

## TP N° 2 - Econometría

1)

```
Data columns (total 12 columns):
#   Column      Non-Null Count  Dtype
---  -
0   price       506 non-null     int64
1   crime       506 non-null     float64
2   nox         506 non-null     float64
3   rooms       506 non-null     float64
4   dist        506 non-null     float64
5   radial      506 non-null     int64
6   proptax     506 non-null     float64
7   stratio     506 non-null     float64
8   lowstat     506 non-null     float64
9   lprice      506 non-null     float64
10  lnox        506 non-null     float64
11  lproptax    506 non-null     float64
dtypes: float64(10), int64(2)
memory usage: 47.6 KB
```

	min	q25	q50	q75	max	mean	range	std	skew	kurt
price	5000.000000	16850.0000	21200.0000	24999.000000	50001.000000	22511.509881	45001.000000	9208.856171	1.109568	1.485846
crime	0.006000	0.0820	0.2565	3.677000	88.975998	3.611536	88.969998	8.590247	5.233601	37.305912
nox	3.850000	4.4900	5.3800	6.240000	8.710000	5.549783	4.860000	1.158395	0.724734	-0.069637
rooms	3.560000	5.8825	6.2100	6.620000	8.780000	6.284051	5.220000	0.702594	0.404004	1.889982
dist	1.130000	2.1000	3.2100	5.187500	12.130000	3.795751	11.000000	2.106137	1.012032	0.487790
radial	1.000000	4.0000	5.0000	24.000000	24.000000	9.549407	23.000000	8.707259	1.004815	-0.867232
proptax	18.700001	27.9000	33.0000	66.599998	71.099998	40.823715	52.399997	16.853711	0.669956	-1.142408
stratio	12.600000	17.4000	19.1000	20.200001	22.000000	18.459289	9.400000	2.165820	-0.806780	-0.282833
lowstat	1.730000	6.9225	11.3600	17.057500	39.070000	12.701482	37.340000	7.238066	0.947746	0.640760

A partir de las estadísticas descriptivas del cuadro anterior se puede observar en cada variable que:

\* Price: la media es de \$22511.51 y la desviación respecto a la media es de \$ 9208.86. Es la variable con mayor dispersión en términos absolutos. En cuanto a la distribución de probabilidad, las observaciones están altamente sesgadas hacia la derecha (asimetría positiva) y es leptocurtica, es decir, hay una alta concentración de datos alrededor de la media. Sumado a lo anterior, es muy probable que haya una gran cantidad de outliers ya que la diferencia porcentual entre el cuarto y tercer cuartil es de casi un % 100.

\* Crime: la media es de 3.611536 y la desviación respecto a la media es de 8.59. En cuanto a la distribución de probabilidad, las observaciones están altamente sesgadas hacia la derecha (asimetría positiva) y es leptocurtica. Con sus valores de asimetría, curtosis y diferencia porcentual entre el cuarto y tercer cuartil de % 2319.79 debería de haber una gran cantidad de outliers.

\* Nox: la media es de 5.549783 y la desviación respecto a la media es de 1.158395. En cuanto a la distribución de probabilidad, las observaciones están altamente sesgadas hacia la derecha (asimetría positiva) y platicurtica. En ésta variable no debería de haber muchos outliers ya que su curtosis es casi igual a 0.

\* Rooms: la media es de 6.284051 y la desviación respecto a la media es de 0.702594. En cuanto a la distribución de probabilidad, si bien las observaciones están levemente sesgadas hacia la derecha tiende a una distribución simétrica y leptocurtica. Debido a la gran cantidad de observaciones concentradas alrededor de la media, quizá haya varios outliers.

\* Dist: la media es de 3.795751 y la desviación respecto a la media es de 2.106137. En cuanto a la distribución de probabilidad, las observaciones están altamente sesgadas hacia la derecha (asimetría positiva) y es leptocurtica. Al ser la curtosis no tan alta no debiera de tener muchos valores anormales.

