

Propuesta de un método para cuantificar el comportamiento táctico de los equipos de fútbol

■ JOSÉ MARÍA FERNÁNDEZ PONCE

Profesor Titular de Cálculo de Probabilidades y Estadística Matemática.
Universidad de Sevilla.
Entrenador de Fútbol, Nivel I

■ JOSÉ PINO ORTEGA

Doctor en Ciencias de la Actividad Física y Deporte.
Profesor de la Facultad de Ciencias del Deporte.
Universidad de Extremadura.
Entrenador de Fútbol, Nivel II

■ Palabras clave

*Análisis de regresión logística,
Ley de Poisson, Táctica y estrategia*

Resumen

En este trabajo se presentan algunos métodos estadísticos para cuantificar el comportamiento táctico de los equipos de fútbol. Primero se propone un análisis de la efectividad de un equipo, entendiendo ésta por las posibilidades de conseguir gol en función de otras variables que puedan ser influyentes. También se propone un método para medir la calidad ofensiva y defensiva de los equipos, la cual nos permite definir un criterio para la ordenación de los mismos.

■ Abstract

In this work we present some statistical methods to quantify the tactical behaviour of football teams. First we propose an analysis of the effectivity of a team, taking into account the possibilities of scoring a goal in function of the other variables that could influence it. We also propose a method to measure the offensive and defensive quality of the teams, which permits us to define a criterium for putting them in order.

■ Key words

Analysis of logistic regression, Poisson Law, Tactics and strategy

Introducción

En la actualidad es muy común por parte de todos los medios de comunicación la presentación de ciertas estadísticas para explicar el comportamiento de tal equipo en un determinado partido. Dichas estadísticas suelen poseer algunas variables poco informativas, como por ejemplo puede ser el tanto por ciento del tiempo total de posesión de balón de un equipo. Es obvio que no sería sensato abarcar más allá, puesto que sería incomprendible para una gran mayoría de la audiencia. Pero no es menos cierto que la aplicación de las técnicas estadísticas, es decir, el análisis de datos al fútbol, es muy interesante y relativamente moderno.

Reep y Benjamín (1968) plantean un modelo estadístico para explicar el número de pases en un partido. Pero no es hasta finales de los años ochenta cuando se vuelve a retomar el análisis de datos en el fútbol con los trabajos de Ali (1988) y Franks (1988), aunque desde 1970 el profesor dr. Ian Franks se dedicó al análisis cuantitativo en el fútbol, en concreto en procesos y técnicas de entrenamientos en el equipo olímpico canadiense desde 1980 hasta 1983. A partir de estos años aparecen en la literatura estadística numerosos estudios acerca de las diferentes aplicaciones de análisis de datos en fútbol. En este trabajo presentamos algunas de las aplicaciones más interesantes en táctica.

Uno de los aspectos del fútbol que más pasión levanta entre los aficionados y especialistas es la discusión sobre los sistemas de juego empleados por los entrenadores y las diversas tácticas desarrolladas a lo largo de un encuentro. También es muy dado a la discusión las alineaciones presentadas en un partido, el cuándo y quién debe ser sustituido. En definitiva, cualquier proceso en la toma de decisión de un entrenador de fútbol es fácilmente discutible por cualquier persona que presencie un partido aunque ignore los fundamentos básicos de la táctica en el fútbol. Pero dicha toma de decisión, ¿está basada en alguna conclusión lógica y coherente?, ¿o solamente se basa en un proceso intuitivo y basado en la experiencia del entrenador o del propio/a aficionado/a que presencia el partido? Es evidente que nuestro deporte se hace cada año más científico. Desde los preparadores físicos, pasando por un equipo médico, hasta llegar a un grupo de psicólogos. ¿Qué entendemos por un análisis de datos que pueda ser de gran ayuda al entrenador a la hora de la toma de decisiones? Es evidente que no es suficiente con presentar el tanto por ciento de posesión de balón porque ese dato no es informativo acerca de las tácticas o estrategias que emplean los equipos. Tampoco es un dato interesante el número de tiros a puerta si no se acompaña de otra información como puede ser el número de pases realizados hasta llegar a dicho tiro o desde



dónde y cómo se empezó la jugada, si mediante una recuperación del balón al equipo contrario o mediante una recuperación por jugada de estrategia. Las tarjetas amarillas o rojas tampoco es un dato relevante del juego empleado por un equipo, aunque sí es cierto que el número de tarjetas sacadas a un equipo puede influir negativamente para él en el resultado final del partido. Un trabajo interesante en este sentido fue el desarrollado por Ali (1988). En definitiva, cuáles serían las variables que se deben observar durante un partido para sacar la mayor información posible del estilo de juego de un equipo. En primer lugar debemos plantearnos una variable objetivo, es decir, una variable cuyos valores nos indiquen si un equipo es más o menos eficaz. Para ello, si analizamos el juego del fútbol podemos concluir que el objetivo final y más importante dentro de este deporte es el marcar gol, y por tanto cualquier conjunto de estrategias o tácticas que desarrolle un equipo van destinadas a este fin, o bien a impedir que te marquen un gol. A partir de esta premisa podemos establecer que nuestro principal objetivo va a ser cuantificar la efectividad de las tácticas de un equipo con vistas a conseguir gol. El siguiente problema a resolver sería cómo podemos cuantificar esa efectividad.

