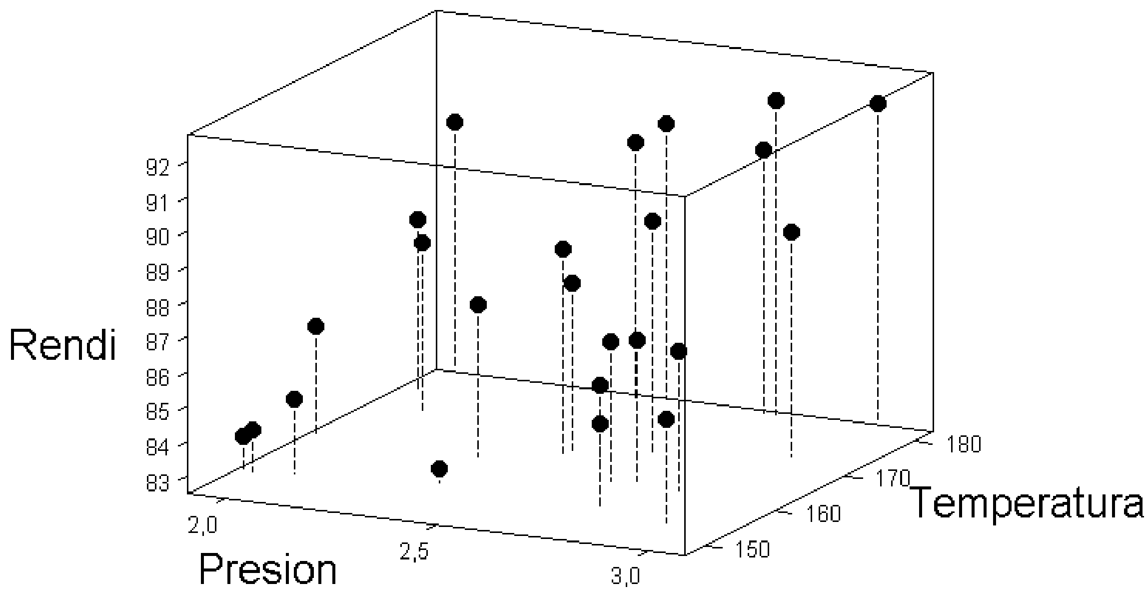



**Els residus contenen informació**

**Model adequat**

# Regressió múltiple

| ↓ | C1 | C2 | C3 |
|---|---|---|---|
| | Presion | Temperatura | Rendi |
| 1 | 2,8 | 152 | 86,0 |
| 2 | 2,8 | 178 | 91,1 |
| 3 | 2,0 | 160 | 86,0 |
| 4 | 2,4 | 153 | 85,4 |
| 5 | 2,5 | 179 | 90,8 |
| 6 | 2,7 | 166 | 88,6 |
| 7 | 2,1 | 169 | 88,0 |
| 8 | 2,0 | 174 | 88,8 |
| 9 | 2,5 | 164 | 87,9 |
| 10 | 2,5 | 163 | 87,7 |
| 11 | 2,9 | 158 | 87,4 |
| 12 | 2,7 | 160 | 87,4 |
| 13 | 2,1 | 151 | 84,4 |
| 14 | 2,7 | 158 | 87,1 |
| 15 | 2,5 | 177 | 90,4 |
| 16 | 3,0 | 169 | 89,8 |
| 17 | 2,8 | 178 | 91,2 |
| 18 | 2,8 | 159 | 87,4 |
| 19 | 2,4 | 160 | 86,8 |
| 20 | 2,0 | 150 | 84,0 |
| 21 | 2,0 | 150 | 84,0 |
| 22 | 3,0 | 150 | 86,0 |
| 23 | 2,0 | 180 | 90,0 |
| 24 | 3,0 | 180 | 92,0 |
| 25 | | | |

