

MODEL 3

1. Cas pràctic 1

Es disposa d'una mostra de 500 empleats d'una empresa i es desitja estimar una equació de Mincer, en la qual s'explica el Salari (Sal) de cada empleat en funció de la seva experiència (Exp) i sexe (Sex=1 Home i Sex=0 Dona) amb la següent especificació:

$$Sal_i = \beta_1 + \beta_2 Exp_i + \beta_3 Exp_i^2 + \beta_4 Sex_i + \beta_5 Sex_i \times Exp_i + u_i$$

Els resultats que s'han obtingut són els següents:

```
lm(formula = Sal ~ Exp + Exp2 + sex + sex_Exp)

Residuals:
    Min       1Q   Median       3Q      Max
-1.450e-13 -5.690e-16  2.800e-16  1.054e-15  9.445e-15

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  1.300e+01  1.015e-15  1.281e+16 <2e-16 ***
Exp           7.000e-02  1.439e-16  4.863e+14 <2e-16 ***
Exp2        -1.000e-03  4.609e-18 -2.170e+14 <2e-16 ***
sex         -6.000e+00  1.302e-15 -4.610e+15 <2e-16 ***
sex_Exp     -1.000e-02  7.799e-17 -1.282e+14 <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 6.613e-15 on 495 degrees of freedom
Multiple R-squared:  1, Adjusted R-squared:  1
F-statistic: 2.332e+31 on 4 and 495 DF, p-value: < 2.2e-16
```

a) A partir dels resultats obtinguts com seria el salari esperat d'un ocupat home i quin seria el salari esperat per a una dona.

b) Si en comptes d'utilitzar com a categoria el fet de ser dona, es decideix utilitzar una variable fictícia associada a la categoria Home i una altra associada a la categoria Dona, com s'hauria de re-especificar el model anterior?

Un altre investigador decideix simplificar el model i planteja (mod2):

```
lm(formula = Sal ~ Exp + sex + sex_Exp)

Residuals:
    Min       1Q   Median       3Q      Max
-0.15364 -0.05035  0.02210  0.05634  0.07189

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) 13.1514141  0.0071746 1833.05  <2e-16 ***
Exp           0.0401131  0.0004070   98.57  <2e-16 ***
sex          -6.0205444  0.0126479 -476.01  <2e-16 ***
sex_Exp      -0.0083278  0.0007561  -11.01  <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.06443 on 496 degrees of freedom
Multiple R-squared:  0.9995, Adjusted R-squared:  0.9995
F-statistic: 3.274e+05 on 3 and 496 DF,  p-value: < 2.2e-16
```

I per a comprovar si el model mod2 té problemes es planteja la següent regressió auxiliar (on Sal_fit2 és igual a l'endògena estimada elevada al quadrat obtinguda del segon model, mod2):

```
lm(formula = Sal ~ Exp + sex + sex_Exp + Sal_fit2)

Residuals:
    Min       1Q   Median       3Q      Max
-0.048759 -0.005455 -0.000688  0.006550  0.021624

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) 129.852528  1.012935  128.2  <2e-16 ***
Exp           0.785510  0.006470  121.4  <2e-16 ***
sex          -88.453042  0.715496 -123.6  <2e-16 ***
sex_Exp      -0.428169  0.003647 -117.4  <2e-16 ***
Sal_fit2     -0.675682  0.005865 -115.2  <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.01223 on 495 degrees of freedom
Multiple R-squared:  1, Adjusted R-squared:  1
F-statistic: 6.819e+06 on 4 and 495 DF,  p-value: < 2.2e-16
```

I una segona regressió auxiliar on resid2 són els residus al quadrat de mod2:

```
lm(formula = resid2 ~ Exp + sex + sex_Exp + Exp2 + sex_Exp2)
```

Residuals:

```
      Min       1Q   Median       3Q      Max
-0.0057419 -0.0031901 -0.0004742  0.0026188  0.0121583
```

Coefficients:

```
      Estimate Std. Error t value Pr(>|t|)
(Intercept)  1.077e-02  5.971e-04  18.033  <2e-16 ***
Exp          -1.241e-03  9.126e-05 -13.597  <2e-16 ***
sex          -1.371e-03  1.014e-03  -1.353   0.1768
sex_Exp       -8.058e-05  1.597e-04  -0.505   0.6140
Exp2          4.027e-05  2.960e-06  13.602  <2e-16 ***
sex_Exp2      9.981e-06  5.365e-06   1.860   0.0634 .
---

```

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.003543 on 494 degrees of freedom
Multiple R-squared: 0.3943, Adjusted R-squared: 0.3881
F-statistic: 64.31 on 5 and 494 DF, p-value: < 2.2e-16

c) Quina hipòtesi bàsica del Model de Regressió Lineal Múltiple s'està contrastant en cada cas, amb quin test i finalment a quina conclusió s'arriba?

d) Amb tota la informació anterior, quin creus que és el vertader problema de mod2?

2. Caso pràctic 2

Es disposa de la següent informació sobre el nombre de naixements ocorreguts l'any 2020 a l'illa d'Eivissa, així com el grup d'edat de la mare (totes residents a l'illa):

Grup d'edat de la mare	Nombre de naixements
De 15 a 19 anys	13
De 20 a 24 anys	77
De 25 a 29 anys	218
De 30 a 34 anys	430
De 35 a 39 anys	370
De 40 a 44 anys	140
De 45 a 49 anys	8

A més es coneix el nombre de dones residents a l'illa d'Eivissa per grups d'edat:

Grup d'edat de la mare	Població a 1/1/2020	Població a 1/1/2021
De 15 a 19 anys	3.200	3.288
De 20 a 24 anys	3.414	3.375
De 25 a 29 anys	5.278	5.122
De 30 a 34 anys	7.054	6.729
De 35 a 39 anys	7.608	7.521
De 40 a 44 anys	7.167	7.382
De 45 a 49 anys	6.241	6.397

Se sol·licita calcular:

a) La Taxa General o Global de Fecunditat (TGF).

b) Taxa global de Natalitat o Taxa Bruta de Natalitat (TBN), sabent que la població resident a Eivissa és la següent:

Sexe	Població a 1/1/2020	Població a 1/1/2021
Home	83.920	83.609
Dona	87.916	87.591

c) La Taxa de Fecunditat Específica per al grup de 15 a 19 anys.

d) La Ràtio de Masculinitat al Naixement a 2020 és de 105 ($RMN^{2020}=105$), interpreta aquest resultat i calcula quantes nines han nascut.

e) L'edat mitjana a la Maternitat (EMM).