

εσταδιστιχ̄

εσταδιστιχ̄

Introducción a la Econometría

UOC

εσταδιστιχ̄

Apuntes



0. RECORDATORIO DEL MODELO DE REGRESIÓN

estadistix



1. MODELO DE REGRESIÓN LINEAL MÚLTIPLE: ESPECIFICACIÓN, ESTIMACIÓN Y CONTRASTE

La Econometría es la rama de la Economía que se ocupa del análisis cuantitativo de los fenómenos económicos reales, siguiendo un enfoque probabilístico. Trata de encontrar, cuantificar y contrastar relaciones económicas entre variables utilizando la modelización matemática y las técnicas basadas en los métodos inferenciales de la estadística. El Econometría se nutre de tres disciplinas: Teoría Económica, Matemáticas y Estadística.

Etapas en la investigación econométrica:

Especificación:

Estimación:

Inferencia o validación:

Predicción:

Estimación del modelo de regresión lineal múltiple

Notación algebraica $y = \beta_1 + \beta_2 x_2 + \beta_3 x_3 + \dots + \beta_k x_k + u_i \quad i = 1, 2 \dots n$

Notación matricial $Y = X\beta + u$

$$Y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} \quad X = \begin{bmatrix} 1 & x_{21} & \dots & x_{k1} \\ 1 & x_{22} & & x_{k2} \\ \vdots & \vdots & & \vdots \\ 1 & x_{2n} & & x_{kn} \end{bmatrix} \quad \beta = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{bmatrix} \quad u = \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{bmatrix}$$

Para encontrar los coeficientes del modelo utilizamos el método de los **Mínimos Cuadrados Ordinarios** (MCO), que busca que la recta obtenida genere residuos lo más pequeños posible. Matricialmente, los coeficientes se calculan:

$$\hat{\beta}_{MCO} = (X'X)^{-1}(X'Y)$$

Modelo poblacional

$$y = \beta_1 + \beta_2 x_2 + \beta_3 x_3 + \dots + \beta_k x_k + u_i$$

Modelo muestral

$$y = \hat{\beta}_1 + \hat{\beta}_2 x_2 + \hat{\beta}_3 x_3 + \dots + \hat{\beta}_k x_k + \hat{u}_i$$

Error de predicción (residual, error o término de perturbación): $\hat{u}_i = e_i = y_i - \hat{y}_i$

Coefficiente de determinación o de bondad del ajuste:

$$R^2 = \frac{SQR}{SQT} = 1 - \frac{SQE}{SQT} \quad \bar{R}^2 = 1 - \left[\frac{N-1}{N-k} \cdot (1 - R^2) \right]$$

Ejemplo R:

$$Nota = \beta_1 + \beta_2 \cdot Asistencia + \beta_3 \cdot Edad + u_i$$

```

      Asistencia  Edad      Notas_final
Asistencia  1.0000000 -0.3601305  0.4851734
Edad        -0.3601305  1.0000000 -0.2902839
Notas_final  0.4851734 -0.2902839  1.0000000

Call:
lm(formula = Notas_final ~ Asistencia, data = Estadistica)

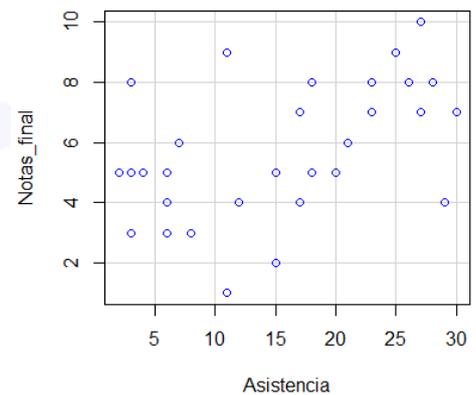
Residuals:
    Min       1Q   Median       3Q      Max
-3.987 -1.109 -0.201  1.149  4.013

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  3.65183    0.66822   5.465 0.00000467 ***
Asistencia   0.12139    0.03808   3.187  0.00314 **
Edad        -0.07256    0.14876  -0.4877  0.63410 **

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

Residual standard error: 1.918 on 33 degrees of freedom
Multiple R-squared:  0.2354, Adjusted R-squared:  0.2122
F-statistic: 10.16 on 1 and 33 DF, p-value: 0.003136

```



INFERENCIA EN EL MODELO DE REGRESIÓN**Contraste de significación individual:**

$$\begin{cases} H_0 \beta_j = 0 & \text{Relación no significativa} \\ H_1 \beta_j \neq 0 & \text{Relación sí significativa} \end{cases}$$

$$EC = \frac{\hat{\beta}_j}{S_{\hat{\beta}_j}} = \sim t_{n-k}$$

Contraste de significación global:

$$\begin{cases} H_0 \beta_1 = \beta_2 = \dots = 0 & \text{Modelo no significativo} \\ H_1 \text{ No } H_0 & \text{Modelo sí significativo} \end{cases}$$

$$EC = \frac{\frac{SCR}{k-1}}{\frac{SCE}{n-k}} = \sim F_{k-1, n-k}$$

Ejemplo: con el modelo anterior, calcula la significación individual y conjunta con un NC del 95%. Interpreta los resultados

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	5.8865	4.3098	1.366	0.402
Ausencias	-0.6965	0.3538	-1.969	0.299
Horas.de.estudio	0.3537	0.2251	1.572	0.361

Residual standard error: 1.168 on 1 degrees of freedom
Multiple R-squared: 0.94, Adjusted R-squared: 0.82
F-statistic: 7.836 on 2 and 1 DF, p-value: 0.2449

ESTADISTIX



EJERCICIOS PEC1

Enunciado

Prueba de evaluación continua 1. Modelo de regresión lineal múltiple

Criterios de evaluación: Cada pregunta vale 1 punto.