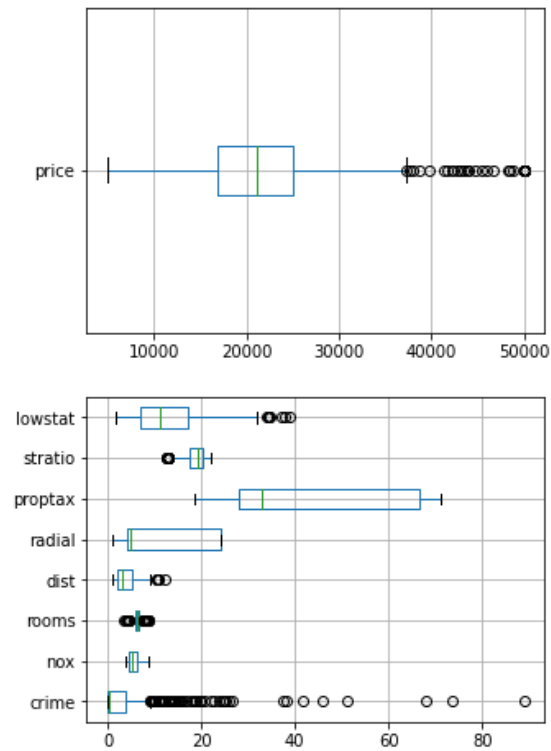
\* Radial: la media es de 9.549407 y la desviación respecto a la media es de 8.707259. En cuanto a la distribución de probabilidad, las observaciones están altamente sesgadas hacia la derecha (asimetría positiva) y es platicurtica por lo que no debiera de haber outliers.

\* Proptax: la media es de 40.823715 y la desviación respecto a la media es de 16.853711. En cuanto a la distribución de probabilidad, las observaciones están levemente sesgadas hacia la derecha (asimetría positiva) y al ser el valor de la curtosis tan pequeño (platicurtica) no hay outliers.

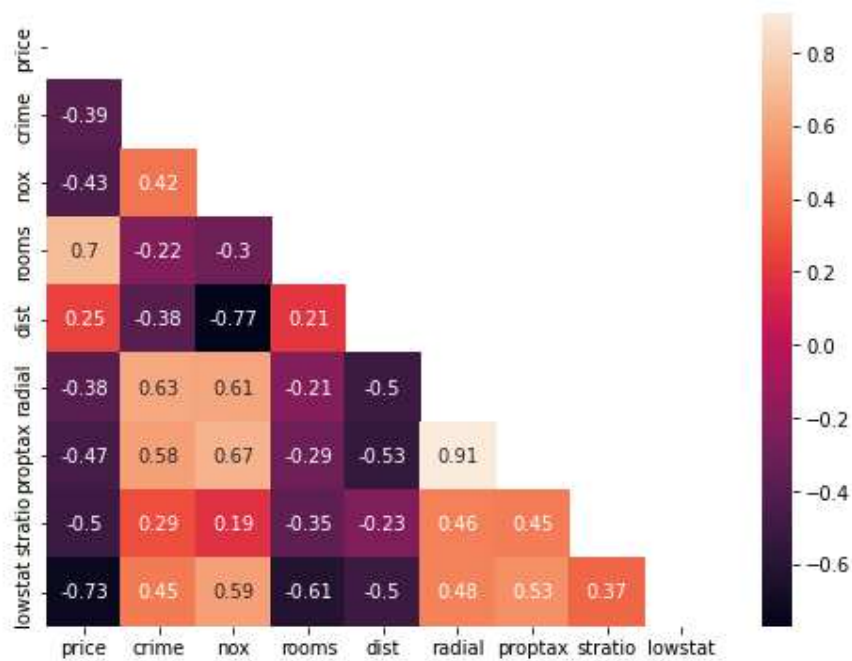
\* Stratio: la media es de 18.459289 y la desviación respecto a la media es de 2.165820. En cuanto a la distribución de probabilidad, las observaciones están levemente sesgadas hacia la izquierda (asimetría negativa) y al ser el valor de la curtosis negativo y no tan lejano al 0 (platicurtica) no debería de haber muchos outliers.

\* Lowstat: la media es de 12.701482 y la desviación respecto a la media es de 7.238066. En cuanto a la distribución de probabilidad, las observaciones están levemente sesgadas hacia la derecha (asimetría positiva) y es leptocurtica. Quizá nos encontremos con outliers.

Como siguiente paso vamos a verificar con diagramas de caja-bigote para verificar las distribuciones de nuestros datos y corroborar variable por variable los outliers que pueden influir los coeficientes de las regresiones que vayamos a estimar.



2)



Al observar el heatmap, la variable price tiene una correlación negativa con lowstat, stratio, nox y crime tal como esperaba; en cambio proptax, si bien tiene correlación negativa, no esperaba la misma ya que al ser el precio de una vivienda mayor los impuestos abonados debieran de ser mayores. En cuanto a las relaciones positivas con el precio nos encontramos con rooms y dist, dichas relaciones tienen sentido ya que al haber mas cuartos implica que en

promedio hay más metros cuadrados cubiertos, y al haber una mayor distancia con zonas productivas implica que las viviendas se encuentren en zonas residenciales por lo que hace aumentar el precio de las mismas.

