

Inferencia en modelos de regresión lineal múltiple

Seguimos caminando

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, interprete los coeficientes de las variables independientes que son significativamente distintos de cero e indique si el modelo es globalmente válido 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

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, responda de forma razonada a las siguientes cuestiones:

- Indique qué coeficientes son significativamente distintos de cero usando los intervalos de confianza de los coeficientes de las variables independientes: $\hat{\beta}_i \pm t_{n-k} \cdot \sqrt{\widehat{var}(\hat{\beta}_i)}$. Interpretelos.
- Indique qué coeficientes son significativamente distintos de cero usando los contrastes de significación individual de los coeficientes de las variables independientes: $t_{exp} = \left| \frac{\hat{\beta}_i}{\sqrt{\widehat{var}(\hat{\beta}_i)}} \right|$.
- ¿Es el modelo válido? $F_{exp} = \frac{SCE/(k-1)}{SCR/(n-k)}$.

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	0.00000000	0.00000000
X	0.00000000	0.003571429	0.00000000
Z	0.00000000	0.00000000	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	0.00000000	0.00
X	0.00000000	0.003571429	0.00
Z	0.00000000	0.00000000	0.01

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

	y
cte	41
X	544
Z	-299

3.- Un concesionario de coches desea analizar la cuantía de las ventas realizadas (en euros) a partir del número de ventas realizadas y si la venta se ha realizado en metálico. ¿Qué variables independientes influyen en la cuantía de las ventas realizadas? ¿Cómo influyen? ¿Cómo interpretaría el término independiente? ¿El modelo es válido?

Responda de forma razonada teniendo en cuenta que a partir de 210 observaciones se ha obtenido que $\beta = (66'85, 66'1, -89'97)$, $\hat{\sigma} = 164'27$, $R^2 = 0'6248$ y

$$\widehat{var}(\hat{\beta}) = \begin{pmatrix} 23'396^2 & \diamond & \diamond \\ \diamond & 3'814^2 & \diamond \\ \diamond & \diamond & 22'96^2 \end{pmatrix}.$$

4.- Se desea lanzar una campaña de publicidad sobre cierto vehículo, por lo que se analiza qué factores influyen en su consumo. Concretamente se dispone de 150 observaciones sobre el gasto promedio (**G**, en litros cada 100 km), la portencia del motor (**M**, en caballos de vapor) y velocidad máxima (**V**, en km/h). Conteste de forma razonada a las siguientes cuestiones:

- Especifique el modelo econométrico del modelo económico planteado.
- Interprete los coeficientes significativamente distintos de cero sabiendo que $\beta = (-123'72, -2'08, 3'03)$ y

$$\widehat{var}(\hat{\beta}) = \begin{pmatrix} 11'69 & \diamond & \diamond \\ \diamond & 0'00048 & \diamond \\ \diamond & \diamond & 0'0022 \end{pmatrix}.$$

- ¿El modelo es conjuntamente válido? Tenga en cuenta que $\bar{R}^2 = 0'9998$ y $\hat{\sigma}^2 = 0'005$.

5.- El bar de mi barrio desea analizar la relación entre las ventas diarias en metálico (**V**, en euros) a partir del número de compras en metálico diarias realizadas **NC**. A partir de 105 observaciones se tiene que:

$$\widehat{V} = 16'4291 + 54'746 \cdot \text{NC}, \quad \bar{R}^2 = 0'6083, \quad SCE = 2242535.$$

Responda de forma razonada las siguientes cuestiones:

- Sabiendo que las desviaciones típicas estimadas del término independiente y del número de compras son, respectivamente, 18'852 y 4'294, interprete los coeficientes significativamente distintos de cero.
- ¿Se puede afirmar, al 90 % de confianza, que cada compra en metálico suponen unos ingresos de 55 euros?
- ¿Es el modelo conjuntamente válido?

