

TP N° 3

Para comenzar vemos si hay datos faltantes y verificar que el formato de los mismos sea el correcto para poder trabajarlos adecuadamente. En éste dataset no hay valores faltantes en ninguna variable y el tipo de datos está bien en cada una de ellas.

```
Data columns (total 3 columns):
#   Column  Non-Null Count  Dtype
---  -
0    y      50 non-null    float64
1   x1      50 non-null    float64
2   x2      50 non-null    float64
dtypes: float64(3)
memory usage: 1.3 KB
```

Y = autonomía fiscal

X1 = educación

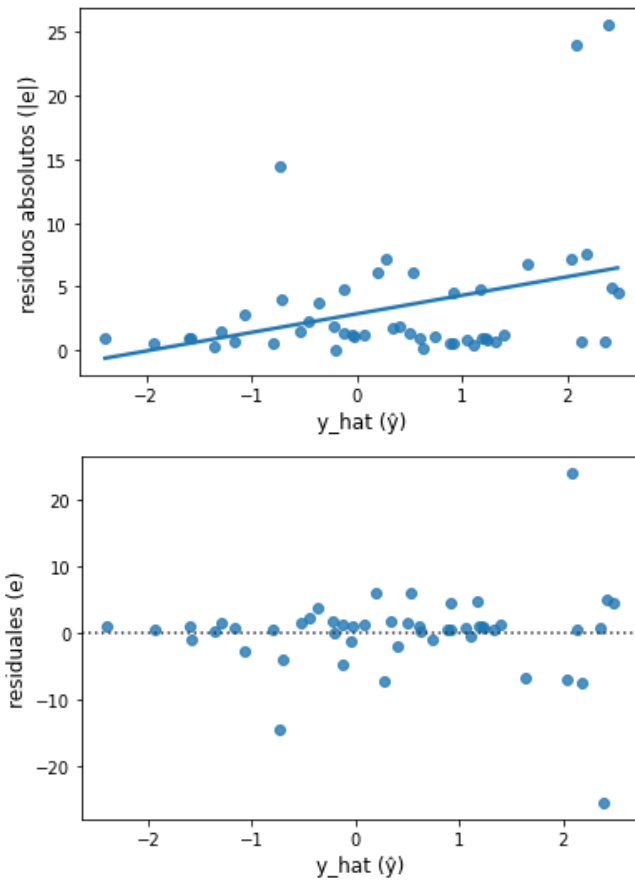
X2 = preferencias políticas

A –

```
OLS Regression Results
=====
Dep. Variable:          y      R-squared:          0.038
Model:                  OLS    Adj. R-squared:      -0.003
Method:                 Least Squares    F-statistic:        0.9258
Date:                   Fri, 10 Dec 2021    Prob (F-statistic):  0.403
Time:                   12:27:23    Log-Likelihood:     -162.04
No. Observations:       50    AIC:                330.1
Df Residuals:           47    BIC:                335.8
Df Model:                2
Covariance Type:        nonrobust
=====
               coef      std err          t      P>|t|      [0.025      0.975]
-----
const          0.1904        0.914        0.208      0.836      -1.649      2.030
x1             1.1311        0.983        1.151      0.255      -0.846      3.108
x2             0.3768        0.440        0.857      0.396      -0.508      1.262
=====
Omnibus:            21.877    Durbin-Watson:       1.856
Prob(Omnibus):      0.000    Jarque-Bera (JB):    144.127
Skew:               -0.552    Prob(JB):            5.05e-32
Kurtosis:           11.244    Cond. No.             2.35
=====
```

Warnings:

[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.



