

Investigación

Estadística multivariante aplicada al análisis y predicción de partidos de fútbol en las principales ligas europeas

Multivariate statistical techniques applied to the analysis and prediction of football matches in the main European leagues

Pedro J. Chocano Feito y Elena Castilla González

Revista de Investigación



Volumen XI, Número 2, pp. 021-030, ISSN 2174-0410
Recepción: 06 May'21; Aceptación: 16 Jun'21

1 de octubre de 2021

Resumen

El propósito de este estudio es analizar las estadísticas de juego en las principales ligas europeas y ver qué factores son más determinantes a la hora de predecir el resultado de un partido. Para ello usaremos técnicas de estadística multivariante incluyendo análisis de componentes principales y regresión logística. Las dos primeras componentes principales explican alrededor del 70 % de precisión obtenida cuando se predicen victorias fuera de casa tomando como variables predictivas las propias componentes. Este estudio también demuestra que en la liga inglesa los partidos son menos equilibrados.

Palabras Clave: componentes principales, desarrollo del juego, regresión logística.

Abstract

The purpose of this study is to analyse main game-related statistics differences between the main European leagues and which factors are more determinant when predicting a match score, by means of multivariate statistical techniques, including principal component analysis and logistic regression. The first two principal components explain around the 70 % of variance, and over a 70 % of accuracy is obtained when predicting away-team wins, with these two principal components as predictive variables. This study also shows that in English Premier League, games are less equilibrated.

Keywords: notational analysis, match performance, principal components, logistic regression.

1. Introducción

El interés global asociado al fútbol, junto con el comienzo de la era del big-data, ha incrementado el profesionalismo del fútbol moderno. En cada partido de las ligas más importantes se recoge una gran cantidad de datos asociados al juego. El principal objetivo es detectar qué indicadores de rendimiento son determinantes en el resultado final. De hecho, este interés ya se manifestó a mediados de los años 70 [14] pero ha explotado en los últimos años. Por ejemplo, véase [4, 1, 3, 5] y sus referencias. Si bien la capacidad económica es, sin duda, uno de los factores principales asociados a un equipo exitoso (véase [9] o [13] para la liga alemana y española respectivamente), invertir en el análisis del juego e identificar los factores más relevantes puede ser determinante a la hora de marcar la diferencia.

El Análisis de Componentes Principales (ACP) es una de las técnicas estadísticas multivariantes más importantes, ya que permite reducir la dimensionalidad y facilita la interpretación de los datos. Fue usado en [2] para proporcionar información táctica sobre un equipo, cuantificando la región de campo en la que varios jugadores actuaban. En [10] también se usó el ACP para analizar los partidos de la fase de grupos del mundial de 2006.

Como se comenta en [8] y [9], se necesita más análisis predictivo para entender mejor los factores que determinan el éxito en el fútbol. El modelo de regresión logística se ha usado en [16] para analizar las estadísticas de juego más determinantes para predecir los resultados de los partidos en la Liga Canadiense de Fútbol. También fue usado en [13] para predecir los resultados en la Primera División española. Una vez que se ajustó el modelo, se mantuvieron las variables significativas atendiendo a sus p-valores. La interpretación de los resultados se hizo, en ambos casos, en base a la significación y símbolo de los estimadores ajustados. Aquí, introducimos el concepto de *odds ratio* (razón de monomios), que cuantifica la influencia de una variable en la variable respuesta.

Un aspecto importante en los modelos de regresión es la determinación del número de variables explicativas. Un modelo con demasiadas variables explicativas podría incluir covariables con una alta intercorrelación. Así, podríamos no darnos cuenta de la importancia que tiene una variable en la respuesta, ya que podría estar “escondida” en el modelo por otras variables correladas. Por otro lado, cuando tratamos con conjuntos de datos de tamaño moderado, no es recomendable usar demasiadas covariables ya que pueden sobre-ajustar el modelo. En este sentido, la conocida *one in ten rule* ([6, 11]) establece que se debe usar un máximo de una covariable por cada 10 eventos. Uno de los métodos más usuales para reducir el número de covariables es el de ajustar el modelo completo y determinar mediante un test o prueba de significación, una por una, qué variables han de ser eliminadas del modelo. Este proceso depende del nivel de significación escogido y puede necesitar muchos pasos. Además, una vez que una variable es eliminada del modelo, toda su información (que no es necesariamente explicada por las otras variables) se pierde. En este artículo, combinamos el ACP con la regresión logística, reduciendo considerablemente el número de covariables en el modelo logístico. En este estudio, usamos ocho variables como predictores, pero este método podría extenderse a más variables. Nótese que, mientras que la mayoría de la literatura citada centra su estudio en una competición en particular, extendemos nuestro estudio a las cuatro ligas europeas más importantes, según los coeficientes UEFA ([15]).

