

## Solución a la práctica 4.2 con Eviews

Se piensa que el apoyo electoral al bipartidismo disminuye cuando aumenta el tamaño de la población de las localidades y cuando se consigue una mayor participación de los electores. También puede influir la situación socioeconómica de los votantes. Para comprobar la veracidad de estos supuestos, se han tomado los siguientes datos sobre los resultados de las elecciones al Congreso de los Diputados celebradas en 2015 para cada municipio de la Región de Murcia (fichero de datos: *practica42.wf1*): porcentaje de votantes que votó a uno cualquiera de los dos partidos mayoritarios hasta ese momento (*bipartidismo*), total de miles de votantes (*votantes*) y total de miles de inscritos en el censo electoral (*censados*). Igualmente, de cada municipio se dispone de los datos de paro registrado en 2015 en porcentaje de su población total (*paro*). Para analizar las cuestiones descritas se plantea el siguiente modelo:

$$\text{bipartidismo} = \beta_0 + \beta_1 \text{votantes} + \beta_2 \text{censados} + \beta_3 \text{paro} + \varepsilon$$

### 1) Estime el modelo propuesto por MCO.

En Eviews, la estimación por MCO del modelo se realiza mediante la siguiente orden y produce los resultados que se observan en la imagen:

*ls bipartidismo votantes censados paro c*

Variable	Coefficient	Std. Error	t-Statistic	Prob.
VOTANTES	0.642473	0.627286	1.024211	0.3117
CENSADOS	-0.528676	0.461143	-1.146447	0.2583
PARO	-0.465602	0.499368	-0.932382	0.3566
C	72.34992	4.791338	15.10015	0.0000

R-squared	0.208337	Mean dependent var	66.54815
Adjusted R-squared	0.150410	S.D. dependent var	6.907453
S.E. of regression	6.366820	Akaike info criterion	6.624765
Sum squared resid	1661.992	Schwarz criterion	6.785357
Log likelihood	-145.0572	Hannan-Quinn criter.	6.684632
F-statistic	3.596572	Durbin-Watson stat	1.932433
Prob(F-statistic)	0.021376		

2) Examine los resultados de la anterior estimación, en particular fíjese en los resultados de los contrastes de significatividad individuales de los coeficientes de las variables explicativas y en el contraste de significatividad conjunta del modelo. ¿Nota algo que le parezca anómalo? ¿Es capaz de aventurar una explicación a dicha anomalía?

Los resultados de la estimación mostrados por Eviews incluyen los resultados de los contrastes de significatividad individual de los coeficientes de cada una de las variables explicativas y el contraste de significatividad conjunta del modelo. Llama la atención el hecho de que cada uno de los p-valores de los tres contrastes individuales (31.2%, 25.8% y 35.7%) indique no significatividad de la explicativa correspondiente y que, por el contrario, el bajo p-valor del contraste conjunto (2.14%) tienda a señalar que, conjuntamente, el grupo de las tres variables sí resulta significativo a la hora de explicar el apoyo del electorado a los partidos mayoritarios. Aun siendo posible que esa contradicción en los resultados se deba al azar, también es factible que sea consecuencia del efecto engañoso provocado por la existencia de fuerte multicolinealidad entre algún grupo de variables explicativas.

3) ¿Qué método(s) sencillo(s) se le ocurre emplear para intentar confirmar si la explicación que ha aventurado para la anomalía es un hecho que realmente se da en los datos? Aplíquelo(s) y obtenga conclusiones.

Para confirmar o descartar la existencia de multicolinealidad fuerte entre algunas de las variables explicativas, se puede comenzar por calcular las diferentes correlaciones entre ellas. Para hacerlo es preciso abrir las tres variables en una misma hoja y desde esa hoja usar *View/Covariance Analysis*. En el menú que se abre debe seleccionarse únicamente la opción *Correlation*:

Correlation			
	CENSADOS	VOTANTES	PARO
CENSADOS	1.000000	0.999157	-0.017073
VOTANTES	0.999157	1.000000	-0.016396
PARO	-0.017073	-0.016396	1.000000

El altísimo valor de uno de los tres coeficientes de correlación (0.9992), el correspondiente a la pareja formada por *centsados* y *votantes*, y el bajo valor de cada uno de los otros dos (-0.0171 y -0.0164), confirman la existencia de multicolinealidad muy fuerte que estaría centrada únicamente en la pareja de variables citada.

