

Puntos equipos de fútbol



Jose Antonio Moreno De Pradas

Econometría II

24-4-2019

Índice

1.	Motivación del trabajo.....	2
2.	Modelo.....	2
3.	Datos.....	3
4.	Análisis de datos de panel.....	4
	a. Efectos fijos.....	4
	b. Efectos aleatorios.....	5
5.	Multicolinealidad.....	6
	a. Matriz de correlaciones.....	6
	b. Factor de inflación de la Varianza (FIV).....	6
	c. MCO.....	7
	d. Omitimos variable con presencia de multicolinealidad.....	7
	e. Diagnóstico de panel.....	8
6.	Estimación del modelo.....	9
	a. Mínimo Cuadrática Ordinaria.....	9
	b. Modelo estimado.....	9
	c. Interpretación de las variables.....	9
	d. Multicolinealidad.....	11
	e. Tabla ANOVA.....	11
	f. Heterocedasticidad.....	11
	g. Contraste de White.....	12
	h. Autocorrelación.....	12
7.	Conclusión.....	12
8.	Bibliografía.....	13



1. Motivación del trabajo.

Decidí realizar este trabajo para intentar resolver una pregunta muy habitual entre los fans del deporte rey, ¿Por qué unos equipos de fútbol tienen más éxito que otros? Aunque muchos ya podemos tener una ligera intuición de cuál puede ser la diferencia en ese éxito (mayor presupuesto, mayor cantidad de goles, etc.), voy a tratar de evidenciarlo mediante datos. Por eso vamos a plantear un modelo el cuál considero que es el más indicado para tratar de explicar esas diferencias y que mediante las conclusiones finales confirmaremos si es el correcto o no. También hay que añadir que otra de las motivaciones para realizar un trabajo relacionado con este deporte, es la facilidad con la podemos encontrar los datos que necesitamos. Como sabemos hay muchos periódicos y páginas webs especializadas en este tema, que pueden incluso llegar a publicar cuantos pases ha realizado un equipo por partido. Algo muy a tener en cuenta sabiendo que nuestro trabajo va a depender de la calidad y cantidad de esos datos.

2. Modelo.

Previamente a la realización del modelo busqué investigaciones ya realizadas en relación con este tema, por si alguien ya había investigado sobre el modelo que voy a plantear. Y aunque sí que encontré investigaciones, estas se planteaban de una manera diferente a la que yo tenía en mente para mi trabajo. Así que finalmente planteo el modelo que había pensado desde un principio.

Lo que voy a tratar de conseguir con este modelo es ver cuántos puntos van a conseguir los equipos de la liga en función de unas variables, las cuales considero que pueden ser influyentes para explicar esas diferencias de puntos y que explicaré ahora:

$$y_{it} = \alpha_{it} + x'_{it}\beta + u_{it} \quad u_{it} \sim N(0, \sigma_u^2)$$

-**Cte**: Constante.

-**GF**: Goles a favor. En un principio suponemos que a más goles, conseguiremos más puntos.

-**GC**: Goles en contra. En este caso pasaría totalmente lo contraria, contra menos goles en contra más puntos conseguirás.

-**Chmps**: Jugar Champions; es una variable Dummy $\rightarrow 1 =$ en el caso de jugar la Champions; $0 =$ en caso de no jugarla. Jugar las Champions debería proporcionarte mayor puntuación al tener supuestamente una mejor plantilla.

-**EL**: Jugar Europa League; es una variable Dummy $\rightarrow 1 =$ en caso de jugar la Europa League; $0 =$ en caso de no jugarla. Ocurría lo mismo que con la Champions, pero en una menor cantidad.

-**Sueldo**: Tope salarial de los equipos (en millones de €). El tope salarial, es lo máximo que se puede gastar cada equipo en los salarios de sus jugadores. La suma del salario de todos sus jugadores no tiene que superar ese máximo. Suponemos que a mayor tope salarial, mayor posibilidad de pagar a mejores jugadores como consecuencia obtendremos mayor puntuación.

-**Presp**: Presupuesto total de todos los equipos en cada temporada (en millones de €). Si tenemos una cuantía presupuestaría mayor, en principio podríamos conseguir más puntos.

-**Market**: Es lo que cada equipo se ha gastado en el mercado de fichajes cada temporada (en millones de €). A mayor gasto en fichajes mejor equipo, por consiguiente más cantidad de puntos.



Aunque en principio el modelo tenía más variables como eran el número de tarjetas amarillas por temporada, gasto en fichajes o tarjetas rojas, hacían al modelo muy complejo para poder realizarlo en un trabajo tan breve. Por eso he decidido dejarlas fuera. También he añadido una pequeña explicación de por qué he introducido cada variable para mi modelo y como considero que van a afectar a priori sobre el modelo, pero eso no tiene por qué ser así, cuando analice los datos puede que estos tengan el efecto sobre los “puntos” que en un principio yo pensaba o al final no lo tengan.