B –

OLS Regression Results						
=====						
Dep. Variable:	y	R-squared:	0.038			
Model:	OLS	Adj. R-squared:	-0.003			
Method:	Least Squares	F-statistic:	2.363			
Date:	Fri, 10 Dec 2021	Prob (F-statistic):	0.105			
Time:	12:27:23	Log-Likelihood:	-162.04			
No. Observations:	50	AIC:	330.1			
Df Residuals:	47	BIC:	335.8			
Df Model:	2					
Covariance Type:	HC1					
=====						
	coef	std err	z	P> z	[0.025	0.975]

const	0.1904	0.747	0.255	0.799	-1.274	1.655
x1	1.1311	0.548	2.064	0.039	0.057	2.205
x2	0.3768	1.103	0.342	0.733	-1.786	2.539
=====						
Omnibus:	21.877	Durbin-Watson:	1.856			
Prob(Omnibus):	0.000	Jarque-Bera (JB):	144.127			
Skew:	-0.552	Prob(JB):	5.05e-32			
Kurtosis:	11.244	Cond. No.	2.35			

Warnings:

[1] Standard Errors are heteroscedasticity robust (HC1)

	OLS se	OLS - rob. se
const	0.9144	0.7470
x1	0.9826	0.5481
x2	0.4399	1.1034

Se puede observar que el error estándar del estimador x1 disminuyó notablemente a tal punto que la educación pasó de ser una variable estadísticamente no significativa a una estadísticamente significativa. Por otro lado, el error estándar de las preferencias políticas aumentó, indicando en ambos modelos que la variable no es estadísticamente significativa para explicar la autonomía fiscal.

La heterocedasticidad en el modelo de regresión afecta principalmente la inferencia estadística ya que los coeficientes de las distintas variables exógenas siguen siendo los mismos.

C –

Prueba de White para heterocedasticidad

H0 = Hay homocedasticidad en el modelo de regresión.

H1 = Hay heterocedasticidad en el modelo de regresión.

'F-Statistic': 31.74641282710611, 'F-Test p-value': 1.5067434695842272e-13

Prueba de Breusch-Pagan para heterocedasticidad

H0 = Hay homocedasticidad en el modelo de regresión.

H1 = Hay heterocedasticidad en el modelo de regresión.

'F-Statistic': 9.329238516126505, 'F-Test p-value': 0.0003872271019027869

Luego de realizados los test de White y de Breush-Pagan para corroborar si hay heteroscedasticidad en nuestro modelo de regresión, podemos verificar que sus p-values son $0.00 < 0.05$ y $0.0004 < 0.05$ respectivamente. Es decir, ambos test indican que rechazamos la hipótesis nula de que hay homocedasticidad en el modelo de regresión ($p < 0.05$).

En consecuencia procederemos a trabajar con OLS con errores estándar robustos (White-Huber) para que nuestras inferencias sobre las variables exógenas del modelo sean confiables ya que las varianzas de los estimadores estarán sesgados afectando al error estándar de los mismos, y por ende a t estadístico y los intervalos de confianza.

D –

La prueba de Goldfeld-Quandt se hizo con cada variable independiente con respecto a la dependiente eliminando 10 observaciones de la muestra, desde la 21 a la 30, luego de ordenar las mismas de forma ascendente dejando de esta manera 20 observaciones en cada sub-muestra.

H0 = La varianza de los residuales en ambas sub-muestras son iguales.

H1 = La varianza de los residuales en ambas sub-muestras no son iguales.

Autonomía fiscal y educación (y, x1):

'F-Statistic': 0.411, 'F-Test p-value': 0.9665

Al ser $p > 0.05$ fallamos en rechazar en H0, es decir que en éste modelo de regresión simple hay homocedasticidad.

Autonomía fiscal y preferencias políticas (y, x2):

'F-Statistic': 4.3372, 'F-Test p-value': 0.0016.

Al ser $p < 0.05$ fallamos en rechazar en H0, es decir que en éste modelo de regresión simple hay heterocedasticidad.

Creo que es importante remarcar que los resultados de ésta prueba de hipótesis son solamente válidos para la metodología previamente explicada. Si cambiásemos el n° de observaciones eliminadas los resultados podrían variar.

E –

Para la regresión de mínimos cuadrados ponderados corrí una regresión OLS sobre los residuales absolutos (provenientes de una OLS de la var. dependiente e independientes originales, punto A) contra las variables independientes originales.

