

Ejercitación 4

El siguiente ejercicio consiste en la valuación de una casa según sus atributos en función a la base de datos de Gary Koop. La misma tiene información sobre 546 casas vendidas en Windsor, Canadá. La variable dependiente (Y) es el precio de venta en dólares canadienses. Las variables explicativas (las x's) son:

- LOTE: el tamaño del lote, medido en pies al cuadrado (*square feet*)
- CUARTOS: número de dormitorios
- BANOS: número de baños
- PISOS: número de pisos, excluido el sótano
- ENTRADA: Dummy = 1 si la casa tiene una entrada para el auto (*driveway*) (= 0 si no tiene)
- REC: Dummy = 1 si la casa tiene un cuarto de recreación (*rec room*) (= 0 si no tiene)
- SOTANO: Dummy = 1 si la casa tiene un sótano (*basement*) (= 0 si no tiene)
- CALEF: Dummy = 1 si la casa tiene calefacción central (gas) (= 0 si no tiene)
- AIRE: Dummy = 1 si la casa tiene aire acondicionado (*air cond*) (= 0 si no tiene)
- GARAGE: mide la cantidad de autos que entran en el *garage* (= 0 si no tiene)
- NBHD: Dummy = 1 si la casa está en un barrio agradable (= 0 si no lo está)
- DB2: Dummy = 1 si la casa tiene exactamente 2 baños (= 0 sino)
- DB3: Dummy = 1 si la casa tiene exactamente 3 baños (= 0 sino)

En este ejercicio las casas tienen 1, 2 ó 3 baños. En las regresiones que se presentan a continuación, se omitió una observación de la base de datos original que satisfacía que tenía 4 baños (la única), de modo que las observaciones usadas en las regresiones a continuación se basan en las restantes 545 observaciones de la base de datos con la que se trabajó en clase.

El output de la regresión es el siguiente:

REGRESION 1

Dependent Variable: PRECIO

Method: Least Squares

Sample: 1 545

Included observations: 545

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-3509.739	3419.390	-1.026423	0.3052
LOTE	3.534066	0.349830	10.10225	0.0000
CUARTOS	1909.324	1046.422	1.824622	0.0686
BANOS	13829.49	1519.347	9.102255	0.0000
PISOS	6492.688	924.6698	7.021628	0.0000
ENTRADA	6722.788	2042.146	3.292022	0.0011
REC	4623.243	1898.206	2.435586	0.0152
SOTANO	5506.627	1585.876	3.472294	0.0006
CALEF	12995.33	3214.103	4.043221	0.0001
AIRE	12622.70	1552.591	8.130086	0.0000
GARAGE	4139.817	841.6709	4.918570	0.0000
NBHD	9469.961	1667.598	5.678803	0.0000
R-squared	0.664891	Mean dependent var		67925.49
Adjusted R-squared	0.657975	S.D. dependent var		26330.72
S.E. of regression	15398.96	Akaike info criterion		22.14376
Sum squared resid	1.26389E+11	Schwarz criterion		22.23846
Log likelihood	-6022.175	Hannan-Quinn criter.		22.18078
F-statistic	96.13890	Durbin-Watson stat		1.697694
Prob(F-statistic)	0.000000			

Genere esta misma regresión usando Excel, Stata y Statsmodels.

- Interprete el coeficiente 4139.8 de la variable GARAGE.
- ¿Para qué sirve el estadístico F (=96.14), cómo está calculado y a qué conclusión puede llegar con el mismo? Indique claramente de qué hipótesis nula y alternativa es este estadístico. Es decir, explique cuál sería el modelo bajo las hipótesis nula y alternativa.
- ¿Puede haber evidencia de multicolinealidad en la Regresión 1? Justifique claramente.
- ¿Cuál es el valor esperado de una casa en un lote de 5100 pies cuadrados, con 3 cuartos, 2 baños, construida en dos pisos, con entrada para auto, cuarto de recreación, calefacción central en un vecindario agradable (NBHD=1), sin garage, ni sótano, ni aire acondicionado central? Muestre cómo obtiene ese valor.

REGRESION 2

A continuación, se presenta otra salida. Esta segunda regresión utiliza las mismas observaciones que la regresión 1. En esta regresión se elimina la variable BANOS y se incorporan las variables DB2 y DB3.

Dependent Variable: PRECIO
Method: Least Squares

Sample: 1 545
Included observations: 545

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	10308.95	3464.573	2.975534	0.0031
LOTE	3.526753	0.350519	10.06152	0.0000
CUARTOS	1902.874	1047.336	1.816871	0.0698
PISOS	6527.316	928.9431	7.026604	0.0000
ENTRADA	6747.292	2044.525	3.300177	0.0010
REC	4623.607	1899.665	2.433906	0.0153
SOTANO	5521.098	1587.458	3.477948	0.0005
CALEF	13020.30	3217.108	4.047207	0.0001
AIRE	12696.44	1563.394	8.121072	0.0000
GARAGE	4141.753	842.3301	4.917019	0.0000
NBHD	9496.899	1670.077	5.686503	0.0000
DB2	13481.18	1726.361	7.809018	0.0000
DB3	29392.54	5079.824	5.786133	0.0000
R-squared	0.665005	Mean dependent var	67925.49	
Adjusted R-squared	0.657449	S.D. dependent var	26330.72	
S.E. of regression	15410.80	Akaike info criterion	22.14709	
Sum squared resid	1.26346E+11	Schwarz criterion	22.24968	
Log likelihood	-6022.082	Hannan-Quinn criter.	22.18719	
F-statistic	88.00712	Durbin-Watson stat	1.700277	
Prob(F-statistic)	0.000000			

- e) ¿Cómo se interpreta el coeficiente 29392.54 correspondiente a la variable DB3? Indique claramente.
- f) Obtenga un intervalo de confianza del 95% para el beta que multiplica a la variable lote a partir de la información de la segunda regresión (REGRESION 2). Puede usar la distribución normal como aproximación. También puede responder usando algún software, en cuyo caso deberá indicar claramente qué hizo.
- g) Suponga que le interesa testear con una probabilidad de error tipo I de 5% si el beta que multiplica a DB3 es el doble al beta que multiplica a DB2. Es decir, desea testear si $\beta_{DB3} = 2\beta_{DB2}$. ¿Puede realizar este test de hipótesis? ¿Tiene alguna intuición testear particularmente esta hipótesis, $\beta_{DB3} = 2\beta_{DB2}$? ¿Qué significa intuitivamente esta hipótesis nula, es decir, qué es lo que se pretende testear con este test de hipótesis?

Explique claramente qué pasos realizó en el software, cómo obtuvo el estadístico (su valor), el p-value y la conclusión a la que llega. Explique cada paso en detalle.