Econometría

Marcos Bujosa

Universidad Complutense de Madrid

20/11/2023

1 Introducción: ¿Por qué modelar?

Modelado consiste en intentar ajustar un modelo matemático (estadístico) a un conjunto de datos ("la muestra").

El modelo es útil cuando (pese a ser *simple*) *capta las características* de los datos que consideramos más interesantes.

Los objetivos por los que se construyen modelos son variados:

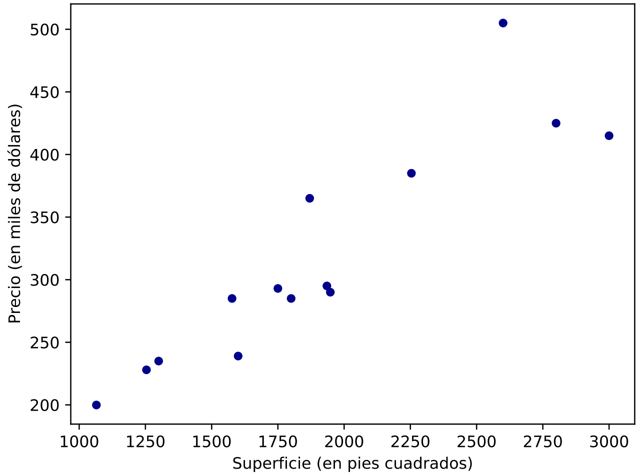
- Estimación
- Previsión
- Simulación
- Control

2 Algunos ejemplos

- **Estimación:**
  sensibilidad de un valor financiero a movimientos de un índice de referencia (evaluación de exposición al riesgo y cobertura con derivados sobre el índice)
- **Previsiones:**
  probabilidad de impago de préstamos (función de las características de la operación y del solicitante)
- **Simulación:**
  rendimiento de una cartera de valores en diferentes escenarios
- **Control:**
  bancos centrales: intervención de tipos para controlar la inflación

Lección 1

1 ¿Hay relación entre tamaño y precio de una vivienda?



Ejemplo: Función de consumo

Suponga que consumo (con) y renta disponible (rd) de las familias siguen la relación:

$$con \;=\; \beta_1 \;+\; \beta_2\; rd \;+\; otrascosas$$

Disponiendo datos de *consumo* y *renta disp.* de N familias como vectores de $\mathbb{R}^N$, *podemos construir una aproximación* ($\widetilde{\boldsymbol{con}}$) *del consumo* con una combinación lineal de la renta disponible ($\boldsymbol{rd}$) y de un término cte. ($\mathbf{1}$) (ignorando las $otrascosas$):

$$\widetilde{\boldsymbol{con}} \;=\; \widetilde{\beta}_1\mathbf{1} \;+\; \widetilde{\beta}_2\boldsymbol{rd} \;=\; \begin{bmatrix}\mathbf{1}; & \boldsymbol{rd};\end{bmatrix}\begin{pmatrix}\widetilde{\beta}_1\\ \widetilde{\beta}_2\end{pmatrix}.$$

Nomenclatura

- *regresando*: vector de datos de *consumo* ($\boldsymbol{con}$)
- *regresores*: vector de unos ($\mathbf{1}$) y de rentas disp. ($\boldsymbol{rd}$): $\mathbf{X} = \begin{bmatrix}\mathbf{1}; & \boldsymbol{rd};\end{bmatrix}$.
- *vector de parámetros*: $\widetilde{\boldsymbol{\beta}} = \begin{pmatrix}\widetilde{\beta}_1\\ \widetilde{\beta}_2\end{pmatrix}$

Otro ejemplo: Un modelo para los salarios

$salario \;=\; \beta_1 \;+\; \beta_2\; educ \;+\; \beta_3\; exper \;+\; \beta_4\; IQ \;+\; otrascosas;$

(disponiendo de datos de N trabajadores) el **ajuste** es

$$\widetilde{\boldsymbol{salario}} \;=\; \widetilde{\beta}_1\mathbf{1} \;+\; \widetilde{\beta}_2\boldsymbol{educ} \;+\; \widetilde{\beta}_3\boldsymbol{exper} \;+\; \widetilde{\beta}_4\boldsymbol{iq}$$

2 Ajuste MCO: función lineal en los parámetros

La aproximación $\widetilde{\boldsymbol{y}}$ es una combinación lineal de los *regresores* $\mathbf{X}_{|j}$:

$$\begin{pmatrix}\widetilde{y_1}\\ \vdots\\ \widetilde{y_N}\end{pmatrix} \;=\; \widetilde{\beta_1}\begin{pmatrix}1\\ \vdots\\ 1\end{pmatrix} + \widetilde{\beta_2}\begin{pmatrix}x_{12}\\ \vdots\\ x_{N2}\end{pmatrix} + \cdots + \widetilde{\beta_k}\begin{pmatrix}x_{1k}\\ \vdots\\ x_{Nk}\end{pmatrix}$$

ó

$$\begin{aligned}\widetilde{\boldsymbol{y}} \;&=\; \widetilde{\beta_1}\,\mathbf{1} \;+\; \widetilde{\beta_2}\,\mathbf{X}_{|2} \;+\; \widetilde{\beta_3}\,\mathbf{X}_{|3} \;+\cdots+\; \widetilde{\beta_k}\,\mathbf{X}_{|k}\\ &=\; \begin{bmatrix}\mathbf{1}; & \mathbf{X}_{|2}; \dots & \mathbf{X}_{|k};\end{bmatrix}\begin{pmatrix}\widetilde{\beta_1}\\ \vdots\\ \widetilde{\beta_k}\end{pmatrix} \;=\; \mathbf{X}\widetilde{\boldsymbol{\beta}};\end{aligned}$$

Así los valores ajustados son $\quad \widetilde{\boldsymbol{y}} = \mathbf{X}\widetilde{\boldsymbol{\beta}} \in \mathbb{R}^N$

Ejemplo

Precio de las viviendas: Precios de venta y Superficie útil de 14 casas unifamiliares en *University City*. San Diego, California. Año 1990. (Ramanathan, 2002, pp. 78).

n	price ($\boldsymbol{y}$)	sqft ($\boldsymbol{x}$)	$\widetilde{\text{price}}$ ($\widetilde{\boldsymbol{y}}$)
1	199.9	1065	?
2	228.0	1254	?
3	235.0	1300	?
4	285.0	1577	?
5	239.0	1600	?
6	293.0	1750	?
7	285.0	1800	?
8	365.0	1870	?
9	295.0	1935	?
10	290.0	1948	?
11	385.0	2254	?
12	505.0	2600	?
13	425.0	2800	?
14	415.0	3000	?

Tabla: Precio (miles de dólares) y superficie (pies al cuadrado). Ramanathan (2002, pp. 78).

Si asumimos que el precio y se relaciona con la superficie x del siguiente modo:

$$y_n = a + b\, x_n + otrascosas_n,$$

podemos "*aproximar*" el vector de precios, $\boldsymbol{y}$, con una combinación lineal de los regresores:

$$\widetilde{\boldsymbol{y}} = \widetilde{\beta_1}\,\mathbf{1} + \widetilde{\beta_2}\,\boldsymbol{x} \quad = \quad \begin{bmatrix}\mathbf{1}; & \boldsymbol{x};\end{bmatrix}\begin{pmatrix}\widetilde{\beta_1}\\ \widetilde{\beta_2}\end{pmatrix} = \quad \mathbf{X}\widetilde{\boldsymbol{\beta}}.$$

De esta manera,

$$\widetilde{\boldsymbol{y}} = (\mathbf{X}_{|1})\widetilde{\beta_1} + (\mathbf{X}_{|2})\widetilde{\beta_2} = \begin{pmatrix}\widetilde{y_1}\\ \widetilde{y_2}\\ \widetilde{y_3}\\ \widetilde{y_4}\\ \widetilde{y_5}\\ \widetilde{y_6}\\ \widetilde{y_7}\\ \widetilde{y_8}\\ \widetilde{y_9}\\ \widetilde{y_{10}}\\ \widetilde{y_{11}}\\ \widetilde{y_{12}}\\ \widetilde{y_{13}}\\ \widetilde{y_{14}}\end{pmatrix} = \begin{pmatrix}1\\1\\1\\1\\1\\1\\1\\1\\1\\1\\1\\1\\1\\1\end{pmatrix}\widetilde{\beta_1} + \begin{pmatrix}1065\\1254\\1300\\1577\\1600\\1750\\1800\\1870\\1935\\1948\\2254\\2600\\2800\\3000\end{pmatrix}\widetilde{\beta_2} = \begin{bmatrix}1 & 1065\\1 & 1254\\1 & 1300\\1 & 1577\\1 & 1600\\1 & 1750\\1 & 1800\\1 & 1870\\1 & 1935\\1 & 1948\\1 & 2254\\1 & 2600\\1 & 2800\\1 & 3000\end{bmatrix}\begin{pmatrix}\widetilde{\beta_1}\\ \widetilde{\beta_2}\end{pmatrix} = \mathbf{X}\widetilde{\boldsymbol{\beta}};$$

así por ejemplo, el precio ajustado para el séptimo piso de la muestra sería

$$\widetilde{y_7} \quad = \quad (1)\widetilde{\beta_1} + (1800)\widetilde{\beta_2} \quad = \quad (1,\ 1800,)\begin{pmatrix}\widetilde{\beta_1}\\ \widetilde{\beta_2}\end{pmatrix} \quad = \quad {}_{7|}\mathbf{X}\widetilde{\boldsymbol{\beta}} \quad = \quad {}_{7|}\widetilde{\boldsymbol{y}}.$$

La cuestión es:

$$\boxed{\text{¿qué criterio empleamos para elegir } \widetilde{\beta_1} \text{ y } \widetilde{\beta_2} \text{ en el ajuste } \widetilde{\boldsymbol{y}} = \mathbf{X}\widetilde{\boldsymbol{\beta}}\text{?}}$$

3 Error de ajuste

Dados $\mathbf{X}$ e $\boldsymbol{y}$, el *"error de ajuste"* al emplear $\widetilde{\boldsymbol{\beta}}$ es

$$\widetilde{\boldsymbol{e}} \quad = \quad \boldsymbol{y} - \mathbf{X}\widetilde{\boldsymbol{\beta}} \quad = \quad \boldsymbol{y} - \widetilde{\boldsymbol{y}};$$

Así, descomponemos los datos observados $\boldsymbol{y}$ en: $\boxed{\boldsymbol{y} = \widetilde{\boldsymbol{y}} + \widetilde{\boldsymbol{e}}}$.

Llamamos "Suma de los Residuos al Cuadrado" del ajuste $\widetilde{\boldsymbol{y}}$ a

$$SRC(\widetilde{\boldsymbol{\beta}}) \quad \equiv \quad \sum_{n=1}^{N} \widetilde{e_n}^2 \quad = \quad \widetilde{\boldsymbol{e}} \cdot \widetilde{\boldsymbol{e}} \quad = \quad \|\widetilde{\boldsymbol{e}}\|^2$$

es decir, al cuadrado de la longitud del vector $\widetilde{\boldsymbol{e}} = (\boldsymbol{y} - \widetilde{\boldsymbol{y}})$.

4 Criterio de ajuste MCO

Suponga $\boldsymbol{y} = \begin{pmatrix}y_1\\ y_2\\ \vdots\\ y_N\end{pmatrix}$ y $\mathbf{X} = \begin{bmatrix}1 & x_1\\ 1 & x_2\\ \vdots & \vdots\\ 1 & x_N\end{bmatrix}$.

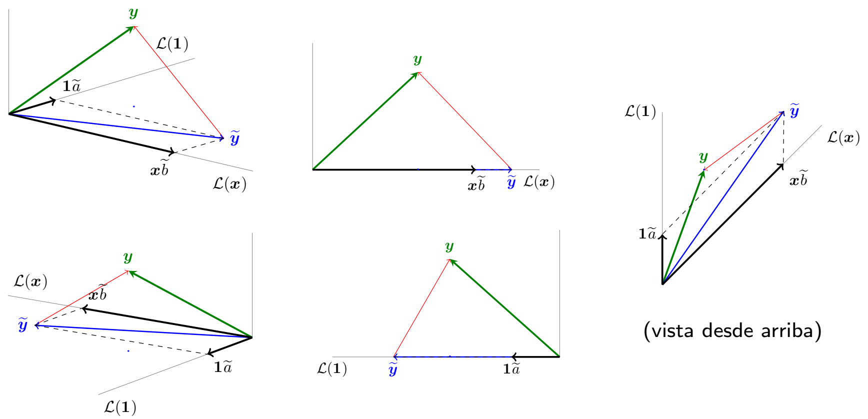
Como "criterio de ajuste" buscaremos un $\widetilde{\boldsymbol{\beta}}$ tal que $\mathbf{X}\widetilde{\boldsymbol{\beta}}$ esté *lo más próximo posible* a $\boldsymbol{y}$; es decir, tal que

la componente $\widetilde{\boldsymbol{e}}$ sea lo más pequeña posible en la descomposición:

$$\begin{aligned}\boldsymbol{y} &= \mathbf{X}\widetilde{\boldsymbol{\beta}} + \widetilde{\boldsymbol{e}}\\ &= \quad \widetilde{\boldsymbol{y}} + \widetilde{\boldsymbol{e}}.\end{aligned}$$

5 Geometría de un mal ajuste lineal

Un $\widetilde{a}$ demasiado pequeño y un $\widetilde{b}$ demasiado grande.



(vista desde arriba)

$$\mathbf{X} = \Big[\mathbf{1};\ \boldsymbol{x};\Big];\quad \widetilde{\boldsymbol{\beta}} = \begin{pmatrix}\widetilde{a}\\ \widetilde{b}\end{pmatrix};\quad \boxed{\widetilde{\boldsymbol{y}} = \mathbf{X}\widetilde{\boldsymbol{\beta}}};\quad \boldsymbol{y} = \widetilde{\boldsymbol{y}} + \widetilde{\boldsymbol{e}};\quad \widetilde{\boldsymbol{e}} = \boldsymbol{y} - \widetilde{\boldsymbol{y}}$$

6 Ecuaciones normales

El vector $\widehat{\boldsymbol{e}}$ es mínimo cuando es *perpendicular* a cada regresor:

$$\widehat{\boldsymbol{e}} \perp \mathbf{X}_{|j} \quad \Leftrightarrow \quad \mathbf{0} = \mathbf{X}^{\intercal}\widehat{\boldsymbol{e}} = \mathbf{X}^{\intercal}(\boldsymbol{y} - \widehat{\boldsymbol{y}}).$$

Consecuentemente

$$\widehat{\boldsymbol{y}} = \mathbf{X}\widehat{\boldsymbol{\beta}} \quad \Leftrightarrow \quad \mathbf{X}^{\intercal}\big(\boldsymbol{y} - \mathbf{X}\widehat{\boldsymbol{\beta}}\big) = \mathbf{0} \quad \Leftrightarrow \quad \mathbf{X}^{\intercal}\boldsymbol{y} - \mathbf{X}^{\intercal}\mathbf{X}\widehat{\boldsymbol{\beta}} = \mathbf{0}$$

Es decir

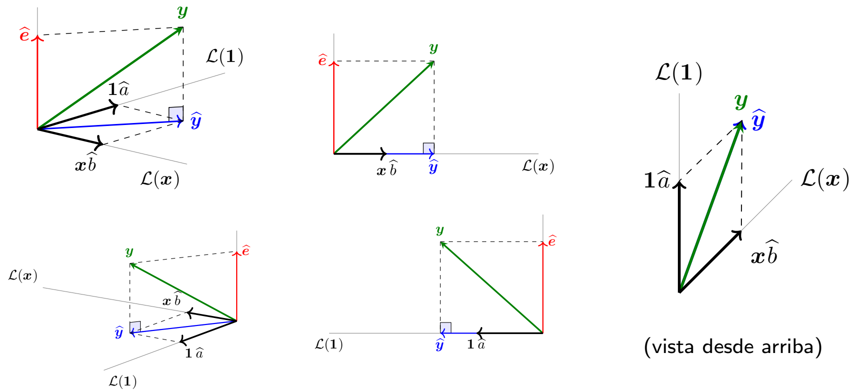
$$\widehat{\boldsymbol{y}} = \mathbf{X}\widehat{\boldsymbol{\beta}} \qquad \text{si y solo si} \qquad \boxed{(\mathbf{X}^{\intercal}\mathbf{X})\widehat{\boldsymbol{\beta}} = \mathbf{X}^{\intercal}\boldsymbol{y}} \tag{1}$$

$$\boxed{\text{Las soluciones } \widehat{\boldsymbol{\beta}} \text{ son los parámetros del ajuste MCO } \widehat{\boldsymbol{y}} = \mathbf{X}\widehat{\boldsymbol{\beta}}}$$

(el ajuste que minimiza la longitud de $\widehat{\boldsymbol{e}}$).

7 Ajuste MCO: geometría de la proyección ortogonal

$$\widehat{\boldsymbol{e}} \perp \mathbf{X} \quad \Longleftrightarrow \quad \widehat{\boldsymbol{\beta}} = \begin{pmatrix}\widehat{a}\\ \widehat{b}\end{pmatrix} \text{ es solución de } \boxed{\mathbf{X}^{\intercal}\mathbf{X}\widehat{\boldsymbol{\beta}} = \mathbf{X}^{\intercal}\boldsymbol{y}.}$$



(vista desde arriba)

$$\mathbf{X} = \Big[\mathbf{1};\ \boldsymbol{x}\Big];\quad \widehat{\boldsymbol{\beta}} = \begin{pmatrix}\widehat{a}\\ \widehat{b}\end{pmatrix};\quad \boxed{\widehat{\boldsymbol{y}} = \mathbf{X}\widehat{\boldsymbol{\beta}}};\quad \boldsymbol{y} = \widehat{\boldsymbol{y}} + \widehat{\boldsymbol{e}};\quad \widehat{\boldsymbol{e}} = \boldsymbol{y} - \widehat{\boldsymbol{y}}$$

8 Condición para que las ecuaciones normales tengan solución única

Puesto que

$$\mathbf{X}\widehat{\boldsymbol{\beta}} = \widehat{\boldsymbol{y}} \quad \Longleftrightarrow \quad (\mathbf{X}^{\intercal}\mathbf{X})\widehat{\boldsymbol{\beta}} = \mathbf{X}^{\intercal}\boldsymbol{y}, \qquad \text{donde } \underset{N\times k}{\mathbf{X}};$$

ambos sistemas tendrán *solución única si y sólo* si sus matrices de coeficientes son de *rango k*.

En tal caso, multiplicando ambos lados de las ecuaciones normales por $\left(\mathbf{X}^{\intercal}\mathbf{X}\right)^{-1}$ tenemos que

$$\boxed{\widehat{\boldsymbol{\beta}} = (\mathbf{X}^{\intercal}\mathbf{X})^{-1}\mathbf{X}^{\intercal}\boldsymbol{y}} \tag{2}$$

es *la única solución*.

Ejemplo

Ecuación de salarios: Supongamos el siguiente modelo (Ejemplo 3.2. Wooldridge, 2006)

$$Salar_n = e^{\big(\beta_1 + \beta_2(educ_n) + \beta_3(antig_n) + \beta_4(exper_n) + otrascosas_n\big)};$$

Tomando logaritmos tenemos un modelo para la nueva variable $\ln(Salar_n)$

$$\ln(Salar_n) = \beta_1 + \beta_2\big(educ_n\big) + \beta_3\big(antig_n\big) + \beta_4\big(exper_n\big) + otrascosas_n.$$

¿Qué pasa si jamás ningún trabajador cambió de empresa?
Como *experiencia* y *antigüedad* coinciden, sólo podemos calcular su efecto conjunto:

$$\ln(Salar_n) = \beta_1 + \beta_2\big(educ_n\big) + (\beta_3 + \beta_4)exper_n + otrascosas_n,$$

Enlace a algunas **prácticas de la Lección 1**

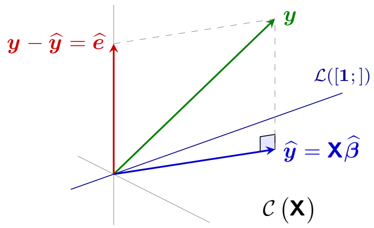
Fin de la lección

Lección 2

1 Geometría MCO

El ajuste de regresión MCO es una descomposición ortogonal:

$$\boldsymbol{y} = \widehat{\boldsymbol{y}} + \widehat{\boldsymbol{e}}; \quad \text{donde} \quad \widehat{\boldsymbol{y}} = \mathbf{X}\widehat{\boldsymbol{\beta}} \perp \widehat{\boldsymbol{e}}$$



donde los parámetros $\widehat{\boldsymbol{\beta}}$ se obtienen resolviendo $\boxed{\mathbf{X}^{\intercal}\mathbf{X}\ \widehat{\boldsymbol{\beta}} = \mathbf{X}^{\intercal}\boldsymbol{y}}$ y donde $\mathbf{1} \in \mathcal{C}\big(\mathbf{X}\big)$.