3. Datos¹

A continuación, vamos a exponer los datos que he utilizado para realizar el modelo planteado. Como se puede apreciar son datos de los diferentes equipos en diferentes temporadas. Son un total de 12 equipos, ya que solo he cogido datos de los equipos que han permanecido en primera durante las 3 temporadas estudiadas.

Team	Year	Pts	GF	GC	Chmps	EL	Sueldo ²	Presp ⁱ	Market ⁱⁱ
Villarreal	2017	67	56	33	0	1	76,9	103	56
Villarreal	2016	64	44	35	1	0	61,5	80,3	48
Villarreal	2018	61	57	50	0	1	87,922	117	47
Valencia	2018	73	54	38	1	0	113,301	91,9	55
Valencia	2017	46	56	65	0	0	129,7	96	36
Valencia	2016	44	46	48	0	0	122,8	129,7	144
Sevilla	2017	72	69	49	1	0	123,8	135	80
Sevilla	2018	58	49	58	0	1	159,636	212	75
Sevilla	2016	52	51	50	0	1	105,1	109	45
Real Sociedad	2017	64	59	53	0	1	56,7	76,3	18
Real Sociedad	2018	49	66	59	0	0	64,835	100,6	22
Real Sociedad	2016	48	45	48	0	0	56,6	68	25
Málaga	2016	48	38	35	0	0	28,7	51,9	9
Málaga	2017	46	49	55	0	0	43,1	69,12	13
Málaga	2018	20	24	61	0	0	53,559	63,5	11
Madrid	2017	93	106	41	1	0	419,3	631,3	30
Madrid	2016	90	110	34	1	0	431,3	515	93
Madrid	2018	76	94	44	1	0	499,698	690,3	41
Las Palmas	2016	44	45	53	0	0	18,5	32	4
Las Palmas	2017	39	53	74	0	0	24,6	55,35	3
Las Palmas	2018	22	24	74	0	0	29,448	75,418	0
Espanyol	2017	56	49	50	0	0	47,4	78,1	19
Espanyol	2018	49	36	42	0	0	60,695	74,6	4
Espanyol	2016	43	40	74	0	0	30,6	62	3
Eibar	2017	54	56	51	0	0	23,5	43	9
Eibar	2018	51	44	50	0	0	33,212	45,3	7
Eibar	2016	43	49	61	0	0	19,1	32,4	4

¹ (BDFUTBOL, 2016)

² (Marca, 2017)



Deportivo	2016	42	45	61	0	0	17,8	55,995	8
Deportivo	2017	36	43	61	0	0	24,7	50,04	6
Deportivo	2018	29	38	76	0	0	40,743	61,56	10
Celta	2016	60	51	59	0	1	22,6	43	17
Celta	2018	49	59	60	0	0	40,936	68	26
Celta	2017	45	53	69	0	0	39,3	65	9
Betis	2018	60	60	61	0	1	65,459	87,67	39
Betis	2016	45	34	52	0	0	39,2	51,05	9
Betis	2017	39	41	64	0	0	44,6	66,51	21
Barcelona	2018	93	99	29	1	0	489,231	897	360
Barcelona	2016	91	112	29	1	0	421,9	679	51
Barcelona	2017	90	116	37	1	0	390,7	695	125
Atl. Bilbao	2017	63	53	43	0	1	61,4	113,9	0
Atl. Bilbao	2016	62	58	45	0	1	53,7	112,9	8
Atl. Bilbao	2018	43	41	49	0	0	68,576	116	32
At. Madrid	2016	88	63	18	1	0	159,6	232	144
At. Madrid	2018	79	58	22	1	0	237,767	347,2	102
At. Madrid	2017	78	70	27	1	0	182,8	266,1	79

4. Análisis de datos de panel.

a. Efectos fijos.