6.- Para 200 empresas se tiene información sobre su producción ($\mathbf{Y}$), trabajo ($\mathbf{L}$) y capital ($\mathbf{K}$), de forma que se obtienen los siguientes resultados:

$$\widehat{\log \mathbf{Y}} = 0'796 + 0'602 \cdot \log \mathbf{L} + 0'348 \cdot \log \mathbf{K}, \quad SCR = 0'051, \quad R^2 = 0'9509.$$

Además, se tiene que:

$$(\mathbf{X}^t \mathbf{X})^{-1} = \begin{pmatrix} 14'31 & \diamond & \diamond \\ \diamond & 0'605 & \diamond \\ \diamond & \diamond & 0'36 \end{pmatrix}.$$

Se pide contestar de forma razonada las siguientes cuestiones:

- Interprete los coeficientes significativamente distintos de cero.
- ¿Es el modelo conjuntamente válido?
- ¿Podría comprobar si los rendimientos son constantes? En caso negativo, ¿qué información le falta?

7.- Para los tres casos del segundo ejercicio contrastar, al 5 % de significación, las siguientes hipótesis:

a) $\beta_2 = \beta_3$, b) $\beta_1 + \beta_2 = 1$ y c) $\beta_1 + 2 \cdot \beta_2 = 0$, $\beta_2 - 0'5 \cdot \beta_3 = 2$.

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

\$`Desviaciones estimadas`

[1] 1.00644786 0.08490888

\$`Intervalo CTE`

[1] -2.912889 2.675806

\$`Intervalo E`

[1] 0.5154678 0.9869575

\$`Valor experimental CTE`

[1] -0.1177819

\$`Valor experimental E`

[1] 8.84728

\$`Valor teórico`

[1] 2.776445

\$`Valor experimental ANOVA`

[1] 78.27436

\$`Valor teórico`

[1] 7.708647

Ejercicio 2

(Caso A)

\$XtX

	cte	V	Z
cte	15	0	0
V	0	280	-12
Z	0	-12	88

\$`inversa XtX`

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

\$Xty

	y
cte	27
V	571
Z	-271

\$beta

	y
cte	1.800000
V	1.918517
Z	-2.817929

\$`Desviaciones estimadas`

[1] 0.4550197 0.1056257

\$`Intervalo CTE`

[1] 0.8085972 2.7914028

\$`Intervalo E`

[1] 1.688379 2.148656

\$`Valor experimental CTE`

[1] 3.955872

\$`Valor experimental E`

[1] 18.16336

\$`Valor teórico`

[1] 2.178813

\$`Valor experimental ANOVA`

[1] 299.3151

```
$`Valor teórico`  
[1] 3.885294
```

(Caso B)

```
$XtX
```

	cte	X	Z
cte	15	0	0
X	0	280	0
Z	0	0	12

```
$`inversa XtX`
```

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

```
$Xty
```

	y
cte	58
X	594
Z	-30

```
$beta
```

	y
cte	3.866667
X	2.121429
Z	-2.500000

```
$`Desviaciones estimadas`
```

```
[1] 0.5302241 0.1227230
```

```
$`Intervalo CTE`
```

```
[1] 2.711408 5.021926
```

```
$`Intervalo E`
```

```
[1] 1.854038 2.388819
```

```
$`Valor experimental CTE`
```

```
[1] 7.292514
```

```
$`Valor experimental E`
```

```
[1] 17.28631
```

```
$`Valor teórico`
```

```
[1] 2.178813
```

\$`Valor experimental ANOVA`

[1] 158.3007

\$`Valor teórico`

[1] 3.885294

(Caso C)

\$XtX

	cte	X	Z
cte	15	0	0
X	0	280	0
Z	0	0	100

\$`inversa XtX`

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

\$Xty

	y
cte	41
X	544
Z	-299

\$beta

	y
cte	2.733333
X	1.942857
Z	-2.990000

\$`Desviaciones estimadas`

[1] 0.4216966 0.0976038

\$`Intervalo CTE`

[1] 1.814535 3.652131

\$`Intervalo E`

[1] 1.730197 2.155518

\$`Valor experimental CTE`

[1] 6.481753

\$`Valor experimental E`

[1] 19.90555

\$`Valor teórico`

[1] 2.178813

\$`Valor experimental ANOVA`

[1] 365.6949

\$`Valor teórico`

[1] 3.885294

Ejercicio 7

(Caso A)