2 Ajuste MCO con una constante como único regresor

¿Qué es el ajuste MCO $\widehat{\boldsymbol{y}}$ si $\mathbf{X} = \begin{bmatrix}\mathbf{1}\end{bmatrix}$? $(Y_n = a + otrascosas_n)$
Las ecuaciones normales

$$\mathbf{X}^\intercal\mathbf{X}\widehat{\boldsymbol{\beta}} = \mathbf{X}^\intercal\boldsymbol{y}$$

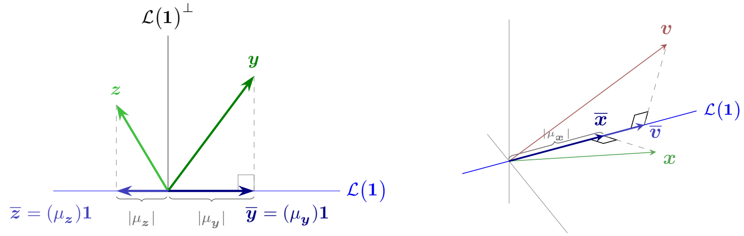
se reducen a una única ecuación

$$\begin{bmatrix}\mathbf{1}\cdot\mathbf{1}\end{bmatrix}\begin{pmatrix}\widehat{a},\end{pmatrix} = \begin{pmatrix}\mathbf{1}\cdot\boldsymbol{y},\end{pmatrix} \quad\Longrightarrow\quad \begin{pmatrix}\widehat{a},\end{pmatrix} = \begin{bmatrix}N\end{bmatrix}^{-1}\begin{pmatrix}\mathbf{1}\cdot\boldsymbol{y},\end{pmatrix}$$

Por tanto $\quad \widehat{a} \;=\; N^{-1}(\mathbf{1}\cdot\boldsymbol{y}) \;=\; \dfrac{1}{N}\sum_{i=1}^{N} y_i \;=\; \mu_{\boldsymbol{y}};$ así que

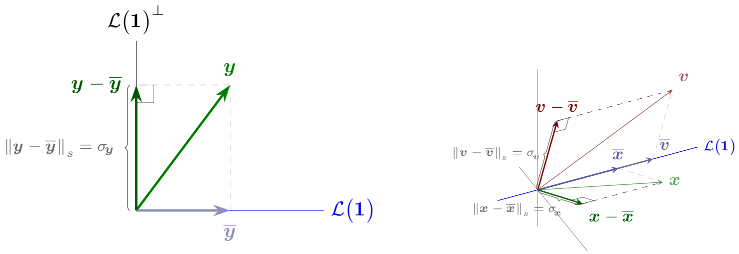
$$\widehat{\boldsymbol{y}} = \mathbf{X}\widehat{\boldsymbol{\beta}} \;=\; \begin{bmatrix}\mathbf{1}\end{bmatrix}\begin{pmatrix}\widehat{a},\end{pmatrix} \;=\; \mathbf{1}\mu_{\boldsymbol{y}} \;\equiv\; \overline{\boldsymbol{y}}\,. \qquad (3)$$

3 El vector de medias $\overline{\boldsymbol{y}}$ y la media aritmética $\mu_{\boldsymbol{y}}$



$$\overline{\boldsymbol{y}} \;=\; (\mu_{\boldsymbol{y}})\mathbf{1} \qquad \text{y} \qquad \mu_{\boldsymbol{y}} \;=\; \langle \boldsymbol{y}|\,\mathbf{1}\rangle_s.$$

4 La desviación típica y el Teorema de Pitágoras

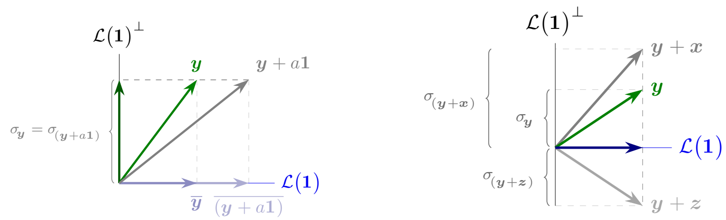


Así, $\quad \sigma_{\boldsymbol{y}}^2 \;=\; \|\boldsymbol{y}-\overline{\boldsymbol{y}}\|_s^2 \;=\; N^{-1}\sum(y_i-\mu_{\boldsymbol{y}})^2 \;=\; \mu_{((\boldsymbol{x}-\overline{\boldsymbol{x}})^2)},$

pero por el T. de Pitágoras, también

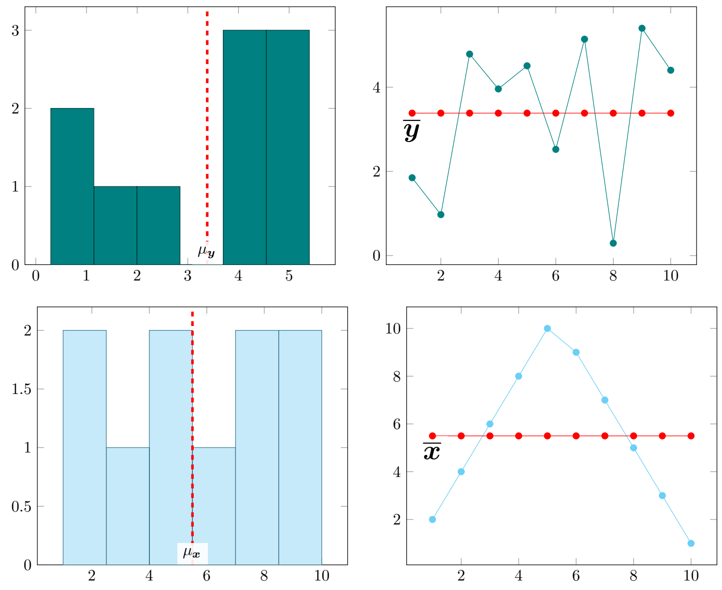
$$\sigma_{\boldsymbol{y}}^2 \;=\; \|\boldsymbol{y}\|_s^2 - \|\overline{\boldsymbol{y}}\|_s^2 \;=\; \mu_{(\boldsymbol{y}^2)} - (\mu_{\boldsymbol{y}})^2.$$

5 Vectores constantes y vectores de media nula



$$\sigma_{\boldsymbol{z}} = 0 \;\Leftrightarrow\; \boldsymbol{z} = a\mathbf{1}; \qquad \mu_{\boldsymbol{z}} = 0 \;\Leftrightarrow\; \boldsymbol{z}\perp\mathbf{1} \qquad (4)$$

6 Media: *valor puntual* y vector de medias



7 Ajuste MCO con un regresor adicional a la constante

$Y_n = a + bX_n + otrascosas_n$ (Modelo Lineal Simple).
Las ecuaciones normales

$$\mathbf{X}^\intercal\mathbf{X}\widehat{\boldsymbol{\beta}} = \mathbf{X}^\intercal\boldsymbol{y},$$

donde ahora

$$\boldsymbol{y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{pmatrix}; \quad \mathbf{X} = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_N \end{bmatrix} = \begin{bmatrix} \mathbf{1}; & \boldsymbol{x}; \end{bmatrix}; \quad \widehat{\boldsymbol{\beta}} = \begin{pmatrix} \widehat{a} \\ \widehat{b} \end{pmatrix};$$

se reducen a

$$\begin{bmatrix} (\mathbf{1}\cdot\mathbf{1}) & (\mathbf{1}\cdot\boldsymbol{x}) \\ (\boldsymbol{x}\cdot\mathbf{1}) & (\boldsymbol{x}\cdot\boldsymbol{x}) \end{bmatrix} \begin{pmatrix} \widehat{a} \\ \widehat{b} \end{pmatrix} = \begin{pmatrix} \mathbf{1}\cdot\boldsymbol{y} \\ \boldsymbol{x}\cdot\boldsymbol{y} \end{pmatrix}.$$

8 Solución para el modelo lineal simple

Para el Modelo Lineal Simple, la solución al sistema de ecuaciones normales es:

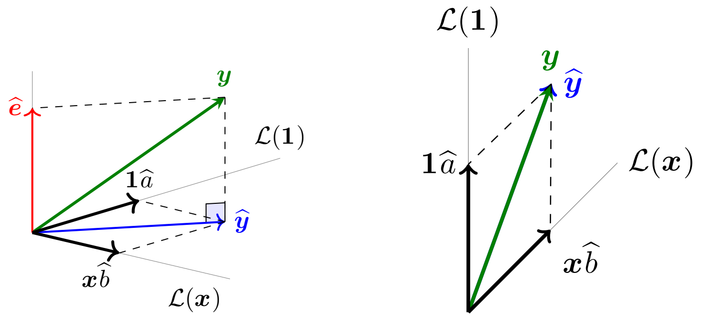
$$\boxed{\widehat{b} = \frac{\sigma_{\boldsymbol{xy}}}{\sigma^2_{\boldsymbol{x}}}} \tag{5}$$

y

$$\boxed{\widehat{a} = \mu_{\boldsymbol{y}} - \widehat{b}\,\mu_{\boldsymbol{x}}} \tag{6}$$

Multiplicando y dividiendo $\widehat{b}$ por $\sigma_{\boldsymbol{y}}$, también tenemos: $\widehat{b} = \rho_{\boldsymbol{xy}} \cdot \frac{\sigma_{\boldsymbol{y}}}{\sigma_{\boldsymbol{x}}}$

9 Ajuste del modelo lineal simple



Ejemplo

Precio de las viviendas: precio de 14 viviendas en *University City*. San Diego, California. Año 1990. (Ramanathan, 2002, pp. 78).

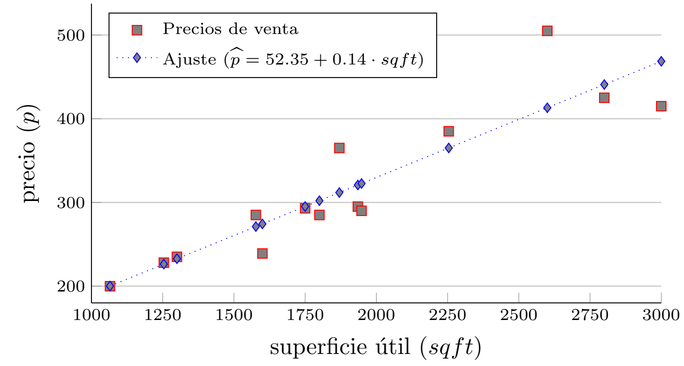
☞**Código:** `EjPvivienda.inp` ....

n	Precio ($\boldsymbol{y}$)	Superficie ($\boldsymbol{x}$)
1	199.9	1065
2	228.0	1254
3	235.0	1300
4	285.0	1577
5	239.0	1600
6	293.0	1750
7	285.0	1800
8	365.0	1870
9	295.0	1935
10	290.0	1948
11	385.0	2254
12	505.0	2600
13	425.0	2800
14	415.0	3000

Tabla: Superficie (pies al cuadrado) y precio de venta (miles de dólares)

10 Recta de regresión

Precio (miles de \$) y superficie útil (pies al cuadrado) de 14 casas unifamiliares

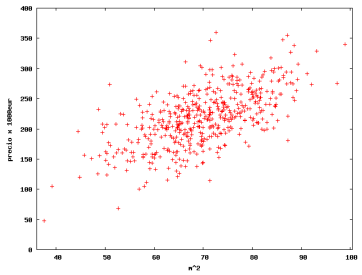


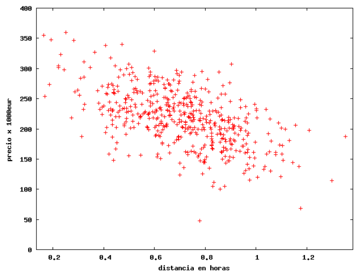
☞**Código:** `EjPvivienda.inp` ....

Ejemplo

Precio de las viviendas simulado (dos regresores):

Modelo simulado: $\boldsymbol{p} = 100 + 3\boldsymbol{s} - 130\boldsymbol{d} + \boldsymbol{u}$





☞**Código:** `SimuladorEjPvivienda.inp` ....

Enlace a algunas **prácticas de la Lección 2**

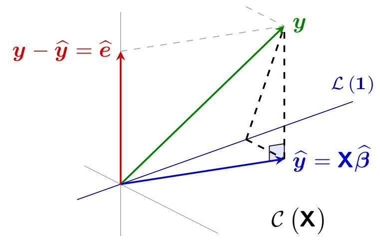
____________________________________Fin de la lección

Lección 3

1 Geometría MCO

El Ajuste por regresión MCO es una descomposición ortogonal:

$$\boldsymbol{y} = \widehat{\boldsymbol{y}} + \widehat{\boldsymbol{e}}; \qquad \text{donde} \qquad \widehat{\boldsymbol{y}} = \mathbf{X}\widehat{\boldsymbol{\beta}} \perp \widehat{\boldsymbol{e}}$$



donde los parámetros $\widehat{\boldsymbol{\beta}}$ satisfacen $\boxed{\mathbf{X}^\intercal\mathbf{X}\ \widehat{\boldsymbol{\beta}} = \mathbf{X}^\intercal\boldsymbol{y}}$ y $\boxed{\mathbf{1} \in \mathcal{C}(\mathbf{X})}$.

2 Mínimos cuadrados ordinarios: Propiedades algebraicas

El cálculo MCO de $\boldsymbol{\beta}$ en $\quad \boldsymbol{y} = \mathbf{X}\widehat{\boldsymbol{\beta}} + \widehat{\boldsymbol{e}} \quad$ implica que (F13)

$$\widehat{\boldsymbol{e}} \perp \mathbf{X}_{|j}, \quad \text{es decir} \quad \boxed{\widehat{\boldsymbol{e}}\mathbf{X} = \mathbf{0}}.$$

Y como $\widehat{\boldsymbol{y}} = \mathbf{X}\widehat{\boldsymbol{\beta}}$, entonces $\widehat{\boldsymbol{e}} \perp \widehat{\boldsymbol{y}}$ (pues $\widehat{\boldsymbol{e}} \cdot \widehat{\boldsymbol{y}} = \widehat{\boldsymbol{e}}\mathbf{X}\widehat{\boldsymbol{\beta}} = \mathbf{0} \cdot \widehat{\boldsymbol{\beta}}$);

$$\text{es decir,} \quad \widehat{\boldsymbol{e}}\mathbf{X} = \mathbf{0} \quad \Rightarrow \quad \boxed{\widehat{\boldsymbol{e}} \cdot \widehat{\boldsymbol{y}} = 0} \tag{7}$$

Y como $\mathbf{1} \in \mathcal{C}(\mathbf{X})$ tenemos que

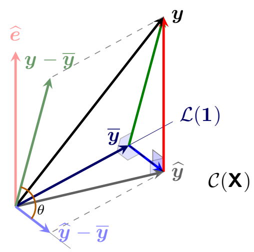
$\boxed{\mu_{\widehat{\boldsymbol{e}}} = 0}$ y por tanto $\boxed{\mu_{\boldsymbol{y}} = \mu_{\widehat{\boldsymbol{y}}}}$. Así que $\boxed{\overline{\boldsymbol{y}} = \overline{\widehat{\boldsymbol{y}}}}$.

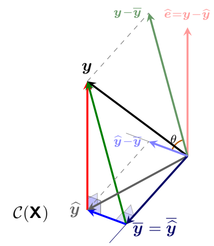
(Véase (F23) y la figura en (F33))

3 MCO: Tª de Pitágoras y sumas de cuadrados

Como $(\widehat{\boldsymbol{y}} - \overline{\boldsymbol{y}}) \perp \widehat{\boldsymbol{e}}$ y su suma es $(\boldsymbol{y} - \overline{\boldsymbol{y}}) = (\widehat{\boldsymbol{y}} - \overline{\boldsymbol{y}}) + \overbrace{\widehat{\boldsymbol{e}}}^{(\boldsymbol{y}-\widehat{\boldsymbol{y}})}$

$$\boxed{\|(\boldsymbol{y} - \overline{\boldsymbol{y}})\|^2 = \|(\widehat{\boldsymbol{y}} - \overline{\boldsymbol{y}})\|^2 + \|\widehat{\boldsymbol{e}}\|^2} \tag{8}$$





Con la norma del producto escalar usual en $\mathbb{R}^N$

$$\underbrace{\|(\boldsymbol{y} - \overline{\boldsymbol{y}})\|_u^2}_{STC} = \underbrace{\|(\widehat{\boldsymbol{y}} - \overline{\boldsymbol{y}})\|_u^2}_{SEC} + \underbrace{\|\widehat{\boldsymbol{e}}\|_u^2}_{SRC}$$

4 Sumas de cuadrados

$$\boxed{STC \;=\; SEC + SRC}$$

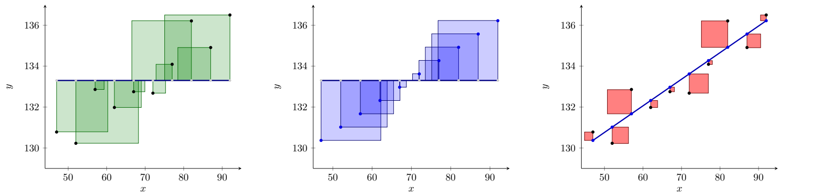
$$STC \equiv (\boldsymbol{y} - \overline{\boldsymbol{y}}) \cdot (\boldsymbol{y} - \overline{\boldsymbol{y}}) \qquad = N\sigma^2_{\boldsymbol{y}}$$

$$SEC \equiv (\widehat{\boldsymbol{y}} - \overline{\boldsymbol{y}}) \cdot (\widehat{\boldsymbol{y}} - \overline{\boldsymbol{y}}) \qquad = N\sigma^2_{\widehat{\boldsymbol{y}}} \;\; \text{(pues } \overline{\boldsymbol{y}} = \overline{\widehat{\boldsymbol{y}}}\text{)}$$

$$SRC \equiv \widehat{\boldsymbol{e}} \cdot \widehat{\boldsymbol{e}} \qquad = N\sigma^2_{\widehat{\boldsymbol{e}}} \;\; \text{(pues } \mu_{\widehat{\boldsymbol{e}}} = 0\text{)}$$

$STC = \sum (y_n - \mu_{\boldsymbol{y}})^2;$ $\quad SEC = \sum (\widehat{y_n} - \mu_{\boldsymbol{y}})^2;$ $\quad SRC = \sum (y_n - \widehat{y_n})^2$

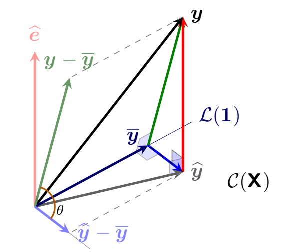
Modelo Lineal Simple

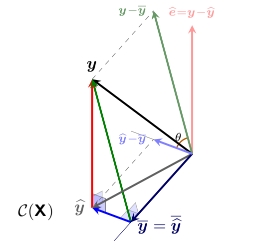


5 Dos varas para medir lo mismo: descomposición de la varianza

Veamos idéntica relación, pero medida con la norma de la estadística

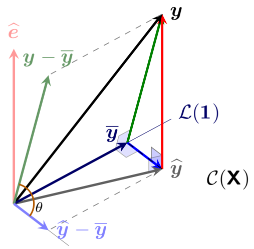
$$\underbrace{\|(\boldsymbol{y} - \overline{\boldsymbol{y}})\|_s^2}_{\sigma^2_{\boldsymbol{y}}} = \underbrace{\|(\widehat{\boldsymbol{y}} - \overline{\boldsymbol{y}})\|_s^2}_{\sigma^2_{\widehat{\boldsymbol{y}}}} + \underbrace{\|\widehat{\boldsymbol{e}}\|_s^2}_{\sigma^2_{\widehat{\boldsymbol{e}}}}$$

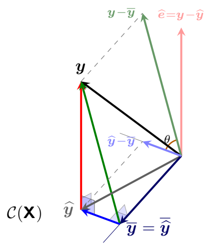




$$\boxed{STC \;=\; SEC + SRC} \xrightarrow{\text{dividiendo por } N} \boxed{\sigma^2_{\boldsymbol{y}} \;=\; \sigma^2_{\widehat{\boldsymbol{y}}} + \sigma^2_{\widehat{\boldsymbol{e}}}}.$$

6 Dos casos extremos





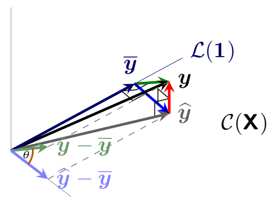
Ajuste perfecto cuando $\boldsymbol{y} \in \mathcal{C}\left(\mathbf{X}\right)$, pues $\widehat{\boldsymbol{y}} = \boldsymbol{y}$

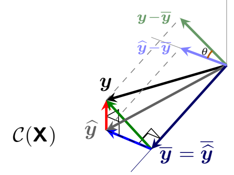
$STC = SEC$; $(SRC = 0)$ es decir $\sigma^2_{\boldsymbol{y}} = \sigma^2_{\widehat{\boldsymbol{y}}}$; $(\sigma^2_{\widehat{\boldsymbol{e}}} = 0)$.

Ajuste nulo cuando $\boldsymbol{y} \perp \mathcal{C}\left(\mathbf{X}\right)$, pues $\widehat{\boldsymbol{y}} = \mathbf{0}$

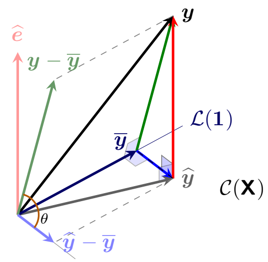
$STC = SRC$; $(SEC = 0)$ es decir $\sigma^2_{\boldsymbol{y}} = \sigma^2_{\widehat{\boldsymbol{e}}}$; $(\sigma^2_{\widehat{\boldsymbol{y}}} = 0)$.

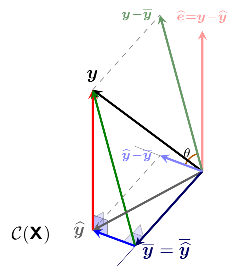
7 ¿Qué ajuste es mejor (donde se parecen más $\boldsymbol{y}$ y $\widehat{\boldsymbol{y}}$)? ¿arriba o abajo?





$$\frac{\|\widehat{\boldsymbol{y}} - \overline{\boldsymbol{y}}\|^2}{\|\boldsymbol{y} - \overline{\boldsymbol{y}}\|^2}$$





8 Medidas de ajuste

Coeficiente de determinación: R^2

$$R^2 = \frac{SEC}{STC} = 1 - \frac{SRC}{STC}; \qquad 0 \leq R^2 \leq 1$$
$$= \frac{\sigma^2_{\widehat{\boldsymbol{y}}}}{\sigma^2_{\boldsymbol{y}}} = 1 - \frac{\sigma^2_{\widehat{\boldsymbol{e}}}}{\sigma^2_{\boldsymbol{y}}} = \left(\rho_{\widehat{\boldsymbol{y}}\boldsymbol{y}}\right)^2.$$

En caso del MLS también tenemos que $R^2 = \left(\rho_{\boldsymbol{x}\boldsymbol{y}}\right)^2$

Coeficiente de determinación corregido o ajustado: $\bar{R}^2$

$$\bar{R}^2 = 1 - \frac{\mathfrak{s}^2_{\widehat{\boldsymbol{e}}}}{\mathfrak{s}^2_{\boldsymbol{y}}} = 1 - \frac{\frac{SRC}{N-k}}{\frac{STC}{N-1}} = 1 - \frac{N-1}{N-k}(1-R^2) \quad \leq 1$$

donde $\frac{SRC}{N-k} \equiv \mathfrak{s}^2_{\widehat{\boldsymbol{e}}}$ es la *cuasi*-varianza de $\widehat{\boldsymbol{e}}$; y $\frac{STC}{N-1} \equiv \mathfrak{s}^2_{\boldsymbol{y}}$ es la *cuasi*-varianza de $\boldsymbol{y}$

9 Ajuste en el ejemplo de las casas

☞Código: `EjPvivienda.inp` ....

