

Estimación en modelos de regresión lineal múltiple

Primeros pasos

Román Salmerón Gómez y Catalina García García

Ejercicios

1.- Dado el modelo de regresión lineal simple $\mathbf{S} = \beta_1 + \beta_2 \cdot \mathbf{E} + \mathbf{u}$, donde $\mathbf{S}$ es el salario medio por hora (medido en euros) y $\mathbf{E}$ son los años de escolarización, obtenga (usando, al menos, cuatro decimales) la estimación de las cantidades constantes del modelo (interpretando las estimaciones obtenidas) y el coeficiente de determinación corregido, a partir de la siguiente información:

Salario medio (euros/hora)	Escolarización (años)
4.45	6
5.77	7
7.31	10
7.81	12
10.67	15
13.61	17

¿Qué interpretación tiene un residuo positivo/negativo?

2.- Dado el modelo de regresión lineal múltiple $\mathbf{y} = \beta_1 + \beta_2 \cdot \mathbf{V} + \beta_3 \cdot \mathbf{Z} + \mathbf{u}$ con $n = 15$ observaciones y $k = 3$ variables independientes, en los siguientes casos, calcule:

- La estimación de los coeficientes de las variables independientes: $\hat{\beta} = (\mathbf{X}^t \mathbf{X})^{-1} \mathbf{X}^t \mathbf{y}$. Interpretación.
- Los errores de la estimación realizada: $\mathbf{e} = \mathbf{y} - \hat{\mathbf{y}} = \mathbf{y} - \mathbf{X}\hat{\beta}$.
- La estimación de la varianza de la perturbación aleatoria: $\hat{\sigma} = \frac{\mathbf{e}^t \mathbf{e}}{n-k}$.
- La suma de cuadrados totales: $SCT = \mathbf{y}^t \mathbf{y} - n \cdot \bar{\mathbf{y}}^2$.
- La suma de cuadrados explicada: $SCE = \hat{\beta}^t \mathbf{X}^t \mathbf{y} - n \cdot \bar{\mathbf{y}}^2$.
- La suma de cuadrados de los residuos: $SCR = \mathbf{y}^t \mathbf{y} - \hat{\beta}^t \mathbf{X}^t \mathbf{y}$.
- El coeficiente de determinación: $R^2 = \frac{SCE}{SCT} = 1 - \frac{SCR}{SCT}$. ¿Se verifica la igualdad? ¿Por qué? Interpretación.
- El coeficiente de determinación corregido: $\bar{R}^2 = 1 - (1 - R^2) \cdot \frac{n-1}{n-k}$.

Responda de forma razonada todas las cuestiones planteadas usando, al menos, cuatro decimales.

(Caso A)

$(\mathbf{y}, \mathbf{X}) =$

	y	cte	V	Z
1	-16	1	-7	2
2	-2	1	-6	-3
3	-10	1	-5	1
4	1	1	-4	-2
5	-16	1	-3	4
6	-3	1	-2	-1
7	-9	1	-1	3
8	1	1	0	0
9	14	1	1	-3
10	3	1	2	1
11	20	1	3	-4
12	4	1	4	2
13	16	1	5	-1
14	6	1	6	3
15	18	1	7	-2

$$(\mathbf{X}^t \mathbf{X})^{-1} =$$

	cte	V	Z
cte	0.06666667	0.0000000000	0.0000000000
V	0.00000000	0.0035924233	0.0004898759
Z	0.00000000	0.0004898759	0.0114304376

$$\mathbf{X}^t \mathbf{y} =$$

	y
cte	27
V	571
Z	-271

(Caso B)

$$(\mathbf{y}, \mathbf{X}) =$$

	y	cte	X	Z
1	-8	1	-7	0
2	-5	1	-6	-1
3	-10	1	-5	1
4	-4	1	-4	-1
5	-5	1	-3	1
6	1	1	-2	-1
7	-2	1	-1	1
8	0	1	0	0
9	5	1	1	1
10	10	1	2	-1

```

11  10   1  3  1
12  17   1  4 -1
13  10   1  5  1
14  19   1  6 -1
15  20   1  7  0

