Métodos Cuantitativos I - Máster Universitario de Economía

Marcos Bujosa
mbujosab@ucm.es
https://github.com/mbujosab

29 de mayo de 2023

Índice

1. Lección-00 Naturaleza y objetivos de la econometría

Sin duda ésta es una sección importante e interesante, pero pude prepararla sin ayuda del profesor, así que **léase alguno de los capítulos introductorios de las referencias de la bibliografía recomendada**, por ejemplo los de Wooldridge[4] o Gujarati[2]. .

Las ideas centrales son:

¿Por qué modelizar? slide

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

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

Ejemplos de objetivos por los que construir modelos

- 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).
- Previsión: 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.

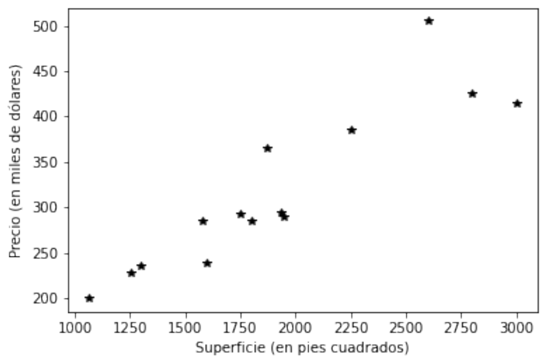
2. Lección-01 Geometría del ajuste MCO

Referencias: Wooldridge[4] (Capítulos 2 y 3 y Apéndice E1)

2.1. Introducción

En ocasiones se logra una mejor comprensión de una variable si se relaciona con otras.

¿Hay relación entre tamaño y precio de una vivienda? **slide**




```python
import numpy as np
import matplotlib.pyplot as plt
data = np.loadtxt('./datos/data3-1.csv', delimiter=',', skiprows=1)
S = data[:, 0]
P = data[:, 1]
plt.plot(P, S, 'k*')
plt.xlabel('Superficie (en pies cuadrados)')
plt.ylabel('Precio (en miles de dólares)')
plt.savefig('./img/scatter-data3-1.png')
```

```
open data3-1
scatters price; sqft --output=img/ScatterEjPvivienda.png
```

Con este ejemplo (sacado del libro de Ramanathan[3]) puede pensarse que los pares de datos (*precio*, *superficie*) están dispuestos de manera *"aproximada"* a lo largo de una linea recta

$$y = a + bx,$$

de manera que podríamos describir su relación aproximadamente como

$$Precio = a + b(Superficie) + OtrasCosas;$$

donde $OtrasCosas$ es lo que desplaza los puntos por encima y por debajo de esa hipotética recta.

La cuestión sobre la que se ocupa esta primera lección es: ¿podemos encontrar "la recta que mejor se ajusta" a esta nube de puntos? es decir ¿podemos *"aproximar"* los datos de *precios* con una recta de la forma

$$\widetilde{\boldsymbol{precio}} = \widetilde{a}\ \mathbf{1} + \widetilde{b}\ \boldsymbol{superficie}$$

donde $\boldsymbol{superficie}$ es el vector con los datos de superficies y $\mathbf{1}$ es un vector cuyas componentes son unos? ¿Cómo calculamos las magnitudes de los parámetros $\widetilde{a}$ y $\widetilde{b}$? es decir, ¿cuáles son los valores de la constante $\widetilde{a}$ y de la pendiente $\widetilde{b}$ de esta aproximación lineal? ¿Y cómo evaluamos la bondad de ajuste de esta aproximación? (bueno... esto ultimo lo abordaremos dentro de un par de lecciones).

Antes de continuar, recordemos el concepto de combinación lineal:

Definición 1 (Combinación lineal de vectores de $\mathbb{R}^n$)**.** *Sean los vectores* $\boldsymbol{b}_1, \ldots, \boldsymbol{b}_n$*. Llamamos combinación lineal a cualquier suma de múltiplos de dichos vectores:*

$$a_1\boldsymbol{b}_1 + a_2\boldsymbol{b}_2 + \cdots + a_n\boldsymbol{b}_n$$

donde los números "a_i" son los coeficientes de la combinación lineal.

Otro ejemplo **slide**

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

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

donde *otrascosas* son otros aspectos distintos de la renta (activos financieros, estado de ánimo, edad, lugar de residencia, etc.).

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

$$\widetilde{\boldsymbol{con}} \;=\; \widetilde{\beta_1}\mathbf{1} \;+\; \widetilde{\beta_2}\boldsymbol{rd} \;=\; \Big[\mathbf{1};\;\; \boldsymbol{rd};\Big]\begin{pmatrix}\widetilde{\beta_1}\\ \widetilde{\beta_2}\end{pmatrix}.$$

Nomenclatura y notación **slide**

- *regresando*: vector de datos de *consumo* ($\boldsymbol{con}$)
- *regresores*: vector de unos ($\mathbf{1}$) y de rentas disp. ($\boldsymbol{rd}$):

  $\mathbf{X} = \big[\mathbf{1};\; \boldsymbol{rd};\big]$, donde $\mathbf{X}_{|1} = \mathbf{1}$ y $\mathbf{X}_{|2} = \boldsymbol{rd}$.
- *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

$$\widetilde{\boldsymbol{salario}} = \widetilde{\beta_1}\mathbf{1} + \widetilde{\beta_2}\boldsymbol{educ} + \widetilde{\beta_3}\boldsymbol{exper} + \widetilde{\beta_4}\boldsymbol{iq}$$

donde *Educ* son los años de formación del trabajador, *Exper* son sus años de experiencia laboral y *IQ* su coeficiente intelectual (una medida de la habilidad del trabajador).