Modelo 6: Efectos fijos, utilizando 45 observaciones
 Se han incluido 15 unidades de sección cruzada
 Largura de la serie temporal = 3
 Variable dependiente: Pts

	Coefficiente	Desv. típica	Estadístico t	valor p	
const	45.5006	7.19829	6.321	1.89e-06	***
GF	0.527489	0.0782823	6.738	7.13e-07	***
GC	-0.485616	0.0807381	-6.015	3.92e-06	***
Chmps	16.0784	3.08932	5.205	2.81e-05	***
EL	11.1538	2.13858	5.216	2.74e-05	***
Sueldo	-0.0949482	0.0516534	-1.838	0.0790	*
Presp	0.0570651	0.0292468	1.951	0.0633	*
Market	0.0102725	0.0176635	0.5816	0.5665	
Media de la vble. dep.	56.97778	D.T. de la vble. dep.	18.78163		
Suma de cuad. residuos	290.8933	D.T. de la regresión	3.556337		
R-cuadrado MCVF (LSDV)	0.981258	R-cuadrado 'intra'	0.893860		
F(21, 23) MCVF	57.34257	Valor p (de F)	3.83e-15		
Log-verosimilitud	-105.8438	Criterio de Akaike	255.6877		
Criterio de Schwarz	295.4343	Crit. de Hannan-Quinn	270.5048		
rho	-0.433359	Durbin-Watson	1.656035		

Contraste conjunto de los regresores (excepto la constante) -
 Estadístico de contraste: $F(7, 23) = 27.6708$
 con valor p = $P(F(7, 23) > 27.6708) = 9.30466e-010$

Contraste de diferentes interceptos por grupos -
 Hipótesis nula: [Los grupos tienen un intercepto común]
 Estadístico de contraste: $F(14, 23) = 2.22486$
 con valor p = $P(F(14, 23) > 2.22486) = 0.0429977$



Podemos apreciar que casi todas las variables de nuestro modelo son significativas a excepción del mercado de fichajes (market). Aunque también encontramos diferencias dentro de las significaciones de las variables. Mientras que la Const, GF, GC, Chmps y EL son bastante significativas, el Sueldo y Presp lo hacen en menor medida. Además, realizando el test de Chow (interceptos por grupos), al ser este menor de 0,05 ($p = 0.0429977$) Rechazamos H_0 , por que rechazamos la hipótesis de pooled (datos agrupados). Como consecuencia el modelo más adecuado sería el de Efectos Fijos (EF).

b. Efectos aleatorios

Modelo 7: Efectos aleatorios (MCG), utilizando 45 observaciones
 Se han incluido 15 unidades de sección cruzada
 Largura de la serie temporal = 3
 Variable dependiente: Pts

	Coeficiente	Desv. típica	z	valor p
const	47.1139	4.79320	9.829	8.42e-023 ***
GF	0.495000	0.0617714	8.013	1.12e-015 ***
GC	-0.441132	0.0666866	-6.615	3.72e-011 ***
Chmps	15.6681	2.75610	5.685	1.31e-08 ***
EL	10.8286	1.80859	5.987	2.13e-09 ***
Sueldo	-0.0490080	0.0289675	-1.692	0.0907 *
Presp	0.0135552	0.0184567	0.7344	0.4627
Market	0.0181987	0.0147617	1.233	0.2176

Media de la vble. dep.	56.97778	D.T. de la vble. dep.	18.78163
Suma de cuad. residuos	714.7095	D.T. de la regresión	4.336836
Log-verosimilitud	-126.0695	Criterio de Akaike	268.1391
Criterio de Schwarz	282.5924	Crit. de Hannan-Quinn	273.5271
rho	-0.433359	Durbin-Watson	1.656035

Varianza 'entre' (between) = 8.64706
 Varianza 'dentro' (Within) = 12.6475
 theta usado para quasi-demeaning (cuasi-centrado de los datos) = 0.427504
 corr(y, yhat)^2 = 0.954095

Contraste conjunto de los regresores (excepto la constante) -
 Estadístico de contraste asintótico: Chi-cuadrado(7) = 527.507
 con valor p = 9.74472e-110

Contraste de Breusch-Pagan -
 Hipótesis nula: [Varianza del error específico a la unidad = 0]
 Estadístico de contraste asintótico: Chi-cuadrado(1) = 1.78826
 con valor p = 0.181138

Contraste de Hausman -
 Hipótesis nula: [Los estimadores de MCG son consistentes]
 Estadístico de contraste asintótico: Chi-cuadrado(7) = 8.86611
 con valor p = 0.262412

Para el caso de los Efectos Aleatorios (EA) tienen la misma cantidad de variables significativas que para el caso de EF a excepción de los Presp, que para el caso de EA ha dejado de serlo.

El contraste de Breusch-Pagan nos da como resultado una $p = 0.181138$, por lo que No se Rechaza H_0 , provocando que en este caso el modelo más adecuado sea Pooled.

Las conclusiones del test de Chow y el contraste de Breusch-Pagan nos llevan a contradicciones, ya que o ambas deben Rechaza H_0 (modelo pooled) o ambas lo tienen que No Rechazar. Y en el caso de Rechazar en ambas ocasiones usar el Contraste de Hausman para elegir entre EF o EA. Por esta contradicción podemos sospechar que hay algún tipo de relación entre los datos (conectividad) que la provoca.