Estimaciones MCO utilizando las 14 observaciones 1–14

Variable dependiente: price

	Coeficiente	Desv. Típica	Estadístico t	Valor p
const	$52,35$	37,29	$1,40$	0,19
sqft	$0,14$	0,02	$7,41$	0,00

Media de la vble. dep.	317,4929	D.T. de la vble. dep.	88,49816
Suma de cuad. residuos	18273,57	D.T. de la regresión	39,02304
R^2	0,820522	R^2 corregido	0,805565
$F(1,12)$	54,86051	Valor p (de F)	8,20e–06
Log-verosimilitud	−70,08421	Criterio de Akaike	144,1684
Criterio de Schwarz	145,4465	Hannan–Quinn	144,0501

Ejemplo

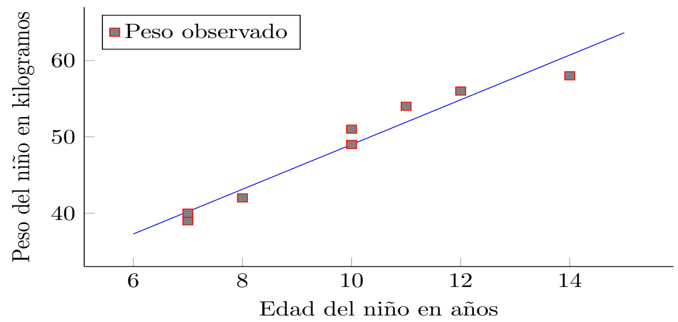
Peso de niños según su edad:

☞Código: `PesoEdad.inp` ....

n	Peso Kg	Edad
1	39	7
2	40	7
3	42	8
4	49	10
5	51	10
6	54	11
7	56	12
8	58	14

Tabla: Peso (en kilogramos) y edad (en años)

Mod 1: $\boldsymbol{peso} = \beta_1\mathbf{1} + \beta_2\boldsymbol{edad} + \boldsymbol{otrascosas}$

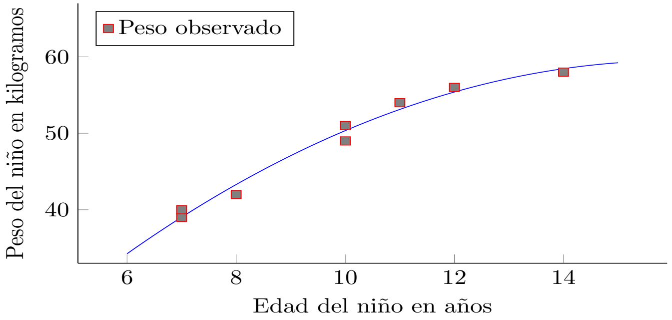


$$\widehat{\text{Peso\_Kg}} = \underset{(6,999)}{19,6910} + \underset{(10,564)}{2,93003}\,\text{Edad}$$

$T = 8 \quad \bar{R}^2 = 0,9405 \quad F(1,6) = 111,6 \quad \hat{\sigma} = 1,8161$

(entre paréntesis, los estadísticos t)

Mod 2: $\boldsymbol{peso} = \beta_1 \mathbf{1} + \beta_2 \boldsymbol{edad} + \beta_3 \boldsymbol{edad}^2 + \boldsymbol{otrascosas}$

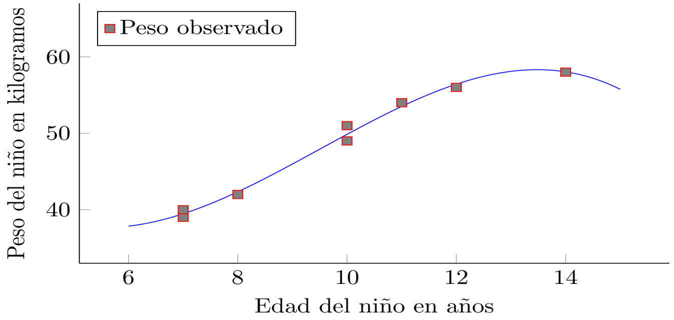


$$\widehat{\text{Peso\_Kg}} = \underset{(-0{,}664)}{-5,11497} + \underset{(5{,}159)}{8,06835}\,\text{Edad} - \underset{(-3{,}305)}{0,252102}\,\text{Edad2}$$

$$T = 8 \quad \bar{R}^2 = 0,9776 \quad F(2,5) = 153,57 \quad \hat{\sigma} = 1,1148$$

(entre paréntesis, los estadísticos t)

Mod 3: $\boldsymbol{peso} = \beta_1 \mathbf{1} + \beta_2 \boldsymbol{edad} + \beta_3 \boldsymbol{edad}^2 + \beta_4 \boldsymbol{edad}^3 + \boldsymbol{otrascosas}$



$$\widehat{\text{Peso\_Kg}} = \underset{(1{,}904)}{81,7714} - \underset{(-1{,}419)}{18,5964}\,\text{Edad} + \underset{(1{,}845)}{2,37778}\,\text{Edad2} - \underset{(-2{,}043)}{0,0836541}\,\text{Edad3}$$

$$T = 8 \quad \bar{R}^2 = 0,9863 \quad F(3,4) = 168,75 \quad \hat{\sigma} = 0,87188$$

(entre paréntesis, los estadísticos t)

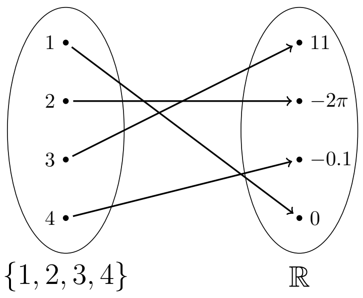
Enlace a algunas **prácticas de la Lección 3**

Fin de la lección

Lección 4

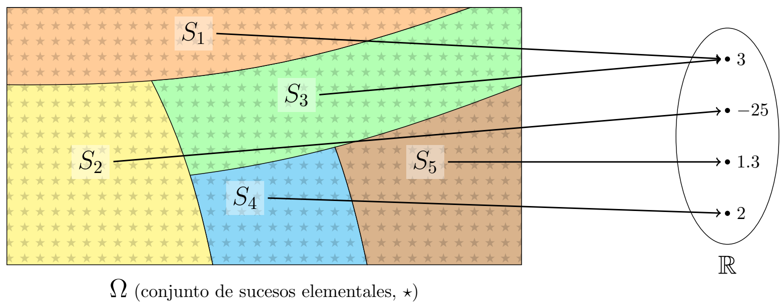
1 Los elementos de $\mathbb{R}^N$ son funciones

Vector de $\mathbb{R}^4$: $(0, \quad -2\pi, \quad 11, \quad -0.1)$



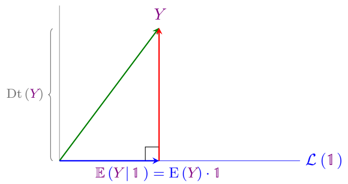
- El conjunto de vectores de $\mathbb{R}^4$ es un espacio vectorial
- Cada vector es una función que va de $\{1, 2, 3, 4\}$ a $\mathbb{R}$

2 Las variables aleatorias (VA) también son funciones



- El conjunto de VAs con varianza es un espacio vectorial
- Cada vector (VA) es una función que va del conjunto de sucesos elementales Ω a $\mathbb{R}$
- Sobre ciertos subconjuntos (S_i) se define una medida de probabilidad.

3 Geometría de los momentos teóricos: Esperanza y varianza



$\mathbb{E}(Y|\mathbb{1}) = \mathrm{E}(Y) \cdot \mathbb{1}$

$\mathrm{Var}(Y) \;=\; \mathrm{E}\left((Y - \mathrm{E}(Y)\,\mathbb{1})^2\right) \;=\; \left\|(Y - \mathrm{E}(Y)\,\mathbb{1})\right\|_\eta^2$

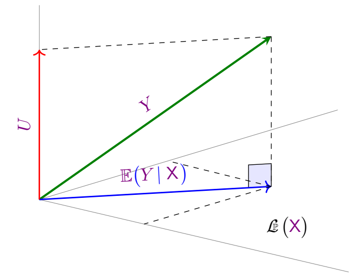
Por Pitágoras:

$$\mathrm{E}\left(Y^2\right) \;=\; \mathrm{Var}(Y) + \mathrm{E}\left(\mathrm{E}(Y)^2 \cdot (\mathbb{1})^2\right) \;=\; \mathrm{Var}(Y) + \mathrm{E}(Y)^2$$

y por tanto $\boxed{\mathrm{Var}(Y) = \mathrm{E}\left(Y^2\right) - \mathrm{E}(Y)^2}$

4 La esperanza condicional también es una proyección ortogonal

$$Y = \mathbb{E}(Y|\mathsf{X}) + U$$



Teorema de las esperanzas iteradas

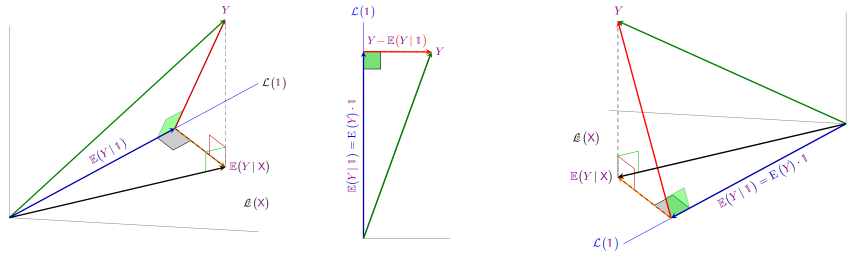
Como $\mathbb{1} \perp \left(Y - \mathbb{E}(Y|\mathsf{X})\right)$, pues $\mathbb{1} \in \mathcal{L}(\mathsf{X})$

$$\mathrm{E}\left(\mathbb{1} \cdot \left(Y - \mathbb{E}(Y|\mathsf{X})\right)\right) = 0$$

$$\mathrm{E}\left(Y - \mathbb{E}(Y|\mathsf{X})\right) = 0$$

$$\mathrm{E}(Y) - \mathrm{E}\left(\mathbb{E}(Y|\mathsf{X})\right) = 0 \quad \Rightarrow \quad \boxed{\mathrm{E}(Y) = \mathrm{E}\left(\mathbb{E}(Y|\mathsf{X})\right)}$$

5 Geometría de los momentos teóricos: Esperanza Condicional



$$\mathbb{E}\Big(\mathbb{E}\left(Y|\mathsf{X}\right)\Big|\,\mathbb{1}\Big) = \mathbb{E}\left(Y|\,\mathbb{1}\right) = \mathrm{E}\left(Y\right)\cdot\mathbb{1}$$

$$\mathrm{Var}\left(Y\right) = \mathrm{E}\Big(\left(Y - \mathrm{E}\left(Y\right)\mathbb{1}\right)^2\Big)$$

$$\mathrm{Var}\left(\mathbb{E}\left(Y|\mathsf{X}\right)\right) = \mathrm{E}\Big(\left(\mathbb{E}\left(Y|\mathsf{X}\right) - \mathrm{E}\left(Y\right)\mathbb{1}\right)^2\Big)$$

$$\mathrm{E}\left(Y^2\right) = \mathrm{Var}\left(Y\right) + \mathrm{E}\left(\mathbb{E}\left(Y|\,\mathbb{1}\right)^2\right) = \mathrm{Var}\left(Y\right) + \mathrm{E}\left(Y\right)^2$$

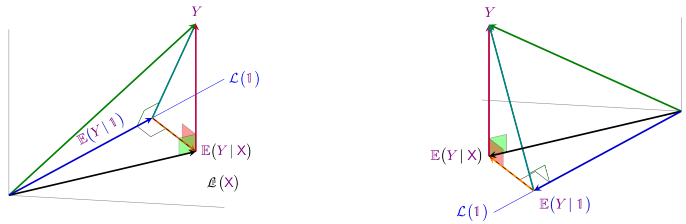
$$\mathrm{E}\left(\mathbb{E}\left(Y|\mathsf{X}\right)^2\right) = \mathrm{Var}\left(\mathbb{E}\left(Y|\mathsf{X}\right)\right) + \mathrm{E}\left(Y\right)^2$$

6 Varianza condicional

Si $\mathrm{E}\left(\left(Y - \mathbb{E}\left(Y|\mathsf{X}\right)\right)^2\right) < \infty$, entonces:

$$\mathbb{V}ar\left(Y\,|\,\mathsf{X}\right) = \mathbb{E}\Big(\left(Y - \mathbb{E}\left(Y|\mathsf{X}\right)\right)^2\Big|\,\mathsf{X}\Big)$$

por tanto $\mathrm{E}\left(\mathbb{V}ar\left(Y|\mathsf{X}\right)\right) = \mathrm{E}\left(\left(Y - \mathbb{E}\left(Y|\mathsf{X}\right)\right)^2\right)$



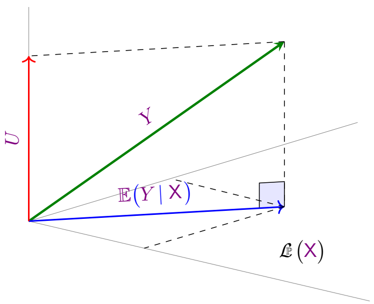
$$\mathrm{E}\left(Y^2\right) = \mathrm{E}\left(\mathbb{E}\left(Y|\mathsf{X}\right)^2\right) + \mathrm{E}\left(\mathbb{V}ar\left(Y|\mathsf{X}\right)\right)$$

Ley de la varianza total

$$\mathrm{Var}\left(Y\right) = \mathrm{Var}\left(\mathbb{E}\left(Y|\mathsf{X}\right)\right) + \mathrm{E}\left(\mathbb{V}ar\left(Y|\mathsf{X}\right)\right)$$

7 La regresión es una descomposición ortogonal (que no implica causalidad)

$$Y = \mathbb{E}\left(Y|\mathsf{X}\right) + U$$



donde $\mathbb{E}\left(Y|\mathsf{X}\right)$ es la proyección ortogonal sobre $\mathcal{L}(\mathsf{X})$.

como relación estadística: siempre es cierta. No implica causalidad ni conclusiones teóricas

como lectura teórica: su interpretación puede carecer de sentido (regresiones espurias)

8 Modelo de regresión: Nombres de las variables

En la expresión

$$Y = \mathbb{E}(Y\,|\,\mathsf{X}) + U$$

usamos los siguientes nombres:

- Y: vble. endógena, objetivo, explicada (o regresando)
- $\mathsf{X} = \left[X_1;\ X_2;\ \dots X_k\right]$: vbles. exógenas, de control, explicativas (o regresores)
- U: factor desconocido o perturbación (a mi me gusta llamarlo *"otras cosas"*)

9 Modelo Clásico de Regresión Lineal

Modelo especial en el que la descomposición ortogonal

$$Y = \mathbb{E}(Y \mid \mathsf{X}) + U$$

es tal que

- $\mathbb{E}(Y \mid \mathsf{X}) = \mathsf{X}\boldsymbol{\beta} \in \mathcal{L}(\mathsf{X})$ (función lineal)
- $\mathbb{V}ar(Y \mid \mathsf{X})$ está definida y es cte.

¿QUÉ HACE FALTA PARA QUE ESTO SE CUMPLA?

¿En qué condiciones es la recta de regresión una estimación insesgada de la esperanza condicional $\mathbb{E}(Y \mid \mathsf{X})$?

Lección 5

1 Modelo Clásico de Regresión Lineal

Modelo especial en el que la descomposición ortogonal

$$Y = \mathbb{E}(Y \mid \mathsf{X}) + U$$

es tal que

- $\mathbb{E}\left(Y \middle| \mathsf{X}\right) = \mathsf{X}\boldsymbol{\beta}$ (Comb. lin. regresores) (Sup. 1 y 2)
- $\mathbb{V}ar\left(Y \middle| \mathsf{X}\right) = \sigma^2 \mathbb{1}$ (v.a. cte.) (Sup. 2 y 3)

¿QUÉ CONDICIÓN ES SUFICIENTE PARA ESTO?

2 Supuesto 1: linealidad en los parámetros $\boldsymbol{\beta}$

Supuesto 1

$$\boxed{Y = \mathsf{X}\boldsymbol{\beta} + U}$$

donde $\mathsf{X} = \left[\mathbb{1};\quad X_2;\quad X_3;\quad \cdots\quad X_k;\right]$ y $\boldsymbol{\beta} = \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_k \end{pmatrix}$

es decir,

$$Y = \underbrace{\beta_1 \mathbb{1} + \beta_2 X_2 + \beta_3 X_3 + \cdots + \beta_k X_k}_{\mathsf{X}\boldsymbol{\beta}} + U$$

3 Supuesto 2: Esperanza condicional de $\boldsymbol{U}$ – Estricta exogeneidad

Supuesto 2 y sus implicaciones

$$\boxed{\mathbb{E}(\boldsymbol{U} \mid \mathsf{X}) = \mathbb{0}} \Rightarrow \begin{cases} \mathrm{E}\left(X_j U\right) = 0 \text{ para } j = 1:k. & \boxed{U \perp X_j} \\ \mathrm{E}\left(U\right) = 0 \\ \mathrm{Cov}\left(U, X_j\right) = 0 \quad \text{para } j = 1:k. \end{cases}$$

Implicación conjunta de los supuestos 1 y 2

$$\left.\begin{array}{l} \boldsymbol{Y} = \mathsf{X}\boldsymbol{\beta} + \boldsymbol{U} \\ \mathbb{E}(\boldsymbol{U} \mid \mathsf{X}) = \mathbb{0} \end{array}\right\} \Rightarrow \mathbb{E}(\boldsymbol{Y} \mid \mathsf{X}) = \mathsf{X}\boldsymbol{\beta} \qquad (F58)$$

4 Supuesto 3: Homocedasticidad

Supuesto 3

$$\boxed{\mathbb{E}\left(\boldsymbol{U}^2 \,\middle|\, \mathsf{X}\right) = \sigma^2 \mathbb{1}}$$

Junto con $\mathbb{E}(\boldsymbol{U} \mid \mathsf{X}) = \mathbb{0}$ es equivalente a: $\quad \mathbb{V}ar(\boldsymbol{U} \mid \mathsf{X}) = \sigma^2 \mathbb{1}$

Implicación de los supuestos 2 y 3

$$\begin{aligned} \sigma^2 \mathbb{1} &= \mathbb{E}\left(\boldsymbol{U}^2 \mid \mathsf{X}\right) \\ &= \mathbb{E}\left(\left(\boldsymbol{Y} - \mathbb{E}\left(\boldsymbol{Y} \mid \mathsf{X}\right)\right)^2 \,\middle|\, \mathsf{X}\right) \\ &= \mathbb{V}ar\left(\boldsymbol{Y} \mid \mathsf{X}\right). \qquad (F58) \end{aligned}$$

5 Supuesto 4 y la identificación de los parámetros $\boldsymbol{\beta}$

$$\begin{aligned} \boldsymbol{Y} &= \mathsf{X}\boldsymbol{\beta} + \boldsymbol{U} && \text{Por Sup. 1} \\ \mathsf{X}^\intercal \boldsymbol{Y} &= \mathsf{X}^\intercal \mathsf{X}\boldsymbol{\beta} + \mathsf{X}^\intercal \boldsymbol{U} && \text{premultiplicando por } \mathsf{X}^\intercal \\ \mathrm{E}\left(\mathsf{X}^\intercal \boldsymbol{Y}\right) &= \mathrm{E}\left(\mathsf{X}^\intercal \mathsf{X}\right)\boldsymbol{\beta} + \mathrm{E}\left(\mathsf{X}^\intercal \boldsymbol{U}\right) && \text{tomando esperanzas} \\ \mathrm{E}\left(\mathsf{X}^\intercal \boldsymbol{Y}\right) &= \mathrm{E}\left(\mathsf{X}^\intercal \mathsf{X}\right)\boldsymbol{\beta} && \mathrm{E}\left(\mathsf{X}^\intercal \boldsymbol{U}\right) = \boldsymbol{0} \text{ (Sup. 2)} \end{aligned}$$

donde ${}_{i|}\left(\mathrm{E}\left(\mathsf{X}^\intercal \mathsf{X}\right)\right)_{|j}$ es $\mathrm{E}\left(X_i X_j\right)$.

Supuesto 4

$$\boxed{\text{La matriz } \mathrm{E}\left(\mathsf{X}^\intercal \mathsf{X}\right) \text{ es de rango completo}}$$

entonces $\boldsymbol{\beta}$ está identificado: $\quad \boldsymbol{\beta} = \left(\mathrm{E}\left(\mathsf{X}^\intercal \mathsf{X}\right)\right)^{-1} \mathrm{E}\left(\mathsf{X}^\intercal \boldsymbol{Y}\right)$

6 Regresión cuando $\mathcal{E}$ es $\mathbb{R}^N$

$$\mathrm{E}\left(\mathsf{X}^\intercal \mathsf{X}\right)\boldsymbol{\beta} = \mathrm{E}\left(\mathsf{X}^\intercal \boldsymbol{Y}\right) \quad \text{se reduce a} \quad \left(\frac{1}{N}\mathbf{X}^\intercal \mathbf{X}\right)\boldsymbol{\beta} = \frac{1}{N}\mathbf{X}^\intercal \boldsymbol{y}$$

Ausencia de multicolinealidad exacta implica que

$$\boldsymbol{\beta} = \left(\frac{1}{N}\mathbf{X}^\intercal \mathbf{X}\right)^{-1} \frac{1}{N}\mathbf{X}^\intercal \boldsymbol{y} = \left(\mathbf{X}^\intercal \mathbf{X}\right)^{-1} \mathbf{X}^\intercal \boldsymbol{y}$$

donde

- ${}_{i|}\left(\frac{1}{N}\left(\mathbf{X}^\intercal \mathbf{X}\right)\right)_{|j} = \mu_{\left(\mathbf{x}_{|i} \odot \mathbf{x}_{|j}\right)}$
- ${}_{i|}\left(\frac{1}{N}\left(\mathbf{X}^\intercal \boldsymbol{y}\right)\right) = \mu_{\left(\mathbf{x}_{|i} \odot \boldsymbol{y}\right)}$

7 Cte. como único regresor

$$\mathsf{X}=[\mathbb{1};] \quad\longrightarrow\quad Y=\mathbb{E}(Y|\mathbb{1})+U;$$

$$\mathrm{E}(\mathsf{X}^\intercal\mathsf{X})\boldsymbol{\beta}=\mathrm{E}(\mathsf{X}^\intercal Y) \qquad (\text{donde } \mathsf{X}=[\mathbb{1};]);$$
$$\mathrm{E}(\mathbb{1}\cdot\mathbb{1})\boldsymbol{\beta}=\mathrm{E}(\mathbb{1}\cdot Y) \quad\Rightarrow\quad \boldsymbol{\beta}=\mathrm{E}(Y)$$

Cuando el EEP es $\mathbb{R}^N$ con $\langle\_|\_\rangle_s$ tenemos

$$\frac{1}{N}\mathbf{X}^\intercal\mathbf{X}\beta=\frac{1}{N}\mathbf{X}^\intercal\boldsymbol{y} \qquad (\text{donde } \mathbf{X}=[\mathbf{1};]);$$

así

$$\frac{1}{N}[\mathbf{1}\cdot\mathbf{1}]\boldsymbol{\beta}=\frac{1}{N}(\mathbf{1}\cdot\boldsymbol{y},) \quad\Rightarrow\quad \boldsymbol{\beta}=\mu_{\boldsymbol{y}}$$

8 Modelo Lineal Simple (Modelo teórico)

$$\mathsf{X}=[\mathbb{1};\ X;] \quad\longrightarrow\quad Y=\mathbb{E}(Y|X)+U;$$

$$\mathrm{E}(\mathsf{X}^\intercal\mathsf{X})\boldsymbol{\beta}=\mathrm{E}(\mathsf{X}^\intercal Y)$$
$$\begin{bmatrix}\mathrm{E}(\mathbb{1}) & \mathrm{E}(X)\\ \mathrm{E}(X) & \mathrm{E}(X^2)\end{bmatrix}\begin{pmatrix}\beta_1\\ \beta_2\end{pmatrix}=\begin{pmatrix}\mathrm{E}(Y)\\ \mathrm{E}(XY)\end{pmatrix}$$

cuya solución es

$$\beta_1=\mathrm{E}(Y)-\beta_2\,\mathrm{E}(X) \qquad \text{y} \qquad \beta_2=\frac{\mathrm{Cov}(X,Y)}{\mathrm{Var}(X)} \tag{9}$$

En $\mathbb{R}^N$ con $\langle\_|\_\rangle_s$:

$$\beta_1=\mu_{\boldsymbol{y}}-\beta_2\mu_{\boldsymbol{x}} \qquad \text{y} \qquad \beta_2=\frac{\sigma_{\boldsymbol{xy}}}{\sigma_{\boldsymbol{x}}^2}$$

Supuesto 4 (indep. lineal de regresores) garantiza $\rightarrow \sigma_{\boldsymbol{x}}^2\neq 0$

9 Estimación MCO con una muestra

Si $\boldsymbol{y}$ es una *muestra* de Y y $\mathbf{X}$ una *muestra* de X; y si se asume que $Y=\mathsf{X}\boldsymbol{\beta}+U$ es un modelo clásico de regresión que cumple los supuestos (y si $\mathbf{X}^\intercal\mathbf{X}$ es invertible)

$$\widehat{\boldsymbol{\beta}} = (\mathbf{X}^\intercal\mathbf{X})^{-1}\mathbf{X}^\intercal\boldsymbol{y}$$

donde

- ${}_{i|}\left(\frac{1}{N}(\mathbf{X}^\intercal\mathbf{X})\right)_{|j}=m_{(\mathbf{X}_{|i}\odot\mathbf{X}_{|j})}$
- ${}_{i|}\left(\frac{1}{N}(\mathbf{X}^\intercal\boldsymbol{y})\right)=m_{(\mathbf{X}_{|i}\odot\boldsymbol{y})}$

Método de los Momentos F62

10 Estimación por MCO del Modelo Lineal Simple

Sea $\quad Y=a\mathbb{1}+bX+U;\quad$ si disponemos de una muestra

$$\boldsymbol{y}\in\mathbb{R}^N, \qquad \underset{N\times 2}{\mathbf{X}}=[\mathbf{1};\ \boldsymbol{x}]$$

resolviendo $\mathbf{X}^\intercal\boldsymbol{y}=\mathbf{X}^\intercal\mathbf{X}\widehat{\boldsymbol{\beta}}$ con $\widehat{\boldsymbol{\beta}}=\begin{pmatrix}\widehat{a}\\ \widehat{b}\end{pmatrix}$, obtenemos cfr. F26

$$\widehat{b} = \frac{s_{\boldsymbol{xy}}}{s_{\boldsymbol{x}}^2} \qquad \text{y} \qquad \widehat{a} = m_{\boldsymbol{y}}-\widehat{b}\,m_{\boldsymbol{x}}$$

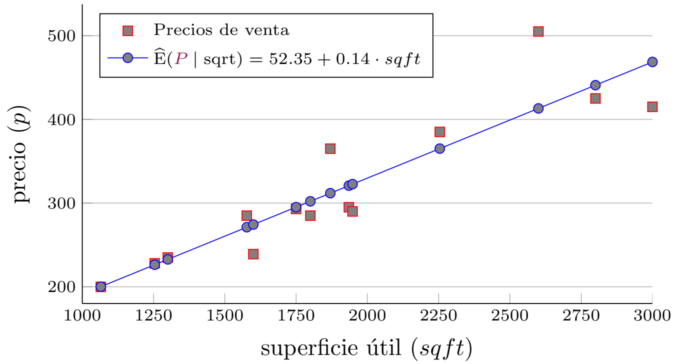
La estimación MCO sustituye los momentos teóricos por los muestrales (*método de los momentos*)

La indep. lineal de regresores garantiza $\rightarrow s_{\boldsymbol{x}}^2\neq 0$

11 Recta de regresión como estimación de la Esp. Cond.

☞**Código:** `EjPvivienda2.inp` ....