En este ejemplo, ¿quienes son el *regresando*, los *regresores* y el *vector de parámetros*?

Ajuste MCO como función lineal en los parámetros **slide**

- La aproximación o ajuste $\widetilde{\boldsymbol{y}}$ 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

$$\begin{aligned}\widetilde{\boldsymbol{y}} \;&=\; \widetilde{\beta_1}\mathbf{1} \;+\; \widetilde{\beta_2}\mathbf{X}_{|2} \;+\; \widetilde{\beta_3}\mathbf{X}_{|3} \;+\cdots+\; \widetilde{\beta_k}\mathbf{X}_{|k}\\ &=\; \Big[\mathbf{1};\; \mathbf{X}_{|2};\ldots\; \mathbf{X}_{|k};\Big]\begin{pmatrix}\widetilde{\beta_1}\\ \vdots\\ \widetilde{\beta_k}\end{pmatrix} \;=\; \mathbf{X}\widetilde{\boldsymbol{\beta}}.\end{aligned}$$

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

donde $\widehat{\boldsymbol{\beta}} = \begin{pmatrix}\widehat{\beta_1}\\ \vdots\\ \widehat{\beta_k}\end{pmatrix}$; $\qquad \mathbf{X} = \Big[\mathbf{1};\; \mathbf{X}_{|2};\ldots\; \mathbf{X}_{|k};\Big]$.

2.1.1. Una nota sobre la linealidad

La Econometría tiene una importante componente algebraica. Hay una enorme cantidad de excelentes referencias donde puede consultar sus dudas sobre Álgebra Lineal, pero aquí citaré de manera recurrente el *Curso de Álgebra Lineal con notación asociativa y un módulo para Python*[1], ya que usaré la misma notación que en el citado curso[1].

No obstante, en este texto sí introduciré algunas definiciones o resultados del Álgebra Lineal que son centrales para un curso de Econometría... Y dado que en la transparencia de más arriba acabo de aludir al concepto de *función lineal* (aunque aparentemente solo se destaca el hecho de que el ajuste es una combinación lineal de los regresores), vayamos con la definición de *función lineal*, y una nota sobre el producto "matriz por vector", que aclarará el uso del término *"función lineal"* en el título de la citada transparencia.

Definición 2 (Función lineal)**.** *Sean $\mathcal{D}$ y $\mathcal{V}$ dos espacios vectoriales. Decimos que la función $f\colon \mathcal{D} \to \mathcal{V}$ es* lineal *si satisface las siguientes propiedades:*

1. *Para todo $\vec{x}, \vec{y} \in \mathcal{D}$, $\quad f(\vec{x} + \vec{y}) = f(\vec{x}) + f(\vec{y})$.*
2. *Para todo $\vec{x} \in \mathcal{D}$ y para todo $\alpha \in \mathbb{R}$, $\; f(\alpha\vec{x}) = \alpha f(\vec{x})$.*

Nota 1 (sobre la expresión matriz por vector $\mathsf{A}\boldsymbol{x}$.)**.** *Por una parte, el producto de una matriz A de orden N por k por un vector $\boldsymbol{x}$ de k componentes es el vector de $\mathbb{R}^N$ que se obtiene tomando la combinación lineal de las k columnas de A cuyos parámetros son los elementos de $\boldsymbol{x}$. Es decir, si A tiene N filas y k columnas, entonces*

$$\mathsf{A}\boldsymbol{x} = x_1\big(\mathsf{A}_{|1}\big) + x_2\big(\mathsf{A}_{|2}\big) + \cdots + x_k\big(\mathsf{A}_{|k}\big) \in \mathbb{R}^N.$$

Pero por otra, como $\mathsf{A}\boldsymbol{x}$ es lineal por la derecha, es decir, como $\mathsf{A}(\boldsymbol{x}+\boldsymbol{y}) = \mathsf{A}\boldsymbol{x} + \mathsf{A}\boldsymbol{y}$, y como $\mathsf{A}(\alpha\boldsymbol{x}) = \alpha\mathsf{A}\boldsymbol{x}$, el producto $\mathsf{A}\boldsymbol{x}$ es una transformación lineal del vector $\boldsymbol{x} \in \mathbb{R}^k$ en el vector $(\mathsf{A}\boldsymbol{x}) \in \mathbb{R}^N$.

Así pues, decir que el ajuste $\widetilde{\boldsymbol{y}}$ *es una transformación lineal en los parámetros* $\widetilde{\boldsymbol{\beta}}$ es sencillamente decir que $\widetilde{\boldsymbol{y}}$ es de la forma: *"una matriz por"* el vector $\widetilde{\boldsymbol{\beta}}$; y decir que *es combinación lineal de los regresores* es decir que es de la forma: la matriz X *"por un vector"*. Consecuentemente, la expresión $\widetilde{\boldsymbol{y}} = \mathsf{X}\widetilde{\boldsymbol{\beta}}$ significa ambas cosas (ahora puede volver a releer la última transparencia y entenderá su título).