2. Métodos

2.1. Descripción de los Datos y uso de Software

Los datos usados en este estudio han sido recogidos de todos los partidos desde la temporada 2016/2017 a la temporada 2018/2019 en las principales ligas europeas: LaLiga (España),

Premier League (Inglaterra), Serie A (Italia) y Bundesliga (Alemania). Esto es, 380 partidos por temporada en las tres primeras ligas y 307 partidos por temporada en la Bundesliga. Mientras que las temporadas 2016/2017 y 2017/2018 se usan para el análisis descriptivo y ajuste de los modelos de predicción, la temporada 2018/2019 se usa para testear estos modelos. Los datos se han obtenido de <http://www.football-data.co.uk/data.php>. Esta fuente de datos ha sido usada también en [13], entre otros.

Para analizar los factores que pueden afectar a la probabilidad de ganar un partido consideramos ocho variables numéricas: HS (Home-team shots) disparos del equipo local, AS (Away-team shots) disparos del equipo visitante, HC (Home-team corners) córneres ocasionados por el equipo local, AF (Away-team fouls) faltas cometidas por el equipo visitante, HY (Home-team yellow cards) tarjetas amarillas mostradas al equipo local, AY (Away-team yellow cards) tarjetas amarillas mostradas al equipo visitante, HR (Home-team red cards) tarjetas rojas mostradas al equipo local y AR (Away-team red cards) tarjetas rojas mostradas al equipo visitante. El número total de disparos a portería y faltas por partido son también usados para un análisis descriptivo. En el modelo binomial distinguiremos entre 0: victorias del equipo visitante y 1: victoria ó empate del equipo local. Esta dicotomía fue considerada también en [13], entre otros.

Para el tratamiento de los datos, hemos usado el Software estadístico R. En concreto, las funciones *glm()* y *prcomp()* para el análisis de regresión logística y el ACP. Para la representación de los datos, usamos los paquetes “ggfortify” y “ggplot2”.

2.2. Análisis Estadístico

2.2.1. Análisis de Componentes Principales

La principal motivación del ACP consiste en reducir el número de variables mientras se intenta preservar la máxima varianza. Geométricamente, se intentan buscar unos nuevos ejes, llamados direcciones principales, que expliquen la mayor varianza posible para representar los datos originales, las nuevas variables obtenidas tras este cambio de representación serán las llamadas componentes principales. Veamos ligeramente los rudimentos básicos por completitud. Supongamos que se dispone de los valores de k variables en n elementos de una población dispuestos en una matriz \mathbf{X} de dimensiones $n \times k$. Denotamos por \mathbf{S} a la matriz de varianzas-covarianzas de las variables,

$$\mathbf{S} = \begin{bmatrix} \text{Var}(X_1) & \text{Cov}(X_1, X_2) & \cdots & \text{Cov}(X_1, X_k) \\ \text{Cov}(X_2, X_1) & \text{Var}(X_2) & \cdots & \text{Cov}(X_2, X_k) \\ \vdots & \vdots & \ddots & \vdots \\ \text{Cov}(X_k, X_1) & \text{Cov}(X_k, X_2) & \cdots & \text{Var}(X_k) \end{bmatrix},$$

la cuál es simétrica y definida positiva. Finalmente, denotamos por a_i al autovector asociado al autovalor λ_i de la matriz \mathbf{S} . Las componentes principales vienen dadas en la matriz \mathbf{Z} cuyas columnas son los valores de las k componentes en los n individuos y $\mathbf{Z} = \mathbf{X}\mathbf{A}$ donde \mathbf{A} es la matriz formada por los autovectores de \mathbf{S} . De manera informal, podemos decir que calcular las componentes principales equivale encontrar unos nuevos ejes de coordenadas que coinciden con los “ejes naturales” de los datos. La proporción de variabilidad total explicada por una componente h viene dada por $\frac{\lambda_h}{\sum_i \lambda_i}$. Finalmente, basta con tomar las $r < k$ componentes que tengan mayor proporción de variabilidad. Para más detalles véase [12, Capítulo 5].