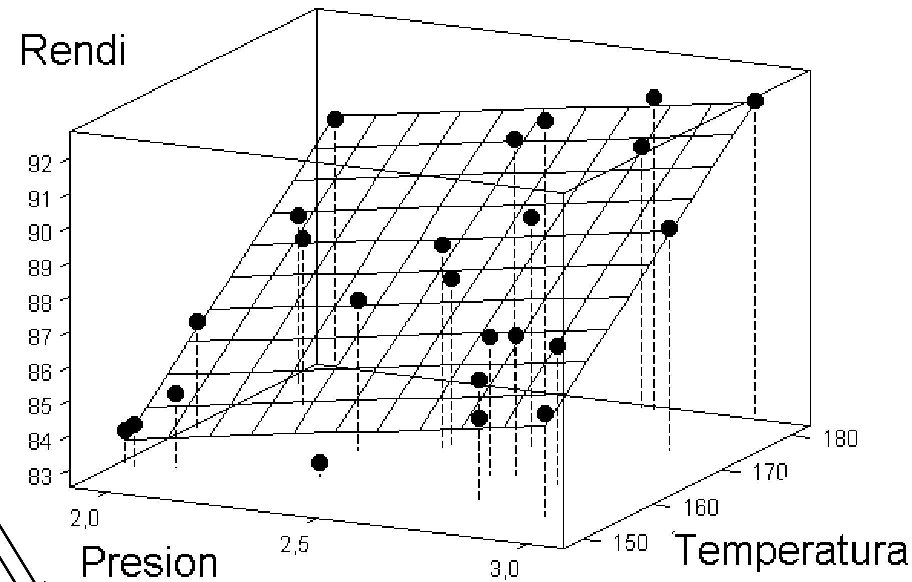


# Regressió múltiple: interpretació dels resultats

```
Regression Analysis: Rendi versus
Presion; Temperatura

The regression equation is
Rendi = 48,9 + 1,84 Presion + 0,208 Temperatura
```

| Predictor | Coef | SE Coef | T | P |
|---|---|---|---|---|
| Constant | 48,941 | 2,709 | 18,07 | 0,000 |
| Presion | 1,8437 | 0,4699 | 3,92 | 0,001 |
| Temperat | 0,20807 | 0,01562 | 13,32 | 0,000 |

```
S = 0,7947     R-Sq = 90,8%     R-Sq(adj) = 89,9%
```