Nótese que la expresión $\mathsf{A}\boldsymbol{x}$ también es lineal por la izquierda, es decir, $(\mathsf{A}+\mathsf{B})\boldsymbol{x} = \mathsf{A}\boldsymbol{x} + \mathsf{B}\boldsymbol{x}$ y $(\alpha\mathsf{A})\boldsymbol{x} = \alpha\mathsf{A}\boldsymbol{x}$. Esto quiere decir que $\mathsf{X}\widetilde{\boldsymbol{\beta}}$ también es una función *lineal en los regresores*, pero lo más relevante en el ajuste MCO es que es lineal en los parámetros $\widetilde{\boldsymbol{\beta}}$.[2]

2.2. Ajuste mínimo cuadrático

Considere los datos correspondientes al ejemplo del precio de las viviendas (sacado del manual de Ramanathan[3]).

Datos del ejemplo del precio de las viviendas **slide**

Precio viviendas (miles de $) y superficie útil (pies al cuadrado) 14 casas unifamiliares en *University City*. San Diego, California. Año 1990.

	price ($\boldsymbol{y}$)	sqft ($\boldsymbol{x}$)	$\widetilde{\text{price}}$ ($\widetilde{\boldsymbol{y}}$)
0	199.9	1065	
1	228	1254	
2	235	1300	
3	285	1577	
4	239	1600	
5	293	1750	
6	285	1800	
7	365	1870	
8	295	1935	
9	290	1948	
10	385	2254	
11	505	2600	
12	425	2800	
13	415	3000	

[1]que además está libremente disponible en la web.
[2]la linealidad en los regresores cobrará interés más adelante, al interpretar los resultados de los modelos estimados por MCO.

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, y en general, dada una muestra $\mathbf{X}$, el ajuste para la observación n-ésima usando el vector $\widetilde{\boldsymbol{\beta}}$ es: $\widetilde{y_n} = {}_{n|}\widetilde{\boldsymbol{y}} = {}_{n|}\mathbf{X}\widetilde{\boldsymbol{\beta}}$, donde ${}_{n|}\mathbf{X} = \begin{pmatrix}1, & x_{n2}, & \cdots, & x_{nk},\end{pmatrix}$ es la fila n-ésima de la matriz $\mathbf{X}$ que contiene los datos del n-ésimo elemento de la muestra. Es decir

$$\widetilde{y_n} \;=\; {}_{n|}\widetilde{\boldsymbol{y}} \;=\; {}_{n|}\mathbf{X}\widetilde{\boldsymbol{\beta}} \;=\; \begin{pmatrix}1, & x_{n2}, & \cdots, & x_{nk},\end{pmatrix}\begin{pmatrix}\widetilde{\beta_1}\\ \widetilde{\beta_2}\\ \vdots\\ \widetilde{\beta_k}\end{pmatrix} = 1\widetilde{\beta_1} + (x_{n2})\widetilde{\beta_2} + \cdots + (x_{nk})\widetilde{\beta_k}.$$

Así tenemos,

Los precios ajustados como combinación lineal de los regresores **slide**

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

El precio ajustado para el séptimo piso de la muestra será

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

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

Para establecer el criterio primero necesitamos definir el:

Término de error de ajuste **slide**

Dados $\mathbf{X}$ e $\boldsymbol{y}$, el *"error de ajuste"* empleando $\widetilde{\boldsymbol{\beta}}$ 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: $\boldsymbol{y} = \widetilde{\boldsymbol{y}} + \widetilde{\boldsymbol{e}}$.

Llamamos "Suma de los Residuos al Cuadrado" (SRC) 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 $\widetilde{\boldsymbol{e}} = (\boldsymbol{y} - \widetilde{\boldsymbol{y}})$.

Así, el error cometido por el ajuste para la observación n-enésima es

$$_{n|}\widetilde{\boldsymbol{e}} \;=\; _{n|}\Big(\boldsymbol{y} - \mathbf{X}\widetilde{\boldsymbol{\beta}}\Big) \;=\; _{n|}\boldsymbol{y} - {}_{n|}\widetilde{\boldsymbol{y}} \;=\; y_n - \widetilde{y_n} \;=\; \widetilde{e_n}.$$

Antes de establecer el criterio de selección de los parámetros de ajuste, vemos un:

2.2.1. Recordatorio sobre la longitud y la perpendicularidad de vectores de $\mathbb{R}^n$ y el Teorema de Pitágoras

Primero recordemos la definición de *producto punto*:[3]

Definición 3. *El producto punto de dos vectores* $\boldsymbol{a}$ *y* $\boldsymbol{b}$ *de* $\mathbb{R}^n$ *es* [4]

$$\boldsymbol{a}\cdot\boldsymbol{b} = a_1b_1 + a_2b_2 + a_3b_3 + \cdots + a_nb_n \;=\; \sum_{i=1}^{n} a_ib_i.$$

Este producto nos permite definir la ortogonalidad o perpendicularidad entre vectores de $\mathbb{R}^n$.

