Lección 2

Marcos Bujosa

29 de septiembre de 2023

Índice

1.	Precio de casas unifamiliares	1
	1.1. Actividad 1 - mostrar datos	2
	1.2. Actividad 2 - diagrama de dispersión	2
	1.3. Actividad 3 - Ajuste por MCO	2
	1.4. Actividad 4 - Recuperar los valores ajustados	3
	1.5. Actividad 5 - Otras formas de recuperar el ajuste	
2.	Plano de regresión	4
3.	Simulación del ejemplo del precio de las viviendas con dos regresores.	5
	3.1. Omisión de regresores	6
	3.2. Proyección omitiendo las constantes	8
	3.3. "Ortogonalizando" los regresores no constantes respecto de las constantes	8
4.	Repetición del ejemplo pero introduciendo más correlación entre regresores	9
5 .	Más sobre la ortogonalización de regresores	11
	5.1. Empleando regresores no constantes que son perpendiculares entre si	11
	5.2. Ajustes aún más curiosos	12
	5.2.1. Una primera proyección auxiliar	12
	5.2.2. Primer ajuste curioso	12
	5.2.3. Segundo ajuste curioso	

1. Precio de casas unifamiliares

Guión: scripts/EjPvivienda.inp

En esta primera práctica con Gretl reproduciremos el ejemplo visto repetidamente en clase.

Los datos corresponden a los precios de venta y superficie útil de 14 casas unifamiliares en *University City*. San Diego, California. Año 1990. (Ramu Ramanathan, 2002)

Veremos como mostrar los datos, generar diagramas de dispersión, realizar un ajuste por MCO, y operar con series de datos y con parámetros estimados. Al final de la práctica aparece el guión completo con el código que evita trabajar con los menús en modo gráfico (que es el peor modo de trabajar con Gretl).

1. Objetivo

- a) Reproducir el primer ejemplo de regresión visto en clase.
- b) Mostrar datos.
- c) Generar gráficos.
- d) Guardar un modelo para poder consultarlo más tarde.
- e) Recuperar valores ajustados, errores estimados, etc.

2. Carga de datos Archivo -->Abrir datos -->Archivo de muestra y en la pestaña Ramanathan seleccione data3-1.

```
o bien teclee en linea de comandos:

open data3-1  # LEEMOS los datos
```

1.1. Actividad 1 - mostrar datos

- 1. Visualice los datos de precios y tamaños de las casas
 - En la ventana principal de Gretl, marque con el ratón ambas variables: price, sqrt.
 - "Pinche" sobre ellas con el botón derecho del ratón.
 - Seleccione mostrar valores del menú desplegable que se ha abierto al pinchar.

```
o bien teclee en linea de comandos:
print -o price sqft  # MOSTRAMOS los datos
```

- 2. Ayuda Para consultar la documentación sobre cualquier comando, puede emplear el menú desplegable *Ayuda* que aparece arriba, a la derecha de la ventana principal de Gretl.
 - Ayuda ->Guía de Instrucciones y "pinche" sobre print

o bien teclee en linea de comandos: help print

1.2. Actividad 2 - diagrama de dispersión

- 1. Scatter plot
 - Marque price y sqrt (pulsando ctrl y pinchando con el botón derecho del ratón sobre ellas). Elija Gráfico de dos variables XY
 - \blacksquare Seleccione ${\tt sqft}$ como variable del eje X

```
o bien teclee en linea de comandos: gnuplot price sqft
```

- 2. Guardar gráfico como icono para editarlo más tarde
 - "Pinche" con el botón derecho sobre la ventana del gráfico.
 - Seleccione Guardar a sesión como icono

```
o bien teclee en linea de comandos:

NombreDelGrafico <- gnuplot price sqft

(NombreDelGrafico es el nombre con el que se guardará el icono)
```

En la zona inferior izquierda de la ventana principal puede ver una serie de iconos. Uno de ellos es la vista de iconos de sesión.