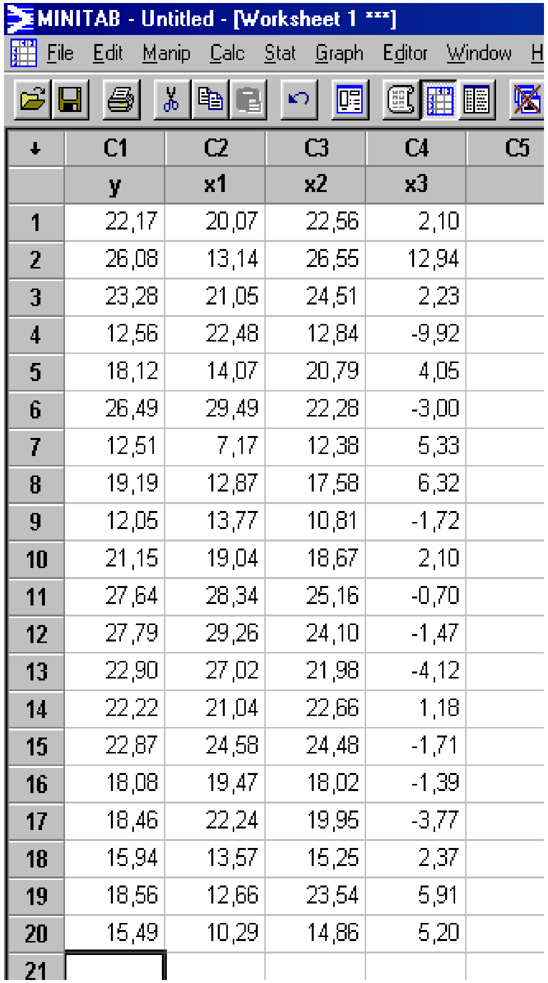



**Desviació tipus dels residus**

**Proves de significació pels coeficients**

**Mesura de qualitat de l'ajust**

# Regressió múltiple: construcció del model




| ↓ | C1 | C2 | C3 | C4 | C5 |
|---|---|---|---|---|---|
| | y | x1 | x2 | x3 | |
| 1 | 22,17 | 20,07 | 22,56 | 2,10 | |
| 2 | 26,08 | 13,14 | 26,55 | 12,94 | |
| 3 | 23,28 | 21,05 | 24,51 | 2,23 | |
| 4 | 12,56 | 22,48 | 12,84 | -9,92 | |
| 5 | 18,12 | 14,07 | 20,79 | 4,05 | |
| 6 | 26,49 | 29,49 | 22,28 | -3,00 | |
| 7 | 12,51 | 7,17 | 12,38 | 5,33 | |
| 8 | 19,19 | 12,87 | 17,58 | 6,32 | |
| 9 | 12,05 | 13,77 | 10,81 | -1,72 | |
| 10 | 21,15 | 19,04 | 18,67 | 2,10 | |
| 11 | 27,64 | 28,34 | 25,16 | -0,70 | |
| 12 | 27,79 | 29,26 | 24,10 | -1,47 | |
| 13 | 22,90 | 27,02 | 21,98 | -4,12 | |
| 14 | 22,22 | 21,04 | 22,66 | 1,18 | |
| 15 | 22,87 | 24,58 | 24,48 | -1,71 | |
| 16 | 18,08 | 19,47 | 18,02 | -1,39 | |
| 17 | 18,46 | 22,24 | 19,95 | -3,77 | |
| 18 | 15,94 | 13,57 | 15,25 | 2,37 | |
| 19 | 18,56 | 12,66 | 23,54 | 5,91 | |
| 20 | 15,49 | 10,29 | 14,86 | 5,20 | |
| 21 | | | | | |

**Construir el millor model de regressió múltiple no és evident…**

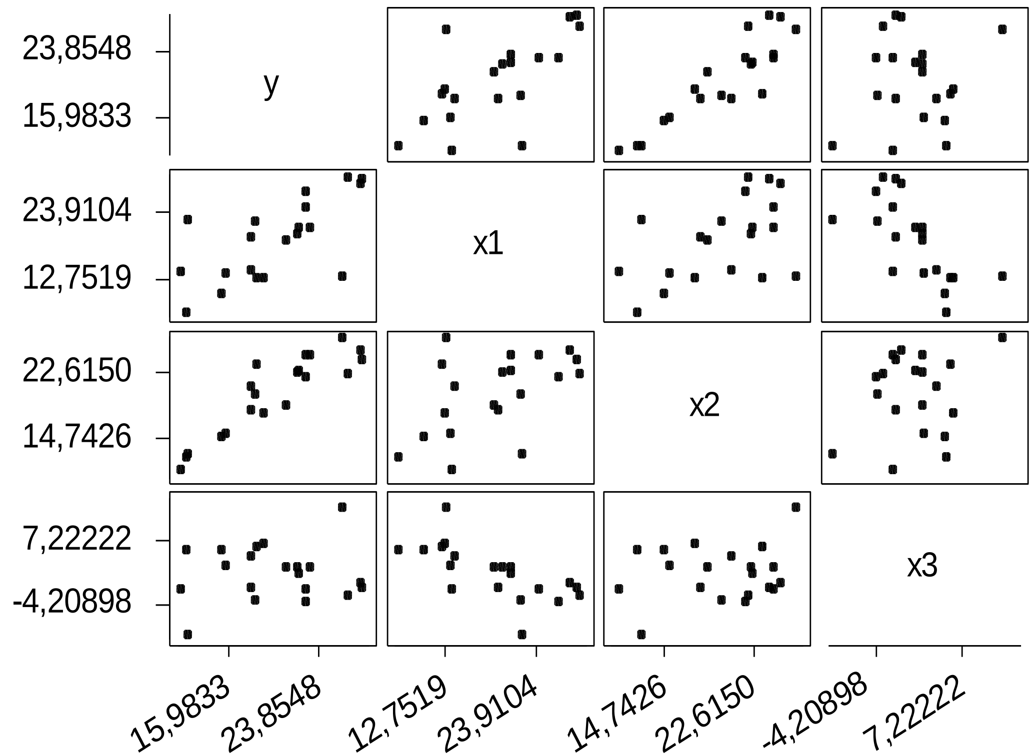
**Volem construir un model per explicar 'y' en funció de només 2 variables regressores (d'entre les 3 disponibles)**
**Quines són les 2 variables que cal fer servir?**
**Obtindrem un bon ajust?**