En cuanto al resto de las variables se puede observar que hay una alta correlación positiva entre radial y crime, proptax y crime, radial y nox, proptax y nox, lowstat y nox, proptax y radial, proptax y lowstat. Las correlaciones altas negativas que muestran los datos son dist y nox, rooms y lowstat, proptax y dist.

Estas altas correlaciones pueden traer problemas de multicolinealidad en nuestras regresiones.

3)

OLS Regression Results						
Dep. Variable:	price		R-squared:	0.566		
Model:	OLS		Adj. R-squared:	0.563		
Method:	Least Squares		F-statistic:	124.8		
Date:	Sat, 13 Nov 2021		Prob (F-statistic):	8.10e-74		
Time:	20:16:53		Log-Likelihood:	-5124.8		
No. Observations:	506		AIC:	1.026e+04		
Df Residuals:	501		BIC:	1.028e+04		
Df Model:	4					
Covariance Type:	HC1					
	coef	std err	z	P> z	[0.025	0.975]
const	-2.033e+04	5162.718	-3.937	0.000	-3.04e+04	-1.02e+04
nox	-1034.1739	260.818	-3.965	0.000	-1545.367	-522.981
rooms	7951.8353	705.330	11.274	0.000	6569.413	9334.258
crime	-161.0131	35.327	-4.558	0.000	-230.252	-91.774
radial	-84.8553	59.029	-1.438	0.151	-200.550	30.839
Omnibus:	242.604		Durbin-Watson:		0.764	
Prob(Omnibus):	0.000		Jarque-Bera (JB):		1891.585	
Skew:	1.933		Prob(JB):		0.00	
Kurtosis:	11.647		Cond. No.		197.	

El modelo ajustado a nuestros datos utiliza las variables polución (cada 100m), número de cuartos, crimen per capita y el acceso a carreteras radiales para explicar la variación en el precio de las viviendas.

Para el análisis de cada coeficiente vamos a suponer que el resto de las variables se mantienen constantes (ceteris paribus). De ésta forma encontramos que el precio de una vivienda:

- Ante un cambio unitario en la variable polución disminuye \$1034.17
- Ante un cambio unitario el N° de cuartos aumenta \$7951.83
- Ante un cambio unitario en el crimen per capita disminuye \$161.01
- Ante un cambio unitario en el acceso a carreteras radiales disminuye \$84.85

Todas los Betas estimados son estadísticamente significativos excepto el del acceso a carreteras radiales ( $p > 0.05$ ), es decir que el valor de éste parámetro puede ser igual a 0 (no influye en nuestra variable dependiente).

El coeficiente de determinación nos indica que el 56.6% de la varianza en el precio de las viviendas es explicado por nuestro modelo.

4)

OLS Regression Results						
Dep. Variable:	price	R-squared:	0.663			
Model:	OLS	Adj. R-squared:	0.660			
Method:	Least Squares	F-statistic:	174.7			
Date:	Sat, 13 Nov 2021	Prob (F-statistic):	3.02e-107			
Time:	20:17:08	Log-Likelihood:	-5061.0			
No. Observations:	506	AIC:	1.013e+04			
Df Residuals:	500	BIC:	1.016e+04			
Df Model:	5					
Covariance Type:	HC1					
	coef	std err	z	P> z	[0.025	0.975]
const	1.088e+05	1.67e+04	6.516	0.000	7.61e+04	1.42e+05
nox	-1417.1975	218.692	-6.480	0.000	-1845.825	-988.570
rooms	-3.181e+04	5398.779	-5.892	0.000	-4.24e+04	-2.12e+04
squared_rooms	3074.0133	435.958	7.051	0.000	2219.552	3928.475
crime	-197.2100	37.242	-5.295	0.000	-270.203	-124.217
radial	-74.0932	54.721	-1.354	0.176	-181.344	33.158
Omnibus:	247.885	Durbin-Watson:	0.888			
Prob(Omnibus):	0.000	Jarque-Bera (JB):	4069.286			
Skew:	1.714	Prob(JB):	0.00			
Kurtosis:	16.463	Cond. No.	2.09e+03			

En ésta regresión podemos observar que el coeficiente de rooms es negativo y el de squared\_rooms es positivo. Esto implica que, para valores bajos de rooms, una habitación más tiene un efecto negativo sobre el precio. En algún punto el efecto se vuelve positivo, y la forma cuadrática significa que la semielasticidad de price respecto a rooms aumenta a medida que rooms aumenta.