1.3. Actividad 3 - Ajuste por MCO

- 1. Ajuste por MCO el modelo de regresión visto en clase
 - Estime el modelo mediante los menús desplegables: Modelo ->Mínimos Cuadrados Ordinarios; indique a Gretl el regresando y los regresores y pulse Aceptar.

```
o bien teclee en linea de comandos:
ols price 0 sqft
(el cero 0 indica el término constante: const)
```

• "Pinche" Archivo en la ventana del modelo estimado y seleccione guardar como un icono y cerrar o bien teclee en linea de comandos:

```
Regresion <- ols price 0 sqft
```

• Recupere el modelo "pinchando" sobre su icono

```
o teclee en linea de comandos el nombre que ha dado al icono seguido de .show, es decir:  \text{Regresion.show}
```

1.4. Actividad 4 - Recuperar los valores ajustados

- 1. Recuperemos los valores ajustados
 - Desde la ventana del modelo ajustado (recupérese con su icono), "pinche" en guardar ->valores estimados. Elija como nombre phat (puede añadir una descripción de la variable). Pulse en Aceptar
 - Repita para guardar los residuos con el nombre ehat

```
o bien teclee en linea de comandos:
series phat = $yhat
series ehat = $uhat
```

- 2. Mostremos las variables price, sqft, phat y ehat
 - Marque las 4 variables (ctrl y "pinchar" con el botón derecho) y elija mostrar valores
 o bien teclee en linea de comandos:
 print -o price sqft phat ehat

1.5. Actividad 5 - Otras formas de recuperar el ajuste

phat2: restar a los precios los errores

Desde la ventana del modelo: Guardar ->Definir una nueva variable y teclee: phat2 = price - ehat o bien teclee en linea de comandos:

```
series phat2 = price - ehat
```

phat2: Cálculo "chapucero": 52.351 + 0.139 sqft Guardar ->Definir una nueva variable y teclee:

```
series phat3 = 52.351 + 0.139*sqft
```

■ phat2: Cálculo correcto: $\widehat{\beta}_1 + \widehat{\beta}_2 \ sqft$ Guardar ->Definir una nueva variable y teclee:

```
series phat4 = $coeff[1] + $coeff[2]*sqft

o bien

series phat5 = $coeff(const) + $coeff(sqft)*sqft
```

Visualice los valores ¿Hay diferencias?

print -o price phat phat2 phat3 phat4

2. Plano de regresión

Guión: scripts/PlanoRegresion.inp

■ Carga de datos. Abra los menús desplegables: Archivo -->Abrir datos -->Archivo de muestra y en la pestaña POE 4th ed. seleccione andy.

```
o bien teclee en linea de comandos:

open andy.gdt
```

Ajuste por MCO las ventas a los precios y gastos en publicidad: Modelo ->Mínimos Cuadrados Ordinarios;
 indique a Gretl el regresando y los regresores y pulse Aceptar.

```
o bien teclee en linea de comandos:

ols sales O price advert

(el cero O indica el término constante: const)
```

• "Pinche" Archivo en la ventana del modelo estimado y seleccione guardar como un icono y cerrar o bien teclee en linea de comandos:

```
Regresion <- ols sales 0 price advert
```

• Observe el plano de regresión: abra la ventana del modelo ajustado y pinche en Gráficos ->Gráfico de variable estimada y observada ->Contra price y advert

Las prácticas siguientes deberían ir en otra lección (no en esta).

3. Simulación del ejemplo del precio de las viviendas con dos regresores.

Guión: scripts/SimuladorEjPvivienda.inp

En este ejercicio Gretl generaremos unos datos simulados para realizar regresiones MCO.

- 1. Simulamos series de datos de 500 observaciones
 - Archivo ->Nuevo conjunto de datos, e indicamos el número de

observaciones: 1500. Marcamos de sección cruzada y continuamos adelante. Dejamos sin marcar empezar a introducir los valores de los datos y pulsamos Aceptar.