Otro modo de examinar la multicolinealidad entre las variables explicativas del modelo es calcular los coeficientes de determinación de las tres regresiones auxiliares de cada una de las variables explicativas sobre las otras dos ( $R_j^2$ ,  $j = 1, 2, 3$ ). Los comandos

*ls centsados votantes paro c*

*ls votantes centsados paro c*

*ls paro votantes centsados c*

estiman esas tres regresiones. Los resultados son

Equation: UNTITLED Workfile: MULTICOLINEALIDAD2015::Multico... - \_ X

View Proc Object Print Name Freeze Estimate Forecast Stats Resids

Dependent Variable: CENSADOS  
 Method: Least Squares  
 Date: 12/23/15 Time: 18:21  
 Sample: 1 45  
 Included observations: 45

Variable	Coefficient	Std. Error	t-Statistic	Prob.
VOTANTES	1.359138	0.008616	157.7420	0.0000
PARO	-0.018224	0.167070	-0.109077	0.9137
C	0.577421	1.600754	0.360718	0.7201

R-squared	0.998315	Mean dependent var	22.52587
Adjusted R-squared	0.998235	S.D. dependent var	50.71220
S.E. of regression	2.130404	Akaike info criterion	4.414841
Sum squared resid	190.6222	Schwarz criterion	4.535285
Log likelihood	-96.33393	Hannan-Quinn criter.	4.459742
F-statistic	12444.90	Durbin-Watson stat	2.064566
Prob(F-statistic)	0.000000		

Equation: UNTITLED Workfile: MULTICOLINEALIDAD2015::Multico... - \_ X

View Proc Object Print Name Freeze Estimate Forecast Stats Resids

Dependent Variable: VOTANTES  
 Method: Least Squares  
 Date: 12/23/15 Time: 18:23  
 Sample: 1 45  
 Included observations: 45

Variable	Coefficient	Std. Error	t-Statistic	Prob.
CENSADOS	0.734521	0.004656	157.7420	0.0000
PARO	0.012850	0.122821	0.104622	0.9172
C	-0.391701	1.177049	-0.332782	0.7410

R-squared	0.998315	Mean dependent var	16.27402
Adjusted R-squared	0.998235	S.D. dependent var	37.28015
S.E. of regression	1.566146	Akaike info criterion	3.799453
Sum squared resid	103.0181	Schwarz criterion	3.919897
Log likelihood	-82.48769	Hannan-Quinn criter.	3.844353
F-statistic	12444.61	Durbin-Watson stat	2.063065
Prob(F-statistic)	0.000000		

Equation: UNTITLED Workfile: MULTICOLINEALIDAD2015::Multico... - \_ X

View Proc Object Print Name Freeze Estimate Forecast Stats Resids

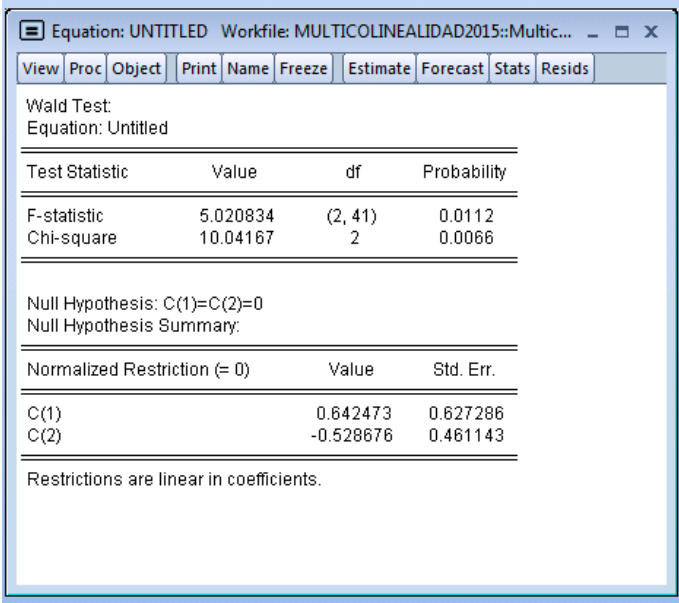
Dependent Variable: PARO  
 Method: Least Squares  
 Date: 12/23/15 Time: 18:24  
 Sample: 1 45  
 Included observations: 45

Variable	Coefficient	Std. Error	t-Statistic	Prob.
VOTANTES	0.020276	0.193804	0.104622	0.9172
CENSADOS	-0.015540	0.142472	-0.109077	0.9137
C	9.359626	0.325780	28.72988	0.0000

R-squared	0.000552	Mean dependent var	9.339541
Adjusted R-squared	-0.047041	S.D. dependent var	1.922628
S.E. of regression	1.967329	Akaike info criterion	4.255571
Sum squared resid	162.5561	Schwarz criterion	4.376015
Log likelihood	-92.75035	Hannan-Quinn criter.	4.300471
F-statistic	0.011598	Durbin-Watson stat	2.206309
Prob(F-statistic)	0.988473		