Precios de venta (miles de dólares) y superficie útil (pies al cuadrado) de 14 casas unifamiliares en la *University City* de la ciudad de San Diego en California en 1990 (Ramanathan, 2002, pp. 78)



Enlace a algunas **prácticas de la Lección 5**

Fin de la lección

Lección 6

1 Estimador MCO $\widehat{\boldsymbol{\beta}}$

Sean $\boldsymbol{Y}$ (vector) y $\mathbf{X}$ (matriz); muestreos aleatorios simples (*m.a.s*) del modelo $Y = \mathrm{X}\boldsymbol{\beta} + U$ que cumple todos los supuestos. Entonces

$$\left[{}_{i|}\boldsymbol{Y};\ {}_{i|}\mathbf{X}\right] \sim \text{iid. } \left[Y;\ \mathrm{X}\right]; \qquad \text{donde (Sup I) } Y = \mathrm{X}\boldsymbol{\beta} + U$$

y (Sup IV) $\mathrm{E}\left(\mathbf{X}^\intercal\mathbf{X}\right)$ es invertible. El *estimador MCO de $\boldsymbol{\beta}$* es

$$\widehat{\boldsymbol{\beta}} = \left(\mathbf{X}^\intercal\mathbf{X}\right)^{-1}\mathbf{X}^\intercal\boldsymbol{Y}$$

Además, el modelo muestral $\boxed{\boldsymbol{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{U}}$ verifica que

(Sup II) $\mathbb{E}\left(\boldsymbol{U} \mid \mathbf{X}\right) = \mathbf{0}$

(Sup III) $\mathbb{V}ar\left(\boldsymbol{U} \mid \mathbf{X}\right) = \sigma^2\mathbf{I}$ (*homocedásticidad*, NO *autocorrelación*)

Dadas las muestras **X** (rango k) e $\boldsymbol{y}$, la *estimación MCO de $\boldsymbol{\beta}$* es:

$$\widehat{\boldsymbol{\beta}} = (\mathbf{X}^\intercal\mathbf{X})^{-1}\mathbf{X}^\intercal\boldsymbol{y}$$

2 Esperanza del estimador MCO $\widehat{\boldsymbol{\beta}}$

En el *m.a.s.*, $\boldsymbol{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{U}$, si $\mathrm{E}\left(\mathbf{X}^\intercal\mathbf{X}\right)$ es invertible y denotamos $(\mathbf{X}^\intercal\mathbf{X})^{-1}\mathbf{X}^\intercal$ por $\underset{k\times N}{\mathbf{A}}$:

$$\widehat{\boldsymbol{\beta}} = \mathbf{A}\boldsymbol{Y} = \mathbf{A}(\mathbf{X}\boldsymbol{\beta} + \boldsymbol{U}) = \mathbf{I}\boldsymbol{\beta} + \mathbf{A}\boldsymbol{U}$$

Así,
$$\begin{aligned}\mathbb{E}(\widehat{\boldsymbol{\beta}} \mid \mathbf{X}) &= \mathbb{E}\left(\mathbf{I}\boldsymbol{\beta} + \mathbf{A}\boldsymbol{U} \mid \mathbf{X}\right)\\ &= \mathbf{I}\boldsymbol{\beta} + \mathbf{A}\mathbb{E}\left(\boldsymbol{U} \mid \mathbf{X}\right) && \mathbf{I}\boldsymbol{\beta}, \mathbf{A} \in \mathcal{L}(\mathbf{X})\\ &= \mathbf{I}\boldsymbol{\beta} + \mathbf{A}\mathbf{0} = \mathbf{I}\boldsymbol{\beta} && (\text{Sup II}).\end{aligned}$$

Por Tª Esperanzas iteradas:

$$\boxed{\mathrm{E}\left(\widehat{\boldsymbol{\beta}}\right) = \mathrm{E}\left(\mathbb{E}(\widehat{\boldsymbol{\beta}} \mid \mathbf{X})\right) = \mathrm{E}\left(\mathbf{I}\boldsymbol{\beta}\right) = \boldsymbol{\beta}}.$$

Por tanto $\widehat{\boldsymbol{\beta}}$ es un estimador insesgado.

3 Varianza del estimador MCO $\widehat{\boldsymbol{\beta}}$

Por los supuestos I, III y IV:

$$\begin{aligned}\mathbb{V}ar(\widehat{\boldsymbol{\beta}} \mid \mathbf{X}) &= \mathbb{V}ar(\widehat{\boldsymbol{\beta}} - \mathbf{I}\boldsymbol{\beta} \mid \mathbf{X}) && = \mathbb{V}ar\left(\mathbf{A}\boldsymbol{U} \mid \mathbf{X}\right)\\ &= \mathbf{A}\mathbb{V}ar\left(\boldsymbol{U} \mid \mathbf{X}\right)\mathbf{A}^\intercal && = \mathbf{A}\sigma^2\mathbf{I}\mathbf{A}^\intercal \quad (\text{Sup III})\\ &= \sigma^2\mathbf{A}\mathbf{A}^\intercal && = \sigma^2(\mathbf{X}^\intercal\mathbf{X})^{-1}.\end{aligned}$$

donde $\mathbf{A} = (\mathbf{X}^\intercal\mathbf{X})^{-1}\mathbf{X}^\intercal$. $\quad$ ($(\mathbf{X}^\intercal\mathbf{X})^{-1}$ es una matriz "llena")

$$\mathrm{Var}\left(\widehat{\boldsymbol{\beta}}\right) = \mathrm{E}\left(\mathbb{V}ar(\widehat{\boldsymbol{\beta}} \mid \mathbf{X})\right) + \mathrm{Var}\left(\overbrace{\mathbb{E}(\widehat{\boldsymbol{\beta}} \mid \mathbf{X})}^{\mathbf{I}\boldsymbol{\beta}\ (\text{Sup II})}\right) = \mathrm{E}\left(\sigma^2(\mathbf{X}^\intercal\mathbf{X})^{-1}\right)$$

Por tanto: $\boxed{\mathrm{Var}\left(\widehat{\boldsymbol{\beta}}\right) = \sigma^2\mathrm{E}\left(\mathbf{X}^\intercal\mathbf{X}\right)^{-1}}$.

Así, $\mathrm{Var}\left(\widehat{\beta_j}\right) = \sigma^2\left({}_{j|}\mathrm{E}\left(\mathbf{X}^\intercal\mathbf{X}\right)^{-1}{}_{|j}\right)$.

Continuación del ejemplo "Precio de las viviendas":

Observe la matriz $(\mathbf{X}^\intercal\mathbf{X})^{-1}$, del ejemplo del "precio de las viviendas".

$$(\mathbf{X}^\intercal\mathbf{X})^{-1} = \begin{bmatrix} 9.1293e-01 & -4.4036e-04\\ -4.4036e-04 & 2.3044e-07\end{bmatrix};$$

¿Qué estimación es más fiable, la pendiente o la constante?

☞**Código:** `EjPvivienda3.inp` ....

Repita la regresión para "precio de las viviendas" con las siguientes modificaciones en la muestra:

1. con todos los datos excepto los de la última vivienda
2. con todos los datos excepto los de las últimas dos viviendas
3. con todos los datos excepto los de la primera y la última viviendas

¿Confirman estos resultados su respuesta a la primera pregunta?

4 Eficiencia del estimador MCO $\widehat{\boldsymbol{\beta}}$: Tª de Gauss-Markov

Gracias a los supuestos I a IV,

$\widehat{\boldsymbol{\beta}}$ **eficiente** *entre estimadores lineales e insesgados*

es decir, para cualquier estimador lineal[1] insesgado $\widetilde{\boldsymbol{\beta}}$

$$\mathbb{V}ar(\widetilde{\boldsymbol{\beta}} \mid \mathbf{X}) \geq \mathbb{V}ar(\widehat{\boldsymbol{\beta}} \mid \mathbf{X}).$$

Entonces se dice ELIO (BLUE en inglés).

5 Consistencia del estimador MCO $\widehat{\boldsymbol{\beta}}$

Además, $\widehat{\boldsymbol{\beta}}$ es **consistente**, es decir,

- es *insesgado*

- la *varianza tiende a cero* cuando la muestra crece

$$\lim_{N\to\infty} \mathbb{V}ar\left(\widehat{\beta}_j \mid \mathbf{X}\right) = \mathbb{0}$$

6 Primeros momentos de $\widehat{\boldsymbol{y}}$ (valores ajustados por MCO)

Denotemos $\mathbf{X}\mathbf{A} = \mathbf{X}(\mathbf{X}^\intercal\mathbf{X})^{-1}\mathbf{X}^\intercal$ por $\mathbf{P}$, entonces $\mathbf{P} \in \mathcal{L}(\mathbf{X})$ y

$$\widehat{\boldsymbol{y}} = \underbrace{\mathbf{X}\widehat{\boldsymbol{\beta}}}_{\text{(Requiere Sup IV)}} = \mathbf{X}(\mathbf{I}\boldsymbol{\beta} + \mathbf{A}\boldsymbol{U}) = \mathbf{X}\boldsymbol{\beta} + \mathbf{P}\boldsymbol{U};$$

$$\begin{aligned}\mathbb{E}(\widehat{\boldsymbol{y}} \mid \mathbf{X}) =& \mathbb{E}(\mathbf{X}\boldsymbol{\beta} + \mathbf{P}\boldsymbol{U} \mid \mathbf{X}) = \mathbb{E}(\mathbf{X}\boldsymbol{\beta} \mid \mathbf{X}) + \mathbf{P}\cdot\mathbb{E}(\boldsymbol{U} \mid \mathbf{X})\\ =& \mathbf{X}\boldsymbol{\beta} \qquad \text{(por Sup. II)}\end{aligned}$$

$$\text{Así,}\quad \mathrm{E}\left(\widehat{\boldsymbol{y}}\right) = \mathrm{E}\left(\mathbf{X}\boldsymbol{\beta}\right) \quad\Rightarrow\quad \mathrm{E}\left(\widehat{Y_n}\right) = \mathrm{E}\left({}_{n|}\mathbf{X}\boldsymbol{\beta}\right)$$

$$\begin{aligned}\mathbb{V}ar(\widehat{\boldsymbol{y}} \mid \mathbf{X}) =& \mathbb{V}ar(\mathbf{X}\boldsymbol{\beta} + \mathbf{P}\boldsymbol{U} \mid \mathbf{X}) = \mathbf{P}\,\mathbb{V}ar(\boldsymbol{U} \mid \mathbf{X})\mathbf{P}^\intercal\\ =& \sigma^2\mathbf{P}\mathbf{I}\mathbf{P}^\intercal = \sigma^2\mathbf{P} \qquad \text{(por Sup. III)}\end{aligned} \tag{10}$$

(matriz "llena")

7 Primeros momentos de los errores MCO

Denotemos $\mathbf{I} - \mathbf{P}$ por $\mathbf{M}$, entonces $\mathbf{M} \in \mathcal{L}(\mathbf{X})$ y

$$\widehat{\boldsymbol{e}} = \underbrace{\boldsymbol{Y} - \widehat{\boldsymbol{y}}}_{\text{(Requiere Sup IV)}} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{U} - (\mathbf{X}\boldsymbol{\beta} + \mathbf{P}\boldsymbol{U}) = \mathbf{M}\boldsymbol{U}$$

$$\begin{aligned}\mathbb{E}(\widehat{\boldsymbol{e}} \mid \mathbf{X}) = \mathbb{E}(\mathbf{M}\boldsymbol{U} \mid \mathbf{X}) =& \mathbf{M}\cdot\mathbb{E}(\boldsymbol{U} \mid \mathbf{X})\\ =& \mathbf{0} \qquad \text{(por Sup. II)}\end{aligned}$$

$$\text{Así,}\quad \mathrm{E}\left(\widehat{\boldsymbol{e}}\right) = \mathbf{0}$$

$$\begin{aligned}\mathbb{V}ar(\widehat{\boldsymbol{e}} \mid \mathbf{X}) =& \mathbf{M}\,\mathbb{V}ar(\boldsymbol{U} \mid \mathbf{X})\mathbf{M}^\intercal\\ =& \sigma^2\mathbf{M}\mathbf{M}^\intercal = \sigma^2\mathbf{M} \qquad \text{(por Sup. III)}\end{aligned} \tag{11}$$

(matriz "llena")

8 Supuesto 5: Distribución Normal de las perturbaciones

La inferencia es muy sencilla bajo el siguiente supuesto sobre la distribución conjunta de $\boldsymbol{U}$:

$$\boldsymbol{U} \sim N\left(\mathbf{0},\, \sigma^2\mathbf{I}\right) \quad\Rightarrow\quad \boldsymbol{Y} \sim N\left(\mathrm{E}\left(\mathbf{X}\boldsymbol{\beta}\right),\, \sigma^2\mathbf{I}\right)$$

donde $\mathbf{I}$ es la matriz identidad de orden $N \times N$. Puesto que

$$\widehat{\boldsymbol{\beta}} = \mathbf{I}\boldsymbol{\beta} + \left(\mathbf{X}^\intercal\mathbf{X}\right)^{-1}\mathbf{X}^\intercal\boldsymbol{U} = \mathbf{I}\boldsymbol{\beta} + \mathbf{A}\boldsymbol{U}$$

entonces $\widehat{\boldsymbol{\beta}}$ tiene distribución normal multivariante.

$$\widehat{\boldsymbol{\beta}} \sim N\left(\boldsymbol{\beta},\, \sigma^2\mathrm{E}\left(\mathbf{X}^\intercal\mathbf{X}\right)^{-1}\right)$$

9 Distribución del estimador MCO $\widehat{\boldsymbol{\beta}}$

Así pues,

$$\widehat{\beta_j} \sim N\left(\beta_j,\, \operatorname{Var}\left(\widehat{\beta_j}\right)\right)$$

donde $\operatorname{Var}\left(\widehat{\beta_j}\right) = \operatorname{E}\left(\mathbb{V}ar\left(\widehat{\beta_j} \mid \mathbf{X}\right)\right) = \sigma^2\,\operatorname{E}\left({}_{j|}(\mathbf{X}^\intercal\mathbf{X})^{-1}{}_{|j}\right)$
(el j-ésimo elemento de la diagonal) y

$$\frac{\widehat{\beta_j} - \beta_j}{\mathbb{D}t\left(\widehat{\beta_j} \mid \mathbf{X}\right)} \sim N\,(0,\,1)$$

10 Estimación de la varianza residual

El parámetro σ^2 es desconocido F61

Pero la cuasivarianza de $\widehat{\boldsymbol{e}}$

$$\widehat{\mathfrak{s}^2_{\boldsymbol{e}}} \equiv (\widehat{\boldsymbol{e}}\cdot\widehat{\boldsymbol{e}})/(N-k)$$

es un estimador *insesgado* de σ^2

Así, el estimador insesgado de la matriz de varianzas condicionada de $\widehat{\boldsymbol{\beta}}$ es

$$\widehat{\operatorname{Var}}\left(\widehat{\boldsymbol{\beta}} \mid \mathbf{X}\right) = \widehat{\mathfrak{s}^2_{\boldsymbol{e}}}\cdot(\mathbf{X}^\intercal\mathbf{X})^{-1}. \qquad (12)$$

11 Más sobre medidas de ajuste

Los criterios de información de Akaike y de Schwartz permiten seleccionar entre modelos alternativos.
(Están calculados bajo el supuesto de normalidad).
Aquí *es preferido el modelo que arroja un resultado más bajo*

- **Akaike** (AIC) Premia la bondad de ajuste, pero penaliza la complejidad del modelo (aunque tiende a sobre-parametrizar)
- **Schwartz** (BIC) Basado en el criterio de Akaike, la penalización por el número de parámetros es mayor que en el AIC para evitar una posible sobre-parametrización.
- **Hannan-Quinn** (HQC) Basado en el criterio de Akaike, la penalización por el número de parámetros es mayor que en el AIC para evitar una posible sobre-parametrización.

Véase los resultados de estimación para el precio de las viviendas.

Enlace a algunas **prácticas de la Lección 6**

Fin de la lección

Lección 7

1 Contrastes de hipótesis paramétricas

Hipótesis afirmación sobre uno o varios parámetros

- H_0: hipótesis *nula*
- H_1: hipótesis complementaria (*alternativa*)

Contraste de hipótesis es una regla que establece

- para que valores muestrales **X** se rechaza H_0 (*región crítica*, RC)
- para que valores muestrales **X** no se rechaza H_0 (*región de no rechazo* ($\neq$ *aceptación*), RA)

Toma de decisión sobre el rechazo o no de H_0

2 Contrastes de hipótesis paramétricas

Caracterizamos RC mediante un estadístico $g(\mathbf{X})$.
Ejemplo

- Tren sale cada hora en punto (tardo 10' en llegar al andén)
- H_0: me da tiempo
  H_1: NO me da tiempo
- g($\mathbf{X}$): hora media de los relojes de los presentes
- $RC = \{$**X** tales que: $g(\mathbf{X}) = m_{\boldsymbol{x}} \geq hh : 40'\}$ (*nivel significación* α)
- Pregunto la hora, y decido si voy al andén

Pero el estadístico podría ser

- $g^*(\mathbf{X})$: hora media de los relojes de más de 60 euros.
- $RC^* = \{$**X** tales que: $g^*(\mathbf{X}) \geq hh : 45'\}$ (*nivel de significación* α)

3 Etapas de un contraste de hipótesis paramétricas

1. Establecimiento de la hipótesis nula H_0 sobre $\boldsymbol{\theta}$
$$H_0: \quad X \sim f_X(x;\boldsymbol{\theta}); \qquad \boldsymbol{\theta} \in \Theta_0$$
   y la hipótesis complementaria (*alternativa*)
$$H_1: \quad X \sim f_X(x;\boldsymbol{\theta}); \qquad \boldsymbol{\theta} \in \Theta_1$$
   donde $\quad \Theta = \Theta_0 \cup \Theta_1 \quad$, y $\quad \Theta_0 \cap \Theta_1 = \emptyset$
2. Elección del estadístico $g(\mathbf{X})$
3. División del espacio muestral en dos regiones: RC y RA (dado un nivel de significación α)
   $RC \cap RA = \emptyset; \qquad RC \cup RA =$ espacio muestral
   - ¿Donde está mi muestra **X**?
   - Cálculo del estadístico: $g(\mathbf{X})$ para decidir si **X** $\in RC$.
   - En consecuencia, Rechazo o no rechazo H_0 (toma de decisión)

4 Estadístico t de Student ($\mathcal{T}$) para los parámetros β_j

Bajo los supuestos muestrales:

$$\frac{\widehat{\beta_j} - \beta_j \mathbb{1}}{\sqrt{\sigma^2 \left[\left(\mathbf{X}^\intercal \mathbf{X} \right)^{-1} \right]_{jj}}} = \frac{\widehat{\beta_j} - \beta_j \mathbb{1}}{\mathbb{D}t\left(\widehat{\beta_j} \mid \mathbf{X}\right)} \sim N\left(0,\, 1\right)$$

y sustituyendo σ^2 por su estimador, $\mathfrak{s}^2_{\widehat{e}} = \frac{\widehat{e} \cdot \widehat{e}}{N-k}$, obtenemos el estadístico $\mathcal{T}$

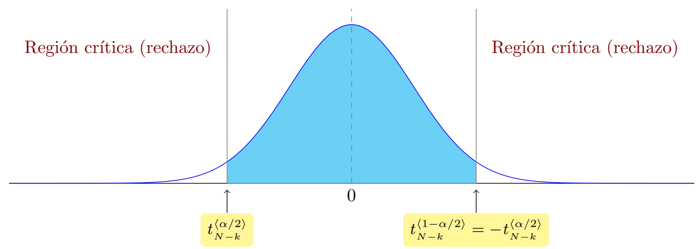
$$\frac{\widehat{\beta_j} - \beta_j \mathbb{1}}{\sqrt{\mathfrak{s}^2_{\widehat{e}} \left[\left(\mathbf{X}^\intercal \mathbf{X} \right)^{-1} \right]_{jj}}} = \frac{\widehat{\beta_j} - \beta_j \mathbb{1}}{\widehat{\mathrm{Dt}}\left(\widehat{\beta_j} \mid \mathbf{X}\right)} \equiv \mathcal{T}_j \underset{\mathrm{E}\left(\widehat{\beta_j}\right) = \beta_j}{\sim} t_{N-k}, \tag{13}$$

Nótese que β_j es desconocido.