```

$$(\mathbf{X}^t \mathbf{X})^{-1} =$$

```

           cte          X          Z
cte 0.06666667 0.000000000 0.00000000
X   0.00000000 0.003571429 0.00000000
Z   0.00000000 0.000000000 0.08333333

```

$$\mathbf{X}^t \mathbf{y} =$$

```

           y
cte  58
X   594
Z   -30

```

(Caso C)

$$(\mathbf{y}, \mathbf{X}) =$$

```

           y cte  X  Z
1  -15   1 -7  1
2   -9   1 -6  0
3   -8   1 -5  0
4   -4   1 -4  0
5   -3   1 -3  0
6    1   1 -2  0
7   23   1 -1 -7
8    1   1  0  0
9  -15   1  1  7
10   5   1  2  0
11   7   1  3  0
12  13   1  4  0
13  11   1  5  0
14  16   1  6  0
15  18   1  7 -1

```

$$(\mathbf{X}^t \mathbf{X})^{-1} =$$

```

           cte          X          Z
cte 0.06666667 0.000000000 0.00
X   0.00000000 0.003571429 0.00
Z   0.00000000 0.000000000 0.01

```

$$\mathbf{X}^t \mathbf{y} =$$

	y
cte	41
X	544
Z	-299

3.- Dado el modelo de regresión lineal múltiple $\mathbf{PA} = \beta_1 + \beta_2 \cdot \mathbf{P} + \beta_3 \cdot \mathbf{A} + \mathbf{u}$ donde $\mathbf{PA}$ es la presión arterial (medida en milímetros de mercurio), $\mathbf{P}$ el peso (medido en kilogramos) y $\mathbf{A}$ la altura (medida en metros). Interprete los coeficientes de las variables independientes. ¿Es intuitivo que la altura cambie y el peso no varíe?

4.- Dado el modelo de regresión lineal múltiple $\mathbf{S} = \beta_1 + \beta_2 \cdot \mathbf{E} + \beta_3 \cdot \mathbf{G} + \mathbf{u}$ donde $\mathbf{S}$ es el salario medio por hora (medido en euros), $\mathbf{E}$ son los años de escolarización y $\mathbf{E}$ es el género (donde masculino se codifica como 1 y en cualquier otro caso se usa el 0), interprete β_1 , β_2 , β_3 y $\beta_1 + \beta_3$

5.- Dado el modelo de regresión lineal múltiple:

$$\mathbf{S} = \beta_1 + \beta_2 \cdot \mathbf{E} + \beta_3 \cdot \mathbf{G} + \beta_4 \cdot \mathbf{G} \times \mathbf{E} + \mathbf{u},$$

donde $\mathbf{S}$ es el salario medio por hora (medido en euros), $\mathbf{E}$ son los años de escolarización y $\mathbf{E}$ es el género (donde masculino se codifica como 1 y en cualquier otro caso se usa el 0). Se pide contestar de forma razonada las siguientes cuestiones:

- Obtenga el efecto de la escolarización sobre el salario medio si $(\hat{\beta}_2, \hat{\beta}_4) = (4, 2)$ ó $(\hat{\beta}_2, \hat{\beta}_4) = (4, -3)$.
- ¿Cuál sería la interpretación de β_1 ? ¿Esta interpretación cambia si se centra la variable referente a los años de escolarización?
- ¿Cómo se obtendría el efecto del género sobre el salario medio?

6.- Para cada una de las Comunidades Autónomas de España se dispone de información sobre la tasa de paro, el partido que gobierna y su producto interior bruto (PIB). Se pide contestar de forma razonada las siguientes cuestiones:

- Especifique un modelo económico y econométrico que permita determinar si el PIB y el gobierno del partido NOX influye en la tasa de paro. Interprete el término independiente del modelo econométrico especificado.
- ¿Qué modelo especificaría si sospecha que el efecto del PIB sobre la tasa de paro depende de si gobierna (o no) NOX en una Comunidad Autónoma?