# Regressió múltiple: construcció del model

**Correlations: y; x1; x2; x3**

| | y | x1 | x2 |
|---|---|---|---|
| x1 | 0,675 | | |
| | 0,001 | | |
| x2 | 0,901 | 0,516 | |
| | 0,000 | 0,020 | |
| x3 | 0,088 | -0,676 | 0,204 |
| | 0,714 | 0,001 | 0,388 |

**Cell Contents: Pearson correlation**
**P-Value**



# Regressió múltiple: construcció del model

MINITAB - Untitled - [Worksheet 1 ***]

| ↓ | C1 | C2 | C3 | C4 | C5 |
|---|---|---|---|---|---|
| | y | x1 | x2 | x3 | |
| 1 | 22,17 | 20,07 | 22,56 | 2,10 | |
| 2 | 26,08 | 13,14 | 26,55 | 12,94 | |
| 3 | 23,28 | 21,05 | 24,51 | 2,23 | |
| 4 | 12,56 | 22,48 | 12,84 | -9,92 | |
| 5 | 18,12 | 14,07 | 20,79 | 4,05 | |
| 6 | 26,49 | 29,49 | 22,28 | -3,00 | |
| 7 | 12,51 | 7,17 | 12,38 | 5,33 | |
| 8 | 19,19 | 12,87 | 17,58 | 6,32 | |
| 9 | 12,05 | 13,77 | 10,81 | -1,72 | |
| 10 | 21,15 | 19,04 | 18,67 | 2,10 | |
| 11 | 27,64 | 28,34 | 25,16 | -0,70 | |
| 12 | 27,79 | 29,26 | 24,10 | -1,47 | |
| 13 | 22,90 | 27,02 | 21,98 | -4,12 | |
| 14 | 22,22 | 21,04 | 22,66 | 1,18 | |
| 15 | 22,87 | 24,58 | 24,48 | -1,71 | |
| 16 | 18,08 | 19,47 | 18,02 | -1,39 | |
| 17 | 18,46 | 22,24 | 19,95 | -3,77 | |
| 18 | 15,94 | 13,57 | 15,25 | 2,37 | |
| 19 | 18,56 | 12,66 | 23,54 | 5,91 | |
| 20 | 15,49 | 10,29 | 14,86 | 5,20 | |
| 21 | | | | | |

**Les variables més adequades NO SON x2 y x1**

**$y = x1 + x3$**

**$R^2 = 100$ %**

**Quan es tenen moltes variables, trobar el millor model no és tan fàcil.**

**Hi ha mètodes que ajuden a trobar-lo.**

# L’exemple de la molla

**Anàlisi com a disseny factorial:**

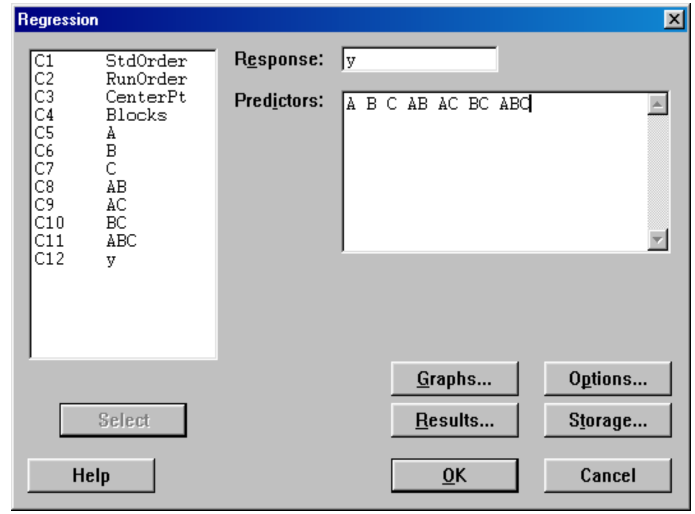
| | C5 | C6 | C7 | C8 | C9 |
|---|---|---|---|---|---|
| ks | A | B | C | y | |
| 1 | -1 | -1 | -1 | 79 | |
| 1 | 1 | -1 | -1 | 97 | |
| 1 | -1 | 1 | -1 | 75 | |
| 1 | 1 | 1 | -1 | 92 | |
| 1 | -1 | -1 | 1 | 64 | |
| 1 | 1 | -1 | 1 | 84 | |
| 1 | -1 | 1 | 1 | 73 | |
| 1 | 1 | 1 | 1 | 90 | |