5 Contraste de la t: de dos colas

1. $H_0 : \beta_j = b; \qquad H_1 : \beta_j \neq b$
2. (De Ec. 13) $\quad \dfrac{\widehat{\beta_j} - b\mathbb{1}}{\widehat{\mathrm{Dt}}\left(\widehat{\beta_j} \mid \mathbf{X}\right)} \equiv \mathcal{T} \underset{H_0}{\sim} t_{N-k}$
3. Se rechaza H_0 cuando $|\widehat{\mathcal{T}}| > t_{N-k}^{\langle 1-\alpha \rangle}$ $\qquad$ (α determina RC)

Distribución t con $(N-k)$ grados de libertad



$t_{N-k}^{\langle \alpha/2 \rangle}$ y $t_{N-k}^{\langle 1-\alpha/2 \rangle}$ son los valores críticos

Ejemplo

Continuación de "precio de las viviendas": Contraste de significación individual de a:

$$H_0 :\ a = 0; \qquad H_1 :\ a \neq 0$$

En este caso la región crítica debe ser

$$RC = \left\{ \mathbf{X} \text{ tales que } \left| \frac{\widehat{a} - 0}{\widehat{\mathrm{Dt}}(\widehat{a} \mid \mathbf{X})} \right| > k_2 \right\}, \text{ donde } \frac{\widehat{a}}{\widehat{\mathrm{Dt}}\left(\widehat{a} \mid \mathbf{X}\right)} \equiv \mathcal{T}_a \underset{H_0}{\sim} t_{12}\,.$$

Si $\alpha = 0.05$, el valor crítico es $k_2 = 2.18 = -k_1 = t_{12}^{\langle 0.025 \rangle}$:

$$\widehat{\mathcal{T}_a} = \frac{52.351}{37.285} = 1.4041 < k_2 \qquad \text{no rechazamos } H_0 \text{ para } \alpha \text{ del 5\%.}$$

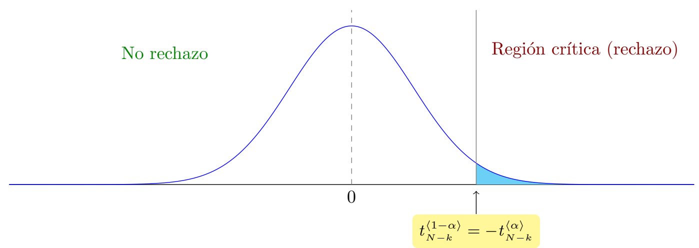
Véase los resultados de estimación del ejemplo del precio de las viviendas.
Para $\alpha = 0.1$, el valor critico es $k_2 = 1.78 = -k_1 = t_{12}^{\langle 0.05 \rangle}$.
¿?

¿Deberíamos quitar el término constante del modelo?

6 Contraste de la t: de una sola cola (derecha)

1. $H_0 : \beta_j = b; \qquad H_1 : \beta_j > b$
2. (De Ec. 13) $\quad \dfrac{\widehat{\beta_j} - b\mathbb{1}}{\widehat{\mathrm{Dt}}\left(\widehat{\beta_j} \mid \mathbf{X}\right)} \equiv \mathcal{T} \underset{H_0}{\sim} t_{N-k}$
3. Se rechaza H_0 cuando $\widehat{\mathcal{T}} > t_{N-k}^{\langle 1-\alpha \rangle}$ $\qquad$ (α determina RC)

Distribución t con $(T-k)$ grados de libertad

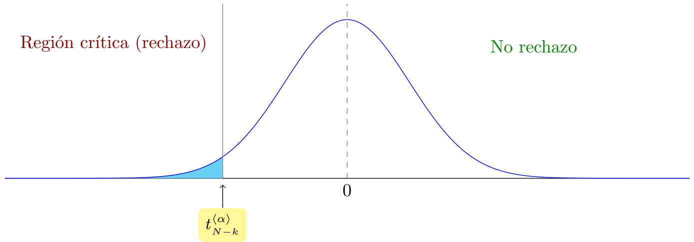


$t_{N-k}^{\langle 1-\alpha \rangle}$ es el valor crítico

7 Contraste de la t: de una sola cola (izquierda)

1. $H_0 : \beta_j = b; \qquad H_1 : \beta_j < b$
2. (De Ec. 13) $\dfrac{\widehat{\beta_j} - b\mathbb{1}}{\widehat{\mathrm{Dt}}\big(\widehat{\beta_j} \mid \mathbf{X}\big)} \equiv \mathcal{T} \underset{H_0}{\sim} t_{N-k}$
3. Se rechaza H_0 cuando $\widehat{\mathcal{T}} < t_{N-k}^{\langle\alpha\rangle}$ $\qquad$ (α determina RC)

Distribución t con $(T-k)$ grados de libertad



$t_{N-k}^{\langle\alpha\rangle}$ es el valor crítico

Ejemplo

Continuación de "precio de las viviendas": Un experto del mercado de la vivienda afirma que un pie cuadrado adicional en la superficie supone un incremento de (*como poco*) 150 dolares, pero *nunca menos*. ¿Podemos creer al experto con una significación del 2.5%?

$$H_0 : \; b = 0.15; \quad H_1 : \; b < 0.15$$

La región critica de cola izquierda

$$RC = \left\{ \mathbf{X} \;\middle|\; \frac{\widehat{b} - 0.15}{\widehat{\mathrm{Dt}}\big(\widehat{b} \mid \mathbf{X}\big)} < k \right\}$$

sustituyendo valores estimados, tenemos que

$$\widehat{\mathcal{T}_b} = \frac{0.139 - 0.15}{0.01873} = -0.58729 > t_{12}^{\langle 0.025\rangle} = -2.18$$

¿?

8 p-valor y regla de decisión

El p-valor es la probabilidad (bajo H_0) de obtener un resultado (igual o) "más extremo" que el observado.

El significado de "más extremo" depende de H_1

- p-valor $= \mathbb{P}_{H_0}\big(\mathcal{T}_j > \widehat{\mathcal{T}_j}\big)$ (cola derecha)
- p-valor $= \mathbb{P}_{H_0}\big(\mathcal{T}_j < \widehat{\mathcal{T}_j}\big)$ (cola izquierda)
- p-valor $= 2 \times \min\Big\{\mathbb{P}_{H_0}\big(\mathcal{T}_j > \widehat{\mathcal{T}_j}\big)H_0, \; \mathbb{P}_{H_0}\big(\mathcal{T}_j < \widehat{\mathcal{T}_j}\big)\Big\}$ (bilateral)

Cuando el p-valor es "*pequeño*" se rechaza H_0

Véase los resultados de estimación del ejemplo del precio de las viviendas

Enlace a algunas **prácticas de la Lección 7**

Fin de la lección

Lección 8

1 Hipótesis lineales

$$H_0 : \mathbf{R}\boldsymbol{\beta} = \boldsymbol{r},$$

$\underset{r\times k}{\mathbf{R}}$ es matriz con $\operatorname{rg}\left(\mathbf{R}\right) = r, \quad (r \leq k);$ y $\boldsymbol{r} \in \mathbb{R}^r$ es vector.

Las r ecuaciones son hipótesis sobre valores de los coeficientes.

Condición $\operatorname{rg}\left(\mathbf{R}\right) = r$, garantiza:

- no hipótesis redundantes
- no hipótesis incompatibles

Ejemplo

Ecuación de salarios (continuación Ejemplo 2 en la página 19)**:**

$$\ln(SALAR) = \beta_1\mathbb{1} + \beta_2 EDUC + \beta_3 ANTIG + \beta_4 EXPER + U$$

Supongamos que queremos contrastar si educación y antigüedad tienen el mismo efecto en el incremento del salario, y que además, la experiencia no tiene ningún efecto (por tanto $r = 2$)

$$\beta_2 = \beta_3 \qquad \text{y} \qquad \beta_4 = 0.$$

En forma matricial, $H_0 : \mathbf{R}\boldsymbol{\beta} = \boldsymbol{r}$, donde

$$\mathbf{R} = \begin{bmatrix} 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \quad \boldsymbol{r} = \begin{pmatrix} 0 \\ 0 \end{pmatrix};$$

donde $\mathbf{R}$ cumple la condición de rango completo.

Añadiendo restricciones que no cumplen la condición de rango:

- Supongamos que adicionalmente imponemos que

$$\beta_2 - \beta_3 = \beta_4.$$

Esta es una restricción redundante, pues ya se cumple con las dos primeras restricciones; en forma matricial

$$\mathbf{R} = \begin{bmatrix} 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 1 & -1 & -1 \end{bmatrix}, \quad \boldsymbol{r} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix};$$

- Supongamos que imponemos una condición incompatible con las dos primeras:

$$\beta_4 = 0.5,$$

que evidentemente es incompatible con $\beta_4 = 0$. Matricialmente

$$\mathbf{R} = \begin{bmatrix} 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \quad \boldsymbol{r} = \begin{pmatrix} 0 \\ 0 \\ 0.5 \end{pmatrix}.$$

De nuevo la condición de rango se incumple.

2 Estadístico F

Bajo supuestos 1 a 5; y si H_0: $\mathbf{R}\boldsymbol{\beta} = \boldsymbol{r}$ cierta, donde $\operatorname{rg}\big(\underset{r\times k}{\mathbf{R}}\big) = r$, definimos el **Estadístico F**:

$$\mathcal{F} = \left(\mathbf{R}\widehat{\boldsymbol{\beta}} - \mathbf{I}\boldsymbol{r}\right)\left[\widehat{\operatorname{Var}}\left(\mathbf{R}\widehat{\boldsymbol{\beta}} \mid \mathbf{X}\right)\right]^{-1}\left(\mathbf{R}\widehat{\boldsymbol{\beta}} - \mathbf{I}\boldsymbol{r}\right)/r \underset{H_0}{\sim} F_{r,N-k} \tag{14}$$

$$\tag{15}$$

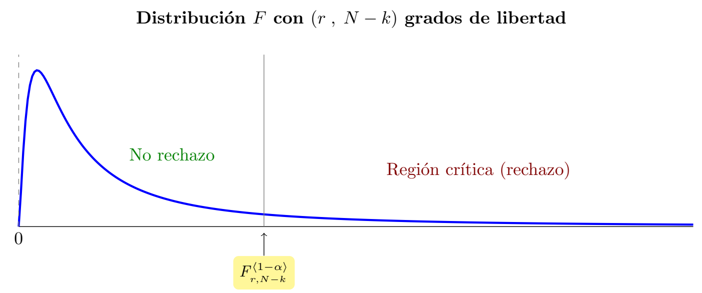
$$= \left(\mathbf{R}\widehat{\boldsymbol{\beta}} - \mathbf{I}\boldsymbol{r}\right)\left[\mathbf{R}\widehat{\operatorname{Var}}\left(\widehat{\boldsymbol{\beta}} \mid \mathbf{X}\right)\mathbf{R}^{\intercal}\right]^{-1}\left(\mathbf{R}\widehat{\boldsymbol{\beta}} - \mathbf{I}\boldsymbol{r}\right)/r$$

(de la Ecuación 12) sustituyendo $\widehat{\operatorname{Var}}\left(\widehat{\boldsymbol{\beta}} \mid \mathbf{X}\right) = \widehat{s^2}\cdot\left(\mathbf{X}^{\intercal}\mathbf{X}\right)^{-1}$

$$= \frac{1}{\widehat{s^2}}\left(\mathbf{R}\widehat{\boldsymbol{\beta}} - \mathbf{I}\boldsymbol{r}\right)\left[\mathbf{R}\left(\mathbf{X}^{\intercal}\mathbf{X}\right)^{-1}\mathbf{R}^{\intercal}\right]^{-1}\left(\mathbf{R}\widehat{\boldsymbol{\beta}} - \mathbf{I}\boldsymbol{r}\right)/r \tag{16}$$

3 Contraste de la F

1. $H_0 : \mathbf{R}\boldsymbol{\beta} = \boldsymbol{r}$; $\quad H_1 : \mathbf{R}\boldsymbol{\beta} \neq \boldsymbol{r}$
2. $\left(\mathbf{R}\widehat{\boldsymbol{\beta}} - \mathbf{I}\boldsymbol{r}\right)\left[\widehat{\operatorname{Var}}\left(\mathbf{R}\widehat{\boldsymbol{\beta}} \mid \mathbf{X}\right)\right]^{-1}\left(\mathbf{R}\widehat{\boldsymbol{\beta}} - \mathbf{I}\boldsymbol{r}\right)/r \underset{H_0}{\sim} F_{r,N-k}$
3. Cuando $\widehat{\mathcal{F}} \in RC$ se rechaza H_0 $\qquad$ (α determina RC)



…o bien: cuando p-valor se considera pequeño, se rechaza H_0

4 t versus F

Contrastación de hipótesis individual es caso particular, donde $r = 1$ y

$$\underset{1\times k}{\mathbf{R}} = \begin{bmatrix} 0 & \cdots & 0 & \underset{(j)}{1} & 0 & \cdots & 0 \end{bmatrix}, \quad \boldsymbol{r} = b_j$$

(14) se reduce a

$$\mathcal{F} = \left(\mathbf{R}\widehat{\boldsymbol{\beta}} - \mathbf{I}\boldsymbol{r}\right)\left[\widehat{\operatorname{Var}}\left(\mathbf{R}\widehat{\boldsymbol{\beta}} \mid \mathbf{X}\right)\right]^{-1}\left(\mathbf{R}\widehat{\boldsymbol{\beta}} - \mathbf{I}\boldsymbol{r}\right)/1$$

$$= \left(\widehat{\beta_j} - b_j\mathbb{1},\right)\left[\widehat{\operatorname{Var}}\left(\widehat{\beta_j} \mid \mathbf{X}\right)\right]^{-1}\left(\widehat{\beta_j} - b_j\mathbb{1},\right) \underset{H_0:\,\beta_j=b_j}{\sim} F_{1,N-k} \tag{17}$$

que es cuadrado[2] del estadístico $\mathcal{T}$ de (13), página 91.

Nota

No solo el contraste de significación individual tiene una distribución $(\mathcal{T})^2$. Si $\underset{1\times k}{\mathbf{R}} = \begin{bmatrix} r_1, & r_2, & \cdots & r_k \end{bmatrix}$ y, consecuentemente, $\boldsymbol{r}$ tiene una única componente (es decir, si hay una única restricción lineal), el estadístico resultante siempre es $\mathcal{F} = (\mathcal{T})^2$; veámoslo:

$$\mathcal{F} = \left(\mathbf{R}\widehat{\boldsymbol{\beta}} - \mathbf{I}\boldsymbol{r}\right)\left[\widehat{\operatorname{Var}}\left(\mathbf{R}\widehat{\boldsymbol{\beta}} \mid \mathbf{X}\right)\right]^{-1}\left(\mathbf{R}\widehat{\boldsymbol{\beta}} - \mathbf{I}\boldsymbol{r}\right)/1$$

operando tenemos:

$$= \left(r_1\widehat{\beta_1} + \cdots + r_k\widehat{\beta_k} - b\mathbb{1},\right)\left[\widehat{\operatorname{Var}}\left(r_1\widehat{\beta_1} + \cdots + r_k\widehat{\beta_k} \mid \mathbf{X}\right)\right]^{-1}\left(r_1\widehat{\beta_1} + \cdots + r_k\widehat{\beta_k} - b\mathbb{1},\right)$$

y por ser una expresión escalar:

$$= \frac{\left(r_1\widehat{\beta_1} + \cdots + r_k\widehat{\beta_k} - b\mathbb{1}\right)^2}{\widehat{\operatorname{Var}}\left(r_1\widehat{\beta_1} + \cdots + r_k\widehat{\beta_k} \mid \mathbf{X}\right)} = \left(\frac{r_1\widehat{\beta_1} + \cdots + r_k\widehat{\beta_k} - b\mathbb{1}}{\widehat{\operatorname{Dt}}\left(r_1\widehat{\beta_1} + \cdots + r_k\widehat{\beta_k} \mid \mathbf{X}\right)}\right)^2 = (\mathcal{T})^2,$$

ya que $r_1\widehat{\beta_1} + \cdots + r_k\widehat{\beta_k}$ es una combinación lineal de Normales, es una variable aleatoria escalar con distribución Normal.

5 Contraste t para una combinación lineal de betas

Si $\underset{1\times k}{\mathbf{R}} = \begin{bmatrix} r_1, & r_2, & \cdots & r_k \end{bmatrix}$ y $b = \mathbf{R}\boldsymbol{\beta}$, entonces

$$\frac{\left(r_1\widehat{\beta_1} + \cdots + r_k\widehat{\beta_k} - b\mathbb{1}\right)}{\widehat{\mathrm{Dt}}\left(\, r_1\widehat{\beta_1} + \cdots + r_k\widehat{\beta_k} \mid \mathbf{X}\right)} = \frac{\mathbf{R}\left(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}\mathbb{1}\right)}{\widehat{\mathrm{Dt}}\left(\mathbf{R}\widehat{\boldsymbol{\beta}} \mid \mathbf{X}\right)} = \mathcal{T} \underset{H_0}{\sim} t_{N-k}\,.$$

6 Significación conjunta del modelo

- En este contraste las hipótesis son

  H_0 : todos los coeficientes (excepto el de la constante) son nulos;
  H_1 : al menos uno es distinto de cero.

- Este contraste no es equivalente a realizar $k-1$ contrastes individuales *por separado.*

- Es un contraste $\mathcal{F}$ y su valor y p-valor se muestran en las regresiones por MCO.

Véase los resultados de estimación del ejemplo del precio de las viviendas (con esto ya sabe que significan casi todos los números del cuadro de resultados).

7 Contrastes de hipótesis e intervalos de confianza

El test t-Student bilateral rechaza H_0 : $\underset{1\times k}{\mathbf{R}}\ \boldsymbol{\beta} = r$ si

$$|\mathcal{T}| = \frac{|\mathbf{R}\widehat{\boldsymbol{\beta}} - r\mathbb{1}\,|}{\widehat{\mathrm{Dt}}\left(\mathbf{R}\widehat{\boldsymbol{\beta}} \mid \mathbf{X}\right)} > t^{\langle 1-\alpha/2\rangle}, \qquad \boxed{\text{F89}}$$

donde α es el nivel de significación; por tanto

$$\begin{aligned}
|\mathcal{T}| > t^{\langle 1-\alpha/2\rangle} \quad &\Leftrightarrow |\mathbf{R}\widehat{\boldsymbol{\beta}} - r\mathbb{1}\,| > t^{\langle 1-\alpha/2\rangle}\cdot\widehat{\mathrm{Dt}}\left(\mathbf{R}\widehat{\boldsymbol{\beta}} \mid \mathbf{X}\right)\\
&\Leftrightarrow |r\mathbb{1} - \mathbf{R}\widehat{\boldsymbol{\beta}}| > t^{\langle 1-\alpha/2\rangle}\cdot\widehat{\mathrm{Dt}}\left(\mathbf{R}\widehat{\boldsymbol{\beta}} \mid \mathbf{X}\right)\\
&\Leftrightarrow (r\mathbb{1} - \mathbf{R}\widehat{\boldsymbol{\beta}}) \notin \left[\pm t^{\langle \alpha/2\rangle}\cdot\widehat{\mathrm{Dt}}\left(\mathbf{R}\widehat{\boldsymbol{\beta}} \mid \mathbf{X}\right)\right]\\
&\Leftrightarrow r\mathbb{1} \notin \left[\mathbf{R}\widehat{\boldsymbol{\beta}} \pm t^{\langle \alpha/2\rangle}\cdot\widehat{\mathrm{Dt}}\left(\mathbf{R}\widehat{\boldsymbol{\beta}} \mid \mathbf{X}\right)\right] \qquad (18)
\end{aligned}$$

No se rechaza H_0 si y solo si: $r\mathbb{1} \in \left[\mathbf{R}\widehat{\boldsymbol{\beta}} \pm t^{\langle \alpha/2\rangle}\cdot\widehat{\mathrm{Dt}}\left(\mathbf{R}\widehat{\boldsymbol{\beta}} \mid \mathbf{X}\right)\right] = \widehat{\mathrm{IC}}^{\mathbf{R}\widehat{\boldsymbol{\beta}}}_{1-\alpha}$.

8 Contrastes de hipótesis e intervalos de confianza para un solo parámetro

Si $\underset{1\times k}{\mathbf{R}} = \begin{bmatrix} 0 & \cdots & 0 & \underset{(j)}{1} & 0 & \cdots & 0 \end{bmatrix}$: el test t-student bilateral rechaza $H_0 : \mathbf{R}\boldsymbol{\beta} = \beta_j = b$ si

$$|\mathcal{T}_j| = \frac{|\widehat{\beta_j} - b\mathbb{1}\,|}{\widehat{\mathrm{Dt}}\left(\widehat{\beta_j} \mid \mathbf{X}\right)} > t^{\langle 1-\alpha/2\rangle}$$

$$\begin{aligned}
|\mathcal{T}_j| > t^{\langle 1-\alpha/2\rangle} \quad &\Leftrightarrow |\widehat{\beta_j} - b\mathbb{1}\,| > t^{\langle 1-\alpha/2\rangle}\cdot\widehat{\mathrm{Dt}}\left(\widehat{\beta_j} \mid \mathbf{X}\right)\\
&\Leftrightarrow |b\mathbb{1} - \widehat{\beta_j}| > t^{\langle 1-\alpha/2\rangle}\cdot\widehat{\mathrm{Dt}}\left(\widehat{\beta_j} \mid \mathbf{X}\right)\\
&\Leftrightarrow (b\mathbb{1} - \widehat{\beta_j}) \notin \left[\pm t^{\langle \alpha/2\rangle}\cdot\widehat{\mathrm{Dt}}\left(\widehat{\beta_j} \mid \mathbf{X}\right)\right]\\
&\Leftrightarrow b\mathbb{1} \notin \left[\widehat{\beta_j} \pm t^{\langle \alpha/2\rangle}\cdot\widehat{\mathrm{Dt}}\left(\widehat{\beta_j} \mid \mathbf{X}\right)\right] \qquad (19)
\end{aligned}$$

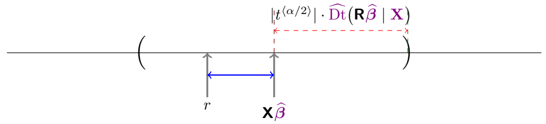
No se rechaza H_0 si y solo si: $b\mathbb{1} \in \left[\widehat{\beta_j} \pm t^{\langle \alpha/2\rangle}\cdot\widehat{\mathrm{Dt}}\left(\widehat{\beta_j} \mid \mathbf{X}\right)\right] = \widehat{\mathrm{IC}}^{\widehat{\beta_j}}_{1-\alpha}$.

9 Intervalos y contrastes

Denominamos *intervalo de confianza* a:

$$\widehat{\mathsf{IC}}^{\mathbf{R}\widehat{\boldsymbol\beta}}_{1-\alpha} \equiv \Big[\mathbf{R}\widehat{\boldsymbol\beta} \pm t^{\langle\alpha/2\rangle}\cdot\widehat{\mathrm{Dt}}\big(\mathbf{R}\widehat{\boldsymbol\beta}\mid \mathbf{X}\big)\Big].$$

$$\boxed{\widehat{\mathsf{IC}}^{\mathbf{R}\widehat{\boldsymbol\beta}}_{1-\alpha} \;=\; \{\text{Hipótesis no rechazables para } \mathbf{R}\boldsymbol\beta \text{ con significación } \alpha\}}$$



$H_0:\ \mathbf{R}\boldsymbol\beta = r$ no se rechaza si: $r\mathbb{1} \in \widehat{\mathsf{IC}}^{\mathbf{R}\widehat{\boldsymbol\beta}}_{1-\alpha}$.