La forma más sencilla sería mediante porcentajes, es decir, qué tanto por ciento de tiros a puerta desarrollados por un equipo acaban en gol. Pero en realidad el problema es más complicado ya que debemos someter a un análisis si existen otros tipos de variables que influyen significativamente en nuestra variable objetivo o dependiente. De momento, decir que nuestra variable objetivo va a ser dicotómica, esto es, que solamente toma dos valores. Cuando un equipo posee el balón solamente pueden ocurrir dos sucesos, que consigan el gol o que no lo consigan. La efectividad la cuantificaremos usando la probabilidad de esos dos sucesos. Recordemos que una probabilidad se puede interpretar como un valor entre cero y uno que multiplicado por cien nos da un tanto por ciento, y que en definitiva nos da el grado de posibilidad de ocurrencia de un suceso del cual poseemos incertidumbre

debido al carácter aleatorio del mismo. Una vez que tenemos claramente identificada nuestra variable objetivo necesitamos determinar cuáles van a ser las variables explicativas, es decir, necesitamos definir un conjunto de variables que *a priori* pensamos que pueden influir y modificar la probabilidad de conseguir gol. Pollard (1997) demuestra cómo un conjunto posible de variables explicativas y por tanto influyentes en el cálculo de esta probabilidad son:

- La distancia desde el punto donde se encuentra la pelota al centro de la portería justo antes de realizarse el tiro a puerta.
- La zona del campo donde se origina la recuperación del balón, Pollard (1997) divide el terreno de juego en seis zonas de igual área a lo ancho del campo.
- Distancia del oponente más cercano al poseedor del balón.
- El ángulo que forma con la línea de meta la recta que une el balón con el poste más cercano antes de ser lanzado,
- El número de toques que le da al balón el lanzador antes de tirar a puerta.

De un análisis realizado sobre 489 posesiones de balón tan sólo en 47 de ellas se consiguió el gol, es decir, un porcentaje del 10 %, comprobándose que dicha efectividad dependía de la zona desde donde se lanzara el tiro a puerta. Por eso, es necesario distinguir la probabilidad de conseguir gol según la zona donde se encuentre el balón cuando es recuperado por un equipo que empieza a realizar sus tácticas ofensivas.

Una vez que se posee una muestra con todo este conjunto de información se realiza un análisis de regresión logística disponible en el paquete informático estadístico SPSS para Windows. Con el conjunto de las probabilidades de conseguir gol de un equipo según la zona donde recupere el balón, un entrenador podrá emplear determinadas tácticas y estrategias defensivas para contrarrestar dichas probabilidades. También puede servir para comprobar qué banda es más fuerte o más débil de un equipo en función de la efectividad de la zona en cuestión. Incluso puede ocurrir que las variables explicativas antes ex-

puestas sean significativas e importantes para un equipo y para otro no, o también que un equipo tenga variables explicativas distintas a otro equipo debido a las diferencias existentes entre las tácticas y estrategias empleadas por los entrenadores. Todo esto será tratado en la sección “Cómo medir la efectividad de la táctica”. En la sección “Parámetros ofensivos y defensivos” se presenta un modelo estadístico para cuantificar la calidad ofensiva y defensiva de los equipos. En función de estos parámetros se propone un criterio para ordenar los equipos al final de liga. Por último, en la sección “Otras aplicaciones estadísticas” se explican brevemente otras aplicaciones interesantes de la estadística.

Cómo medir la efectividad de la táctica

No hay duda que el fútbol es el deporte más famoso del mundo. El fútbol profesional se ha convertido en un gran negocio donde se mueven grandes cantidades de dinero y sin embargo resulta difícil de creer y sorprendente que aún no se hayan tenido demasiado en cuenta los estudios científicos existentes sobre la efectividad de las estrategias y el estilo de juego desde un punto de vista matemático y de análisis de datos provenientes del propio juego. Han sido mucho los intentos por obtener conclusiones fiables de los datos procedentes de un partido de fútbol, pero la mayoría estaban basados en la visión directa del partido (la mayoría de la toma de decisiones realizadas por un entrenador durante un partido son de este tipo). Dichas conclusiones se han presentado en multitud de congresos y conferencias intentando aportar soluciones a los entrenadores de los equipos de fútbol. Ali (1988) propone la solución más rigurosa desde el punto de vista matemático y que continúa con el trabajo de Pollard y Reep (1997).