En esos resultados se observa que de los tres  $R_j^2$ , dos resultan tener valores muy altos, superiores al 99%, indicando que sus respectivas regresiones explican muy bien el comportamiento de las correspondientes variables dependientes (*censados* y *votantes*). Todo lo contrario ocurre con la tercera regresión auxiliar, cuyo  $R^2$  es casi 0. Además, en las regresiones de alto  $R_j^2$ , *paro* es siempre una variable explicativa no significativa, siendo relevante únicamente su pareja correspondiente. Todo ello confirma que la multicolinealidad fuerte existe y se limita a la pareja *censados* y *votantes*.

Por otra parte, nótese también que, en el modelo original, el contraste de significatividad conjunta de las variables *votantes* y *censados* (menú **View/Coefficient Diagnostics/Wald Test-Coefficient Restrictions**) tiene un p-valor de 1.12%:



Wald Test:			
Equation: Untitled			
Test Statistic	Value	df	Probability
F-statistic	5.020834	(2, 41)	0.0112
Chi-square	10.04167	2	0.0066
Null Hypothesis: C(1)=C(2)=0			
Null Hypothesis Summary:			
Normalized Restriction (= 0)	Value	Std. Err.	
C(1)	0.642473	0.627286	
C(2)	-0.528676	0.461143	
Restrictions are linear in coefficients.			

Ese bajo valor sería señal de significatividad conjunta de ambas, en contradicción con lo que señalaban sus contrastes individuales. Esta aparente contradicción seguramente será de nuevo un efecto engañoso provocado por la multicolinealidad fuerte detectada.

**4) ¿Qué se puede aplicar a esta situación con el fin de conseguir hacer desaparecer el efecto anómalo detectado en el apartado 3? Hágalo. Interprete los resultados, señalando los pros y contras del modelo alternativo que haya planteado. ¿Servirá ese modelo para analizar las cuestiones que motivaron este análisis?**

La multicolinealidad fuerte y sus consecuencias indeseadas desaparecen si se elimina de la ecuación una de las variables explicativas involucradas en la relación multicolineal. Empleando como guía los valores de los estadísticos t de significatividad individual, se llega a la conclusión de que la mejor candidata a ser eliminada es *votantes* pues el valor absoluto de su estadístico t es menor que el de la variable *censados* y solo ligeramente superior a 1. Todo esto lleva a plantear el modelo simplificado:

$$\text{bipartidismo} = \beta_0 + \beta_2 \text{censados} + \beta_3 \text{paro} + \varepsilon$$

Los resultados de su estimación son:

*ls bipartidismo censados paro c*

The screenshot shows the EViews software interface with the following data:

Variable	Coefficient	Std. Error	t-Statistic	Prob.
CENSADOS	-0.056766	0.018941	-2.997028	0.0046
PARO	-0.457346	0.499594	-0.915435	0.3652
C	72.09826	4.787825	15.05867	0.0000

R-squared	0.188082	Mean dependent var	66.54815
Adjusted R-squared	0.149419	S.D. dependent var	6.907453
S.E. of regression	6.370533	Akaike info criterion	6.605584
Sum squared resid	1704.515	Schwarz criterion	6.726028
Log likelihood	-145.6256	Hannan-Quinn criter.	6.650484
F-statistic	4.864677	Durbin-Watson stat	1.956244
Prob(F-statistic)	0.012582		

Con este procedimiento se consigue una reducción en la varianza de los estimadores que hace desaparecer el engañoso resultado de la no significatividad individual en los regresores multicolineales no suprimidos. Obsérvese que ahora el estadístico  $t$  para el contraste de significatividad individual de la variable *censados* tiene un p-valor igual a 0.0046. Este método tiene la desventaja de introducir sesgo en las estimaciones de los coeficientes de esos mismos regresores por implicar la eliminación de una variable relevante, aunque ese efecto negativo se palía, en la medida de lo posible, al haber seguido las indicaciones acerca de cómo elegir la variable explicativa a eliminar según los valores de los estadísticos  $t$ .

En todo caso hay que tener en cuenta que la estimación de  $\beta_2$  en el modelo simplificado es sesgada y, de hecho, recoge una combinación de los efectos *ceteris paribus* reales de *votantes* y *censados*.

Además, al no aparecer la explicativa *votantes* explicitada en la ecuación, resulta imposible estudiar su efecto *ceteris paribus*, que era uno de los objetivos pretendidos inicialmente.

