

Práctica de Aula PA-H.1:

Se dispone de una base de datos sobre el precio de venta y distintas características de 224 viviendas pertenecientes a dos áreas residenciales del condado de *Orange* en California (USA), *Dove Canyon* y *Coto de Caza*¹. Dove Canyon es una zona de viviendas relativamente pequeñas construidas alrededor de un campo de golf. Coto de Caza es un área de mayor nivel de vida, aunque más rural con viviendas más grandes. Las variables que se consideran son:

salepric : precio de venta de la vivienda en miles de dólares
 sqft : tamaño de la vivienda en pies cuadrados
 age : edad de la vivienda en años
 city : 1 si está en Coto de Caza, 0 si está en Dove Canyon

A continuación se muestran los resultados de la estimación por Mínimos Cuadrados Ordinarios (MCO) de un modelo para el precio de venta de la vivienda utilizando esa base de datos:

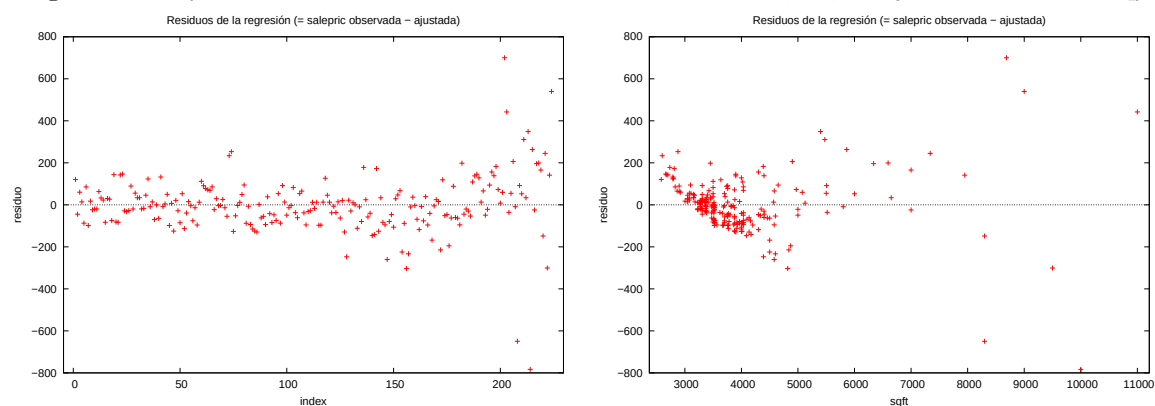
RESULTADOS A: Estimaciones MCO utilizando las 224 observaciones 1–224 Variable dependiente: salepric

	Coefficiente	Desv. típica	estadístico t	valor p
const	−440,31	35,3203	−12,4663	0,0000
sqft	0,2520	0,0081	30,9047	0,0000
age	3,6980	3,0241	1,2228	0,2227
city	91,8038	21,7494	4,2210	0,0000
Media de la var. dependiente			642,929	
D.T. de la variable dependiente			371,376	
Suma de cuadrados de los residuos			4,27804e+06	
Desviación típica de la regresión ($\hat{\sigma}$)			139,4480	
R^2			0,8609	
$\bar{R}^2$ corregido			0,8590	
$F(3, 220)$			453,8840	

1. Escribe el modelo teórico que se ha estimado y comenta los resultados obtenidos en términos de bondad de ajuste, significatividad y signos de los coeficientes estimados.
2. Analiza de forma razonada la información que te proporcionan los gráficos recogidos en la Figura 1 y la regresión auxiliar. Si realizas algún contraste, indica todos los elementos del mismo. ¿Cuál de los gráficos es más informativo y por qué?

¹Fichero data7-24.gdt, disponible en gretl pestaña Ramanathan. Recogida en Ramanathan, Ramu (2002), *Introductory econometrics with applications*, 5th. edn., South-Western.

Figura 1: Gráfico de residuos MCO sobre las observaciones $i = 1, \dots, 224$ y sobre la variable $sqft$



$$\frac{\widehat{u_i^2}}{SCR_A/224} = -5,94184 + 0,00172457 \text{ sqft}_i$$

$$N = 224 \quad R^2 = 0,421826 \quad SCR = 1478,52$$

A continuación se muestran los resultados de la estimación por MCO utilizando un estimador de la matriz de varianzas y covarianzas de los coeficientes consistente, aunque exista heterocedasticidad.

RESULTADOS B: Estimaciones MCO utilizando las 224 observaciones 1–224

Variable dependiente: salepric

Desviaciones típicas robustas ante heterocedasticidad, variante HC3

	Coefficiente	Desv. típica	estadístico t	valor p
const	−440,31	110,8800	−3,9711	0,0001
sqft	0,2520	0,0279	9,0120	0,0000
age	3,6980	5,1672	0,7157	0,4750
city	91,8038	26,3997	3,4774	0,0006

Media de la var. dependiente	642,9290
D.T. de la variable dependiente	371,3760
Suma de cuadrados de los residuos	4,27804e+06
Desviación típica de la regresión ($\hat{\sigma}$)	139,4480
R^2	0,8609
$\bar{R}^2$ corregido	0,8590
$F(3, 220)$	161,8190

1. ¿En qué varían los resultados mostrados ahora (RESULTADOS B) con los primeros (RESULTADOS A)? ¿Por qué? ¿Cuáles son fiables y para qué? Explica razonadamente.

Por último, se muestran los resultados de la estimación por Mínimos Cuadrados Generalizados o Ponderados utilizando como variable de ponderación el inverso del cuadrado del tamaño de la vivienda esto es, $\frac{1}{sqft^2}$.

RESULTADOS C: Estimaciones MC.Ponderados utilizando las 224 observaciones 1–224

Variable dependiente: salepric

Variable utilizada como ponderación: $\frac{1}{sqft^2}$

	Coefficiente	Desv. típica	estadístico t	valor p
const	−285,20	37,2121	−7,6643	0,0000
sqft	0,2155	0,0095	22,4752	0,0000
age	−0,5492	2,2800	−0,2409	0,8098
city	110,7800	15,6896	7,0607	0,0000

Estadísticos basados en los datos ponderados:

Suma de cuadrados de los residuos	0,15074
Desviación típica de la regresión ($\hat{\sigma}$)	0,02617
R^2	0,79881
$\bar{R}^2$ corregido	0,79607
$F(3, 220)$	291,177

Estadísticos basados en los datos originales:

Media de la var. dependiente	642,929
D.T. de la variable dependiente	371,376
Suma de cuadrados de los residuos	4,73514e+06
Desviación típica de la regresión ($\hat{\sigma}$)	146,708

1. ¿Qué se quiere decir con datos ponderados y datos originales? ¿Por qué se utiliza como variable de ponderación el inverso de $sqft^2$? Explica razonadamente.
2. ¿Qué resultados de los tres A,B, ó C te parecen mejores? ¿Por qué?