• o bien:

```
nulldata 1500
```

- 2. Generamos tres variables: S con distribución uniforme (35, 120); D con distribución Chi cuadrado con 5 grados de libertad que vamos a multiplicar por 3 y U con distribución Normal de media 0 y desviación típica 40
 - Añadir -> Variable aleatoria y se elige para cada variable el tipo de distribución, los valores de los parámetros y el nombre de la variable
 - o bien

```
series S = randgen(U, 35, 120)
series D = randgen(X, 5) * 3
series U = randgen(N, 0, 40)
```

- 3. Simulamos los precios según el modelo p = 3001 + 5s 2d + u
 - Añadir ->Definir nueva variable y tecleamos: P = 300 + 5*S 2*D + U
 - O bien

```
series P = 300 + 5*S - 2*D + U
```

- 4. Observe los estadísticos de las variables de modelo. En particular, ¿son ortogonales S y D respecto al regresor constante 1?
 - Ver ->Estadísticos principales con el ratón marcamos D, S y P.
 - O bien

```
summary P S D
```

- 5. El valor absoluto de la correlación entre D y S ¿es grande o pequeño?
 - Ver ->Matriz de correlación y selecciones D y S.
 - O bien

```
corr S D
```

- 6. Observe los diagramas de dispersión de P con S, el de P con D, y el de los regresores S y D:
 - Ver ->Gráficos ->Gráfico X-Y (Scatter) y marcamos S como variable del eje x, y P como variable del eje y.

- Ver ->Gráficos ->Gráfico X-Y (Scatter) y marcamos D como variable del eje x, y P como variable del eje y.
- Ver -> Gráficos -> Gráficos X-Y (Scatter) y marcamos S como variable del eje x, y D como variable del eje y.
- O bien

```
scaterPS <- gnuplot P S
scaterPD <- gnuplot P D
scaterDS <- gnuplot D S</pre>
```

Nótese como dichos diagramas reflejan las correlaciones entre las distintas variables del modelo.

- 7. Ajuste por MCO P empleando S y D:
 - pinche en Modelo ->Mínimos Cuadrados Ordinarios y selecciones D y S.
 - Elija P como regresando o "variable dependiente" (marque la opción Selección por defecto).
 - Elija S y D como regresores o "Variables independientes". Pinche en Aceptar.
 - En la ventana del modelo pinche en Archivo ->Guardar como icono y cerrar.
 - Con el botón derecho del ratón, pinche sobre el icono del modelo estimado y seleccione Añadir a la tabla de modelos.
 - O bien

```
ModeloCompleto <- ols P O S D modeltab add
```

Observe cómo el ajuste recupera de manera muy aproximada los verdaderos valores de los parámetros del modelo simulado.

Observe el plano de regresión: abra la ventana del modelo ajustado y pinche en $Gráficos \rightarrow Gráfico de variable estimada y observada \rightarrow Contra <math>S$ y D

■ Tarea adicional Observe que el papel del parámetro β_1 que acompaña al término constante es equilibrar los valores medios a ambos lados de la ecuación; es decir, asegurar que

$$\mu_{\boldsymbol{p}} = \widehat{\beta}_1 + \widehat{\beta}_2 \mu_{\boldsymbol{s}} + \widehat{\beta}_3 \mu_{\boldsymbol{d}}$$

Verifique que efectivamente

$$\widehat{\beta}_1 = \mu_{\mathbf{p}} - \widehat{\beta}_2 \mu_{\mathbf{s}} - \widehat{\beta}_3 \mu_{\mathbf{d}}$$

- Hágalo empleando una calculadora
- O bien

Otra manera de aludir a los betas estimados es generar una matriz columna con los betas: beta = \$coeff, de manera que beta[1,1], beta[2,1] y beta[3,1] son $\widehat{\beta}_1$, $\widehat{\beta}_2$ y $\widehat{\beta}_3$ respectivamente.

3.1. Omisión de regresores

- 1. Si se omite el regresor D del ajuste; qué esperaría que ocurra con el parámetro asociado a la cte?
 - pinche en Modelo ->Mínimos Cuadrados Ordinarios

- En la lista de variables dependientes, pinche sobre D y pulse la flecha roja para eliminar dicha variable del modelo. Pinche en Aceptar.
- la ventana del modelo pinche en Archivo ->Guardar como icono y cerrar.
- el botón derecho del ratón, pinche sobre el icono del modelo estimado y seleccione Añadir a la tabla de modelos.
- O bien

```
ModeloSinD <- ols P 0 S modeltab add
```

¿Confirman los resultados lo esperado respecto a los parámetros estimados?