Ejemplo

Continuación de del ejemplo del precio de las viviendas

Los intervalos de confianza de los parámetros a y b son de la forma

$$\widehat{\mathsf{IC}}^{\widehat{\beta_j}}_{1-\alpha} = \Big[\ \widehat{\beta_j}\ \pm\ t^{\langle\alpha/2\rangle}_{N-k}\cdot\widehat{\mathrm{Dt}}\big(\widehat{\beta_j}\mid\mathbf{X}\big)\Big]$$

por tanto, en el caso del efecto marginal de la superficie sobre el precio y de la constante sus estimaciones son respectivamente

$$\widehat{\mathsf{IC}}^{\widehat{b}}_{1-\alpha} = \Big[0.139 \pm (t^{\langle\alpha/2\rangle}_{12})\cdot 0.01873\Big];$$

$$\widehat{\mathsf{IC}}^{\widehat{a}}_{1-\alpha} = \Big[52.3509 \pm (t^{\langle\alpha/2\rangle}_{12})\cdot 37.285\Big];$$

☞**Código:** `EjPvivienda2.inp` ....

10 Estimación por intervalos de confianza (de una combinación lineal de betas)

Si se cumplen los supuestos: $\dfrac{\mathbf{R}(\widehat{\boldsymbol\beta}-\boldsymbol\beta\mathbb{1})}{\widehat{\mathrm{Dt}}(\mathbf{R}\widehat{\boldsymbol\beta}|\mathbf{X})} \underset{H_0}{\sim} t_{N-k}$, donde $\underset{1\times k}{\mathbf{R}}$:

$$\mathbb{P}_{H_0}\left(t^{\langle\alpha/2\rangle} < \frac{\mathbf{R}\big(\widehat{\boldsymbol\beta}-\boldsymbol\beta\mathbb{1}\big)}{\widehat{\mathrm{Dt}}\big(\mathbf{R}\widehat{\boldsymbol\beta}\mid\mathbf{X}\big)} < t^{\langle 1-\alpha/2\rangle}\right) = 1-\alpha$$

$$\mathbb{P}_{H_0}\Big(|\mathbf{R}\widehat{\boldsymbol\beta}-\mathbf{R}\boldsymbol\beta\mathbb{1}| < t^{\langle 1-\alpha/2\rangle}\cdot\widehat{\mathrm{Dt}}\big(\mathbf{R}\widehat{\boldsymbol\beta}\mid\mathbf{X}\big)\Big) = 1-\alpha$$

$$\mathbb{P}_{H_0}\Big(|\mathbf{R}\boldsymbol\beta\mathbb{1}-\mathbf{R}\widehat{\boldsymbol\beta}| < t^{\langle 1-\alpha/2\rangle}\cdot\widehat{\mathrm{Dt}}\big(\mathbf{R}\widehat{\boldsymbol\beta}\mid\mathbf{X}\big)\Big) = 1-\alpha$$

$$\mathbb{P}_{H_0}\Big(\mathbf{R}\boldsymbol\beta\mathbb{1} \in \Big[\mathbf{R}\widehat{\boldsymbol\beta} \pm t^{\langle\alpha/2\rangle}\cdot\widehat{\mathrm{Dt}}\big(\mathbf{R}\widehat{\boldsymbol\beta}\mid\mathbf{X}\big)\Big]\Big) = 1-\alpha$$

$$\mathbb{P}_{H_0}\Big(\mathbf{R}\boldsymbol\beta\mathbb{1} \in \widehat{\mathsf{IC}}^{\mathbf{R}\widehat{\boldsymbol\beta}}_{1-\alpha}\Big) = 1-\alpha. \qquad (20)$$

$\widehat{\mathsf{IC}}^{\mathbf{R}\widehat{\boldsymbol\beta}}_{1-\alpha}$ se denomina *estimador por intervalo* de $\mathbf{R}\boldsymbol\beta$ y $1-\alpha$ es el *nivel de confianza* del intervalo.

11 Regiones de confianza

Si $\underset{r\times k}{\mathbf{R}}$ es de rango r, la condición

$$\mathcal{F} = \big(\mathbf{R}\widehat{\boldsymbol\beta}-\mathbf{R}\boldsymbol\beta\mathbb{1}\big)\Big[\widehat{\mathrm{Var}}\big(\mathbf{R}\widehat{\boldsymbol\beta}\mid\mathbf{X}\big)\Big]^{-1}\big(\mathbf{R}\widehat{\boldsymbol\beta}-\mathbf{R}\boldsymbol\beta\mathbb{1}\big)/r \le c$$

define un elipsoide en $\mathbb{R}^k$. De esta manera, de (14) se deduce que

$$\mathbb{P}_{H_0}\big(\mathcal{F} < F^{\langle 1-\alpha\rangle}\big) = 1-\alpha \quad \text{(operando como para el test-}t\text{)}$$

$$\mathbb{P}_{H_0}\big(\ \mathbf{R}\widehat{\boldsymbol\beta}\ \in\ \widehat{\mathsf{IC}}^{\mathbf{R}\widehat{\boldsymbol\beta}}_{1-\alpha}\ \big) = 1-\alpha,$$

donde $\widehat{\mathsf{IC}}^{\mathbf{R}\widehat{\boldsymbol\beta}}_{1-\alpha} \subset \mathbb{R}^r$ se denomina *elipse (o elipsoide) de confianza*.

$\widehat{\mathsf{IC}}^{\mathbf{R}\widehat{\boldsymbol\beta}}_{1-\alpha}$ contiene los vectores $\boldsymbol{r}\mathbf{I} \in \mathbb{R}^r$ tales que $H_0:\ \mathbf{R}\boldsymbol\beta = \boldsymbol{r}$ no se rechaza con un nivel de significación α.

Ejemplo

Región de confianza de dos parámetros:

$H_0: \ \beta_1 = a$, y $\beta_2 = b$; $\quad k=2$; $\quad \mathbf{R}\boldsymbol{\beta}{=}\boldsymbol{r}$; $\quad \mathbf{R} = \left(\begin{smallmatrix}1&0\\0&1\end{smallmatrix}\right)$; $\quad \boldsymbol{r} = \left(\begin{smallmatrix}a\\b\end{smallmatrix}\right)$.

solución tentativa pero incorrecta

No rechazar si

$$\begin{pmatrix}a\\b\end{pmatrix} \in \text{región tal que} \begin{cases} a \in \left[\widehat{\beta_1} \pm t^{\langle\alpha/2\rangle}\cdot \widehat{\mathrm{Dt}}\left(\widehat{\beta_1} \mid \mathbf{X}\right)\right] \\ b \in \left[\widehat{\beta_2} \pm t^{\langle\alpha/2\rangle}\cdot \widehat{\mathrm{Dt}}\left(\widehat{\beta_2} \mid \mathbf{X}\right)\right] \end{cases}$$

que es un rectángulo (formado por el producto cartesiano de los intervalos de confianza individuales).

solución correcta

No rechazar si

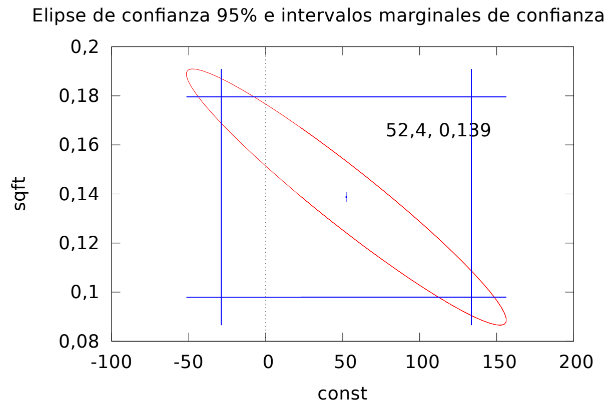
$$\begin{pmatrix}a\\b\end{pmatrix} \in \left\{ \begin{pmatrix}\beta_1\\\beta_2\end{pmatrix} \middle| (\widehat{\boldsymbol{\beta}} - \boldsymbol{r})^\intercal \left[\widehat{\mathrm{Var}}(\widehat{\boldsymbol{\beta}} \mid \mathbf{X})\right]^{-1} (\widehat{\boldsymbol{\beta}} - \boldsymbol{r}) < 2 \cdot F_{2,N-2}^{\langle 1-\alpha\rangle} \right\}$$

que es una elipse.

Ejemplo

Continuación de "precio de las viviendas"

☞**Código:** `EjPvivienda2.inp` . . . .



`Análisis -> Elipse de Confianza`

Enlace a algunas **prácticas de la Lección 8**

Fin de la lección

Lección 9

1 Estimación restringida

Motivos:

- análisis previo $\rightarrow$ restricciones plausibles
  (restricciones correctas $\rightarrow$ estimación más precisa)
- la comparación entre la estimación restringida y la no restringida permite contrastar hipótesis

Ejecución:

- por sustitución
- método de mínimos cuadrados restringidos linealmente (MCR)

Ejemplo

Estimación restringida vía sustitución Suponga el modelo en logaritmos (de una función de Cobb-Douglas):

$$\ln Y = \beta_1 \mathbb{1} + \beta_2 \ln K + \beta_3 \ln L + U$$

Considere la restricción: $\beta_2 + \beta_3 = 1$. La estimación imponiendo *rendimientos constantes a escala* se logra re-escribiendo el modelo:

$$\begin{aligned} \ln Y &= \beta_1 \mathbb{1} + \beta_2 \ln K + (1 - \beta_2) \ln L + U \\ \ln Y - \ln L &= \beta_1 \mathbb{1} + \beta_2 (\ln K - \ln L) + U \\ \ln \frac{Y}{L} &= \beta_1 \mathbb{1} + \beta_2 \ln \frac{K}{L} + U \end{aligned}$$

y estimando por MCO el modelo con los nuevos regresores.

. . . pero hay otra forma de lograrlo. . .

2 Mínimos cuadrados restringidos (MCR)

Bajo los supuestos habituales, buscamos un estimador $\widehat{\boldsymbol{\beta}^*}$ que cumpla el conjunto de restricciones lineales:

$$\mathbf{R}\widehat{\boldsymbol{\beta}^*} = \mathbf{I}\boldsymbol{r}; \qquad \operatorname{rg}\Big(\underset{r\times k}{\mathbf{R}}\Big) = r.$$

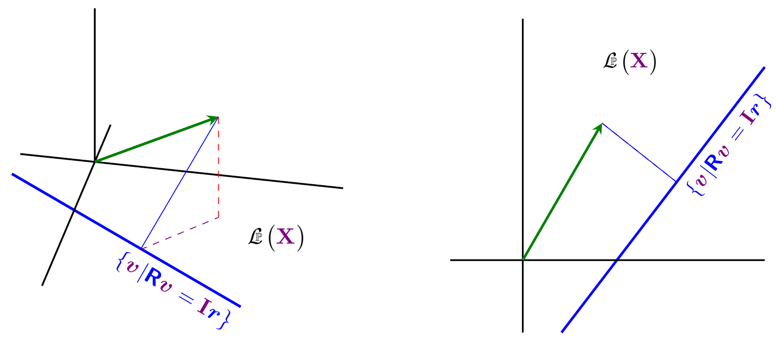
El estimador de Mínimos Cuadrados con Restricciones Lineales

$$\widehat{\boldsymbol{\beta}^*} = \widehat{\boldsymbol{\beta}} - \left(\mathbf{X}^\intercal\mathbf{X}\right)^{-1}\mathbf{R}^\intercal\left[\mathbf{R}\left(\mathbf{X}^\intercal\mathbf{X}\right)^{-1}\mathbf{R}^\intercal\right]^{-1}\left(\mathbf{R}\widehat{\boldsymbol{\beta}} - \mathbf{I}\boldsymbol{r}\right) \qquad (21)$$

La estimación correspondiente a la muestra $\mathbf{X}$ es

$$\widehat{\boldsymbol{\beta}^*} = \widehat{\boldsymbol{\beta}} - \left(\mathbf{X}^\intercal\mathbf{X}\right)^{-1}\mathbf{R}^\intercal\left[\mathbf{R}\left(\mathbf{X}^\intercal\mathbf{X}\right)^{-1}\mathbf{R}^\intercal\right]^{-1}\left(\mathbf{R}\widehat{\boldsymbol{\beta}} - \boldsymbol{r}\right) \qquad (22)$$

3 Mínimos cuadrados restringidos



Nótese que $\widehat{\boldsymbol{e}^*}$, $\widetilde{\boldsymbol{e}}$ y $\mathbf{X}(\widehat{\boldsymbol{\beta}^*} - \widetilde{\boldsymbol{\beta}})$ forman un triángulo rectángulo (MCRL).

4 Estimador MCRL

El estimador **siempre verifica** la condición: $\quad \mathbf{R}\widehat{\boldsymbol{\beta}^*} = \mathbf{I}\boldsymbol{r}$

Si $\mathbf{R}\boldsymbol{\beta} = \boldsymbol{r}$ se cumple (restricción es cierta), de (21)

$\mathrm{E}\left(\widehat{\boldsymbol{\beta}^*}\right) = \boldsymbol{\beta}$ ¡sólo cuando se cumple restricción!... ($\boldsymbol{\beta}$ es *desconocido*)

y además, tanto si la restricción es cierta como si no

$$\mathbb{V}ar\left(\widehat{\boldsymbol{\beta}} \mid \mathbf{X}\right) \geq \mathbb{V}ar\left(\widehat{\boldsymbol{\beta}^*} \mid \mathbf{X}\right)$$

ya que

$$\mathbb{V}ar\left(\widehat{\boldsymbol{\beta}^*} \mid \mathbf{X}\right) = \mathbb{V}ar\left(\widehat{\boldsymbol{\beta}} \mid \mathbf{X}\right) - \sigma^2\left(\mathbf{X}^\intercal\mathbf{X}\right)^{-1}\mathbf{R}^\intercal\left[\mathbf{R}\left(\mathbf{X}^\intercal\mathbf{X}\right)^{-1}\mathbf{R}^\intercal\right]^{-1}\mathbf{R}\left(\mathbf{X}^\intercal\mathbf{X}\right)^{-1},$$

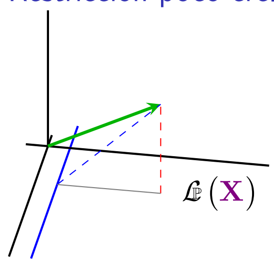
donde las tres matrices son definidas positivas.

5 Contraste de la F mediante sumas residuales

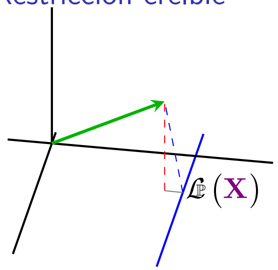
$$\mathcal{F} = \frac{\left(\widehat{\boldsymbol{e}^*} \cdot \widehat{\boldsymbol{e}^*} - \widehat{\boldsymbol{e}} \cdot \widehat{\boldsymbol{e}}\right)/r}{\widehat{\boldsymbol{e}} \cdot \widehat{\boldsymbol{e}}/(N-k)} = \frac{N-k}{r} \cdot \frac{SRC^* - SRC}{SRC} \underset{H_0}{\sim} F_{r,N-k} \tag{23}$$

donde $H_0 : \mathbf{R}\boldsymbol{\beta} = \boldsymbol{r}$

Restricción poco creíble



Restricción creíble



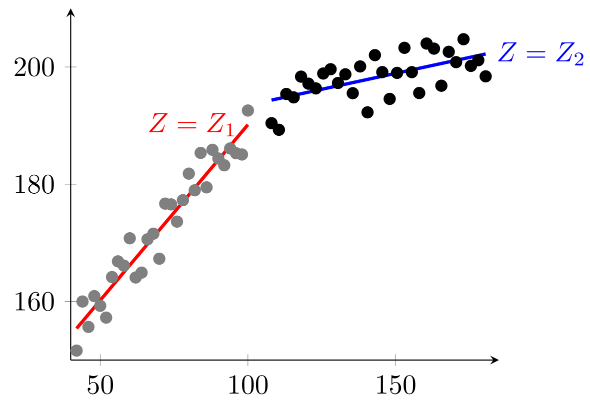
6 Contraste de la F en modelos con constante

$$\mathcal{F} = \frac{N-k}{r} \cdot \frac{R^2 - R^{2*}}{1 - R^2} \underset{H_0}{\sim} F_{r,N-k}$$

Contraste de significación global

$$\mathcal{F} = \frac{N-k}{k-1} \cdot \frac{R^2}{1 - R^2} \underset{H_0}{\sim} F_{k-1,N-k} \quad \text{(caso especial)}$$

7 Cambio estructural del modelo



8 Contrastes de cambio estructural: *Test de Chow*

H_0: parámetros no varían en la muestra (No cambio estructural)
H_1: σ^2 cte., pero betas toman dos conjuntos de valores.

Modelo sin restringir

$$\begin{aligned} Y_n &= {}_{n|}\mathbf{X}\boldsymbol{\beta}_A + U_n \quad n \in \{\text{índices correspondientes al caso } A\} \\ Y_n &= {}_{n|}\mathbf{X}\boldsymbol{\beta}_B + U_n \quad n \in \{\text{índices correspondientes al caso } B\} \end{aligned},$$

Modelo restringido $H_0: \boldsymbol{\beta}_A = \boldsymbol{\beta}_B$, es decir,

$$Y_n = {}_{n|}\mathbf{X}\boldsymbol{\beta} + U_n \quad n = 1 : N \ ,$$

$U_n \sim N(0, \sigma^2)$ para $n = 1, \dots, N$ en ambos modelos.

9 Contrastes de cambio estructural: *Test de Chow*

Modelo sin restringir $2k$ parámetros estimados $(\boldsymbol{\beta}_A,\ \boldsymbol{\beta}_B)$;
y además $SRC = SRC_A + SRC_B$.

Modelo restringido k restricciones lineales:
$(\boldsymbol{\beta}_A)_{|j} = (\boldsymbol{\beta}_B)_{|j};\ j = 1 : k.$

Por lo tanto,

$$\begin{aligned} \mathcal{F} &= \tfrac{N-2k}{k}\tfrac{SRC^* - SRC}{SRC} \\ &= \tfrac{N-2k}{k}\tfrac{SRC^* - (SRC_A + SRC_B)}{(SRC_A + SRC_B)} \end{aligned}$$

10 Contraste de Jarque-Bera

$$JB = \frac{N-k}{6}\left(S^2 + \frac{1}{4}(K-3)^2\right)$$

donde S es el coeficiente de asimetría muestral, y K el coeficiente de curtosis

Si la muestra proviene de una distribución normal, el contraste JB se distribuye asintóticamente como una χ^2_2

(Gretl dispone de varios contrastes de normalidad, entre ellos el JB)

Enlace a algunas **prácticas de la Lección 9**

________________________________Fin de la lección

Lección 10

1 Elasticidad

$$\frac{\partial \ln z}{\partial z} = \frac{1}{z} \quad \Rightarrow \quad \partial \ln z = \frac{\partial z}{z} = \text{cambio relativo (infinitesimal) de } z$$

La elasticidad η de y respecto a x se define cómo:

$$\eta = \frac{\text{cambio relativo infinitesimal de } y}{\text{cambio relativo infinitesimal de } x} = \frac{\partial \ln y}{\partial \ln x} = \frac{\partial y / y}{\partial x / x} = \frac{x}{y} \cdot \frac{\partial y}{\partial x}.$$

Relacionemos esto con distintas formas funcionales de los modelos (¡todos lineales en los parámetros!).

Efectos marginales y elasticidades para distintas funciones lineales en los parámetros

Nombre	**Forma Funcional**	**Efecto Marginal:** $\frac{dy}{dx}$	**Elasticidad:** $\frac{x}{y}\frac{dy}{dx}$
Lineal	$y = \alpha + \beta x$	β	$\beta x / y$
Lin-Log	$y = \alpha + \beta \ln x$	β / x	β / y
Reciproco	$y = \alpha + \beta 1/x$	$-\beta / x^2$	$-\beta/(xy)$
Cuadrático	$y = \alpha + \beta x + \gamma x^2$	$\beta + 2\gamma x$	$(\beta + 2\gamma x)x/y$
Interacción	$y = \alpha + \beta x + \gamma x z$	$\beta + \gamma z$	$(\beta + \gamma z)x/y$
Log-Lin	$\ln y = \alpha + \beta x$	βy	βx
Log-Reciproco	$\ln y = \alpha + \beta(1/x)$	$-\beta y / x^2$	$-\beta / x$
Log-Cuadrático	$\ln y = \alpha + \beta x + \gamma x^2$	$y(\beta + 2\gamma x)$	$x(\beta + 2\gamma x)$
Log-Log	$\ln y = \alpha + \beta \ln x$	$\beta y / x$	β
Logístico	$\ln \left[\frac{y}{1-y}\right] y = \alpha + \beta x$	$\beta y (1-y)$	$\beta(1-y)x$

Tabla: Efectos marginales y elasticidades para distintas formas funcionales

2 Interpretación de coeficientes en modelos con logs

Modelo	Interpretación	
$y = \alpha + \beta x$	$\beta = \frac{\partial y}{\partial x}$	Cambio esperado en nivel de y si x aumenta una unidad
$\ln(y) = \alpha + \beta \ln(x)$	$\beta = \frac{x}{y}\frac{\partial y}{\partial x}$	*(Aprox.)* Cambio *porcentual* esperado de y si x aumenta un uno por ciento (en tanto por uno, i.e., 0.01)
$\ln(y) = \alpha + \beta x$	$\beta = \frac{1}{y}\frac{\partial y}{\partial x}$	*(Aprox.)* Cambio relativo esperado de y (en tanto por uno) si x aumenta una unidad
$y = \alpha + \beta \ln(x)$	$\beta = x\frac{\partial y}{\partial x}$	*(Aprox.)* Cambio esperado en el nivel de y si x aumenta un uno por ciento (en tanto por uno)

(derivando respecto a x, sustituyendo $\partial \ln z$ por $\frac{\partial z}{z}$ y despejando)

Ejemplo

Función de consumo (lin-lin):

$$CON = \beta_1 \mathbb{1} + \beta_2 RD + U$$

Ejemplo

Ecuación de salarios (log-lin):

$$SALAR = e^{(\beta_1 \mathbb{1} + \beta_2 EDUC + \beta_3 ANTIG + \beta_4 EXPER + U)};$$

Al tomar logaritmos tenemos un nuevo modelo para $\ln(SALAR)$ que es lineal en los parámetros:

$$\ln(SALAR) = \beta_1 \mathbb{1} + \beta_2 EDUC + \beta_3 ANTIG + \beta_4 EXPER + U$$

Ejemplo

Precio de la vivienda (lin-log):

$$PRICE = \beta_1 \mathbb{1} + \beta_2 \ln SQFT + U.$$

Ejemplo

Función de producción Cobb-Douglas (log-log):

$$Q = cK^{\beta_2} L^{\beta_3} \nu;$$

Tomando logaritmos tenemos

$$\ln Q = \beta_1 \mathbb{1} + \beta_2 \ln K + \beta_3 \ln L + U,$$

donde $\beta_1 = \ln c$, y $U = \ln \nu$.

(Lección 10) Ejercicio en clase. N-1.

☞**Código:** `POE2-4.inp` ....

Cargue los datos `food.gdt` del libro POE, sobre los gastos en alimentación *food_exp* de las familias y la renta disponible *income*.