5. Multicolinealidad.

a. Matriz de correlaciones

Coefficientes de correlación, usando las observaciones 1:1 - 15:3
 Valor crítico al 5% (a dos colas) = 0.2940 para n = 45

Pts	GF	GC	Chmps	EL	
1.0000	0.8524	-0.7947	0.8206	0.1023	Pts
	1.0000	-0.5053	0.7172	-0.0473	GF
		1.0000	-0.6988	0.0102	GC
			1.0000	-0.3015	Chmps
				1.0000	EL
Sueldo	Presp	Market			
0.7826	0.7671	0.5937	Pts		
0.8856	0.8789	0.5092	GF		
-0.5536	-0.5388	-0.5419	GC		
0.7795	0.7472	0.5704	Chmps		
-0.1531	-0.1550	-0.0773	EL		
1.0000	0.9774	0.6289	Sueldo		
	1.0000	0.6680	Presp		
		1.0000	Market		

La matriz de coeficientes de correlaciones nos indica que es posible que haya multicolinealidad entre las variables del Presupuesto y Sueldo, algo que tiene bastante sentido, porque ante un mayor presupuesto en principio podrás gastar una mayor cantidad de dinero, como por ejemplo en sueldos. Pero también podemos apreciar que los GF tienen un coeficiente de correlación muy alto con las otras dos variables mencionadas anteriormente.

Aunque tengamos indicios de que puedes existir multicolinealidad entre las variables todavía debemos de realizar algunas medidas para estar seguro de ello:

b. Factor de inflación de la Varianza (FIV)

Factores de inflación de varianza (VIF)

Mínimo valor posible = 1.0

Valores mayores que 10.0 pueden indicar un problema de colinealidad

GF	5.433
GC	2.278
Chmps	4.434
EL	1.337
Sueldo	27.721
Presp	27.711
Market	2.194

$VIF(j) = 1/(1 - R(j)^2)$, donde $R(j)$ es el coeficiente de correlación múltiple entre la variable j y las demás variables independientes

El FIV nos muestra que existe multicolinealidad elevada entre las variables Sueldo y Presp. Ahora debemos de tratar de darle una solución.



c. MCO.

Modelo 1: MCO combinados, utilizando 45 observaciones
 Se han incluido 15 unidades de sección cruzada
 Largura de la serie temporal = 3
 Variable dependiente: Pts

	Coefficiente	Desv. típica	Estadístico t	valor p	
const	46.4334	4.83916	9.595	1.40e-011	***
GF	0.510512	0.0684778	7.455	7.08e-09	***
GC	-0.437225	0.0686126	-6.372	1.97e-07	***
Chmps	14.8978	3.05390	4.878	2.05e-05	***
EL	9.87503	1.85390	5.327	5.12e-06	***
Sueldo	-0.0239026	0.0248067	-0.9636	0.3415	
Presp	-0.00342597	0.0158457	-0.2162	0.8300	
Market	0.0179527	0.0156551	1.147	0.2588	
Media de la vble. dep.	56.97778	D.T. de la vble. dep.	18.78163		
Suma de cuad. residuos	684.8394	D.T. de la regresión	4.302229		
R-cuadrado	0.955877	R-cuadrado corregido	0.947529		
F(7, 37)	114.5080	Valor p (de F)	4.21e-23		
Log-verosimilitud	-125.1090	Criterio de Akaike	266.2180		
Criterio de Schwarz	280.6713	Crit. de Hannan-Quinn	271.6060		
rho	0.273218	Durbin-Watson	0.912487		

Sin considerar la constante, el valor p más alto fue el de la variable 9 (Presp)

Contraste de omisión de variables -

Hipótesis nula: [Los parámetros son cero para las variables]

Sueldo

Estadístico de contraste: $F(1, 37) = 0.928437$

con valor p = $P(F(1, 37) > 0.928437) = 0.341525$

Contraste de omisión de variables -

Hipótesis nula: [Los parámetros son cero para las variables]

Presp

Estadístico de contraste: $F(1, 37) = 0.0467461$

con valor p = $P(F(1, 37) > 0.0467461) = 0.830013$

d. Omitimos variable con presencia de multicolinealidad

Hacemos la prueba de la Variable Omitida Condicional (VOC) $\rightarrow (-0,9636)^2 = 0,9285 < 1$ se omite la variable

Contraste sobre el Modelo 1:

Hipótesis nula: el parámetro de regresión es cero para Sueldo

Estadístico de contraste: $F(1, 37) = 0.928437$, valor p 0.341525

Al omitir variables mejoraron 3 de 3 criterios de información.