■ Tarea adicional Recuerde que el papel del parámetro β_1 que acompaña al término constante es equilibrar los valores medios a ambos lados de la ecuación; en este caso,

$$\mu_{\mathbf{p}} = \widehat{\beta}_1 + \widehat{\beta}_2 \mu_{\mathbf{s}}$$

Verifique que efectivamente

$$\widehat{\beta}_1 = \mu_{\boldsymbol{p}} - \widehat{\beta}_2 \mu_{\boldsymbol{s}}$$

- Hágalo empleando una calculadora
- O bien

- 2. Repita el experimento, pero esta vez incluyendo D y excluyendo S
 - pinche en Modelo ->Mínimos Cuadrados Ordinarios
 - Elimine de la lista de regresores la variable S, e incluya D como regresor. Pinche en Aceptar.
 - la ventana del modelo pinche en Archivo ->Guardar como icono y cerrar.
 - el botón derecho del ratón, pinche sobre el icono del modelo estimado y seleccione Añadir a la tabla de modelos.
 - O bien

ModeloSinS <- ols P O D modeltab add

donde modeltab show muestra la tabla de modelos ajustados.

¿Confirman los resultados lo esperado respecto a los parámetros estimados?

■ Tarea adicional Aquí β_1 asegura que

$$\mu_{\mathbf{p}} = \widehat{\beta}_1 + \widehat{\beta}_2 \mu_{\mathbf{d}}$$

Verifique que efectivamente

$$\widehat{\beta}_1 = \mu_{\mathbf{p}} - \widehat{\beta}_2 \mu_{\mathbf{d}}$$

- Hágalo empleando una calculadora
- O bien

3.2. Proyección omitiendo las constantes

- ¿Y si calculamos la proyección ortogonal de los precios sobre S y D omitiendo la constante? ¿Qué espera que ocurra con los valores medios de P y del ajuste? Verifique su respuesta con el ordenador.
 - pinche en Modelo ->Mínimos Cuadrados Ordinarios
 - Elija S y D como regresores, pero elimine la constante de la lista (¡que algo que jamás debe de hacer es sus estimaciones!). Pinche en Aceptar.
 - o la ventana del modelo pinche en Archivo ->Guardar como icono y cerrar.
 - o el botón derecho del ratón, pinche sobre el icono del modelo estimado y seleccione Añadir a la tabla de modelos.
 - O bien

```
ModeloSin1 <- ols P S D modeltab add
```

¿Confirman los resultados lo esperado respecto a los parámetros estimados?

• Tarea adicional Aquí no hay un término constante, así que

$$\mu_{\mathbf{p}} \neq \widehat{\beta}_1 \mu_{\mathbf{s}} + \widehat{\beta}_2 \mu_{\mathbf{d}}$$

El equilibro se da añadiendo la media de los errores (que por no existir término constante es distinta de cero). Verifique que efectivamente

$$\mu_{\mathbf{p}} = \widehat{\beta}_1 \mu_{\mathbf{s}} + \widehat{\beta}_2 \mu_{\mathbf{d}} + \mu_{\widehat{\mathbf{e}}}$$

- o Hágalo empleando una calculadora
- o O bien

donde \$uhat son los residuos de la última regresión realizada.

Este ajuste sin constante (y que por tanto yo no denominaría regresión MCO) no logra equiparar correctamente los valores medios de ambos lados de la ecuación. Los únicos elementos disponibles para intentar dicho ajuste han sido los vectores de datos S y D, que tienen medias distintas de cero. Pero simultáneamente dichos vectores son necesarios para lograr el ajuste de las pendientes. Por ello los resultados procediendo así pueden ser desastrosos.