Utilizamos ACP para las temporadas 2016/2017 y 2017/2018 en cada principal liga europea. Los autovectores, que corresponden a los factores de carga de cada variable original para obtener el conjunto de datos modificado, de la primera y segunda componente (PC1 y PC2 respectivamente) se presentan en la Tabla 3. El tamaño y signo de estos factores indica la relación

entre las variables originales y las componentes principales. Los gráficos que incluyen tanto la posición de cada muestra en términos de PC1 y PC2 además de la proyección de las variables originales sobre ellos pueden verse en la Figura 3. Los círculos rojos representan aquellos partidos en los que el equipo visitante ha ganado mientras que los triángulos azules representan una victoria o empate del equipo local.

2.2.2. Regresión Logística

La regresión logística es un método de clasificación donde la variable dependiente y solo puede tomar valor 0 ó 1. Este método es usado para predecir la probabilidad de que ocurra un evento. Dicha probabilidad se obtiene como sigue:

$$\pi(\mathbf{x}) = \frac{\exp(\beta_0 + \beta_1 x_1 + \dots + \beta_k x_k)}{1 + \exp(\beta_0 + \beta_1 x_1 + \dots + \beta_k x_k)},$$

donde $\pi(\mathbf{x}) = P(y = 1|x_1, \dots, x_k)$, β_0 es la ordenada en el origen y β_1, \dots, β_k son los coeficientes asociados a las k variables x_1, \dots, x_k . Estos coeficientes son estimados a partir de la muestra y pueden ser usados para la predicción de la respuesta de nuevas muestras. Así, sea una muestra de tamaño n , donde la observación i -ésima, y_i , tiene asociado el vector de variables explicativas $\mathbf{x}_i = (x_{i1}, \dots, x_{ik})^T$, el estimador de máxima verosimilitud, $\hat{\boldsymbol{\beta}} = (\hat{\beta}_0, \dots, \hat{\beta}_k)$, se obtiene como solución al sistema de ecuaciones

$$\tilde{\mathbf{X}}^T (\mathbf{y} - \boldsymbol{\pi}(\mathbf{x})) = \mathbf{0}_{k+1}, \tag{1}$$

siendo $\mathbf{y} = (y_1, \dots, y_n)^T$, $\boldsymbol{\pi}(\mathbf{x}) = (\pi(x_1), \dots, \pi(x_n))^T$ y $\tilde{\mathbf{X}}$ la matriz de dimensión $n \times (k + 1)$

$$\tilde{\mathbf{X}} = \begin{pmatrix} 1 & x_{11} & \dots & x_{1k} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & \dots & x_{nk} \end{pmatrix}.$$

Ya que excepto casos específicos no hay una solución explícita del sistema (1), se suele hacer uso de métodos iterativos, tales como el de Newton-Raphson, para aproximar la solución. Nótese que no todas las variables son necesarias en el modelo. Estas variables, que no aportan suficiente información, deben ser ignoradas.

Un concepto importante en la teoría de regresión logística es el de odds ratio. Para la variable x_j el odds ratio puede definirse como:

$$\exp(\beta_j), \quad j = 1, \dots, k,$$

y representa cuánto incrementa la probabilidad de un evento por cada unidad que incrementa x_j .

Usamos el modelo de regresión logística para cada una de las principales ligas europeas. Los parámetros estimados de las variables significativas y sus odd ratios son representados en la Tabla 4. Se plantean modelos logísticos mediante el uso de las dos componentes principales. Estos modelos son usados para predecir resultados en la temporada 2018/2019. El ratio de éxitos en las predicciones de las victorias del equipo visitante son recogidos en la Tabla 2.