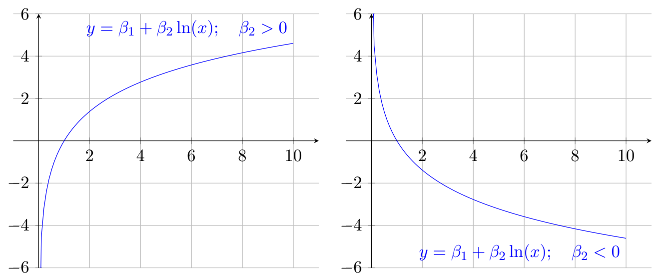
(a) Ajuste por MCO el gasto en comida en función de la renta disponible
(b) Observe los estadísticos principales de ambas variables
(c) Grafique un diagrama de dispersión del gasto sobre la renta
(d) Muestre los valores de ambas variables
(e) Calcule la elasticidad de la demanda de alimentos respecto de la renta en el valor medio muestral de la renta, donde

$$\left(\frac{\text{variación \% de } x}{\text{variación \% de } y}\right) \qquad \approx elasticidad = \frac{\partial y / y}{\partial x / x} = \frac{\partial y}{\partial x}\frac{x}{y} \approx \widehat{\beta_2}\frac{m_{\boldsymbol{x}}}{m_{\boldsymbol{y}}}$$

(f) ¿Qué gasto prevé este modelo para una familia cuya renta asciende a 20?
(g) Realice un contraste de normalidad para los residuos ¿Puede rechazar que la distribución es normal?
(h) Grafique los residuos de la regresión ¿Le parece que la varianza de los residuos es independiente de la renta? ¿Es creíble que se cumple el supuesto de homocedasticidad en este modelo?

3 Modelo Lin-Log

$$y = \beta_1 + \beta_2 \ln x$$



Pendiente

$\frac{dy}{dx} = \beta_2/x \qquad \Longrightarrow \qquad \Delta y \approx \beta_2 \times \frac{\Delta x}{x}$ (si es pequeña)

1% incremento de x $\left(\frac{\Delta x}{x} = 0.01\right) \Rightarrow$ Incremento $Y = \frac{\beta_2}{100}$ unid.

Elasticidad $\eta = \frac{x}{y}\frac{dy}{dx} = \beta_2/y$ (decreciente en valor absoluto)

(Lección 10) Ejercicio en clase. N-2.

☞**Código:** `RamanathanEX6-1.inp` ....

Precio de casas unifamiliares Use `data4-1.gdt`.

(a) Estime por MCO: $PRICE = \beta_1 \mathbb{1} + \beta_2 SQFT + U$; y añádalo a la tabla de modelos.
(b) Estime después el siguiente modelo Lin-Log

$$PRICE = \beta_1 \mathbb{1} + \beta_2 \ln SQFT + \beta_3 \ln BEDRMS + \beta_4 \ln BATH + U;$$

(c) Decida si es necesario quitar alguna variable del modelo. Opere secuencialmente (añadiendo a la tabla de modelos aquellos que le parezcan interesantes) hasta quedarse con un modelo definitivo.
(d) Compare los resultados de los distintos modelos ajustados.
(e) Calcule las elasticidades del modelo lineal y del siguiente modelo Lin-Log:

$$PRICE = \beta_1 \mathbb{1} + \beta_2 \ln SQFT + U;$$

para casas con superficies de 1500, 2000 y 2500 pies al cuadrado respectivamente.
(f) ¿Cuanto aumenta el precio de las casa por un aumento del 1% de su superficie (nótese que este aumento es independiente del tamaño de la casa (lin-log)).

4 Modelo en semi-logaritmos (Log-Lin)

Ejemplo

Modelo de crecimiento constante: Suponga que la variable P crece a una tasa constante g:

$$P_t = P_{t-1} \cdot (1+g).$$

Mediante sustituciones sucesivas, llegamos a

$$P_t = P_0(1+g)^t.$$

Este modelo se puede linealizar tomando logaritmos:

$$\underbrace{\ln P_t}_{Y} = \underbrace{\ln P_0}_{\beta_1} + \underbrace{\ln(1+g)}_{\beta_2} \cdot \underbrace{t}_{X} \qquad \Rightarrow \qquad g = \exp(\beta_2) - 1 \quad (24)$$

5 Ejemplo de modelo en semi-logaritmos (Log-Lin)

Si el retorno de un año adicional de estudios es g, entonces, $w_1 = (1+g)w_0$, y $w_2 = (1+g)^2 w_0$, En general

$$w_t = (1+g)^t w_0.$$

Tomando logs: $\ln w_t = \ln w_0 + \ln(1+g) \cdot t = \beta_1 + \beta_2 \cdot t$.

Ejemplo

$$SALAR = e^{(\beta_1 \mathbb{1} + \beta_2 EDUC + \beta_3 ANTIG + \beta_4 EXPER + U)};$$

Tomando logaritmos → modelo para $\ln(SALAR)$

$$\ln(SALAR) = \beta_1 \mathbb{1} + \beta_2 EDUC + \beta_3 ANTIG + \beta_4 EXPER + U$$

Si $\beta_2 = .03$; *cada año educ → incremen. esperado (aprox.) salario 3%*
(mejor $g = \exp(\beta_2) - 1$) → $g = \exp(0.03) - 1 = 0.030455$.

(Lección 10) Ejercicio en clase. N-3.

☞**Código:** `RamanathanEX6-5.inp` ....

Modelo para los salarios. Abra el conjunto de datos `data6-4.gdt`, del libro de Ramanathan, con datos del salarios mensuales (*wage*), años de educación (*educ*) y de experiencia (*exper*), y la edad (*age*) de 49 trabajadores.

(a) Estime el modelo

$$\ln W = \beta_1 \mathbb{1} + \beta_2 \cdot EDUC + \beta_3 \cdot EDUC^2 + \beta_4 \cdot AGE + \beta_5 \cdot AGE^2 + \beta_6 \cdot EXPER + \beta_7 \cdot EXPER^2 + U$$

(b) Vaya eliminando variables no significativas hasta obtener un modelo final.
(c) ¿Qué efecto estimado tiene un año adicional de experiencia?
(d) Recordando que

$$\widehat{W} = \exp\left(\mathsf{X}\widehat{\boldsymbol{\beta}} + \widehat{s^2}/2\right),$$

calcule los salarios estimados por el modelo y compárelos con los salarios de la muestra. Con el diagrama de dispersión de salarios observados y ajustados podrá comprobar que este modelo no funciona muy bien.
(e) Pese a ello calcule el efecto estimado que tiene un año adicional de educación en el salario de trabajadores con 1 y 7 años de formación respectivamente.

6 Comparación de coeficientes de determinación entre modelos

R^2 de modelos Lin-Lin y Log-Lin no son comparables (distinto regresando)

- Una forma de intentar comparar ajustes es calcular el cuadrado de la correlación entre $\boldsymbol{y}$ y $\widehat{\boldsymbol{y}}$; donde

$$\widehat{Y} = \exp\left(\widehat{\ln Y} + \widehat{\sigma^2}/2\right)$$

- O calcular los estadísticos de selección empleando la suma de errores al cuadrado y la varianza estimada:

$$SRC = \sum(Y - \widehat{Y})^2; \qquad \widetilde{\sigma^2} = \frac{SRC}{n-k}$$

(Lección 10) Ejercicio en clase. N-4.

☞**Código:** `RamanathanEX6-6.inp` ....

Modelo para los salarios. Abra el conjunto de datos `data6-4.gdt`, del libro de Ramanathan, con datos del salarios mensuales (*wage*), así como años de educación (*educ*), años de experiencia (*exper*) y edad (*age*) de 49 trabajadores.

(a) Estime los modelos

$$W = \beta_1 \mathbb{1} + \beta_2 \cdot EDUC^2 + \beta_3 \cdot EXPER + U$$

$$\ln W = \beta_1 \mathbb{1} + \beta_2 \cdot EDUC^2 + \beta_3 \cdot EXPER + U$$

Aunque los R^2 parecen semejantes, no son comparables.

(b) Guarde los salarios estimados por el segundo modelo, así como los errores y la varianza estimada de los errores.

(c) Calcule el cuadrado de la correlación entre los salarios observados y los estimados (o predichos). ¿Qué modelo presenta un mejor ajuste? ¿El primero o el segundo?

(d) Cargando la función `criteria.gfn`, calcule los criterios de selección de modelo (mire el guión adjunto). A la luz de los resultados, ¿qué modelo parece preferible?

7 Ejemplo de modelo Log-Log

Ejemplo

Función de producción Cobb-Douglas

Tomando logaritmos en $Q = cK^{\beta_2}L^{\beta_3}\nu$, tenemos

$$\ln Q = \beta_1 \mathbb{1} + \beta_2 \ln K + \beta_3 \ln L + U,$$

donde $\beta_1 = \ln c$, y $U = \ln \nu$

En los modelos Log-Log los parámetros β_j son elasticidades constantes. . .

Si $\beta_2 = 5$; *un ΔK de 1% $\rightarrow$ incremento esperado producción 5%.*

(Lección 10) Ejercicio en clase. N-5.

☞**Código:** `RamanathanAp6-11.inp` ....

Elasticidades en la demanda del transporte en autobús. Abra el conjunto de datos `data4-4.gdt`, del libro de Ramanathan.

(a) Estime un modelo de regresión entre el logaritmo de *bustravl* y el resto de variables, también en logaritmos.

(b) Elimine secuencialmente del modelo las variables no significativas al 10% ni individual ni conjuntamente.

(c) Decimos que la demanda es inelastica cuando el valor absoluto de la elasticidad es menor que 1 (elástica en caso contrario). Contraste si la elasticidad de la demanda de viajes de autobús con respecto a las distintas variables explicativas es 1.

Enlace a algunas **prácticas de la Lección 10**

Fin de la lección

Lección 11

1 Variables ficticias (*dummies*)

Variable discreta que clasifica "categorías"
(Indicador que toma valores 0 ó 1)

Usos:

- inclusión de información cualitativa (empresa, sexo, etc.)
- división de la muestra en dos periodos (contraste cambio estructural)

En este caso los coeficientes β_j tienen otra interpretación (no son pendientes).

Ejemplo

Relación entre salario por hora trabajada percibido por el trabajador n-ésimo (W_n) y su nivel de estudios (variable cualitativa representada por 3 dummies:)

$$W = \text{Salario del trabajador } n\text{-ésimo}$$

$$\mathbb{1}_P = \begin{cases} 1, & \text{sin estudios o sólo estudios primarios } (\mathbf{P}) \\ 0, & \text{en caso contrario} \end{cases}$$

$$\mathbb{1}_M = \begin{cases} 1, & \text{con estudios medios (no superiores) } (\mathbf{M}) \\ 0, & \text{en caso contrario} \end{cases}$$

$$\mathbb{1}_S = \begin{cases} 1, & \text{con estudios superiores } (\mathbf{S}) \\ 0, & \text{en caso contrario} \end{cases}$$

$$W = \alpha_1 \mathbb{1}_P + \alpha_2 \mathbb{1}_M + \alpha_3 \mathbb{1}_S + U \tag{25}$$

donde $\mathbb{1}_P + \mathbb{1}_M + \mathbb{1}_S = \mathbb{1}$.

La matriz de regresores es **X** es

$$\mathbf{X} = \begin{bmatrix} \underset{N_1\times 1}{\mathit{1}} & \mathit{0} & \mathit{0} \\ \mathit{0} & \underset{N_2\times 1}{\mathit{1}} & \mathit{0} \\ \mathit{0} & \mathit{0} & \underset{N3\times 1}{\mathit{1}} \end{bmatrix},$$

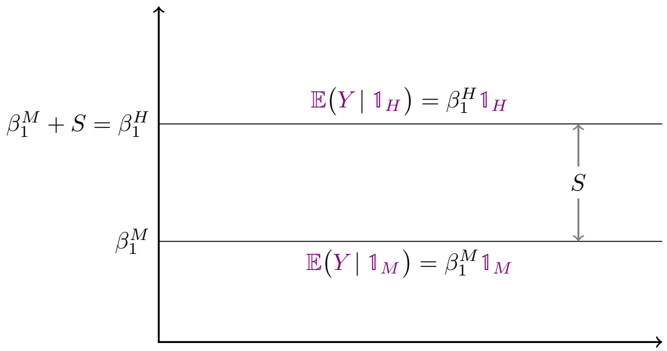
N_j: número de trabajadores del grupo j.
Ecuaciones normales $\mathbf{X}^\intercal\mathbf{X}\widehat{\boldsymbol{\alpha}} = \mathbf{X}^\intercal\boldsymbol{w}$:

$$\begin{bmatrix} N_1 & 0 & 0 \\ 0 & N_2 & 0 \\ 0 & 0 & N_3 \end{bmatrix} \begin{bmatrix} \widehat{\alpha_1} \\ \widehat{\alpha_2} \\ \widehat{\alpha_3} \end{bmatrix} = \begin{bmatrix} \sum_{n\in P} w_n \\ \sum_{n\in M} w_n \\ \sum_{n\in S} w_n \end{bmatrix},$$

por lo tanto $\widehat{\alpha_j} = N_j^{-1}\sum_{n=1}^{N_j} w_n = m_{\boldsymbol{w}_j}$

Diferentes términos constantes

$$\mathbb{1}_H(\omega) = \begin{cases} 1 & \omega \in H \\ 0 & \omega \notin H \end{cases} \qquad \text{y donde } \mathbb{1}_H + \mathbb{1}_M = \mathbb{1}$$



(Lección 11) Ejercicio en clase. N-1.

☞**Código:** `RamanathanPp7-1.inp` ....

Diferencias salariales entre hombres y mujeres.
Abra el conjunto de datos `data7-1.gdt`, del libro de Ramanathan, con datos sobre 49 trabajadores.

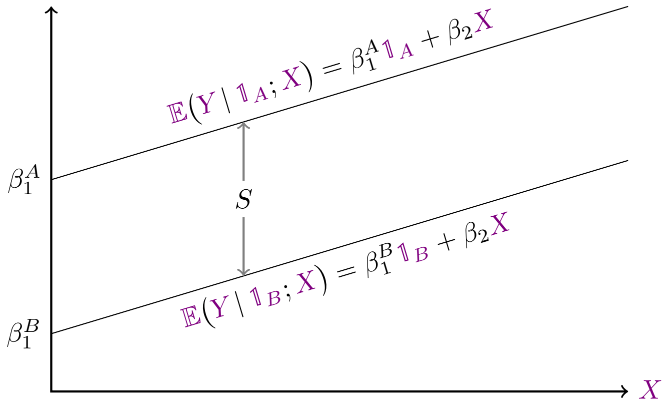
(a) Estime el modelo

$$WAGE = \beta_1\mathbb{1} + \beta_2 D + U$$

donde D es una variable que toma el valor 1 si el trabajador es varón.

(b) Interprete los coeficientes.
Calcule los salarios medios de hombres y mujeres, así como la diferencia de dichas medias. ¿Confirman su interpretación de los coeficientes?

La misma idea se puede generalizar



$$\beta_1^A = \beta_1^B + S$$

2 Modelo Log-lin con variables binarias

Suponga el modelo

$$\ln(Y) = a\mathbb{1} + bX + cD + U$$

donde D solo toma los valores cero o uno.

Calculando la exponencial de esta expresión:

- el crecimiento porcentual $\frac{\Delta Y}{Y}$ al pasar de $D = 0$ a $D = 1$ es
$$100[\exp(c) - 1]$$
- el crecimiento porcentual $\frac{\Delta Y}{Y}$ al pasar de $D = 1$ a $D = 0$ es
$$100[\exp(-c) - 1]$$

(Lección 11) Ejercicio en clase. N-2.

☞**Código:** `RamanathanEX7-1.inp` ....

Diferencias salariales entre hombres y mujeres. Abra el conjunto de datos `data7-2.gdt`, del libro de Ramanathan, con datos sobre 49 trabajadores.

(a) Estime el modelo

$$WAGE = \beta_1\mathbb{1} + \beta_2 D + \beta_3 EXPER + U$$

donde D es una variable ficticia que toma el valor 1 si el trabajador es varón. Interprete los coeficientes.

(b) Estime el modelo

$$\ln WAGE = \beta_1\mathbb{1} + \beta_2 D + \beta_3 EXPER + U$$

Interprete los coeficientes.

(c) Estime el modelo

$$\ln WAGE = \beta_1\mathbb{1} + \beta_2 D + \beta_3 EXPER + \beta_4 EDUC + U.$$

Interprete los coeficientes y compare los resultados de los modelos.

Ejemplo

Un modelo de salarios más completo: Contemplemos además las variables *antigüedad en la empresa* (A), los años de *experiencia* en el sector (X)

$$W = \beta_1\mathbb{1} + \beta_2 A + \beta_3 X + \alpha_1\mathbb{1}_P + \alpha_2\mathbb{1}_M + \alpha_3\mathbb{1}_S + U \qquad (26)$$

Aquí

- β_1 salario "autónomo" común a todos los trabajadores
- β_2 efecto antigüedad
- β_3 efecto experiencia
- α_j efecto del nivel de estudios j

Pero $\mathbb{1} = \mathbb{1}_P + \mathbb{1}_M + \mathbb{1}_S$

Posibles soluciones:

- Reemplazar la constante $\mathbb{1}$ por $\mathbb{1}_P + \mathbb{1}_M + \mathbb{1}_S$. Operando:
$$W = \beta_2 A + \beta_3 X + \delta_1\mathbb{1}_P + \delta_2\mathbb{1}_M + \delta_3\mathbb{1}_S + U \qquad (27)$$
  - $\delta_j = (\beta_1 + \alpha_j)$ combinación: salario "autónomo" y Niv. Est. j
- Reemplazar $\mathbb{1}_M$ por $(\mathbb{1} - \mathbb{1}_P - \mathbb{1}_S)$. Operando:
$$W = \theta_0\mathbb{1} + \beta_2 A + \beta_3 X + \theta_1\mathbb{1}_P + \theta_3\mathbb{1}_S + U \qquad (28)$$
  - $\theta_0 = (\beta_1 + \alpha_2)$ es como δ_2 de (27) (autónomo + Est. **M**)
  - $\theta_1 = (\alpha_1 - \alpha_2)$ pérdida por tener estudios **P** en lugar de **M**
  - $\theta_3 = (\alpha_3 - \alpha_2)$ ganancia por tener estudios **S** en lugar de **M**

  (el referente es la categoría eliminada: Estudios **M**)

Piense en la interpretación con otras soluciones alternativas.

3 Contrastes de homogeneidad entre grupos

¿Difiere el salario de trabajadores con distinto nivel de educación?

- Modelo original (26)

$$W = \beta_1 \mathbb{1} + \beta_2 A + \beta_3 X + \alpha_1 \mathbb{1}_P + \alpha_2 \mathbb{1}_M + \alpha_3 \mathbb{1}_S + U$$

$$H_0: \ \alpha_1 = \alpha_2 = \alpha_3$$

  No se puede contrastar debido a la multicolinealidad

- Modelo transformado (28) *(quitando $\mathbb{1}_M$)*

$$W = \theta_0 \mathbb{1} + \beta_2 A + \beta_3 X + \theta_1 \mathbb{1}_P + \theta_3 \mathbb{1}_S + U$$

$$H_0: \ \theta_1 = 0 \ \text{ y } \ \theta_3 = 0$$

4 Variables ficticias: contrastes de homogeneidad (constante)

$$Y = \beta_1 \mathbb{1} + \beta_2 X + U^*. \tag{29}$$

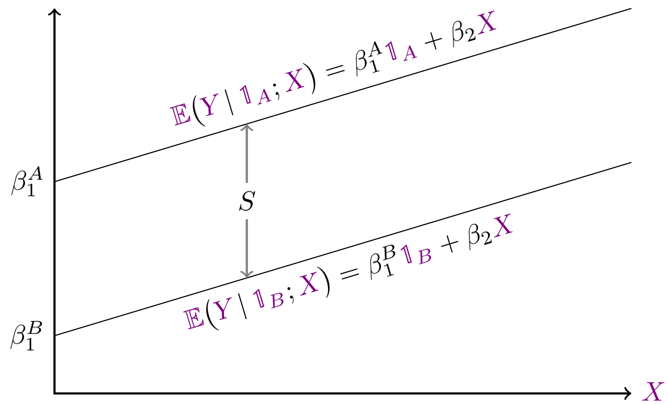
Partición en sub-muestras A y B.

Si sospechamos que β_1 cambia $\to$ Modelo no restringido:

$$Y = \beta_1^A \mathbb{1}_A + \beta_1^B \mathbb{1}_B + \beta_2 X + U, \tag{30}$$

donde

$$\mathbb{1}_A(\omega) = \begin{cases} 1 & \omega \in A \\ 0 & \omega \notin A \end{cases}, \quad \text{y donde } \ \mathbb{1}_A + \mathbb{1}_B = \mathbb{1}.$$



5 Variables ficticias: contrastes de homogeneidad (constante)

Contraste $H_0: \ \beta_1^A = \beta_1^B$. Dos opciones:

1. Contraste F de sumas residuales $\boxed{\text{F121}}$:
   Estimando (29) y (30) $\quad (H_1: \ \beta_1^A \neq \beta_1^B)$
2. Por sustitución: $\quad \mathbb{1}_B = \mathbb{1} - \mathbb{1}_A$ en (30);

$$Y = \beta_1^B \mathbb{1} + \alpha \mathbb{1}_A + \beta_2 X + U, \tag{31}$$

   donde $\alpha \equiv \beta_1^A - \beta_1^B$ $\quad$ (29 y 31 idénticas bajo H_0).

   Ahora $H_0: \ \alpha = 0$;
   (basta contraste de signif. individual; *uni o bilateral*).

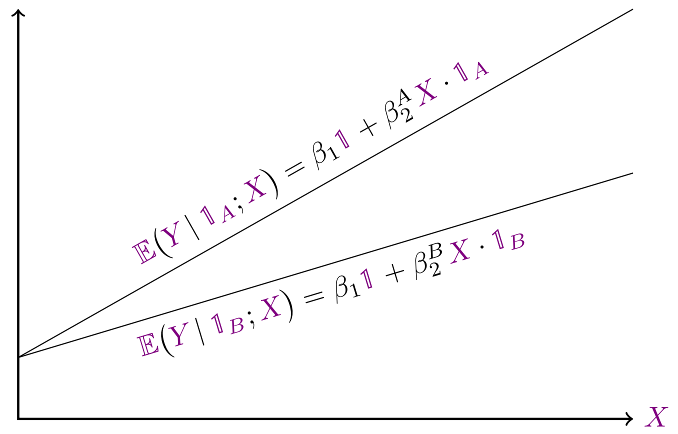
6 Variables ficticias: contrastes de homogeneidad (pendiente)

$$Y = \beta_1 \mathbb{1} + \beta_2 X + U^*.$$

Partición en sub-muestras A y B.