Análisis de la posesión de balón

La unidad básica que vamos a utilizar para realizar el análisis de un equipo es la posesión de balón. Una posesión comienza cuando un jugador recupera el balón

independientemente de la forma y se lo pasa a otro jugador del mismo equipo. El jugador debe tener el suficiente control sobre el balón que le permita enviarlo en cualquier dirección. La posesión del equipo continuará con una serie de pases entre jugadores del mismo equipo y finalizará cuando uno de los siguientes sucesos ocurra:

- El balón está fuera de juego.
- Un jugador del equipo contrario toca el balón (por una entrada, interceptación o una parada). Se excluye el caso de que la dirección del balón sea cambiada por el toque momentáneo de un jugador contrario.
- Si se incumple alguna de las reglas del juego.

Cada posesión de un equipo está formada por varias componentes. Por ejemplo, puede ser debido a un pase en largo hacia delante desde el medio campo. Para medir la efectividad de estas jugadas es necesario cuantificarlas. En cada posesión de balón pueden darse diferentes circunstancias, veamos algunas de ellas.

Los goles

El objetivo primordial del fútbol y, por tanto, de la posesión del balón es conseguir gol. Pollar y Reep (1997) comprobaron que de 6.000 posesiones de equipo solamente en 47 de ellas acabaron en gol. Por tanto, si los goles se codifican como una variable binaria, alrededor del 99 % de las posesiones de balón no acaban en gol.

Los tiros a puerta

Normalmente, un gol es precedido por un tiro a puerta, entendiendo por tiro a puerta como un intento directo de marcar un gol por un jugador que golpea el balón hacia la portería del contrario. De igual forma, sólo el 8 % de las posesiones de balón acaba en tiro a puerta; por tanto, tampoco sería una variable apta para cuantificar la efectividad. También ocurre que la probabilidad de conseguir gol varía considerablemente dependiendo desde dónde se realice el tiro a puerta. Por ejemplo, si el tiro a puerta se realiza desde dentro del área de penalty es 15 veces más probable

que se obtenga gol que si el tiro es desde afuera. Por tanto, parece lógico ponderar cada tiro a puerta según la probabilidad de conseguir gol desde dicha posición.

Tiros a puerta ponderados

En este caso codificaremos los tiros a puerta mediante dos valores: 0 en el caso que la posesión de balón no produzca un tiro a puerta y valdrá p si la posesión de balón termina en tiro a puerta. Evidentemente, p sería un valor aproximado de la probabilidad de conseguir gol. Una estimación para el valor de p se puede obtener mediante un análisis de regresión logística basado en las posesiones de equipo que acaban en tiros a puerta. Incluso con esta codificación tendríamos que, de las posesiones de balón, en el 92 % de los casos tendríamos un valor de 0. Por tanto, parece lógico codificar las posesiones de balón que no acaban en tiro a puerta. Por ejemplo, podemos distinguir entre las posesiones de balón que son relativamente exitosas (aquellas que terminan en un córner) y las que son menos exitosas (como puede ser la pérdida de la posesión sin haber sobrepasado la línea de medio campo).

La efectividad de la posesión de balón

Para cuantificar la variable efectividad, primero se clasifica la posesión de cada equipo en dos variables, la zona donde se origine y el tipo de posesión. El terreno de juego se divide en seis zonas de igual anchura. Un análisis previo de los tiros a puerta sugiere que la probabilidad de marcar gol depende de si la posesión de balón se originó como jugada de estrategia (por ejemplo, un libre directo) o de juego directo; esta información binaria se representa como "tipo de posesión". Para una posesión de tipo j que comienza en la zona i , la probabilidad de marcar un gol p_{ij} se puede aproximar por

$$p_{ij} = \sum p_{ijk}/n_{ij}.$$

Donde $i = 1, \dots, 6$ dependiendo de la zona (ver figura 1), $j = 1$ (si es juego directo) o $j = 2$ (jugada de estrategia), p_{ijk} denota la k -ésima posesión de balón de

tipo j originada en la zona i y es igual a la probabilidad aproximada de marcar gol p de que la posesión de balón si finaliza en un tiro o 0 en caso contrario. n_{ij} es el número total de posesiones de balón que se originan en la zona i de tipo j .

Por tanto, $p_{51} = 0,014$, significaría que de 1.000 posesiones de balón que se originan en la zona 5 en juego directo el equipo esperaría obtener 14 goles.