\$`Varianza de la perturbación aleatoria`

[,1]

[1,] 3.105644

\$`inversa XtX`

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

\$`Estimación coeficientes (beta)`

[,1]

cte	1.800000
V	1.918517
Z	-2.817929

\$`R (a)`

[,1] [,2] [,3]

[1,] 0 1 -1

\$`r (a)`

[1] 0

\$`Fexp (a)`

[1] 514.3874

\$`F teórico (a)`

[1] 4.747225

\$`R (b)`

[,1] [,2] [,3]

[1,] 1 1 0

\$`r (b)`

```
[1] 1
```

```
$`Fexp (b)`  
[1] 33.86959
```

```
$`F teórico (b)`  
[1] 4.747225
```

```
$`R (c)`  
      [,1] [,2] [,3]  
[1,]    1    2  0.0  
[2,]    0    1 -0.5
```

```
$`r (c)`  
      [,1]  
[1,]    0  
[2,]    2
```

```
$`Fexp (c)`  
[1] 85.24315
```

```
$`F teórico (c)`  
[1] 3.885294
```

(Caso B)

```
$`Varianza de la perturbación aleatoria`  
      [,1]  
[1,] 4.217063
```

```
$`inversa XtX`  
      cte      X      Z  
cte 0.06666667 0.00000000 0.00000000  
X   0.00000000 0.003571429 0.00000000  
Z   0.00000000 0.00000000 0.08333333
```

```
$`Estiamción coeficientes (beta)`  
      [,1]  
cte 3.866667  
X   2.121429  
Z  -2.500000
```

```
$`R (a)`  
      [,1] [,2] [,3]  
[1,]    0    1  -1
```

```
$`r (a)`
```

```
[1] 0
```

```
$`Fexp (a)`  
[1] 58.27721
```

```
$`F teórico (a)`  
[1] 4.747225
```

```
$`R (b)`  
      [,1] [,2] [,3]  
[1,]    1    1    0
```

```
$`r (b)`  
[1] 1
```

```
$`Fexp (b)`  
[1] 84.00142
```

```
$`F teórico (b)`  
[1] 4.747225
```

```
$`R (c)`  
      [,1] [,2] [,3]  
[1,]    1    2  0.0  
[2,]    0    1 -0.5
```

```
$`r (c)`  
      [,1]  
[1,]    0  
[2,]    2
```

```
$`Fexp (c)`  
[1] 98.46635
```

```
$`F teórico (c)`  
[1] 3.885294
```

(Caso C)

```
$`Varianza de la perturbación aleatoria`  
      [,1]  
[1,] 2.667421
```

```
$`inversa XtX`  
      cte      X      Z  
cte 0.06666667 0.00000000 0.00  
X   0.00000000 0.003571429 0.00
```

```

Z    0.00000000 0.000000000 0.01

$`Estiamción coeficientes (beta)`
      [,1]
cte  2.733333
X    1.942857
Z    -2.990000

$`R (a)`
      [,1] [,2] [,3]
[1,]    0    1   -1

$`r (a)`
[1] 0

$`Fexp (a)`
[1] 672.1714

$`F teórico (a)`
[1] 4.747225

$`R (b)`
      [,1] [,2] [,3]
[1,]    1    1    0

$`r (b)`
[1] 1

$`Fexp (b)`
[1] 72.13263

$`F teórico (b)`
[1] 4.747225

$`R (c)`
      [,1] [,2] [,3]
[1,]    1    2  0.0
[2,]    0    1 -0.5

$`r (c)`
      [,1]
[1,]    0
[2,]    2

$`Fexp (c)`
[1] 126.5615

$`F teórico (c)`

```

[1] 3.885294