3. Resultados

Comparamos los disparos y faltas por partido en cada liga (véase Figura 1). El test de Kolmogorov–Smirnov para la distribución normal muestra que la variable disparos por partido se distribuye como una normal solo para la liga italiana, mientras que la variable faltas por partido solamente se distribuye como una normal para la liga alemana. Esto puede explicarse debido al gran número de datos atípicos (outliers) presentados. Parece no haber una diferencia significativa de disparos por partido entre ligas, mientras que las faltas por partido en la liga inglesa parecen inferiores al resto de ligas.

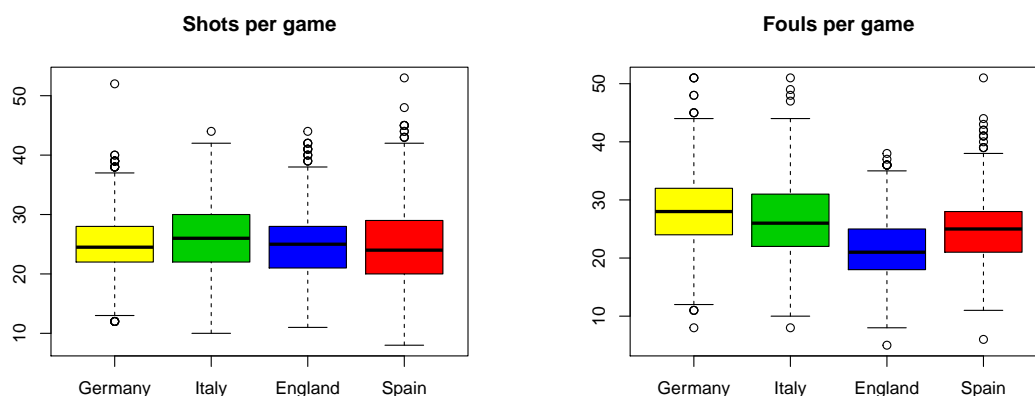


Figura 1. Disparos (*Shots per game*) y faltas (*Fouls per game*) por partido en las diferentes ligas europeas.

Las correlaciones entre HS (disparos equipo local) y AS (disparos equipo visitante) se pueden observar en la Tabla 1. Todas las ligas presentan una correlación negativa, esto es, cuando el número de disparos del equipo visitante incrementa el número de disparos del equipo local disminuye y al revés. Este comportamiento es más evidente para la liga inglesa, pero no lo es para la española. Esto puede observarse en la Figura 2, donde los partidos de liga inglesa y española han sido representados en un diagrama de dispersión. El conjunto de datos está codificado mostrando la variable de victorias del equipo visitante. Discernir entre victorias y derrotas es más evidente para la liga inglesa debido a su alta correlación.

Tabla 1. Correlaciones entre HS y AS.

Germany	Italy	England	Spain
-0.333	-0.449	-0.492	-0.252

El ACP muestra que las ocho variables pueden ser reducidas a dos componentes principales explicando una gran parte de la varianza. La primera componente explica el 41,52 %, 57,4 %, 50,68 % y 44,58 % de la varianza para las ligas española, inglesa, italiana y alemana respectivamente. Las primeras componentes enfrentan los disparos del equipo local y visitante respectivamente. Las segundas componentes explican el 26,22 %, 17,55 %, 23,92 % y 24,33 % de la varianza en las respectivas ligas. En esta segunda componente observamos más diferencias entre las ligas. En las ligas alemana, italiana y española, las faltas del equipo visitante presentan los factores (o *loadings*) de valor más alto. Las faltas del equipo visitante se enfrentan a los disparos de ambos equipos, los cuáles presentan de nuevo el factor más alto en la segunda componente de la liga inglesa.