Si se tienen calculados los valores para cada p_{ij} sería posible asignar cada uno de ellos a cada posesión de balón observada, dependiendo del tipo de posesión y de la zona donde se origine la siguiente posesión de balón. Por ejemplo, si al final de una posesión de balón inmediatamente es continuada por el mismo equipo recuperando el balón en juego directo en su zona 4, entonces el valor del suceso para la primera posesión de balón debería ser p_{42} . Sin embargo, si la posesión no se recupera, es decir el balón pasa al equipo contrario en juego directo, entonces el valor del suceso para la primera posesión debería ser $-p_{31}$, puesto que la siguiente posesión sería con el equipo contrario en su propia zona 3. El signo negativo indica, por tanto que la posesión inicial del balón finaliza con la posesión de los contrarios, el valor p_{ij} sería la probabilidad esperada que el equipo contrario marque desde la situación de la cual recuperaron el balón. En este momento, para cada posesión de balón se podría asignar el valor p o cualquiera de los anteriormente calculados, a los cuales nos referiremos como la efectividad preliminar. Usando estos nuevos valores, el valor medio del suceso para las posesiones de balón originadas en cada zona, y según el tipo se podrían calcular usando las efectividades preliminares en lugar de los tiros a puerta ponderados. Los nuevos valores medios se llamarán efectividades y_{ij} y serán la nueva estimación del valor correspondiente a una posesión de balón de tipo j originada en la zona i . Cada nueva posesión de balón se podrá calcular según un valor determinado como antes, mediante cómo y dónde comienza la siguiente posesión del equipo. Este valor será la efectividad definido anteriormente, sustituyendo la efectividad preliminar. Este proceso iterativo se conti-



núa en cada estado y el nuevo efectividad medio se calcula basado en las efectividades desde la iteración anterior. Cuando ninguno de los valores y_{ij} varía mas de 0,001 en una iteración entonces el proceso termina y la efectividad permanece constante.

La efectividad será nuestro valor final. Vamos a especificar los dos usos distintos que posee esta variable. En primer lugar, la efectividad se puede utilizar para cuantificar el suceso esperado de una posesión de balón de tipo j originada en la zona i . Y, en segundo lugar, se puede usar para medir el suceso actual de la posesión de balón según el equipo, la zona y el tipo de posesión siguiente.

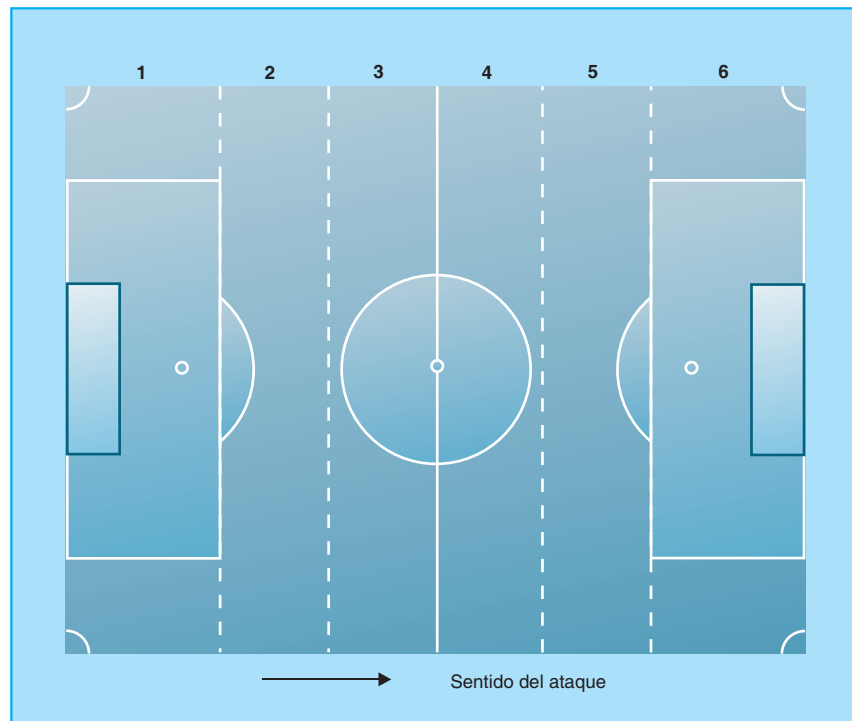
Por ejemplo, si la efectividad $y_{51} = 0,025$, entonces por cada 1.000 posesiones de balón originadas en juego directo en la zona 5 se espera conseguir 25 goles más que los concedidos. Alternativamente, una posesión de balón que finaliza y es seguida por una recuperación de balón en juego directo en la zona 5 se le asignaría un peso de $y_{51} = 0,025$.