Modelo 2: MCO combinados, utilizando 45 observaciones

Se han incluido 15 unidades de sección cruzada

Largura de la serie temporal = 3

Variable dependiente: Pts

	Coefficiente	Desv. típica	Estadístico t	valor p	
const	46.6496	4.82940	9.659	8.86e-012	***
GF	0.501352	0.0677509	7.400	7.12e-09	***
GC	-0.438250	0.0685397	-6.394	1.64e-07	***
Chmps	14.0038	2.90679	4.818	2.34e-05	***
EL	9.73219	1.84622	5.271	5.67e-06	***
Presp	-0.0167248	0.00777686	-2.151	0.0379	**
Market	0.0206154	0.0153947	1.339	0.1885	
Media de la vble. dep.	56.97778	D.T. de la vble. dep.	18.78163		
Suma de cuad. residuos	702.0240	D.T. de la regresión	4.298176		
R-cuadrado	0.954769	R-cuadrado corregido	0.947628		
F(6, 38)	133.6897	Valor p (de F)	5.46e-24		
Log-verosimilitud	-125.6666	Criterio de Akaike	265.3332		
Criterio de Schwarz	277.9798	Crit. de Hannan-Quinn	270.0477		
rho	0.193462	Durbin-Watson	0.984251		

Sin considerar la constante, el valor p más alto fue el de la variable 10 (Market)



e. Diagnóstico de panel.

Diagnósticos: utilizando 15 unidades de sección cruzada

Estimador de efectos fijos

permite interceptos distintos para las unidades de sección cruzada

	Coefficiente	Desv. típica	Estadístico t	valor p
const	40.3128	6.94238	5.807	5.49e-06 ***
GF	0.539859	0.0817666	6.602	7.90e-07 ***
GC	-0.481959	0.0846193	-5.696	7.23e-06 ***
Chmps	16.4608	3.23145	5.094	3.28e-05 ***
EL	11.2580	2.24127	5.023	3.92e-05 ***
Presp	0.0156349	0.0195401	0.8001	0.4315
Market	0.0142295	0.0183802	0.7742	0.4464

Varianza de los residuos: $333.628 / (45 - 21) = 13.9012$

Significatividad conjunta de las medias de los diferentes grupos:

$F(14, 24) = 1.89293$ con valor p 0.0819059

(Un valor p bajo es una indicación en contra de la hipótesis nula de que el modelo de MCO combinados es el adecuado, en favor de la alternativa de efectos fijos.)

Variance estimators:

between = 6.69889

within = 13.9012

theta used for quasi-demeaning = 0.36056

Estimador de efectos aleatorios

permite un componente específico de la unidad en el término de error

	Coefficiente	Desv. típica	Estadístico t	valor p
const	46.7215	4.88200	9.570	1.14e-011 ***
GF	0.482674	0.0633435	7.620	3.62e-09 ***
GC	-0.432672	0.0681486	-6.349	1.89e-07 ***
Chmps	14.7215	2.80009	5.257	5.92e-06 ***
EL	10.5823	1.85016	5.720	1.38e-06 ***
Presp	-0.0152073	0.00771965	-1.970	0.0562 *
Market	0.0226358	0.0149164	1.518	0.1374

Estadístico de contraste de Breusch-Pagan:

$LM = 1.07766$ con valor p = $\text{prob}(\text{chi-cuadrado}(1) > 1.07766) = 0.299222$

(Un valor p bajo es una indicación en contra de la hipótesis nula de que el modelo de MCO combinados es el adecuado, en favor de la alternativa de efectos aleatorios.)

Estadístico de contraste de Hausman:

$H = 7.20464$ con valor p = $\text{prob}(\text{Chi-cuadrado}(6) > 7.20464) = 0.302336$

(Un valor p bajo es una indicación en contra de la hipótesis nula de que el modelo de efectos aleatorios es consistente, en favor del modelo de efectos fijos.)

Tras omitir la variable que generaba multicolinealidad y realizar un diagnóstico de panel de nuevo, podemos apreciar que ahora No Rechazamos H_0 en ninguna de las ocasiones, por lo que el modelo que mejor se ajusta a estos datos es el pooled. Esto provoca que, aunque tengamos datos de panel podamos estimar los parámetros por MCO.