Los modelos de regresión logística ajustados para las diferentes ligas muestran las variables más significativas a la hora de predecir las victorias del equipo visitante. Observamos que

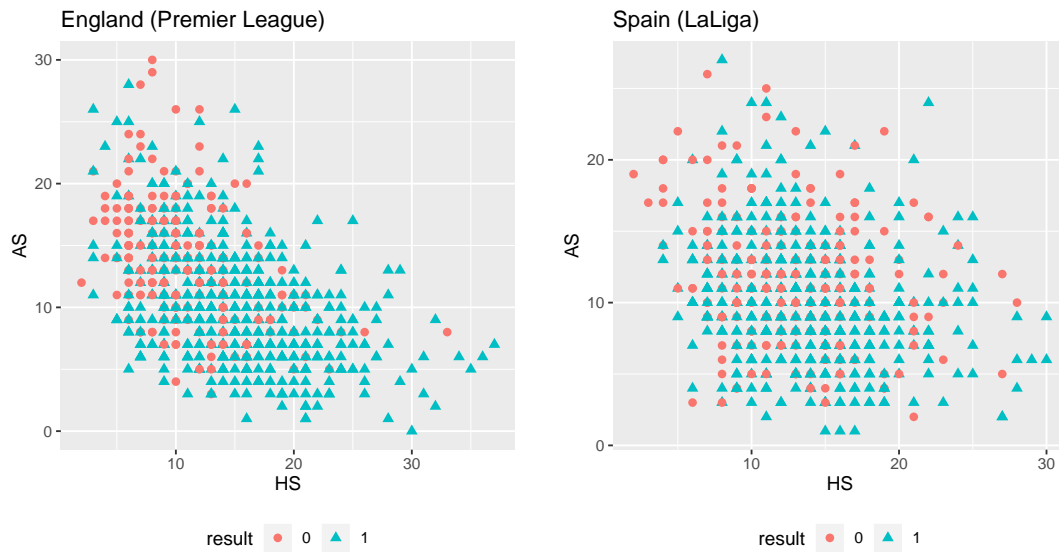


Figura 2. HS contra AS, liga inglesa y española. Aquí, 1 representa victoria o empate del equipo local y 0 victoria del equipo visitante.

los coeficientes estimados difieren entre ligas, aunque los signos se mantienen igual. Las faltas visitantes no son significativas en ninguna de las ligas, mientras que las tarjetas amarillas visitantes y las rojas locales son sólo significativas en las ligas española e italiana respectivamente. El odds ratio de las tarjetas rojas locales en la liga italiana muestra que una tarjeta roja multiplica la probabilidad de que el equipo local puntúe por 0,38, es decir, esta probabilidad decrece aproximadamente un 62 %. De la misma manera, una tarjeta roja del equipo visitante en la liga alemana incrementa la probabilidad de puntuar del equipo local un 280 %. Curiosamente, más córners locales está asociado con menos probabilidad de victoria local. Nótese que, en la liga inglesa, las únicas variables suficientemente significativas en el resultado son HS y AS, lo que concuerda con las Figuras 2 y 3.

Tabla 2. Proporción de ajustes correctos.

	Logit		PCA+Logit			
	par.	training	test	par.	training	test
Germany	6	0.7418	0.6797	2	0.7467	0.7059
Italy	6	0.7013	0.7158	2	0.7013	0.7289
England	2	0.7474	0.6974	2	0.7474	0.6789
Spain	7	0.7092	0.7289	2	0.7026	0.7263

En la Tabla 2 vemos cómo el índice de acierto de clasificación de los modelos está en torno a un 70 % tanto para los conjuntos de entrenamiento y de test. En particular, el uso de ACP para reducir el número de variables en la regresión logística incrementa el acierto en las ligas italiana y alemana, y sólo resulta innecesario en la liga inglesa, donde obtenemos resultados ligeramente mejores usando un modelo de regresión logística con sólo dos variables.