En términos de probabilidad, la efectividad de la posesión de balón sería la probabilidad estimada de marcar gol menos la probabilidad estimada que nos marquen un gol, basada en el suceso de la posesión. Aunque la efectividad es menos fácil de interpretar que un gol o un tiro, tiene la ventaja de cuantificar y distinguir entre la gran variedad de posesiones fallidas. De hecho, el desarrollo de la efectividad tiene una conexión directa con marcar gol. La efectividad se basa en la estimación de la probabilidad de marcar gol, la cual está basada en marcar un gol mediante un análisis de regresión logística.

La *tabla 1* proporciona la efectividad en goles marcados por cada 1.000 posesiones según en la zona donde se origine y el tipo de juego basados en 5.844 posesiones de balón del Mundial de Fútbol celebrado en México en 1986.

Para cualquier zona, las posesiones originadas en juego directo tienen mayor efectividad que las jugadas a estrategias. Esto es presumiblemente un reflejo del tiempo extra que una jugada de estrategia proporciona al equipo contrario a situarse en posiciones defensivas.

■ FIGURA 1.
División del terreno de juego en seis parcelas.



■ TABLA 1.
Posesiones y efectividad de posesiones.

ZONA DE ORIGEN	N.º POSESIONES JUEGO ABIERTO	N.º POSESIONES EN ESTRATEGIAS	EFFECTIVIDAD EN JUEGO ABIERTO	EFFECTIVIDAD EN ESTRATEGIAS
1	865	651	5,9	2,2
2	822	244	8,5	0,5
3	837	321	6,2	2,2
4	473	450	10,9	8,5
5	318	336	24,8	12,6
6	111	416	78,3	18,0

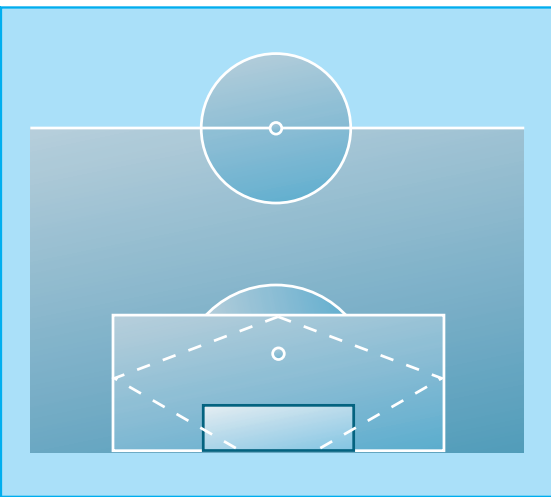
La alta efectividad de la recuperación del balón en la zona 6 (1 gol por cada 13 posesiones) indica la importancia que los equipos deberían darle a la búsqueda de la posesión en esta zona mediante la presión en dicha área de penalty. Ésta es una de las estrategias fundamentales de un estilo de juego directo y agresivo.

Quizás sea sorprendente que la efectividad, la variable del suceso de la posesión

de balón de un equipo, debería ser tan dependiente del tipo de posesión y de la zona de origen, los cuales son componentes iniciales de la posesión.

Es muy posible que para una posesión se obtenga el efecto de estas componentes iniciales sea poco importante. La mayoría de las posesiones son de muy corta duración. De 23.000 posesiones analizadas en 1958 en la primera división inglesa tan

■ FIGURA 2.
Zona del área con mayor probabilidad de gol.



sólo el 5 % se realizaron a más de 4 pases completos. Incluso en el Mundial de 1986 se llegó a un 15 %, donde éste se caracterizó por la conservación de balón por parte de los equipos.

La efectividad de los tiros a puerta

La variable 'tiros ponderados' requiere valores para la probabilidad de marcar un gol según las circunstancias. Para estimar estas probabilidades se realizó un análisis a partir de 489 posesiones de balón que terminaron en tiro a puerta. Se consiguió gol en tan sólo 47 de ellas, dando un porcentaje de acierto del 9,6 %, el cual es consistente con una razón de 1 sobre 10 ya establecido por varios autores anteriormente, donde se relacionaba la probabilidad de conseguir gol en función de la localización del tiro a puerta.

La figura 2 dispone una zona marcada dentro del área de penalty donde son marcados la mayoría de los goles. Esta zona está comprendida entre dos líneas imaginarias que forman un ángulo de 45° con la línea de meta y tienen como origen cada poste y una longitud de 16,38 metros. La zona queda cerrada uniendo los puntos finales de estas líneas en el borde del área de meta. La probabilidad de marcar de un tiro en esa zona es de 0,189 ($n = 206$), comparada con 0,014 desde fuera ($n = 278$). Se excluyen los penalties.