5)

OLS Regression Results						
Dep. Variable:	lprice		R-squared:	0.527		
Model:	OLS		Adj. R-squared:	0.523		
Method:	Least Squares		F-statistic:	87.29		
Date:	Sat, 13 Nov 2021		Prob (F-statistic):	3.05e-56		
Time:	20:17:20		Log-Likelihood:	-75.960		
No. Observations:	506		AIC:	161.9		
Df Residuals:	501		BIC:	183.1		
Df Model:	4					
Covariance Type:	HC1					
	coef	std err	z	P> z	[0.025	0.975]
const	7.0039	0.428	16.382	0.000	6.166	7.842
lnox	-0.1676	0.096	-1.745	0.081	-0.356	0.021
lrooms	1.7171	0.219	7.833	0.000	1.287	2.147
lcrime	-0.0694	0.015	-4.485	0.000	-0.100	-0.039
lradial	0.0115	0.026	0.433	0.665	-0.040	0.063
Omnibus:	89.947	Durbin-Watson:	0.644			
Prob(Omnibus):	0.000	Jarque-Bera (JB):	509.557			
Skew:	0.626	Prob(JB):	2.24e-111			
Kurtosis:	7.754	Cond. No.	88.7			

La interpretación de éste modelo log-log es que a partir de un cambio del 1% de la variable que estemos analizando, suponiendo que el resto de las variables independientes se mantienen constantes, el precio de la vivienda variará en la misma proporción que el Beta de la variable en cuestión. Ejemplo: Un cambio unitario en *lrooms* genera, en promedio, un incremento de 1,71 en *lprice*. Por propiedades del logaritmo natural, esto es equivalente a decir que un cambio porcentual unitario en *rooms* genera, en promedio, un incremento de 1,71% en *price*.

6) El modelo que mejor ajusta a nuestros datos es el log-log ya que el AIC ajustado es menor que el AIC del modelo lineal.

AIC log-log transformado =  $\sum (\log y) * 2 + \text{AIC modelo log-log} = 10222.2692$

AIC modelo lineal = 10259.6311

AIC log-log transformado < AIC modelo lineal.

7)

La disponibilidad marginal a pagar por vivir en un vecindario más seguro es del % 0.0694. Es decir, ante un incremento del 1 % del crimen en un vecindario dado, los individuos de la muestra están dispuestos a pagar % -0.0694 menos por la vivienda.

8)

```

                                OLS Regression Results
=====
Dep. Variable:                  price    R-squared:                  0.557
Model:                          OLS      Adj. R-squared:            0.554
Method:                        Least Squares  F-statistic:              86.66
Date:                          Fri, 01 Apr 2022  Prob (F-statistic):    6.42e-56
Time:                          10:21:59    Log-Likelihood:           -5130.1
No. Observations:              506      AIC:                      1.027e+04
Df Residuals:                  501      BIC:                      1.029e+04
Df Model:                      4
Covariance Type:               HC1
=====
                                coef      std err          z      P>|z|      [0.025      0.975]
-----
Intercept      -2.701e+04    4255.889     -6.346     0.000    -3.54e+04    -1.87e+04
C(nox_2)[T.1]  -758.1286     719.706     -1.053     0.292    -2168.727     652.470
rooms          8225.0098     678.376     12.125     0.000     6895.417     9554.603
crime          -162.2183     37.583     -4.316     0.000     -235.880     -88.557
radial         -139.3654     51.056     -2.730     0.006     -239.434     -39.297
=====
Omnibus:                242.965    Durbin-Watson:           0.774
Prob(Omnibus):           0.000    Jarque-Bera (JB):        1989.917
Skew:                    1.918    Prob(JB):                0.00
Kurtosis:                11.926    Cond. No.:               149.
=====

Warnings:
[1] Standard Errors are heteroscedasticity robust (HC1)

```

Para poder captar que ocurre con el precio de los inmuebles ubicados en zonas con 6 o más partes de óxido nitroso vamos crear una variable dummy con un valor igual a 1. El coeficiente de dicha variable ( $C(nox\_2[T.1])$ ) es negativa con un valor de -758.13, siendo estadísticamente no significativa. Es decir, que cuando nox cambia de 5 a 6, éste suceso afecta negativamente al precio de un inmueble.