Los valores ajustados de esta regresión son una estimación del error de cada observación que luego serán utilizados para calcular la inversa de la varianza ($1/(\sigma_i)^2$) utilizada como peso en WLS.

```

=====
WLS Regression Results
=====
Dep. Variable:          y      R-squared:          0.052
Model:                  WLS    Adj. R-squared:       0.011
Method:                 Least Squares    F-statistic:      1.282
Date:                   Fri, 10 Dec 2021    Prob (F-statistic): 0.287
Time:                   18:03:10    Log-Likelihood:   -186.83
No. Observations:       50    AIC:              379.7
Df Residuals:           47    BIC:              385.4
Df Model:                2
Covariance Type:        nonrobust
=====
               coef      std err          t      P>|t|      [0.025      0.975]
-----
const          1.4813      1.817      0.815      0.419     -2.174      5.136
x1             1.4465      0.972      1.489      0.143     -0.508      3.401
x2             0.8946      0.745      1.201      0.236     -0.604      2.394
=====
Omnibus:                 56.599    Durbin-Watson:       2.029
Prob(Omnibus):           0.000    Jarque-Bera (JB):    514.550
Skew:                    -2.665    Prob(JB):             1.85e-112
Kurtosis:                 17.784    Cond. No.              9.38
=====

```

Warnings:

[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

Comparación de errores estándar (OLS, OLS con errores robustos a la het. y WLS)

	OLS se	OLS robust se	WLS se
const	0.9144	0.7470	1.8169
x1	0.9826	0.5481	0.9716
x2	0.4399	1.1034	0.7451

Si bien hay un comportamiento similar en los errores estándar del modelo OLS con errores estándar robustos y el de WLS en el sentido de que ambos tienden a disminuir en x1 y a aumentar en x2, WLS no modifica el error de manera significativa como para cambiar la interpretación de alguna variable explicativa con respecto al modelo base OLS, en cambio OLS con errores estándares robustos a la heterocedasticidad, convierte a la variable x1 en estadísticamente significativa para explicar la autonomía fiscal.

Otro punto importante es que luego de correr el test de White para la heterocedasticidad, el p-value del F estadístico me dio $4.479391855012567e-25$ ($p < 0.05$), es decir que se rechaza la hipótesis nula de que hay homocedasticidad en el modelo WLS. En consecuencia, tanto OLS como WLS no se pueden utilizar para hacer inferencia estadística.

BONUS

1 –

OLS Regression Results						
Dep. Variable:	inf	R-squared:	0.053			
Model:	OLS	Adj. R-squared:	0.033			
Method:	Least Squares	F-statistic:	2.616			
Date:	Mon, 06 Dec 2021	Prob (F-statistic):	0.112			
Time:	09:32:15	Log-Likelihood:	-124.43			
No. Observations:	49	AIC:	252.9			
Df Residuals:	47	BIC:	256.6			
Df Model:	1					
Covariance Type:	nonrobust					
	coef	std err	t	P> t	[0.025	0.975]
const	1.4236	1.719	0.828	0.412	-2.035	4.882
unem	0.4676	0.289	1.617	0.112	-0.114	1.049
Omnibus:	8.905	Durbin-Watson:	0.803			
Prob(Omnibus):	0.012	Jarque-Bera (JB):	8.336			
Skew:	0.979	Prob(JB):	0.0155			
Kurtosis:	3.502	Cond. No.	23.5			

Warnings:

[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

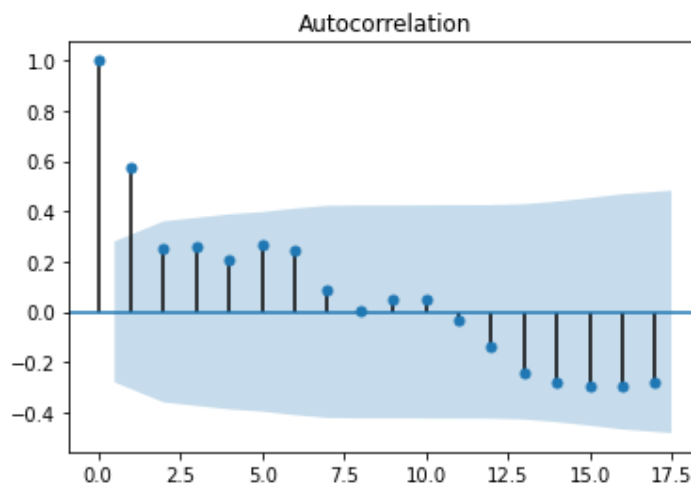
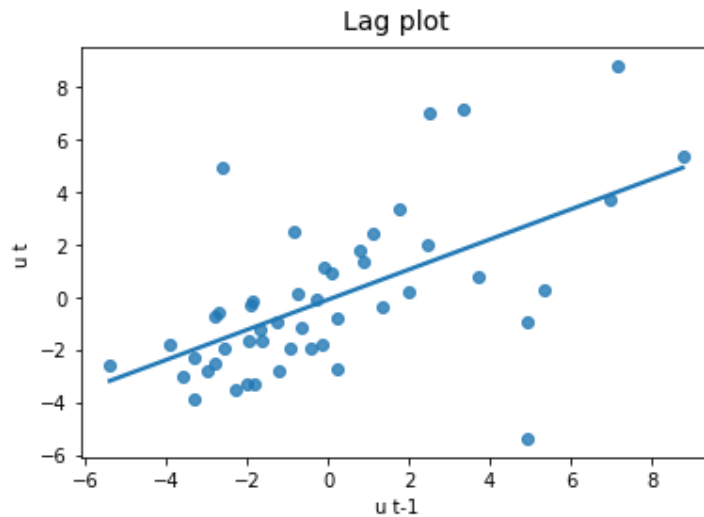
2 –

OLS Regression Results						
=====						
Dep. Variable:	u t	R-squared:	0.346			
Model:	OLS	Adj. R-squared:	0.332			
Method:	Least Squares	F-statistic:	24.34			
Date:	Mon, 06 Dec 2021	Prob (F-statistic):	1.10e-05			
Time:	09:32:59	Log-Likelihood:	-110.88			
No. Observations:	48	AIC:	225.8			
Df Residuals:	46	BIC:	229.5			
Df Model:	1					
Covariance Type:	nonrobust					
=====						
	coef	std err	t	P> t	[0.025	0.975]

const	-0.1134	0.359	-0.316	0.754	-0.837	0.610
u t-1	0.5730	0.116	4.934	0.000	0.339	0.807
=====						
Omnibus:	6.807	Durbin-Watson:	1.354			
Prob(Omnibus):	0.033	Jarque-Bera (JB):	10.853			
Skew:	0.158	Prob(JB):	0.00440			
Kurtosis:	5.308	Cond. No.	3.09			
=====						

Warnings:

[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.



En el primer gráfico se puede observar que hay una correlación lineal positiva entre el residual t y el residual rezagado $t-1$, es decir hay correlación serial de primer orden lo cual implica una violación al supuesto de errores no correlacionados.

En el segundo gráfico intentamos verificar si hay auto correlación para aquellos rezagos mayores a 1. Al parecer sólo hay correlación serial de primer orden ya que la misma se encuentra por fuera del intervalo de confianza del 95%.

La violación al supuesto de errores independientes/no correlacionados implica que los estimadores de OLS ya no son MELI y que los errores estándar y los estadísticos de prueba ya no son más válidos, los últimos no se podrán usar más para pruebas de hipótesis. En aquellos casos cuando $p > 0$, se estará subestimando la varianza del estimador de OLS y, en consecuencia, un error estándar más chico significa un estadístico t mayor y un intervalo de confianza más acotado.

3 –

H_0 = No hay correlación serial en los residuos.

H_1 = Hay correlación serial en los residuos.

El valor de éste estadístico siempre se encontrará entre 0 y 4. Mientras más cercano al 0, más evidencia habrá de correlación serial positiva. Si el estadístico es igual a 2 entonces no hay auto correlación. Entre más cercano al 4, más evidencia habrá de correlación serial negativa.

El valor del estadístico de la prueba Durbin-Watson es de 0.8027. Al encontrarse dicho valor cerca del 0 implica que rechazamos la hipótesis nula debido a que hay suficiente evidencia estadística de que hay correlación serial positiva en los residuales de primer orden.