Es evidente que la localización del tiro tiene mucha influencia en la probabilidad de conseguir gol. Se observó que existían otras variables que podrían influir en la probabilidad de conseguir gol. Así se plantea el siguiente modelo: la variable dependiente toma dos valores, 1 si se consigue gol y 0 si no se consigue. Las variables explicativas son:

- X_1 : sería la distancia entre el punto medio de la portería situado sobre la línea de gol y el punto del tiro a puerta;
- X_2 : sería el ángulo en radianes que forman la recta que une el punto desde donde se realiza el tiro y el poste más cercano al balón y la línea de gol;
- X_3 : vale 0 si el jugador realiza el tiro cuando ha tocado solamente una vez el balón y vale 1 si lo ha tocado más de una vez antes de realizar el tiro a puerta;
- X_4 : vale 0 si realiza el disparo a una distancia inferior a una yarda del defensor más próximo y cero en caso contrario;
- X_5 : vale 0 si la posesión de balón se originó en juego directo y valdrá 1 si es mediante una jugada de estrategia.

Bajo este modelo es apropiado usar un modelo de regresión logístico. Se realizó separadamente para remates de cabeza y tiros a puerta con el pie.

Tiros a puerta con el pie

Analizando 410 tiros a puerta se obtiene que la variable x_3 no es significativa, luego el modelo final queda como:

$$Y = 1,245 - 0,219x_1 - 1,578x_2 + 0,947x_4 - 1,069x_5.$$

Para lo cual la probabilidad de marcar se puede estimar mediante la fórmula:

$$P(Y = 1) = e^y / (1 + e^y).$$

Por tanto, para cada posesión de balón de un equipo con valores particulares para las variables x 's se puede estimar la probabilidad de conseguir gol. Por ejemplo, supongamos un tiro desde 14,56 metros, frontal a portería con un oponente a menos de una yarda y cuya posesión se originó en juego directo, entonces el valor de

$y = -3,328$ y la estimación de la probabilidad sería $p = 0,035$.

Este modelo también permite otras interpretaciones. Por ejemplo, observando el coeficiente de la primera variable, $\exp(0,219) = 1,24$, indica que por cada yarda que nos acerquemos a puerta la probabilidad de marcar gol aumenta en un 24 %. Similarmente, $\exp(0,947) = 2,58$, significa que un jugador que tiene su oponente a más de una yarda aumentará más del doble su probabilidad de meter gol.

Análisis similares se pueden realizar para los remates y los lanzamientos desde el punto de pena máxima.

Aplicación en táctica

La efectividad también se puede utilizar para cuantificar el actual suceso de posesión de balón de un equipo como se ha descrito en la sección "Análisis de la posesión de balón". Por tanto, para asignar la efectividad de una estrategia particular, se puede calcular la media de las efectividades de todas las posesiones basadas en dichas estrategias. Pudiéndose así comparar diferentes estrategias. Como ejemplo simple, supongamos un equipo que está en la zona 6. Existen dos estrategias básicas: tirar a puerta o realizar un pase en corto. Utilizando los datos del Mundial de 1986, los tiros a puerta tuvieron una efectividad de 21,7 y los pases en corto 3,5. Usando un test no paramétrico (ya que la distribución de los valores de efectividad no sigue una normal) estas diferencias correspondientes a la zona 6 eran claramente significativas ($p < 0,01$).

Parámetros ofensivos y defensivos

Características del modelo

Una de las medidas aceptada por la RFEF y la LFP para clasificar los equipos empataados a puntos es la diferencia entre los goles conseguidos y los encajados, favoreciendo en caso de igualdad al que más goles haya conseguido. La justificación de este criterio de toma de decisión es para premiar el juego ofensivo con vistas a mejorar la calidad del espectáculo. Pero realmente, ¿está fun-



damentado este criterio con cierta lógica? Desde el punto de vista matemático se puede realizar una ordenación de elementos si nos basamos en un número real para la comparación. Por ejemplo, los componentes de un equipo de fútbol pueden ser ordenados por su altura desde el más bajo hasta el más alto. La dificultad radica cuando queremos ordenar elementos de un conjunto comparando al mismo tiempo dos o más características. La ordenación actual que establece el reglamento oficial del fútbol se denomina lexicográfica, es decir, primero ordenamos por la característica que consideramos más importante (en nuestro caso el número de puntos conseguido) y en caso de empates se ordenan los equipos según la diferencia de goles a favor, y si de nuevo hay empate el que más goles haya conseguido. No sería extraño encontrar equipos que no podamos ordenarlos, es decir, decidir cuál es mejor, porque empatan en todos estos criterios. Incluso este criterio puede presentar ciertas dudas sobre la calidad de los equipos en cuanto a sus referencias ofensivas y defensivas. No olvidemos que el fútbol se basa en dos bloques tácticos, principios ofensivos y principios defensivos, y que realizarlos a la perfección redundaría en la mejora de la calidad del espectáculo. Por tanto, sería interesante crear otro criterio de ordenación donde los empates sean “casi imposibles” y que además sea un criterio que tenga en cuenta y valore en justa medida tanto el juego ofensivo como el juego defensivo.