3.3. "Ortogonalizando" los regresores no constantes respecto de las constantes

- 1. ¿Hay alguna forma de "ortogonalizar" S y D respecto de la constante?... ¡Si!, basta restar las medias (generar nuevos regresores en desviaciones respecto a su media):
 - Añadir -> Definir nueva variable y escribimos series SO = S mean(S)
 - Añadir ->Definir nueva variable y escribimos series DO = D mean(D)
 - O bien

```
series S0 = S - mean(S)
series D0 = D - mean(D)
```

2. Verifique que la media de S0 y D0 es cero. Ajuste por MCO los siguientes modelos:

$$P = \beta_1 + \beta_2 S0 + \beta_3 D0$$

$$P = \beta_2 S0 + \beta_3 D0$$

Compare los resultados entre ambos y con el modelo completo original.

```
ModeloCompEnDesviaciones <- ols P 0 S0 D0 modeltab add
ModeloEnDesviaciones <- ols P S0 D0 modeltab add
modeltab show
```

Compare la estimación del parámetro β_1 que acompaña al término constante en el primer caso con la media de P.

- Compare también la estimación del parámetro β_1 que acompaña al término constante en el primer caso con la media de P.
- ¿Se le ocurre alguna explicación para este resultado?
- Observe en la tabla de modelos todas las analogías en los resultados de estimación del primer modelo en desviaciones con las obtenidas con el modelo completo original... ¡Todo es idéntico salvo la estimación de β_1 !
- Observe que aunque las pendientes estimadas toman los mismos valores en ambos modelos en desviaciones, la incertidumbre asociada en el caso del modelo en desviaciones sin contante es mucho mayor

4. Repetición del ejemplo pero introduciendo más correlación entre regresores

Guión: scripts/SimuladorEjPvivienda2.inp

En la simulación de la práctica anterior no había una correlación muy fuerte entre S y D. En ejemplos reales podemos encontrar una mayor riqueza de interrelaciones entre regresores. Para experimentar las consecuencias, vamos a repetir todo el ejercicio pero introduciendo un factor común entre S y D que genere correlación entre ambos regresores.

Los datos ■ Genere un factor común C con distribución normal de esperanza 40 y desviación típica 20: Añadir
->Definir nueva variable y escribimos series C = randgen(N, 40, 20)

- Genere un vector de datos de superficies con distribución uniforme entre 20 y 70 y súmele C: Añadir
 ->Definir nueva variable y escribimos series S = randgen(U, 20, 70) + C
- Genere un vector de tiempos de viaje con distribución Chi cuadrado con 8 grados de libertad, y súmele
 100 y réstele C: Añadir ->Definir nueva variable y escribimos series D = randgen(X, 8) + 100
 C
- Genere un vector de perturbaciones con distribución Normal de media 0 y desviación típica 40. Añadir
 ->Definir nueva variable y escribimos series U = randgen(N, 0, 40)
- O bien

```
nulldata 1500
series C = randgen(N, 40, 20)
series S = randgen(U, 20, 70) + C
series D = randgen(X, 8) + 100 - C
series U = randgen(N, 0, 40)
```

El modelo Simulamos los precios según el modelo p = 3001 + 5s - 2d + u

■ Añadir ->Definir nueva variable y tecleamos: P = 300 + 5*S - 2*D + U

O bien

```
series P = 300 + 5*S - 2*D + U
```

Actividades Repita todas las actividades de la práctica anterior pero con estos nuevos datos simulados. Preste especial atención a cuáles son las diferencias entre esta simulación y la anterior sin correlación entre S y D.

```
summary P S D
corr S D
scaterPS <- gnuplot P S
scaterPD <- gnuplot P D
scaterDS <- gnuplot D S
ModeloCompleto <- ols P 0 S D
modeltab add
## TAREA ADICIONAL ####
AjusteMediasC = mean(P) - $coeff(S)*mean(S) - $coeff(D)*mean(D)
beta1HatC
            = $coeff(const)
########################
ModeloSinD <- ols P 0 S
modeltab add
## TAREA ADICIONAL ####
AjusteMediasNoD = mean(P) - $coeff(S)*mean(S)
beta1HatNoD = $coeff(const)
#######################
ModeloSinS <- ols P 0 D
modeltab add
## TAREA ADICIONAL ####
AjusteMediasNoS = mean(P) - $coeff(D)*mean(D)
beta1HatNoD = $coeff(const)
#######################
ModeloSin1 <- ols P S D
modeltab add
## TAREA ADICIONAL ####
MediaLadoIzdo = mean(P)
MediaLadoDcho = $coeff(S)*mean(S) + $coeff(D)*mean(D) + mean($uhat)
########################
series SO = S - mean(S)
series DO = D - mean(D)
ModeloCompEnDesviaciones <- ols P 0 S0 D0
modeltab add
ModeloEnDesviaciones <- ols P S0 D0
modeltab add
```