Definición 4 (Vectores perpendiculares u ortogonales)**.** *Decimos que* $\boldsymbol{a}$ *y* $\boldsymbol{b}$ *son ortogonales* o perpendiculares ($\boldsymbol{a}\perp\boldsymbol{b}$) *cuando* $\boldsymbol{a}\cdot\boldsymbol{b} = 0$.

Consecuentemente, $\mathbf{0}$ es ortogonal a todos los vectores (incluido él mismo).

Además, el producto punto nos dota de una métrica particular que define la longitud o norma de un vector:[5]

Definición 5. *La longitud* (o norma) *de un vector* $\boldsymbol{a}$ *es la raíz cuadrada de* $\boldsymbol{a}\cdot\boldsymbol{a}$*:*

$$\textit{longitud de } \boldsymbol{a} \;=\; \|\boldsymbol{a}\| \;=\; \sqrt{\boldsymbol{a}\cdot\boldsymbol{a}}.$$

Por ejemplo, la longitud de $\boldsymbol{x} = \begin{pmatrix}6\\0\\-2\\3\end{pmatrix}$ es $\sqrt{\boldsymbol{x}\cdot\boldsymbol{x}} = \sqrt{\begin{pmatrix}6\\0\\-2\\3\end{pmatrix}\cdot\begin{pmatrix}6\\0\\-2\\3\end{pmatrix}} = \sqrt{6^2+0^2+(-2)^2+3^2} = \sqrt{49} = 7.$

Para finalizar este recordatorio, veamos el

Teorema 2.1 (Teorema de Pitágoras en $\mathbb{R}^n$)**.** *Sean* $\boldsymbol{x}$ *e* $\boldsymbol{y}$ *dos vectores de* $\mathbb{R}^n$*; entonces* $\boldsymbol{x}\cdot\boldsymbol{y} = 0$ (son perpendiculares) *si y solo si*

$$\|\boldsymbol{x}+\boldsymbol{y}\|^2 = \|\boldsymbol{x}\|^2 + \|\boldsymbol{y}\|^2.$$

Demostración. Si $\boldsymbol{x}$ e $\boldsymbol{y}$ son *perpendiculares*, entonces $\boldsymbol{x}\cdot\boldsymbol{y} = 0$ y

$$\begin{aligned}
\|\boldsymbol{x}+\boldsymbol{y}\|^2 = (\boldsymbol{x}+\boldsymbol{y})\cdot(\boldsymbol{x}+\boldsymbol{y}) =& \boldsymbol{x}\cdot(\boldsymbol{x}+\boldsymbol{y}) + \boldsymbol{y}\cdot(\boldsymbol{x}+\boldsymbol{y})\\
=&(\boldsymbol{x}+\boldsymbol{y})\cdot\boldsymbol{x} + (\boldsymbol{x}+\boldsymbol{y})\cdot\boldsymbol{y}\\
=&\boldsymbol{x}\cdot\boldsymbol{x} + \boldsymbol{y}\cdot\boldsymbol{x} + \boldsymbol{x}\cdot\boldsymbol{y} + \boldsymbol{y}\cdot\boldsymbol{y}\\
=&\boldsymbol{x}\cdot\boldsymbol{x} + \quad 2(\boldsymbol{x}\cdot\boldsymbol{y}) \quad + \boldsymbol{y}\cdot\boldsymbol{y}\\
=&\boldsymbol{x}\cdot\boldsymbol{x} + \quad\;\; 0 \quad\;\; + \boldsymbol{y}\cdot\boldsymbol{y} \qquad\qquad = \|\boldsymbol{x}\|^2 + \|\boldsymbol{y}\|^2.
\end{aligned}$$

□

Fíjese que en el anterior teorema, $\boldsymbol{x}$ e $\boldsymbol{y}$ corresponden a los catetos y el vector suma $\boldsymbol{x}+\boldsymbol{y}$ a la hipotenusa.

Ahora ya estamos en condiciones de establecer el criterio de ajuste, pues consiste en minimizar una distancia.

[3] que es el producto escalar usual en $\mathbb{R}^N$ (aunque ya veremos maś adelante que no es el producto escalar usado en estadística).
[4] En la mayoría de los libros lo denotan con $\boldsymbol{a}^\intercal\boldsymbol{b}$, pero aquí seguiré la notación del curso de álgebra citado más arriba.
[5] Más adelante veremos que esta no es la forma de medir en estadística.

Criterio del ajuste MCO

- **Objetivo**. Encontrar una combinación lineal de los regresores

Sean $\boldsymbol{y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{pmatrix}$ y $\mathbf{X} = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_N \end{bmatrix}$. Buscamos una $\mathbf{X}\widetilde{\boldsymbol{\beta}}$.