El *Bureau of Labour Statistics* de los Estados Unidos de América recopila datos cada trimestre sobre los gastos de consumo de las familias. Podéis encontrar los datos públicos en la siguiente dirección: https://www.bls.gov/cex/pumd_data.htm Esta información tiene mucha complejidad y sirve para aproximar el consumo de todas las familias americanas en general, a partir de una encuesta y un seguimiento diario de un pequeño grupo de familias aleatorio. Para poder extraer conclusiones, hay que tener en cuenta la población de los estados y las zonas donde viven las familias seleccionadas. En este ejercicio, simplificaremos el análisis para ver el comportamiento de tan solo un subconjunto de 4411 familias que participan en la encuesta global. Nosotros en este ejercicio no extrapolaremos los resultados.

Cada familia designa el llamado “cabeza de familia” que es quien se toma como persona de referencia. Estamos interesados en ver **qué determina el salario de la persona de referencia en este grupo de familias observadas en 2020**. Para explicar este salario anual expresado en dólares, usaremos un conjunto de variables que se encuentran a la Tabla 1.

Tabla 1. Descripción de las variables

Tipo	Nombre de las variables	Descripción
Dependiente	SALARY	Salario anual expresado en dólares de la persona de referencia en la familia (puede haber casos con salarios igual a cero, porque los miembros quizás disponen de otras rentas)
Independientes	EDUC	Años de escolarización de la persona de referencia
	AGE	Edad de la persona de referencia
	FEMALE	Vale 1 si la persona de referencia es una mujer, 0 en caso contrario
	CHILDAGE	Indica si en la familia conviven hijos con la siguiente codificación (0=Sin hijos 1=El hijo mayor tiene menos de 6 años 2=El hijo mayor tiene de 6-11 años y al menos un hijo < 6 años 3=Todos los hijos entre 6 y 11 años 4=El hijo mayor de 12-17 años y al menos un hijo < 12 años 5=Todos los hijos de 12-17 años 6=El hijo mayor > 17 años y al menos un hijo < 18 años 7=Todos los hijos > 17 años)
	REGION	Indica la región de residencia con la siguiente codificación (1=Northeast, 2=Midwest, 3=South, 4=West)
	STATE	Indica el estado de residencia de la familia (2 dígitos)
	FINCBTAX	Ingresos anuales totales de la familia antes de impuestos expresados en dólares

PEC1. Modelo de regresión lineal múltiple

Donde sea necesario, suponed un nivel de significación del 5%.

1. Leed los datos que se encuentran en el fichero adjunto (ya os las damos preparados en R **FAMILIES.Rdata** y también los tenéis en Excel por si os va mejor **FAMILIES.xlsx** o bien **FAMILIES.csv**), usas lo programa *R-Commander*.
 - 1.1. Calculad la media de la variable **SALARY**. ¿Cuánto gana anualmente un cabeza de familia en promedio? ¿Y al mes?

2. Estimad el siguiente modelo de regresión lineal múltiple (Modelo 1) y contestad a las siguientes preguntas:

$$SALARY_i = \beta_1 + \beta_2 EDUC_i + u_i, \quad i = 1, \dots, 4411$$

- 2.1. Copiad en la hoja de respuestas los resultados que proporciona lo programa *R-Commander* (también lo podéis hacer con R si usáis la función `lm`).
- 2.2. ¿Tiene la variable explicativa que hemos incluido en el modelo un efecto significativo sobre la variable "SALARY"? Justificad la respuesta.
- 2.3. Interpretad el coeficiente asociado a la variable **EDUC**.
- 2.4. ¿El modelo es globalmente significativo?
- 2.5. ¿Cuál es el coeficiente de determinación del modelo? ¿Qué interpretación tiene? ¿Consideráis que este es un modelo con un buen ajuste?

3. Añadid al modelo las variables **AGE** y **FINCBTAX**. Estimad el siguiente modelo de regresión lineal múltiple (Modelo 2),

$$\text{SALARY}_i = \beta_1 + \beta_2 \text{EDUC}_i + \beta_3 \text{AGE}_i + \beta_4 \text{FINCBTAX}_i + u_i, \quad i = 1, \dots, 4411$$

- 3.1. Copiad en la hoja de respuestas los resultados que proporciona el programa *R-Commander* (también lo podéis hacer con R si usáis la función `lm`).
- 3.2. ¿Cuál de los dos modelos es preferible (Modelo 1 pregunta 2 o Modelo 2 pregunta 3) ? Justificad la respuesta.
- 3.3. ¿Por qué creéis que el signo de la variable **AGE** es negativo?
- 3.4. ¿Hay alguna variable incluida en el modelo que no tenga un efecto significativo sobre el salario de la persona de referencia?



Consulta todos nuestros cursos actualizados en:

www.estadistix.com

**Y si tienes cualquier consulta,
escribenos un whatsapp al 644310902**