| Term | Effect | Coef |
|---|---|---|
| Constant | | 81,750 |
| A | 18,000 | 9,000 |
| B | 1,500 | 0,750 |
| C | -8,000 | -4,000 |
| A*B | -1,000 | -0,500 |
| A*C | 0,500 | 0,250 |
| B*C | 6,000 | 3,000 |
| A*B*C | -0,500 | -0,250 |

# L'exemple de la molla

## Anàlisi per regressió

| C5 | C6 | C7 | C8 | C9 | C10 | C11 | C12 |
|---|---|---|---|---|---|---|---|
| A | B | C | AB | AC | BC | ABC | y |
| -1 | -1 | -1 | 1 | 1 | 1 | -1 | 79 |
| 1 | -1 | -1 | -1 | -1 | 1 | 1 | 97 |
| -1 | 1 | -1 | -1 | 1 | -1 | 1 | 75 |
| 1 | 1 | -1 | 1 | -1 | -1 | -1 | 92 |
| -1 | -1 | 1 | 1 | -1 | -1 | 1 | 64 |
| 1 | -1 | 1 | -1 | 1 | -1 | -1 | 84 |
| -1 | 1 | 1 | -1 | -1 | 1 | -1 | 73 |
| 1 | 1 | 1 | 1 | 1 | 1 | 1 | 90 |




```
The regression equation is
y = 81,8 + 9,00 A + 0,750 B - 4,00 C - 0,500 AB + 0,250 AC + 3,00 BC - 0,250 ABC

Predictor         Coef     SE Coef          T         P
Constant       81,7500      0,0000          *         *
A              9,00000     0,00000          *         *
B             0,750000    0,000000          *         *
C             -4,00000     0,00000          *         *
AB           -0,500000    0,000000          *         *
AC            0,250000    0,000000          *         *
BC             3,00000     0,00000          *         *
ABC          -0,250000    0,000000          *         *
```

# Models de regressió lineals

```
> summary(aov(Y~L*G*T, data=molles.cod))
            Df Sum Sq Mean Sq F value    Pr(>F)
L            1   1296    1296   259.2 2.224e-07 ***
G            1      9       9     1.8 0.2165473
T            1    256     256    51.2 9.658e-05 ***
L:G          1      4       4     0.8 0.3972038
L:T          1      1       1     0.2 0.6665811
G:T          1    144     144    28.8 0.0006724 ***
L:G:T        1      1       1     0.2 0.6665811
Residuals    8     40       5
```

Amb R:

Fent servir aov i lm (sobre la matriu de disseny en unitats codificades)

```
lm(formula = Y ~ L * G * T, data = molles.cod)

Residuals:
       Min         1Q     Median         3Q        Max
-2.000e+00 -1.250e+00 -5.551e-17  1.250e+00  2.000e+00

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   81.750      0.559 146.239 5.35e-15 ***
L              9.000      0.559  16.100 2.22e-07 ***
G              0.750      0.559   1.342 0.216547
T             -4.000      0.559  -7.155 9.66e-05 ***
L:G           -0.500      0.559  -0.894 0.397204
L:T            0.250      0.559   0.447 0.666581
G:T            3.000      0.559   5.367 0.000672 ***
L:G:T         -0.250      0.559  -0.447 0.666581
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.236 on 8 degrees of freedom
Multiple R-squared: 0.9772,     Adjusted R-squared: 0.9572
F-statistic: 48.89 on 7 and 8 DF,  p-value: 6.101e-06
```