6. Estimación del modelo

a. Mínimo Cuadrática Ordinaria

Modelo 4: MCO combinados, utilizando 45 observaciones
 Se han incluido 15 unidades de sección cruzada
 Largura de la serie temporal = 3
 Variable dependiente: Pts

	Coeficiente	Desv. típica	Estadístico t	valor p
const	46.6496	4.82940	9.659	8.86e-012 ***
GF	0.501352	0.0677509	7.400	7.12e-09 ***
GC	-0.438250	0.0685397	-6.394	1.64e-07 ***
Chmps	14.0038	2.90679	4.818	2.34e-05 ***
EL	9.73219	1.84622	5.271	5.67e-06 ***
Presp	-0.0167248	0.00777686	-2.151	0.0379 **
Market	0.0206154	0.0153947	1.339	0.1885
Media de la vble. dep.	56.97778	D.T. de la vble. dep.	18.78163	
Suma de cuad. residuos	702.0240	D.T. de la regresión	4.298176	
R-cuadrado	0.954769	R-cuadrado corregido	0.947628	
F(6, 38)	133.6897	Valor p (de F)	5.46e-24	
Log-verosimilitud	-125.6666	Criterio de Akaike	265.3332	
Criterio de Schwarz	277.9798	Crit. de Hannan-Quinn	270.0477	
rho	0.193462	Durbin-Watson	0.984251	

Sin considerar la constante, el valor p más alto fue el de la variable 10 (Market)

b. Modelo estimado

$$\text{Puntos} = 46,6496 + 0,501352GF - 0,438250GC + 14,0038Champs + 9,73219EL - 0,0167248Presp + 0,0206154Market$$

c. Interpretación de las variables.

Constante: Los puntos mínimos de los equipos de fútbol estudiados son 46,6496 manteniendo el resto de las variables constantes.

Goles a favor (GF): Por cada gol que consigan los equipos podrán conseguir 0,501352 puntos más, manteniendo el resto de las variables constantes.

Goles en contra (GC): Por cada gol en contra que reciban los equipos sus puntos se reducirán en 0,438250 manteniendo el resto de las variables constantes.

Champions (Chmps): Los equipos que juegan la Champions (1) conseguirán 14,0038 puntos más, que los que no la juegan (0), manteniéndose el resto de las variables constantes.

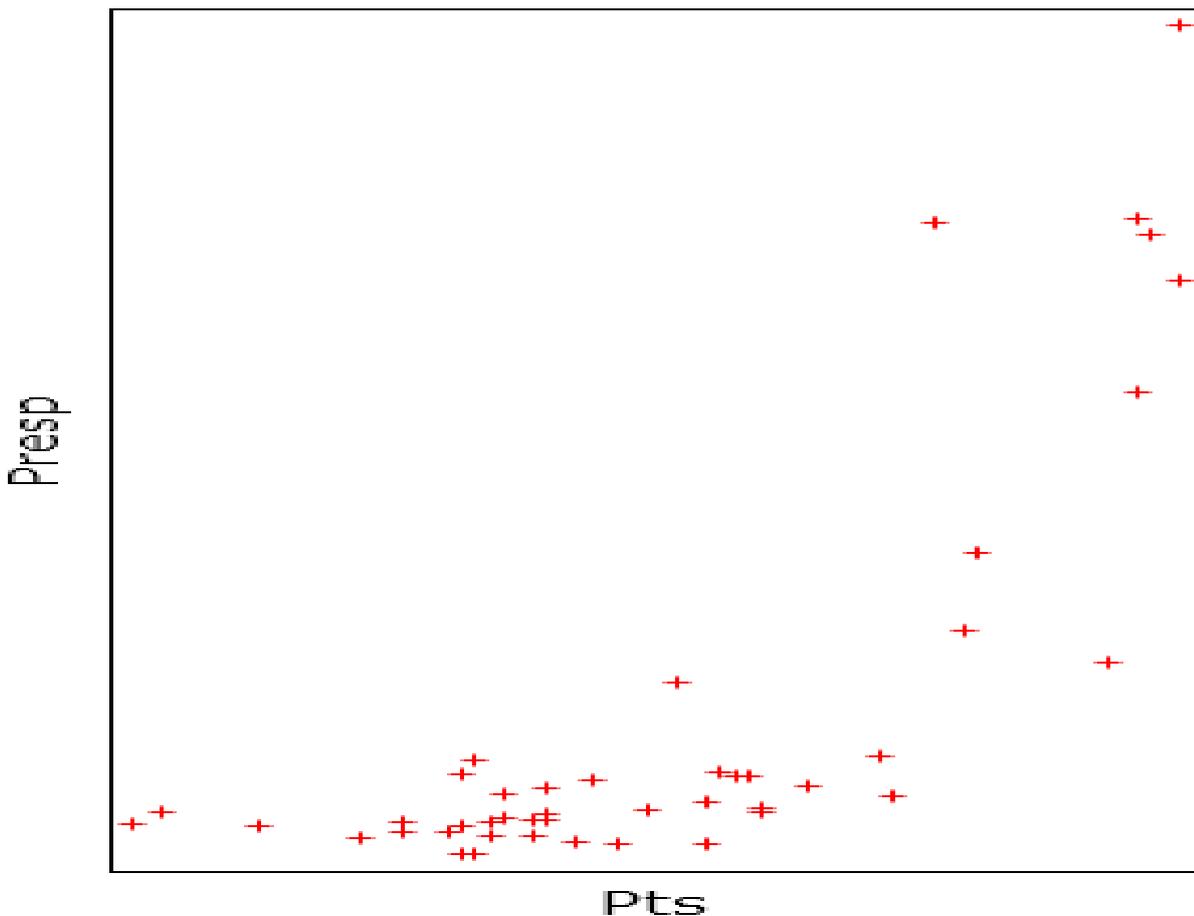
Europa League (EL): Los equipos que juegan la Europa League (1) consiguen 9,73219 puntos más, que los que no la juegan (0), manteniéndose el resto de las variables constantes.