5. Más sobre la ortogonalización de regresores

Guión: scripts/SimuladorEjPvivienda3.inp

En el ejemplo anterior hemos visto que al generar nuevos regresores S0 y P0 que son las componentes de S y P perpendiculares a las constantes, y usarlos en sustitución de los regresores originales, hemos logrado nuevos ajustes en los que incluir u omitir el regresor constante no cambia el valor de las pendientes estimadas.

5.1. Empleando regresores no constantes que son perpendiculares entre si

Veamos si es posible generalizar este resultado, es decir, si al emplear como regresor la componente de S perpendicular a P y la constante, podemos ajustar P con tres regresores perpendiculares entre sí, de tal manera que omitir cualesquiera de los regresores no afecta a los parámetros estimados correspondientes a los regresores que sí se mantienen en el ajuste.

```
nulldata 1500
series C = randgen(N, 40, 20)
series S = randgen(U, 20, 70) + C
series D = randgen(X, 8) + 100 - C
series U = randgen(N, 0, 40)
series P = 300 + 5*S - 2*D + U
ModeloCompleto <- ols P 0 S D
modeltab add
ModeloSinD <- ols P 0 S
modeltab add
ModeloSinS <- ols P 0 D
modeltab add
ModeloSin1 <- ols P S D
modeltab add
# cálculo del coeficiente de determinación
R2Sin1 = 1 - \frac{(T-1) * var(P)}{}
R2AjustadoSin1 = 1 - ($T-1)/($T-2)*(1 - R2Sin1)
# ortogonalizamos D respecto de las constantes
series DO = D - mean(D)
# o de manera equivalente
ols D 0
series D0 = $uhat
# ortogonalizamos S respecto de las constantes y D
ols S 0 D
series SpD = $uhat
```

```
ModeloCompleto2 <- ols P 0 SpD D0
modeltab add

ModeloSolo1 <- ols P 0
modeltab add

ModeloSoloS <- ols P SpD
modeltab add

ModeloSoloD <- ols P D0
modeltab add
```

5.2. Ajustes aún más curiosos

Dos ajustes curiosos

En los siguientes ajustes lo que se mantiene es el valor del parámetro correspondiente la constante del modelo original (aunque no aparezca explícitamente ningún regresor constante en el ajuste... ahí radica la curiosidad).

5.2.1. Una primera proyección auxiliar

Lo primero es obtener la componente del vector 1 que es perpendicular a S y a D mediante una primera proyección auxiliar (no la llamaré regresión, pues el espacio sobre el que proyectamos no contiene los vectores constantes). Llamaré a dicha componente cte (pero obviamente no es un vector constante):

5.2.2. Primer ajuste curioso

En una nueva proyección auxiliar obtenemos la componente de S que es perpendicular a cte y a D, que llamaré SpD (componente de S perpendicular a D y a cte).

Y ahora realizamos la *regresión* de P sobre el mismo subespacio que en modelo original, pero empleando tres regresores que son ortogonales entre si, y tales que el parámetro de cte es igual al correspondiente a la constante del modelo original y el de SpD al del parámetro de S en la *proyección* de P sobre S y D (omitiendo la constante).

```
\# regresión MCO que mantiene la pendiente de S del ajuste sin cte ModeloCompletoAlternativo1 <- ols P cte SpD D modeltab add
```

5.2.3. Segundo ajuste curioso

Alternativamente podemos dar los pasos análogo, pero usando la proyección auxiliar de D sobre "ctez S.

modeltab show

Compare los resultados en la tabla de modelos.	