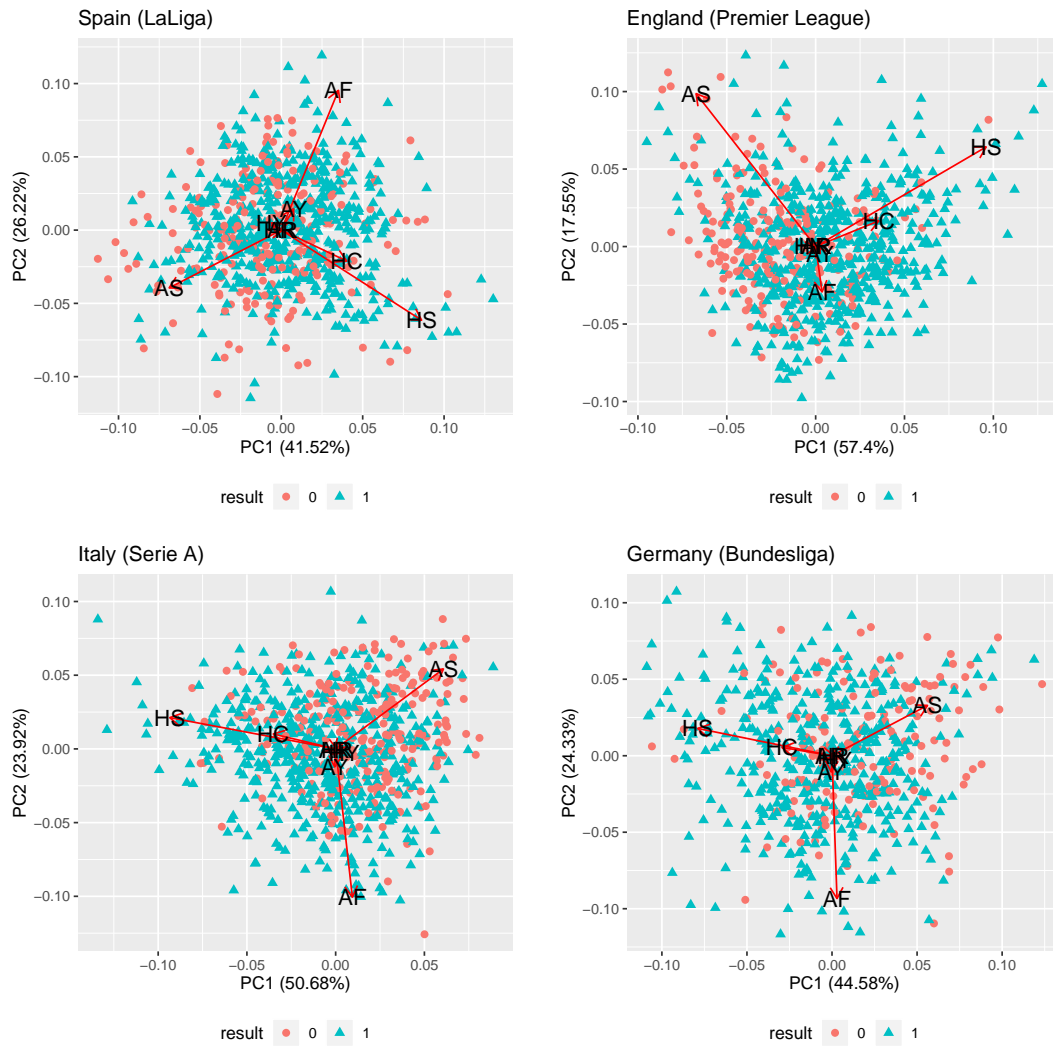


Figura 3. ACP en las principales ligas europeas. Aquí, 1 representa victoria o empate del equipo local y 0 victoria del equipo visitante.

4. Análisis de los resultados

El propósito principal de este estudio es el determinar los factores de éxito en las principales ligas europeas. La interpretación y reducción de las variables explicativas es crucial para identificar los aspectos técnicos más importantes para predecir un resultado. Hemos demostrado como el ACP y la regresión logística son técnicas útiles para clasificar y predecir resultados exitosos, así como para determinar qué variables específicas son más influyentes para este propósito. En particular, la combinación de ambos métodos, usando sólo dos componentes principales en el modelo de regresión logística, nos lleva a predicciones robustas.

Un análisis descriptivo previo nos muestra que el comportamiento general de las variables en las cuatro ligas es muy similar, pero algunas diferencias particulares se observan en las faltas y disparos por partido. Cuando enfrentamos los disparos del equipo local y visitante, se observa una correlación mayor en la liga inglesa, acompañada con una clasificación más clara de los resultados. Esto es, los partidos de la liga inglesa son muy poco igualados, y el equipo que tiene bajo control el partido (disparos favorables) es generalmente el que obtiene un mejor resultado.