Soluciones

Ejercicio 1

\$XtX

	cte	Escolaridad
cte	6	67
Escolaridad	67	843

\$`inversa XtX`

	cte	Escolaridad
cte	1.4815466	-0.11775044
Escolaridad	-0.1177504	0.01054482

\$Xty

	Salario
cte	49.62
Escolaridad	625.33

\$beta

	Salario
cte	-0.1185413
Escolaridad	0.7512127

\$e

	Salario
[1,]	0.06126538
[2,]	0.63005272
[3,]	-0.08358524
[4,]	-1.08601054
[5,]	-0.47964851
[6,]	0.95792619

\$sigma

[1] 0.6837026

\$SCT

[1] 56.2512

\$SCE

[1] 53.51639

\$SCR

[1] 2.734811

\$`R2 = SCE/SCT`

[1] 0.9513822

```
$`R2 = 1 - SCR/SCT`  
[1] 0.9513822
```

```
$`R2 corregido`  
[1] 0.9392277
```

Ejercicio 2

(Caso A)

```
$beta
```

```
          y  
cte  1.800000  
V    1.918517  
Z    -2.817929
```

```
$e
```

```
          y  
[1,]  1.2654801  
[2,] -0.7426845  
[3,]  0.6105160  
[4,]  1.2382103  
[5,] -0.7727302  
[6,] -3.7808948  
[7,] -0.4276943  
[8,] -0.8000000  
[9,]  1.8276943  
[10,] 0.1808948  
[11,] 1.1727302  
[12,] 0.1617897  
[13,] 1.7894840  
[14,] 1.1426845  
[15,] -2.8654801
```

```
$sigma
```

```
[1] 3.105644
```

```
$SCT
```

```
[1] 1896.4
```

```
$SCE
```

```
[1] 1859.132
```

```
$SCR
```

```
[1] 37.26773
```

```
$`R2 = SCE/SCT`
```

```
[1] 0.9803482
```

```
$`R2 = 1 - SCR/SCT`
```

```
[1] 0.9803482
```

```
$`R2 corregido`
```

```
[1] 0.9770729
```

(Caso B)

```
$beta
```

```
          y  
cte  3.866667  
X    2.121429  
Z   -2.500000
```

```
$e
```

```
          y  
[1,]  2.983333333  
[2,]  1.361904762  
[3,] -0.759523810  
[4,] -1.880952381  
[5,] -0.002380952  
[6,] -1.123809524  
[7,] -1.245238095  
[8,] -3.866666667  
[9,]  1.511904762  
[10,] -0.609523810  
[11,]  2.269047619  
[12,]  2.147619048  
[13,] -1.973809524  
[14,] -0.095238095  
[15,]  1.283333333
```

```
$sigma
```

```
[1] 4.217063
```

```
$SCT
```

```
[1] 1385.733
```

```
$SCE
```

```
[1] 1335.129
```

```
$SCR
```

```
[1] 50.60476
```

```
$`R2 = SCE/SCT`
```

```
[1] 0.9634816
```

```
$`R2 = 1 - SCR/SCT`
```

```
[1] 0.9634816
```

```
$`R2 corregido`
```

```
[1] 0.9573952
```

(Caso C)

```
$beta
```

```
          y  
cte  2.733333  
X    1.942857  
Z   -2.990000
```

```
$e
```

```
          y  
[1,] -1.14333333  
[2,] -0.07619048  
[3,] -1.01904762  
[4,]  1.03809524  
[5,]  0.09523810  
[6,]  2.15238095  
[7,]  1.27952381  
[8,] -1.73333333  
[9,]  1.25380952  
[10,] -1.61904762  
[11,] -1.56190476  
[12,]  2.49523810  
[13,] -1.44761905  
[14,]  1.60952381  
[15,] -1.32333333
```

```
$sigma
```

```
[1] 2.667421
```

```
$SCT
```

```
[1] 1982.933
```

```
$SCE
```

```
[1] 1950.924
```

```
$SCR
```

```
[1] 32.00905
```

```
$`R2 = SCE/SCT`
```

```
[1] 0.9838577
```

```
$`R2 = 1 - SCR/SCT`
```

```
[1] 0.9838577
```

```
$`R2 corregido`
```

```
[1] 0.9811674
```