- **Criterio** de búsqueda (ajuste MCO)

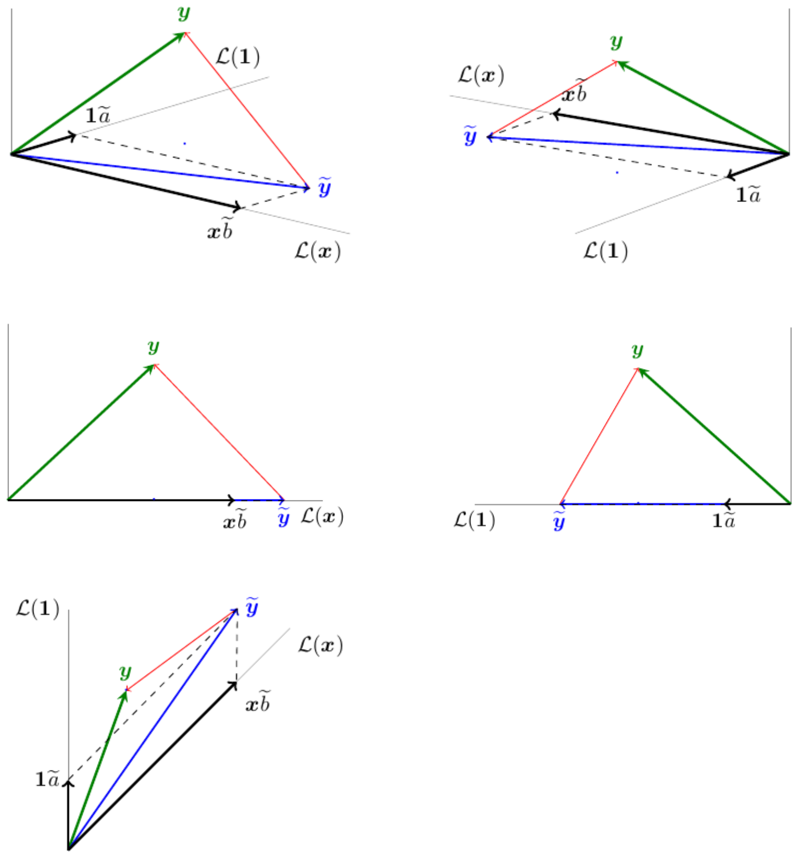
Elegir el vector $\widetilde{\boldsymbol{\beta}}$ tal que $\mathbf{X}\widetilde{\boldsymbol{\beta}}$ esté *lo más cerca* de $\boldsymbol{y}$; es decir, tal que

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

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

Geometría de un mal ajuste lineal

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



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

Las cinco figuras muestran una representación del mismo ejemplo visto desde distintos ángulos (las dos primeras desde cierta altura, las dos siguientes desde el suelo y la última sería la “vista de pájaro” desde la vertical). En cada figura, el *“suelo”* es el plano (el subespacio) generado por los regresores $\mathbf{1}$ y $\boldsymbol{x}$; es decir, el conjunto de todas las combinaciones lineales de los regresores, que en los libros de álgebra también se conoce como espacio columna de la matriz $\mathbf{X}$, y que con frecuencia se denota con $\mathcal{C}\left(\mathbf{X}\right)$; es decir:

$$\mathcal{C}\left(\mathbf{X}\right) \text{ es el conjunto de vectores de la forma } \mathbf{X}\boldsymbol{v}$$

Buscar *la combinación de regresores más próxima a* $\boldsymbol{y}$ es buscar el punto del *"suelo"* más próximo a $\boldsymbol{y}$ (más cercano al extremo superior de la flecha verde). Dicho de otro modo, buscar el vector de $\mathcal{C}\left(\mathbf{X}\right)$ más próximo a $\boldsymbol{y}$ es buscar la combinación lineal $\mathbf{X}\boldsymbol{v}$ que más se acerca a $\boldsymbol{y}$.

Si agudiza la vista verá que dicho punto aparece pintado en las figuras. A partir de ahora a dicho punto lo denotaremos con $\widehat{\boldsymbol{y}}$ y lo denominaremos el ajuste MCO de $\boldsymbol{y}$ (dada una matriz de regresores $\mathbf{X}$).

Se aprecia claramente que la elección de los parámetros $\widetilde{a}$ y $\widetilde{b}$ de las figuras de más arriba es mejorable, pues usar $\widetilde{a}$ veces $\mathbf{1}$ y $\widetilde{b}$ veces $\boldsymbol{x}$ nos conduce a un punto que no es el más próximo a $\boldsymbol{y}$. En esas figuras

$$\widehat{\boldsymbol{y}} \neq [\mathbf{1};\boldsymbol{x};] \begin{pmatrix}\widetilde{a}\\ \widetilde{b}\end{pmatrix}.$$

Para seguir con la exposición, demos el nombre técnico del punto $\widehat{\boldsymbol{y}}$ que estamos buscando y demostremos que efectivamente es el punto del *"suelo"* más próximo a $\boldsymbol{y}$:

Definición 6. *Llamamos* proyección ortogonal de un vector $\vec{y}$ sobre un subespacio $\mathcal{V}$ *a la función lineal* $f(\vec{y})$ *tal que la diferencia* $\vec{y} - f(\vec{y})$ *es ortogonal* $\mathcal{V}$.

En el caso particular que nos ocupa, para cada vector de datos $\boldsymbol{y}$ de $\mathbb{R}^N$, el vector $\widehat{\boldsymbol{y}}$ del subespacio $\mathcal{C}\left(\mathbf{X}\right)$ es la *proyección ortogonal de* $\boldsymbol{y}$ *sobre el subespacio* $\mathcal{C}\left(\mathbf{X}\right) \subset \mathbb{R}^m$ engendrado por las columnas de $\mathbf{X}$ (i.e., por los regresores). Dicho vector existe y es único (véase la Lección 11 de Bujosa[1]). La diferencia $\boldsymbol{y} - \widehat{\boldsymbol{y}}$ es el vector de errores de ajuste $\widehat{\boldsymbol{e}}$.