Tabla 3. Coeficientes de cada variable para las primeras y segundas componentes principales.

	Germany		Italy		England		Spain	
	PC1	PC2	PC1	PC2	PC1	PC2	PC1	PC2
HS	-0.7792	0.1762	-0.7963	0.1807	0.7827	0.5243	0.7010	-0.5000
AS	0.5539	0.3246	0.5175	0.4621	-0.5487	0.8050	-0.5603	-0.3206
HC	-0.2904	0.0535	-0.2996	0.0861	0.2901	0.1355	0.3267	-0.1717
AF	0.0289	-0.9214	0.0804	-0.8572	0.0296	-0.2391	0.2856	0.7759
HY	0.0268	-0.0272	0.0417	-0.0252	-0.0275	0.0045	-0.0486	0.0405
AY	-0.0040	-0.1051	-0.0046	-0.1045	0.0206	-0.0376	0.0617	0.1184
HR	0.0094	-0.0023	0.0074	-0.0037	-0.0028	0.0048	-0.0073	-0.0018
AR	-0.0078	-0.0030	-0.0059	-0.0075	0.0044	0.0008	0.0076	0.0014

Tabla 4. Resultados de la estimación logística y odds ratios en los modelos restringidos.

	Germany		Italy		England		Spain	
	Est.	Odds ratio	Est.	Odds ratio	Est.	Odds ratio	Est.	Odds ratio
Int.	2.6248	-	1.6259	-	-	-	1.7970	-
HS	0.0620	1.0639	0.0732	1.0759	0.1132	1.1471	0.0539	1.0554
AS	-0.1409	0.8686	-0.1133	0.8929	-0.0949	0.9308	-0.0878	0.9160
HC	-0.1026	0.9025	-0.0859	0.9177	-	-	-0.1282	0.8797
AF	-	-	-	-	-	-	-	-
HY	-0.1913	0.8259	-	-	-	-	-0.1570	0.8547
AY	-	-	-	-	-	-	0.1199	1.1274
HR	-	-	-0.9589	0.3833	-	-	-	-
AR	1.0292	2.7989	0.4179	1.5188	-	-	0.9557	2.6004

Los partidos en la liga española son más igualados, siendo más difícil predecir el resultado final. Después de evaluar ocho variables relacionadas con el juego, las dos primeras componentes principales mantienen más del 67 % de la varianza total en la liga española y alemana, y sobre el 74 % en la liga inglesa e italiana. Esto es, con sólo dos variables podemos explicar la mayoría de la varianza (información). La primera componente principal se basa principalmente en los disparos locales y visitantes, lo que lleva a una clasificación más simple en la liga inglesa.

De hecho, los disparos locales y visitantes son tan importantes en la liga inglesa, que cuando ajustamos un modelo de regresión logística, estas son las únicas variables significativas a un nivel de $\alpha = 0,1$. Por otra parte, la liga española necesita un total de 7 parámetros (incluida la ordenada en el origen). Las faltas del equipo visitante es la única variable insignificante en los cuatro modelos.

El porcentaje de ajuste es parecido en las cuatro ligas, con buenos resultados tanto para el modelo de regresión logística clásico como para el combinado con el ACP. La reducción de dimensión en este segundo método es notable excepto en la liga inglesa por los motivos ya comentados. Las buenas predicciones, en torno a un 70 % de aciertos, son comparables a las obtenidas en [10] para la Copa Mundial de 2006 o a las obtenidas mediante métodos frecuentistas en la liga española [13]. Además, mejoran notablemente las obtenidas en [7] también para la liga española, si bien en este artículo se usa un análisis discriminante y se dividen los resultados en tres categorías.

5. Conclusiones

El propósito principal de este estudio era el de analizar las principales diferencias entre las grandes ligas europeas y determinar qué factores son más determinantes al predecir el resultado de un partido usando el análisis de componentes principales y la regresión logística. Estos últimos facilitan la correcta clasificación de la mayor parte de las victorias visitantes, incluso con sólo dos variables, las dos primeras componentes principales. A nivel práctico, este estudio ha mostrado cómo los disparos a favor y en contra son los dos factores más importantes de éxito, especialmente en la Premier League, donde los partidos son menos equilibrados. Esto es, el equipo con más disparos a favor es, con una alta probabilidad, el que obtiene un resultado favorable. Las tarjetas rojas del equipo visitante es un factor especialmente influyente en la Bundesliga alemana y la liga española, donde los partidos son más equilibrados. Curiosamente, los córners a favor del equipo local están asociados a una menor probabilidad de éxito.