El modelo matemático

El criterio que se propone a continuación está basado en el trabajo de Dixon y Coles, 1997. La hipótesis de partida es que el número de goles marcados por un equipo como local y como visitante siguen leyes de probabilidad de Poisson cuyas medias se determinan por cualidades ofensivas y defensivas. La tabla 2 presenta una aproximación de la probabilidad del número de goles conseguidos en casa por los equipos de la Premier League en la temporada 1994-1995. Realizando un contraste no paramétrico de bondad de ajuste no podemos rechazar la hipótesis de que dichos porcentajes son

aproximados a los correspondientes de una ley de Poisson. Los tantos por cientos esperados de la ley de Poisson son calculados utilizando una aproximación a la media con los datos que tenemos; en nuestro caso la media vale:

$$\bar{x} = 1,41 \text{ gol/partido.}$$

De igual forma podemos hacer con los goles de los equipos visitantes, representados en la tabla 3.

En este segundo caso el número medio de goles por partido conseguidos por los equipos visitantes es de:

$$\bar{y} = 1,075.$$

En la tabla 4 se presentan los porcentajes de los diferentes resultados. De esta forma podemos estudiar estadísticamente si el número de goles marcados por los equipos locales es independiente del número de goles marcados por los equipos visitantes. Al final se llega a la conclusión de que no tenemos suficiente evidencia estadística cómo para rechazar la independencia entre esas dos variables.

Así por ejemplo, en el 8,2 % de los partidos el resultado final fue de empate a cero, y en el 4,5 % el resultado final fue de 3 a 1.

En definitiva, y tras este análisis previo, se puede realizar el siguiente modelo esta-

■ **TABLA 2.**
Porcentajes de goles de equipos locales.

NÚMERO DE GOLES	%	% ESPERADOS DE POISSON
0	22,6	24,41
1	33,5	34,42
2	25	24,26
3	13,1	11,40
Más de 4	5,8	5,51

■ **TABLA 3.**
Número de goles conseguidos por los visitantes.

NÚMERO DE GOLES	%	% ESPERADOS DE POISSON
0	33,4	34,10
1	36,4	36,68
2	19,5	19,70
3	7,9	7,00
Más de 4	2,8	2,52

■ **TABLA 4.**
% de resultados en la temporada 1994-1995.

LOCAL	VISITANTE				
	0	1	2	3	Más de 4
0	8,2	7,4	4,5	1,4	0,4
1	10,3	12,7	6,4	2,7	0,6
2	8,2	9,1	4,8	1,9	0,5
3	4,2	4,5	2,3	1,2	0,4
Más de 4	1,6	1,8	1,1	0,6	0,1

■ TABLA 5.
Puntuaciones de la temporada 1994-1995 de la Premier League.

EQUIPO	ATAQUE	DEFENSA	POSICIÓN FINAL	POSICIÓN REAL
Blackburn R.	1,73	0,534	1	1
Manchester U.	1,869	0,402	2	2
Nottingham	1,46	0,658	3	3
Liverpool	1,448	0,561	4	4
Leeds U.	1,51	0,583	5	5
Newcastle U.	1,659	0,578	6	6
Tottenham H.	1,622	0,775	7	7
Queen's P.R.	1,497	0,717	8	8
Wimbledon	1,281	0,732	9	9
Southampton	1,446	0,772	10	11
Chelsea	1,238	0,658	11	10
Arsenal	1,235	0,527	12	12
Sheffield W.	1,387	0,698	13	13
West Ham	1,192	0,649	14	14
Everton	1,177	0,667	15	15
Coventry	1,115	0,669	16	16
Manchester C.	1,232	0,728	17	17
Aston Villa	1,278	0,527	18	18

$$d_{ij} = a_i b_j.$$

Estos parámetros *a priori* son desconocidos y hay que calcularlos según va desarrollándose la competición. Los métodos estadísticos para obtenerlos son bastante complejos como para desarrollarlos aquí, así que hacemos mención a los trabajos de Dixon y Coles (1997) y Dixon y Robinson (1998) para verlos de forma más extendida.

Ordenación paramétrica

Basándonos en los parámetros ofensivos y defensivos calculados anteriormente podemos construir un criterio de ordenación entre equipos. Fijémonos que para un equipo cualquiera lo ideal es que tenga su parámetro ofensivo lo más alto posible y su parámetro defensivo lo más bajo posible. De esta forma podemos definir la razón de calidad del equipo *i* como el cociente entre el parámetro ofensivo y el defensivo:

$$C_i = a_i / b_i.$$

Por tanto, un equipo será mejor que otro cuanto mayor sea su razón de calidad. Resulta curioso que si ordenamos la clasificación final de una liga, por ejemplo la clasificación final de la temporada 1994-1995 de la Premier League, se obtiene otra distinta a la que se obtuvo (ver *tabla 5*). Aunque en esa temporada no modifica de una manera importante la clasificación (en nuestro ejemplo solamente se ven afectadas las posiciones décimas y undécimas) puede darse que en algunas temporadas existan empates a puntos entre el cuarto clasificado y el quinto y decidir cuál de los dos es mejor es una decisión en la que hay en juego muchos millones de dólares ya que el cuarto puesto tiene derecho a jugar la Champions League, mientras que el quinto solamente juega la Copa de la UEFA; de igual forma para los puestos de descenso.