Lo más interesante de $\widehat{\boldsymbol{y}}$ se afirma en la siguiente

Proposición 2.2. *La proyección ortogonal de* $\boldsymbol{y} \in \mathbb{R}^m$ *sobre* $\mathcal{C}\left(\mathbf{X}\right) \subset \mathbb{R}^m$ *es el vector de* $\mathcal{C}\left(\mathbf{X}\right)$ *más próximo a* $\boldsymbol{y}$.

Demostración. Sea $\widehat{\boldsymbol{y}}$ la proyección de $\boldsymbol{y}$ sobre $\mathcal{C}\left(\mathbf{X}\right)$ y tomemos un vector $\boldsymbol{z}$ cualquiera de $\mathcal{C}\left(\mathbf{X}\right)$. Veamos que $\boldsymbol{y}$ está más lejos de $\boldsymbol{z}$ que de su proyección ortogonal $\widehat{\boldsymbol{y}}$.

Como $\boldsymbol{z}$ e $\widehat{\boldsymbol{y}}$ están en $\mathcal{C}\left(\mathbf{X}\right)$, su diferencia $(\widehat{\boldsymbol{y}} - \boldsymbol{z})$ está en $\mathcal{C}\left(\mathbf{X}\right)$ (pues $\mathcal{C}\left(\mathbf{X}\right)$ es subespacio); consecuentemente $(\widehat{\boldsymbol{y}} - \boldsymbol{z})$ es ortogonal a $(\boldsymbol{y} - \widehat{\boldsymbol{y}}) = \widehat{\boldsymbol{e}}$ (por ser $\widehat{\boldsymbol{y}}$ la proyección ortogonal sobre $\mathcal{C}\left(\mathbf{X}\right)$). Y como la suma de ambos vectores perpendiculares es $(\boldsymbol{y} - \widehat{\boldsymbol{y}}) + (\widehat{\boldsymbol{y}} - \boldsymbol{z}) = (\boldsymbol{y} - \boldsymbol{z})$, por el Tma. de Pitágoras concluímos que

$$\|\boldsymbol{y} - \boldsymbol{z}\|^2 = \|\widehat{\boldsymbol{y}} - \boldsymbol{z}\|^2 + \|\boldsymbol{y} - \widehat{\boldsymbol{y}}\|^2 \geq \|\boldsymbol{y} - \widehat{\boldsymbol{y}}\|^2,$$

donde $(\widehat{\boldsymbol{y}} - \boldsymbol{z})$ e $(\boldsymbol{y} - \widehat{\boldsymbol{y}})$ son los catetos y la hipotenusa es $(\boldsymbol{y} - \boldsymbol{z})$. Por tanto, $\boxed{\|\boldsymbol{y} - \boldsymbol{z}\| \geq \|\boldsymbol{y} - \widehat{\boldsymbol{y}}\| = \|\widehat{\boldsymbol{e}}\|}$ para todo $\boldsymbol{z} \in \mathcal{C}\left(\mathbf{X}\right)$. □

2.2.2. Las ecuaciones normales

El modo de encontrar los coeficientes $\widehat{\boldsymbol{\beta}}$ de la combinación de regresores que resulta ser la proyección ortogonal $\widehat{\boldsymbol{y}} = \mathbf{X}\widehat{\boldsymbol{\beta}}$ consiste en resolver un sistema de ecuaciones que se denomina *sistema de ecuaciones normales*.

Para verlo recordemos que dos vectores son perpendiculares si su producto punto es cero

$$\boldsymbol{a} \perp \boldsymbol{b} \iff \boldsymbol{a} \cdot \boldsymbol{b} = \sum_{i=1}^{n} a_i b_i = 0.$$

Consecuentemente, las filas de una matriz $\mathbf{A}$ son perpendiculares a un vector $\boldsymbol{b}$ si $\mathbf{A}\boldsymbol{b} = \mathbf{0}$ y, por tanto, las columnas de una matriz $\mathbf{C}$ son perpendiculares a $\boldsymbol{b}$ si $\mathbf{C}^\intercal\boldsymbol{b} = \mathbf{0}$.

$$\boldsymbol{b} \perp \mathbf{C}_{|j} \iff \mathbf{C}^\intercal\boldsymbol{b} = \mathbf{0}$$

Ecuaciones normales **slide**

Como el vector $\widehat{e} = (\boldsymbol{y} - \widehat{\boldsymbol{y}})$ es mínimo si es *perpendicular* a cada regresor.
Es decir, si:

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

Tenemos que,

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

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

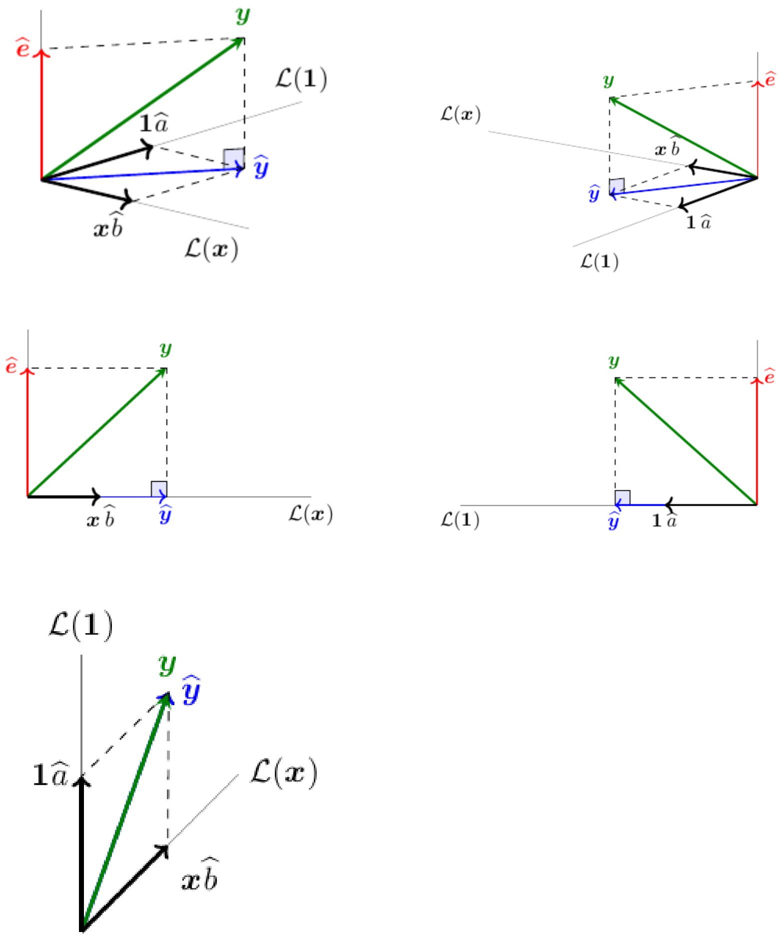
Las soluciones $\widehat{\beta}$ son los parámetros del ajuste MCO: $\widehat{\boldsymbol{y}} = \mathbf{X}\widehat{\beta}$

El sistema de ecuaciones $\big(\mathbf{X}^\intercal\mathbf{X}\big)\widehat{\beta} = \mathbf{X}^\intercal \boldsymbol{y}$ se denomina *sistema de ecuaciones normales*;[6] y para obtener la proyección $\widehat{\boldsymbol{y}}$ de $\boldsymbol{y}$ sobre $\mathcal{C}\left(\mathbf{X}\right)$ basta multiplicar $\mathbf{X}$ por cualquier vector que sea solución de dicho sistema.[7]

Lo llamativo es que para resolver $\mathbf{A}\widehat{\beta} = \widehat{\boldsymbol{y}}$ (para encontrar la combinación lineal de las columnas de $\mathbf{X}$ que buscamos) resolvemos un sistema diferente (el *sistema de ecuaciones normales*) donde no aparece $\widehat{\boldsymbol{y}}$. Este procedimiento indirecto funciona porque el *sistema de ecuaciones normales* y el sistema $\mathbf{A}\widehat{\beta} = \widehat{\boldsymbol{y}}$ tienen el mismo conjunto de soluciones (fíjese en las implicaciones "si y solo si" de (1)).

Ajuste MCO (Geometría de la proyección ortogonal) — slide

$$\widehat{e} \perp \mathbf{X} \quad \Longleftrightarrow \quad \widehat{\beta} \ \text{ es tal que } \quad (\mathbf{X}^\intercal\mathbf{X})\widehat{\beta} = \mathbf{X}^\intercal \boldsymbol{y}$$



[6]en este contexto geométrico, normalidad significa perpendicularidad, de ahí el nombre

[7]Fíjese que aunque el punto $\widehat{\boldsymbol{y}}$ es único, el sistema de ecuaciones normales puede tener infinitas soluciones, es decir, en principio puede haber infinitas combinaciones de los regresores que sean iguales a $\widehat{\boldsymbol{y}}$. En la siguiente siguiente sección veremos la condición sobre $\mathbf{X}$ para que la solución sea única.

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

2.2.3. Ausencia de multicolinealidad exacta (regresores linealmente independientes)

Para que la solución al sistema de ecuaciones normales (1) sea única (para que $\widehat{\boldsymbol{\beta}}$ sea único) es necesario que se verifique una condición sobre la matriz de regresores $\mathbf{X}$:

Condición para que las ecuaciones normales tengan solución única Puesto que

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

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 $(\mathbf{X}^\intercal\mathbf{X})^{-1}$ tenemos que

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

es *la única solución*. El modelo está sacado del manual de Wooldridge [4]

La condición de independencia lineal de las columnas de $\mathbf{X}$ implica que $\mathbf{X}^\intercal\mathbf{X}$ es invertible[8] y viceversa. Dicha condición garantiza la unicidad de las soluciones. Es decir, si hay k regresores, entonces:

$$\widehat{\boldsymbol{\beta}} \text{ es único} \quad \Leftrightarrow \quad \operatorname{rg}(\mathbf{X}) = k \quad \Leftrightarrow \quad (\mathbf{X}^\intercal\mathbf{X})^{-1} \text{ es invertible.}$$

Cuando no se cumple la condición de rango hay infinitos valores para los parámetros que permiten alcanzar $\widehat{\boldsymbol{y}}$. En tal caso se dice que hay *multicolinealidad perfecta* (i.e., *los regresores son linealmente dependientes*). Veamos un ejemplo.