En la estimación del modelo simplificado se aprecia que la variable *paro* no parece ser relevante pues el p-valor asociado a su test de significatividad individual toma un valor elevado igual a 0.3652. Resulta conveniente simplificar el modelo eliminando esa variable explicativa:

### ls bipartidismo censados c

Equation: UNTITLED Workfile: PRACTICA4B::Multicolinealidad2015\

View Proc Object Print Name Freeze Estimate Forecast Stats Resids

Dependent Variable: BIPARTIDISMO  
 Method: Least Squares  
 Date: 06/16/16 Time: 08:16  
 Sample: 1 45  
 Included observations: 45

Variable	Coefficient	Std. Error	t-Statistic	Prob.
CENSADOS	-0.056470	0.018902	-2.987466	0.0046
C	67.82019	1.039116	65.26718	0.0000

R-squared 0.171882 Mean dependent var 66.54815  
 Adjusted R-squared 0.152623 S.D. dependent var 6.907453  
 S.E. of regression 6.358523 Akaike info criterion 6.580896  
 Sum squared resid 1738.525 Schwarz criterion 6.661192  
 Log likelihood -146.0702 Hannan-Quinn criter. 6.610829  
 F-statistic 8.924951 Durbin-Watson stat 1.981024  
 Prob(F-statistic) 0.004633

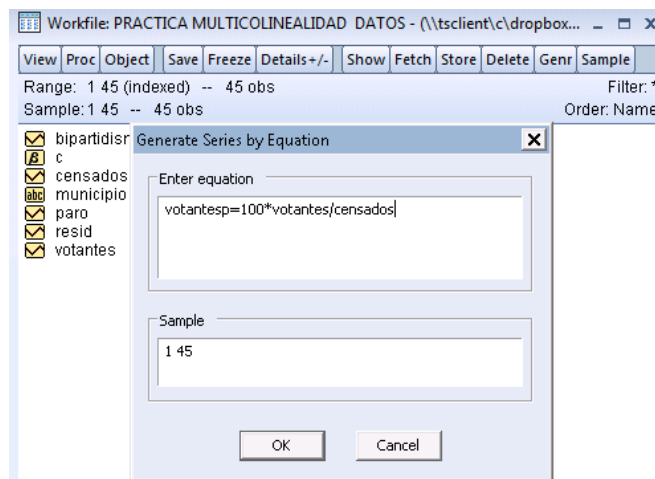
5) Ante las carencias de la solución dada en el apartado 4 a las anomalías detectadas, un investigador sugiere otra solución alternativa. Propone estimar el modelo

$$\text{bipartidismo} = \beta_0^* + \beta_1^* \text{votantesp} + \beta_2^* \text{censados} + \beta_3^* \text{paro} + \varepsilon$$

donde *votantesp* mide ahora la participación electoral mediante el porcentaje de votantes sobre el total del censo. Indique ventajas de esta solución respecto a la del apartado anterior. Compruebe que los problemas detectados han desaparecido. Estime el modelo e interprete los resultados que obtenga.

Lo que el investigador propone es medir la participación en términos relativos mediante el porcentaje de votantes sobre el total del censo, en lugar de en términos absolutos mediante el número total de votantes. A este respecto, lo que se haría es equiparar la forma de medir la participación, a la que se emplea con el paro, variable que desde el principio ha estado medida en términos porcentuales y no absolutos.

La variable  $\text{votantesp} = 100 * \text{votantes} / \text{censados}$  se puede generar pinchando sobre la pestaña *Genr*:



Se dispone ahora de tres regresores, *votantesp*, *censados* y *paro*, entre los que no se da la multicolinealidad como se puede comprobar, por ejemplo, en la matriz de correlaciones (menú *View/CovarianceAnalysis*):

	VOTANTESP	CENSADOS	PARO
VOTANTESP	1.000000	-0.153856	0.106343
CENSADOS	-0.153856	1.000000	-0.017073
PARO	0.106343	-0.017073	1.000000

Se debe comprobar que también se consiguen resultados que indican ausencia de multicolinealidad, con las regresiones auxiliares de cada una de esas tres variables sobre las otras dos.

Además de no sufrir del problema de la multicolinealidad, el modelo alternativo

$$bipartidismo = \beta_0^* + \beta_1^*votantesp + \beta_2^*censados + \beta_3^*paro + \varepsilon$$

sí vuelve a ser útil para analizar el efecto *ceteris paribus* de la participación electoral y el tamaño de las poblaciones sobre el apoyo a los dos partidos tradicionalmente mayoritarios. Estas deben ser las razones que seguramente han llevado al investigador a proponer este modelo alternativo.