Si sospechamos β_2 (pendiente) cambia → Modelo no restringido:

$$Y = \beta_1 \mathbb{1} + \ \beta_2^A b X \cdot \mathbb{1}_A \ + \ \beta_2^B X \cdot \mathbb{1}_B \ + U, \qquad (32)$$



7 Variables ficticias: contrastes de homogeneidad (pendiente)

Contraste $H_0: \ \beta_2^A = \beta_2^B$. Dos opciones:

1. Por sumas residuales:
   Estimando (29) y (32) $\quad (H_1: \ \beta_2^A \neq \beta_2^B)$
2. Por sustitución: $\quad \mathbb{1}_B = \mathbb{1} - \mathbb{1}_A$ en (32);

$$Y = \beta_1 \mathbb{1} + \beta_2^B X + \delta X \cdot \mathbb{1}_A + U, \qquad (33)$$

   donde $\delta \equiv \beta_2^A - \beta_2^B$,
   (29 y 33 idénticas bajo H_0).

   Ahora $H_0: \ \delta = 0$;
   (basta contraste de signif. individual; *uni o bilateral*).

8 Términos de interacción

Considere el modelo de consumo

$$C = \alpha \mathbb{1} + \beta Y + U$$

Considere la hipótesis de que la propensión marginal al consumo (β) depende de la posesión de activos A Entonces

$$C = \alpha \mathbb{1} + (\beta_1 + \beta_2 \mathbb{1}_A) Y + U,$$

o

$$C = \alpha \mathbb{1} + \beta_1 Y + \beta_2 (\mathbb{1}_A \cdot Y) + U.$$

El término $(\mathbb{1}_A \cdot Y)$ se llama término de interacción.

Más prácticas para la Lección 11

- A continuación tiene algunos ejercicios adicionales propuestos.

(Lección 11) Ejercicio en clase. N-3.

☞**Código:** `RamanathanPS7-6.inp` ....

Posible cambio estructural en la participación de las mujeres en el mercado laboral
Abra el conjunto de datos `data7-4.gdt`, del libro de Ramanathan, con datos de 50 estados de EEUU sobre la participación de las mujeres en el mercado laboral. Los 50 primeros son del año 1980 y los 50 últimos de 1990. La variable a explicar es *WLFP*, que es el porcentaje de participación de mujeres mayores de 16 años en el mercado laboral. *YF* es el salario mediano de las mujeres (en miles de dólares); *YM* es el salario mediano de los hombres (en miles de dólares); *EDUC* es el porcentaje, de entre las mujeres con 24 o más años, con el título de bachillerato; *UE* es la tasa de desempleo; *MR* es el porcentaje de mujeres mayores de 16 años que están casadas; *DR* es el porcentaje de mujeres divorciadas; *URB* es el porcentaje de población urbana; *WH* es e porcentaje de mujeres mayores de 16 años que son de raza blanca. Por último, la variable ficticia *D90* vale 1 si el dato corresponde al año 1990 y 0 en caso contrario.

(a) Estime un modelo para *WLFP* empleando todas las variables explicativas (excepto *D90*).
(b) Realice un contraste de cambio estructural (Contraste de Chow), para estudiar si ha habido un cambio en la disposición de las mujeres a entrar en el mercado laboral entre los años 1980 y 1990.
(c) Si rechaza H_0 de ausencia de cambio estructural, genere todas las variables de interacción necesarias para captar el cambio (genere todas las variables que han

sido empleadas en el test de cambio estructural). Re-estime el modelo con ellas.
(d) Este último modelo tiene muchos regresores. Si hay variables estadísticamente no significativas, reduzca el modelo como de costumbre.
(e) Interprete los resultados. En particular,
  - ¿Son distintos los efectos del porcentaje de mujeres casadas (*MF*)? ¿Cuales son sus efectos? ¿Es significativo el efecto en los años 90?
  - ¿Son distintos los efectos “desaliento” debidos a la tasa de paro (*UE*)? ¿Cuales son sus efectos? ¿Es significativo el efecto en los años 90?
  - ¿Son distintos los efectos debidos al salario mediano de las mujeres (*YF*)? ¿Cuales son sus efectos? ¿Es significativo el efecto en los años 90? Ramanathan hace notar que este efecto no está justificado y lo atribuye a una difícil identificación de los efectos de ésta variable. ¿Cuál puede ser el problema?

(Lección 11) Ejercicio en clase. N-4.

☞**Código:** `wage1dummiesB.inp` ....

Log-lin con variables ficticias. Estimaremos las diferencias salariales entre cuatro grupos: hombres casados (*marrmale*), mujeres casadas (*marrfem*), hombres solteros y mujeres solteras (*singfem*)

(a) Cargue los datos `wage1.gdt` del libro de texto de Wooldridge (2006, Ejemplo 7.6)
(b) Genere las variables ficticias necesarias para indicar los cuatro grupos.
(c) Estime por MCO el siguiente modelo

$$\log(wage) = \beta_1 + \beta_2 \cdot marrmale + \beta_3 marrfem + \beta_4 singfem$$
$$+\beta_5 educ + \beta_6 exper + \beta_7 exper^2 + \beta_8 tenure + \beta_9 tenure^2 + OtrosFactores$$

(d) ¿Quien es el grupo de referencia? Interprete los coeficientes correspondientes a las variables ficticias que ha generado; en particular, ¿en qué porcentaje varía el salario con cada una de estas variables ficticias? (recuerde que el cálculo es $100 * (\exp(\beta) - 1)$)
(e) ¿Qué pasaría si también incluimos en el modelo la variable ficticia correspondiente a los hombres solteros?
(f) ¿Es significativa la diferencia de salarios entre mujeres solteras y casadas al 5%? Calcule un intervalo de confianza para $\beta_4 - \beta_3$ al 95% para comprobarlo.

(g) A partir del modelo estimado no es fácil ver si esta última diferencia salarial es estadísticamente significativa. Hay una alternativa. Cambiar el grupo de referencia. Estime el siguiente modelo

$$\log(wage) = \beta_1 + \beta_2 \cdot marrmale + \beta_3 singmale + \beta_4 singfem$$
$$+\beta_5 educ + \beta_6 exper + \beta_7 exper^2 + \beta_8 tenure + \beta 9 tenure^2 + OtrosFactores$$

y verifique que la estimación e intervalo de confianza para β_4 (diferencia entre mujer soltera y el grupo de referencia, que ahora es mujer casada) coincide con lo calculado en el apartado anterior.

(h) Calcule la diferencia estimada en el salario (no en el logaritmo del salario) entre mujeres solteras y casadas. Calcule también el intervalo de confianza al 95%.

(Lección 11) Ejercicio en clase. N-5.

☞**Código:** `RamanathanEX7-2.inp` ....

Precio de viviendas unifamiliares Abra el conjunto de datos `data7-3.gdt` del libro de Ramanathan.

(a) Estime un modelo para el precio en función del tamaño.
(b) Estime un modelo para el precio en función de todas las variables explicativas disponibles.
(c) Elimine del último modelo aquellas variables no significativas.
(d) Compare los resultados e interprete los coeficientes de este último modelo.

(e) Repita los pasos anteriores pero usando el logaritmo del *sqft* en lugar de *sqft*
(f) Elimine del último modelo el regresor $\ln$ *sqft*. ¿Empeoran los resultados?

Enlace al documento con el código de las **prácticas de la Lección 11**

Lección 12

1 Dos tipos de multicolinealidad

Estricta o perfecta: $|\mathbf{X}^\intercal\mathbf{X}| = 0 \qquad \operatorname{rg}\underset{k\times k}{\left(\mathbf{X}^\intercal\mathbf{X}\right)} < k.$

- incumplimiento del Sup-IV de independencia lineal
- Infinitas soluciones (β no identificado).

No estricta o de grado: $|\mathbf{X}^\intercal\mathbf{X}| \simeq 0$

- Alta correlación entre regresores.
- A mayor correlación menor determinante $|\mathbf{X}^\intercal\mathbf{X}|$ y mayor gravedad del problema.

Multicolinealidad $\rightarrow$ incertidumbre sobre el valor los β_j.

2 Multicolinealidad No estricta: Causas

Variables explicativas con una fuerte correlación entre ellas

- Varios regresores son series temporales con tendencia
- Inclusión de varios retardos de variables explicativas.
- Regresores cuya información ya está en otras vbles. expl.
- Inclusión de muchos regresores aumenta las posibilidades de multicolinealidad.

Caso opuesto: variables explicativas ortogonales. . . . (Práctica Simulación 1 y 2).

3 Multicolinealidad No estricta: Efectos

Pese a que MCO es ELIO:

- Difícil interpretación coeficientes: varianza de $\widehat{\boldsymbol{\beta}}$ enorme:

$$\operatorname{Var}\left(\widehat{\boldsymbol{\beta}}\right) = \sigma^2 \operatorname{E}\left(\mathbf{X}^\intercal\mathbf{X}\right)^{-1} = \frac{\sigma^2}{\left|\operatorname{E}\left(\mathbf{X}^\intercal\mathbf{X}\right)\right|}\operatorname{Adj}(\operatorname{E}\left(\mathbf{X}^\intercal\mathbf{X}\right))$$

  Poca precisión (pequeñas variaciones muestra $\rightarrow$ grandes variaciones en estimación).
  Alta correlación entre estimadores.
- Propensión a *no rechazar (casi) cualquier hipótesis* $\boxed{\text{F108}}$
  ($H_0: \;\beta_i = a \;\rightarrow\; \mathcal{T} = \frac{\widehat{\beta_j} - a1}{\widehat{\operatorname{Dt}}\left(\widehat{\beta_j}\right)} \sim t_{N-k}\,.$)
- Pero no afecta al contraste de significación conjunta

$$\mathcal{F} = \frac{N-k}{k-1}\cdot\frac{\widehat{R^2}}{1 - \widehat{R^2}} \sim F_{k-1,\;N-k,}.$$

Ningún problema si sólo nos preocupa la predicción.

4 Multicolinealidad No estricta: Detección

- Elevados R^2 con parámetros no significativos individualmente.
- Examen de la correlación de las vbles. expl.
  - Cálculo de correlaciones simples (sólo entre pares)
  - Regresiones entre vbles. expl.
- Añadir o quitar regresores al modelo ocasiona grandes cambios en los coeficientes estimados.
- Tests de colinealidad y análisis del "tamaño" de $\mathbf{X}^\intercal\mathbf{X}$

5 Multicolinealidad No estricta: Soluciones

¡Difícil solución!
No es un problema de especificación de modelo; es un problema de datos (Columnas de **X**) (insuficiente información para estimar TODOS los parámetros con precisión)

- Ignorar el problema si sólo nos interesa la predicción.
- Obtener más datos.
- Reformular el modelo
  - Imponer restricciones (información extra-muestral)
  - Transformar variables (primeras diferencias, datos per-cápita, etc.)
- Suprimir variables (si NO hay razones teóricas para mantenerlas)
  - ¡Ojo! quitar variables puede suponer incumplimiento del Supuesto 2: $\mathbb{E}(\boldsymbol{U}\,|\,\mathbf{X}) = \mathbf{0}$
- Ortogonalizar regresores con regresiones auxiliares.

Enlace a algunas **prácticas de la Lección 12**

________________________________Fin de la lección

Lección 13

1 Errores de especificación: Omisión de variables

Sea el MLG

$$Y = \mathsf{X}\boldsymbol{\beta} + \mathsf{Z}\boldsymbol{\gamma} + U \qquad \text{con } \boldsymbol{\gamma} \neq \mathbf{0}$$

donde $\mathsf{X} = \left[X_1; \;\ldots X_k\right]$ y $\mathsf{Z} = \left[Z_1; \;\ldots Z_p\right]$

1. Correcto: $Y = \mathsf{X}\boldsymbol{\beta} + \underbrace{\mathsf{Z}\boldsymbol{\gamma} + U}$ ("Completo")
2. Incorrecto: $Y = \mathsf{X}\boldsymbol{\beta}^* + \quad U^*$ ("Incompleto")

2 Estimación MCO con omisión de variables

Sea $\boldsymbol{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\boldsymbol{\gamma} + \boldsymbol{U}$ un *m.a.s* que cumple los supuestos. Si omitimos $\mathbf{Z}$ el estimador de $\boldsymbol{\beta}$ del modelo incompleto es:

$$\begin{aligned}\widehat{\boldsymbol{\beta}^*} &= (\mathbf{X}^\intercal\mathbf{X})^{-1}\mathbf{X}^\intercal\boldsymbol{Y}\\ &= (\mathbf{X}^\intercal\mathbf{X})^{-1}\mathbf{X}^\intercal(\mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\boldsymbol{\gamma} + \boldsymbol{U})\\ &= \mathbf{I}\boldsymbol{\beta} + (\mathbf{X}^\intercal\mathbf{X})^{-1}\mathbf{X}^\intercal\mathbf{Z}\boldsymbol{\gamma} + (\mathbf{X}^\intercal\mathbf{X})^{-1}\mathbf{X}^\intercal\boldsymbol{U}.\end{aligned}$$

Entonces, $\mathbb{E}(\widehat{\boldsymbol{\beta}^*} \mid \mathbf{X}) = \mathbf{I}\boldsymbol{\beta} + (\mathbf{X}^\intercal\mathbf{X})^{-1}\mathbf{X}^\intercal\mathbb{E}(\mathbf{Z} \mid \mathbf{X})\boldsymbol{\gamma}.$

Es decir, $\widehat{\boldsymbol{\beta}^*}$ es insesgado cuando $\mathbf{Z} \perp \mathbf{X}$, pues en tal caso:

$$\mathbb{E}(\mathbf{Z} \mid \mathbf{X}) = \boldsymbol{0}.$$

3 Un ejemplo sencillo

Sea el MLG

$$Y = \beta_1 1 + \beta_2 X + \beta_3 Z + U \qquad \text{con } \beta_3 \neq 0.$$

Si omitimos Z y estimamos el modelo

$$Y = \beta_1^* 1 + \beta_2^* X + U^*$$

con $\mathbf{X} = [1;\ X]$ *m.a.s* de $\mathsf{X} = [1;\ X]$, entonces

$$\mathrm{E}\begin{pmatrix}\widehat{\beta_1^*}\\ \widehat{\beta_2^*}\end{pmatrix} = \mathrm{E}\left(\mathbb{E}\left(\begin{pmatrix}\widehat{\beta_1^*}\\ \widehat{\beta_2^*}\end{pmatrix}\middle|\,\mathbf{X}\right)\right) = \begin{pmatrix}\beta_1\\ \beta_2\end{pmatrix} + \begin{pmatrix}\mathrm{E}(Z)\,\beta_3\\ \frac{\mathrm{Cov}(X,Z)}{\mathrm{Var}(X)}\beta_3\end{pmatrix};$$

es decir

$$\mathrm{E}\left(\widehat{\beta_1^*}\right) = \beta_1 + \mathrm{E}(Z)\,\beta_3. \qquad \text{y} \qquad \mathrm{E}\left(\widehat{\beta_2^*}\right) = \beta_2 + \frac{\mathrm{Cov}(X,Z)}{\mathrm{Var}(X)}\beta_3,$$

Son insesgados solo si...

4 Varianza del estimador MCO con el modelo incompleto

$$\mathbb{V}ar(\widehat{\boldsymbol{\beta}^*} \mid \mathbf{X}) \quad \lesseqqgtr \quad \mathbb{V}ar\left(\widehat{\boldsymbol{\beta}}_{|(1:k)} \mid \mathbf{X};\mathbf{Z}\right)$$

¡Ojo! la mayoría de textos afirman: $\mathbb{V}ar(\widehat{\boldsymbol{\beta}^*} \mid \mathbf{X}) \leq \mathbb{V}ar(\widehat{\boldsymbol{\beta}} \mid \mathbf{X};\mathbf{Z})$.

5 Consecuencias de la omisión de variables del modelo

Si la variable omitida es ortogonal al resto de regresores

- No se introducen sesgos

☞**Código:** `OmisionRegresorOrtogonal.inp` ....

Si la variable omitida no es ortogonal al resto de regresores

- Sesgos: Si no es ortogonal, se incumple el Supuesto 2. Regresores no exógenos ⇒ MCO sesgado.
- Efecto indeterminado en la varianza... pero
- Disminución de la varianza si había elevada correlación con el resto de regresores (multicolinealidad).

☞**Código:** `OmisionRegresorNoOrtogonal.inp` ....

☞**Código:** `OmisionRegresorMulticol.inp` ....

6 Errores de especificación: Inclusión errónea variables

Sea el MLG
$$Y = \mathsf{X}\boldsymbol{\beta}^{*} + U^{*}$$
donde $\mathsf{X} = \left[X_1; \; \dots X_k\right]$ y sea $\mathsf{Z} = \left[Z_1; \; \dots Z_p\right]$

1. Correcto: $\quad Y = \mathsf{X}\boldsymbol{\beta}^{*} + U^{*}$
2. Incorrecto: $\quad Y = \mathsf{X}\boldsymbol{\beta} \; + \mathsf{Z}\boldsymbol{\gamma} + U$

Sea $\boldsymbol{Y} = \mathbf{X}\boldsymbol{\beta}^{*} + \boldsymbol{U}^{*}$ un *m.a.s* que cumple los supuestos

¿Qué cabe esperar respecto a la esperanza y varianza de los estimadores en el modelo incorrecto?

Si se cumplen los supuestos en el modelo incorrecto: $\widehat{\boldsymbol{\gamma}}$ insesgado.
¿Pero si $\mathbb{E}\big(\boldsymbol{U} \mid \mathbf{X};\mathbf{Z}\big) \neq \boldsymbol{0}$?

7 Un ejemplo sencillo de inclusión errónea de variables

1. Correcto: $\quad Y = \beta_1^{*}\mathit{1} + \beta_2^{*}X + U^{*}$
2. Incorrecto: $\quad Y = \beta_1\mathit{1} + \beta_2 X + \beta_3 Z + U$

¿Qué cabe esperar respecto a $\mathbb{E}\big(\widehat{\boldsymbol{\beta}} \mid \mathbf{X};\mathbf{Z}\big)$ y a $\mathbb{V}ar\big(\widehat{\boldsymbol{\beta}} \mid \mathbf{X};\mathbf{Z}\big)$ en el modelo incorrecto?

Si se cumplen los supuestos: $\quad \widehat{\boldsymbol{\beta}}$ insesgado
(en concreto ¿$\mathrm{E}\left(\widehat{\beta_3}\right)$?) $\quad$ ¿Pero si $\mathbb{E}\big(\boldsymbol{U} \mid \mathbf{Z}\big) \neq \boldsymbol{0}$?

8 Un ejemplo sencillo de inclusión errónea de variables

Del modelo lineal simple, sabemos que, si se cumplen los supuestos, para el modelo correcto:

$$\mathrm{Var}\left(\widehat{\beta_2^{*}}\right) = \frac{\sigma_*^2}{\mathrm{Var}\left(X\right)} \quad \text{donde } \sigma_*^2 = \mathbb{V}ar\big(U^{*} \mid X\big),$$

para el incorrecto (asumiendo $Z \perp U$) se puede demostrar que

$$\mathrm{Var}\left(\widehat{\beta_2}\right) = \frac{\sigma^2}{(1 - R^2_{XZ})\,\mathrm{Var}\left(X\right)} \quad \text{donde } \sigma^2 = \mathbb{V}ar\big(U \mid X; Z\big),$$

donde R^2_{XZ} es el R^2 de la regresión lineal simple de $\boldsymbol{X}$ sobre $\boldsymbol{Z}$ (el cuadrado de $\mathrm{Corr}\left(X, Z\right)$).

- ¿Qué pasa cuando $R^2_{XZ} = 0$?
- ¿Qué pasa cuando R^2_{XZ} es casi 1?
- ¿Qué varianza es mayor?

¿Y si no es cierto que $Z \perp U$? (y entonces $\sigma^2 \leq \sigma_*^2$)

9 Consecuencias por la inclusión de regresores ortogonales a $\mathbf{X}$

$Y = \mathsf{X}\boldsymbol{\beta} + U$ cumple supuestos y $\mathbb{V}ar\big(U \mid \mathsf{X}\big) = \sigma_*^2$
Pero estimamos $Y = \mathsf{X}\boldsymbol{\beta} + \mathsf{Z}\boldsymbol{\gamma} + U$ $\quad$ con $\boxed{\mathbf{Z} \perp \mathbf{X}}$

Si $\mathbb{E}\big(\boldsymbol{U} \mid \mathbf{X}, \mathbf{Z}\big) = \boldsymbol{0}$ (es decir $\boxed{\mathbf{Z} \perp \boldsymbol{U}}$):

- $\mathrm{E}\left(\widehat{\boldsymbol{\beta}}\right) = \boldsymbol{\beta} \quad$ y $\quad \mathrm{E}\left(\widehat{\boldsymbol{\gamma}}\right) = \mathbf{0}.$
- No hay efectos en la varianza

Si $\mathbb{E}\big(\boldsymbol{U} \mid \mathbf{Z}\big) \neq \boldsymbol{0}$ (es decir $\boxed{\mathbf{Z} \not\perp \boldsymbol{U}}$):

- $\mathrm{E}\left(\widehat{\boldsymbol{\beta}}\right) = \boldsymbol{\beta} \quad$ y $\quad \mathrm{E}\left(\widehat{\boldsymbol{\gamma}}\right) \neq \mathbf{0}.$
- Disminución de la varianza por $\mathbb{V}ar\big(U \mid \mathsf{X}, \mathsf{Z}\big) \;\leq\; \sigma_*^2$
(Test de hipótesis)

10 Consecuencias por la inclusión de regresores NO ortogonales a $\mathbf{X}$

$Y = \mathsf{X}\boldsymbol{\beta} + U$ cumple supuestos y $\mathbb{V}ar\big(U \mid \mathsf{X}\big) = \sigma^2_*$

Pero estimamos $Y = \mathsf{X}\boldsymbol{\beta} + \mathsf{Z}\boldsymbol{\gamma} + U$ con $\boxed{\mathbf{Z} \not\perp \mathbf{X}}$

Si $\mathbb{E}\big(\boldsymbol{U} \mid \mathbf{X}, \mathbf{Z}\big) = \boldsymbol{0}$ (es decir $\boxed{\mathbf{Z} \perp \boldsymbol{U}}$):

- $\operatorname{E}\big(\widehat{\boldsymbol{\beta}}\big) = \boldsymbol{\beta}$ y $\operatorname{E}\big(\widehat{\boldsymbol{\gamma}}\big) = \mathbf{0}$.
- Incremento de la varianza por $\mathbf{Z} \not\perp \mathbf{X}$.
- Este caso se denomina *inclusión de variables irrelevantes*

Si $\mathbb{E}\big(\boldsymbol{U} \mid \mathbf{Z}\big) \neq \boldsymbol{0}$ (es decir $\boxed{\mathbf{Z} \not\perp \boldsymbol{U}}$):

- $\operatorname{E}\big(\widehat{\boldsymbol{\beta}}\big) \neq \boldsymbol{\beta}$ y $\operatorname{E}\big(\widehat{\boldsymbol{\gamma}}\big) \neq \mathbf{0}$.
- Disminución de la varianza por $\mathbb{V}ar\big(U \mid \mathsf{X}, \mathsf{Z}\big) \leq \sigma^2_*$
  Incremento de la varianza por $\mathbf{Z} \not\perp \mathbf{X}$. ¿Efecto final?

☞**Código:** `ErrorDeEspecificacionPorInclusion.inp` ......

Enlace a algunas **prácticas de la Lección 13**

Ramanathan, R. (2002). *Introductory Econometrics with applications*. South-Western, Mason, Ohio, fifth ed. ISBN 0-03-034186-8.

Wooldridge, J. M. (2006). *Introducción a la econometría. Un enfoque moderno*. Thomson Learning, Inc., second ed.