Modelo para los salarios **slide**

Ecuación de salarios: Supongamos el siguiente modelo

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

Tomando logaritmos tenemos una nueva variable $\ln(Salar_n)$ que podremos ajustar con un modelo lineal en los parámetros pues,

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

Pero ¿qué pasará si jamás ningún trabajador cambió de empresa?

La matriz $\mathbf{X}^\intercal\mathbf{X}$ será singular. De hecho, como *experiencia* y *antigüedad* coinciden, como mucho sólo podemos calcular su efecto conjunto:

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

donde $Salar_n$ es el salario del individuo n-ésimo, $educ_n$ son sus años de educación, $antig_n$ sus años de antigüedad en la empresa, y $exper_n$ sus años de experiencia en el sector y $otros\ factores_n$ son otros factores distintos de los anteriores. (el modelo está sacado del manual de Wooldridge [4]).

Fíjese que si una de las columnas de $\mathbf{X}$ es un múltiplo de otra; es decir, si la correlación entre dos regresores es *1 en valor absoluto*[9] (por ejemplo, si la tercera columna es a veces la segunda) habrá multicolinealidad exacta. Pero no es necesario que haya correlación *uno en valor absoluto* entre algunos regresores para que haya multicolinealidad exacta; por ejemplo, si la tercera columna es a veces la primera más b veces la segunda, evidentemente las columnas de $\mathbf{X}$ serán linealmente dependientes (aunque no haya correlación uno en valor absoluto entre dichas columnas).

[8] lo que requiere que para cada regresor hay al menos tantas observaciones como regresores tiene el modelo ($N \geq k$).

[9] Véase la correlación entre vectores alineados en la Página~**??**.

2.2.4. ¡Ojo! El lenguaje habitual en econometría se toma ciertas licencias

El modelo inicial para los salarios no es lineal en los parámetros, y una vez se ha transformado logarítmicamente dicho modelo… *tampoco* pues lo podemos escribir como

$$\ln(Salar_n) = \left[1;\ educ;\ antig;\ exper;\right]\begin{pmatrix}\beta_1\\ \beta_2\\ \beta_3\\ \beta_4\end{pmatrix} + otrascosas$$

o de manera más compacta

$$\boldsymbol{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{otrascosas}.$$

Dicho modelo no cumple las condiciones de la Definición 2. No obstante, los libros de econometría se refieren a un modelo así como lineal en los parámetros porque se puede aproximar mediante el modelo lineal $\widehat{\boldsymbol{y}} = \mathbf{X}\widehat{\boldsymbol{\beta}}$. Por ello se dice que tomando logaritmos "se *linealiza* el modelo" inicial, pero la expresión no es estrictamente correcta. Como dicha expresión está completamente extendida en econometría, aquí también diremos que un modelo de la forma $\boldsymbol{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{u}$ es lineal en los parámetros, aunque realmente solo sería lineal si $\boldsymbol{u}$ también estuviera multiplicado por un parámetro. Así pues, en econometría, al decir que el modelo $\boldsymbol{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{u}$ es lineal en los parámetros lo que se quiere decir es que se puede aproximar por MCO mediante el modelo lineal $\widehat{\boldsymbol{y}} = \mathbf{X}\widehat{\boldsymbol{\beta}}$.

2.3. Prácticas

A continuación aparecen algunas prácticas con Gretl. Pero también las puede reproducir con Python usando los Notebooks accesibles desde aquí

2.3.1. Visualización de los datos del ejemplo del *Precio de casas unifamiliares*

Veamos como mostrar los datos, generar diagramas de dispersión.

Mostrar datos.

1. Leamos el fichero de datos data3-1 del manual de Ramanathan

```
open data3-1
scatters price; sqft --output=img/ScatterEjPvivienda.png
```

Generar gráficos.

2.3.2. Gráficos

- Descargue los datos simulados `ventas.txt` que puede encontrar en el subdirectorio `datos` del material de clase (puede acceder a él desde el campus virtual o mi página web. Guarde el fichero en el disco de su ordenador).

  Puede cargar los datos seleccionando el menú `Archivo ->Abrir archivo de datos ->Archivo de usuario` y buscando el archivo `*.txt` en el directorio donde guardó los datos.

  Gretl preguntará si deseamos interpretar los datos como series temporales o como datos de panel; contestaremos "No".

  - También podemos cargar los datos tecleando en la consola de comando (donde `./datos/` es la ruta hasta el directorio donde está el fichero)

    ```
    open ./datos/ventas.txt
    ```

    (o escribiéndolo en un guión de instrucciones con extensión `*.inp` y ejecutándolo después).

3. Lección 19

Bibliografía

[1] Bujosa, Marcos. 2022. *Un Curso de Álgebra Lineal con notación asociativa y un módulo para Python.* https://github.com/mbujosab (self-published).

[2] Gujarati, Damodar N. 2003. *Basic Econometrics*. Fourth edn. International edition.

[3] Ramanathan, Ramu. 2002. *Introductory Econometrics with applications*. Fifth edn. Mason, Ohio: South-Western.

[4] Wooldridge, Jeffrey M. 2006. *Introducción a la econometría. Un enfoque moderno*. Thomson Learning, Inc.