Su estimación por MCO da los siguientes resultados:

*ls bipartidismo votantesp censados paro c*

Variable	Coefficient	Std. Error	t-Statistic	Prob.
VOTANTESP	0.601999	0.171011	3.520228	0.0011
CENSADOS	-0.047615	0.016999	-2.801028	0.0077
PARO	-0.622009	0.445565	-1.396002	0.1702
C	28.90454	12.98418	2.226136	0.0316

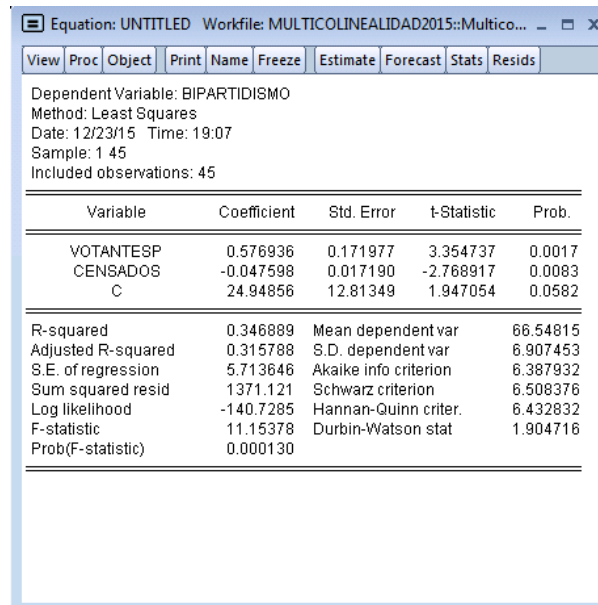
  

R-squared	0.376524	Mean dependent var	66.54815
Adjusted R-squared	0.330904	S.D. dependent var	6.907453
S.E. of regression	5.650180	Akaike info criterion	6.385939
Sum squared resid	1308.906	Schwarz criterion	6.546532
Log likelihood	-139.6836	Hannan-Quinn criter.	6.445807
F-statistic	8.253442	Durbin-Watson stat	1.851100
Prob(F-statistic)	0.000206		

Nótese, primero, que los resultados contradictorios entre sí observados en el modelo inicial al examinar los test de significatividad, desaparecen. Los contrastes de significatividad individual indican que tanto *votantesp* como *censados* son variables relevantes en el modelo.

Claramente, además, en el nuevo modelo se aprecia que el nivel de paro registrado no resulta significativo a la hora de explicar el bipartidismo (p-valor de su test individual de significatividad igual a 17.0%), por lo que *paro* es una variable explicativa irrelevante. Ello sugiere mejorar las estimaciones eliminando esa variable explicativa:

***ls bipartidismo votantesp censados c***



Equation: UNTITLED Workfile: MULTICOLINEALIDAD2015::Multico... - □ ×

View Proc Object Print Name Freeze Estimate Forecast Stats Resids

Dependent Variable: BIPARTIDISMO  
 Method: Least Squares  
 Date: 12/23/15 Time: 19:07  
 Sample: 1 45  
 Included observations: 45

Variable	Coefficient	Std. Error	t-Statistic	Prob.
VOTANTESP	0.576936	0.171977	3.354737	0.0017
CENSADOS	-0.047598	0.017190	-2.768917	0.0083
C	24.94856	12.81349	1.947054	0.0582

R-squared	0.346889	Mean dependent var	66.54815
Adjusted R-squared	0.315788	S.D. dependent var	6.907453
S.E. of regression	5.713646	Akaike info criterion	6.387932
Sum squared resid	1371.121	Schwarz criterion	6.508376
Log likelihood	-140.7285	Hannan-Quinn criter.	6.432832
F-statistic	11.15378	Durbin-Watson stat	1.904716
Prob(F-statistic)	0.000130		

La interpretación de los coeficientes de las dos variables explicativas restantes es clara: en el apoyo al bipartidismo, el tamaño de la población tiene un efecto negativo si permanece constante la participación medida en porcentaje de votantes sobre el censo. En otras palabras, según estos resultados, los pueblos pequeños son más proclives al bipartidismo. Pero ocurre al revés con el porcentaje de participación, que al aumentar, hace también aumentar el apoyo al bipartidismo si no varía el tamaño del censo de las localidades. Estas conclusiones discrepan en parte de las ideas preconcebidas de las que se partía.

La muestra usada se reduce a municipios de la Región de Murcia. Sería interesante comprobar qué ocurre si la muestra se centra en otras comunidades o se hace extensiva al conjunto del país.