Referencias

- [1] ARAYA, J., & LARKIN, P. *Key performance variables between the top 10 and bottom 10 teams in the English premier league 2012/13 season*. Hum Mov Health Coach Edu, 1,17–29, 2014.
- [2] BARROS, R., CUNHA, S., MAGALHAES, W., GUIMARAES, M., et al. *Representation and analysis of soccer players' actions using principal components*. Journal of Human Movement Studies, 2006.
- [3] BROICH, H., MESTER, J., SEIFRIZ, F., & YUE, Z. *Statistical analysis for the first Bundesliga in the current soccer season*. Progress in Applied Mathematics, 7(2), 1–8, 2014.
- [4] DOBSON, S., & GODDARD, J. *Modelling and forecasting match results in the English premier league and football league*. In Economics, management and optimization in sports (pp. 59–77). Springer, 2004.
- [5] ELYAKIM, E., MORGULEV, E., LIDOR, R., MECKEL, Y., ARNON, M., & BEN-SIRA, D. *Comparative analysis of game parameters between Italian league and Israeli league football matches*. International Journal of Performance Analysis in Sport, 20(2), 165–179, 2020.
- [6] HARRELL JR, F. E., LEE, K. L., CALIFF, R. M., PRYOR, D. B., & ROSATI, R. A. *Regression modeling strategies for improved prognostic prediction*. Statistics in medicine, 3(2), 143–152, 1984.
- [7] LAGO-PEÑAS, C., LAGO-BALLESTEROS, J., DELLAL, A., & GÓMEZ, M. *Game-related statistics that discriminated winning, drawing and losing teams from the Spanish soccer league*. Journal of sports science & medicine, 9(2), 288, 2010.
- [8] LEPSCHY, H., WÄSCHE, H., & WOLL, A. *How to be successful in football: a systematic review*. The Open Sports Sciences Journal, 11(1), 2018.
- [9] LEPSCHY, H., WÄSCHE, H., & WOLL, A. *Success factors in football: an analysis of the German Bundesliga*. International Journal of Performance Analysis in Sport, 1–15, 2020.
- [10] MOURA, F. A., MARTINS, L. E. B., & CUNHA, S. A. *Analysis of football game-related statistics using multivariate techniques*. Journal of sports sciences, 32(20), 1881–1887, 2014.
- [11] PEDUZZI, P., CONCATO, J., KEMPER, E., HOLFORD, T. R., & FEINSTEIN, A. R. *A simulation study of the number of events per variable in logistic regression analysis*. Journal of clinical epidemiology, 49(12), 1373–1379, 1996.
- [12] PEÑA, D. *Análisis de datos multivariante*. Mc Graw Hill, 2002.

- [13] PEREZ-SÁNCHEZ, J. M., GÓMEZ-DENIZ, E., & DAVILA-CÁRDENES, N. *A comparative study of logistic models using an asymmetric link: Modelling the away victories in football*. *Symmetry*, 10(6), 224, 2018.
- [14] REILLY, T. *A motion analysis of work-rate in different positional roles in professional football match-play*. *J. Human Movement Studies*, 2, 87–97, 1976.
- [15] UEFA. *Association club coefficients*, 2019.
- [16] WILLOUGHBY, K. A. *Winning games in canadian football: A logistic regression analysis*. *The College Mathematics Journal*, 33(3), 215–220, 2002.

Sobre los autores:

Nombre: Pedro J. Chocano Feito

Correo electrónico: pedrocho@ucm.es

Institución: Becario FPI en el Departamento de Álgebra, Geometría y Topología, Universidad Complutense de Madrid.

Nombre: Elena Castilla González

Correo electrónico: elecasti@ucm.es

Institución: Profesora Ayudante en el Departamento de Estadística e Investigación Operativa I, Universidad Complutense de Madrid.