Presupuesto (Presp): Por cada millón de euros invertido los puntos de ese equipo se reducirán en 0,0167248 manteniéndose el resto de las variables constantes.



Market: Por cada millón de euros invertido en fichajes se conseguirán 0,0206154 puntos más, manteniendo el resto las de variables constantes.

La mayoría de las variables de nuestro modelo son significativas a excepción del “Market” que, aunque la hemos interpretado, en principio esta variable no es significativa para interpretar el modelo. Otra cosa que me llama la atención es la negatividad del presupuesto, ya que en principio a mayor cantidad de dinero se espera que se consigan más puntos, pero no menos. Esto puede ser debido a que equipos con un gran presupuesto no alcancen una puntuación muy elevada en ocasiones y equipos con menor presupuesto consigan una mayor cantidad de puntos.



Aunque se puede apreciar que los equipos con un mayor presupuesto son los equipos que mayores puntos consiguen, podemos ver que los que tienen un presupuesto mediano consiguen hacer casi la misma cantidad de puntos (pero con menos recursos) llegando incluso a igualar o superar en algunas ocasiones. Además, si nos fijamos en la parte izquierda de la gráfica en los equipos con un menor presupuesto, podemos observar que equipos con un mayor presupuesto hacen los mismos puntos o menos en muchas ocasiones que los equipos que tienen uno menor (dentro de este intervalo para que el presupuesto no es tan importante). Por lo que podemos deducir que el presupuesto es influyente a partir de una determinada cantidad, pero que hasta que no llega a un nivel intermedio no marca la diferencia y no le asegura tener más cantidad de puntos.



d. Multicolinealidad

Factores de inflación de varianza (VIF)
 Mínimo valor posible = 1.0
 Valores mayores que 10.0 pueden indicar un problema de colinealidad

GF	5.328
GC	2.278
Chmps	4.025
EL	1.328
Presp	6.687
Market	2.126

$VIF(j) = 1/(1 - R(j)^2)$, donde $R(j)$ es el coeficiente de correlación múltiple entre la variable j y las demás variables independientes

No hay presencia de multicolinealidad elevada.

e. Tabla ANOVA

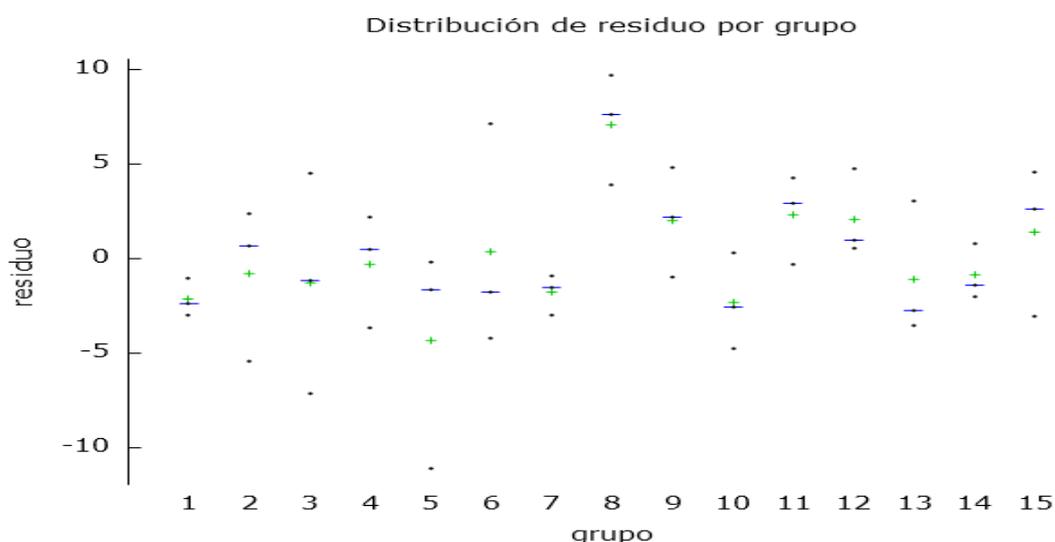
Análisis de Varianza:

	Suma de cuadrados	gl	Media de cuadrados
Regresión	14819	6	2469.83
Residuo	702.024	38	18.4743
Total	15521	44	352.749

$R^2 = 14819 / 15521 = 0.954769$
 $F(6, 38) = 2469.83 / 18.4743 = 133.69$ [valor p 5.46e-024]

Como el valor F-Snedecor es 133,69 y tenemos un valor p de 5,46e-024 al ser menor de 0,05, Rechazamos H_0 , como consecuencia el modelo es significativo conjuntamente, algo que es coherente, porque como hemos comentado anteriormente casi todas las variables son significativas individualmente en el modelo.

f. Heterocedasticidad.



A simple vista no parece plantear un problema de heterocedasticidad.



g. Contraste de White.

```

Contraste de heterocedasticidad de White
MCO, utilizando 45 observaciones
Variable dependiente: uhat^2

```

	Coefficiente	Desv. típica	Estadístico t	valor p
const	33.2704	280.115	0.1188	0.9066
GF	-6.31053	7.69899	-0.8197	0.4221
GC	5.11206	6.25798	0.8169	0.4236
Chmps	56.6011	215.604	0.2625	0.7956
EL	214.387	235.128	0.9118	0.3727
Presp	-0.0790214	0.947013	-0.08344	0.9343
Market	0.689485	3.61848	0.1905	0.8508
sq_GF	0.0336690	0.0550945	0.6111	0.5480
X2_X3	0.0320487	0.0881910	0.3634	0.7201
X2_X4	1.17627	5.84537	0.2012	0.8426
X2_X5	-1.47576	3.23380	-0.4564	0.6531
X2_X6	0.00165293	0.0139719	0.1183	0.9070
X2_X7	-0.0218585	0.0492246	-0.4441	0.6618
sq_GC	-0.0546582	0.0503583	-1.085	0.2907
X3_X4	-3.26092	6.95277	-0.4690	0.6441
X3_X5	-2.01245	2.97267	-0.6770	0.5062
X3_X6	0.00307073	0.0126756	0.2423	0.8110
X3_X7	0.000756771	0.0600621	0.01260	0.9901
X4_X6	0.0588497	0.662494	0.08883	0.9301
X4_X7	0.706086	2.05909	0.3429	0.7352
X5_X6	-0.348074	0.608441	-0.5721	0.5736
X5_X7	0.684716	0.987200	0.6936	0.4959
sq_Presp	-0.000380621	0.00128502	-0.2962	0.7701
X6_X7	0.000216895	0.00453434	0.04783	0.9623
sq_Market	0.00226982	0.00663198	0.3423	0.7357

R-cuadrado = 0.373281

Estadístico de contraste: $TR^2 = 16.797651$,
con valor p = $P(\text{Chi-cuadrado}(24) > 16.797651) = 0.857163$

Como el valor obtenido de p, es mayor de 0,05 podemos decir que, nuestro modelo es homocedástico.

h. Autocorrelación

No tenemos suficientes series temporales para poder estudiar la autocorrelación, ya que solo contamos con tres periodos.

7. Conclusión

Podemos concluir que la mayoría de las variables de nuestro modelo son significativas. Además, no tenemos problemas ni de multicolinealidad, ni de heterocedasticidad, por lo que podríamos decir de que no hay evidencia de que nuestro modelo no sea válido.



8. Bibliografía

BDFUTBOL. (14 de 05 de 2016). *BDFUTBOL*. Obtenido de <https://www.bdfutbol.com/es/t/t2015-16.html>

Marca. (21 de 09 de 2017). *Marca*. Obtenido de Marca: <https://www.marca.com/futbol/primera-division/2017/09/20/59c2d4c922601d4a058b460e.html>

Transfermarkt. (s.f.). *Transfermarkt*. Obtenido de Transfermarkt: <https://www.transfermarkt.es/primera-division/transferbilanz/wettbewerb/ES1>

Trullols, J. (09 de 01 de 2018). *La Jugada Financiera*. Obtenido de La Jugada Financiera: <http://lajugadafinanciera.com/presupuestos-la-liga-2017-2018/>

ⁱ (Trullols, 2018)

ⁱⁱ (Transfermarkt, s.f.)



LaLiga