Otras aplicaciones estadísticas

Existen muchos tópicos dentro del fútbol. Por ejemplo, suele decirse que con diez

dístico. En un partido entre los equipos *i* y *j*, sean:

- X_{ij} = Núm. de goles marcados por el local *i* cuando se enfrenta a *j*.
- Y_{ij} = Núm. de goles marcados por el visitante *j* cuando se enfrenta a *i*.

Los porcentajes de dichas variables pueden ser aproximadas por leyes de probabilidad de Poisson cuya media, que en definitiva sería el número medio de goles por partido, va a depender de los parámetros ofensivos y defensivos de cada equipo. Así, el equipo *i* tendrá dos parámetros, uno ofensivo (a_i) y otro defensivo (b_i). De tal forma que el pará-

metro ofensivo mide la razón de “ataque” y el parámetro defensivo la razón de “defensa”. El número medio de goles marcados por el equipo *i* como local cuando se enfrenta al equipo *j* (m_{ij}) va a depender de los parámetros anteriores según el modelo:

$$m_{ij} = a_i b_j g.$$

Donde $g > 0$ es un parámetro que nos permite tener en cuenta la ventaja de jugar en casa. Usando el mismo razonamiento podemos modelar el número medio de goles marcados por el equipo *j* como visitante cuando se enfrenta al equipo *i* (d_{ij}) de la siguiente forma:



jugadores se juega mejor que con once cuando un jugador es expulsado. Es obvio que este axioma dificulta uno de los principios básicos del balompié para conseguir la victoria en un partido, que no es otro que el conseguir la superioridad numérica allá donde se encuentre el balón. Está demostrado estadísticamente cómo una expulsión afecta negativamente al equipo infractor (ver Ridder *et al.*, 1994). En este trabajo se modela la evolución del marcador de un partido de fútbol en términos de variables estadísticas como el minuto en el que un jugador es expulsado, el número de goles conseguidos por el equipo antes de la expulsión, el número de goles conseguidos después de dicha expulsión. Aparte de estas variables, se estudian algunas relativas al comportamiento de los equipos en diferentes partidos y diferenciando con los equipos que se enfrentan. El modelo se diseña bajo unas determinadas suposiciones, tales como:

- Los dos equipos van consiguiendo goles según la ley de probabilidad de Poisson en términos de tiempo y de manera independiente uno del otro, es decir, el número de goles marcados por un equipo es independiente del número de goles marcados por el otro. Esta suposición, que *a priori* parece irreal está contrastada estadísticamente. Pero aún más, también se ha contrastado que los intervalos de tiempos entre goles son independientes y que las intensidades de marcar un gol no es constante a lo largo de un partido.
- La intensidad de marcar gol de un equipo con 11 jugadores solamente depende del equipo al que se enfrenta pero permanece constante a lo largo de un partido aunque depende de una función del tiempo en el partido.
- Después de una tarjeta roja, la intensidad de marcar gol varía puesto que el equipo pasa a tener diez jugadores, modificándose a partir de ese instante las condiciones impuestas en la hipótesis 2.

Otro de los tópicos usuales en el fútbol es la ventaja inicial que poseen los equipos locales por jugar en casa. Clarke y Norman (1984) analizaron esta situación. En su trabajo modelaron la habilidad de un

equipo y la ventaja de ser local de la siguiente forma:

$$w_{ij} = u_i - u_j + h_i + e_{ij}.$$

donde w_{ij} puede tomar tres valores $\{-1, 0, 1\}$ si el equipo local i pierde, empatata o gana el equipo j , respectivamente. La variable u_i cuantifica la habilidad del equipo local i , la variable u_j cuantifica la habilidad del equipo visitante j . La variable h_i cuantifica el efecto de la ventaja de ser local para el equipo i . Por último, en todos los modelos de regresión se introduce una variable denominada error aleatorio de media nula, que en este caso viene representada por e_{ij} . Una de las suposiciones de este modelo es que la ventaja de ser local y la habilidad de los equipos permanecen constantes a lo largo de una temporada. Este modelo fue aplicado a partidos de fútbol de la liga inglesa desde la temporada 1981-1982 hasta la temporada 1990-1991, siendo en total 920 equipos y 20.306 partidos. Los resultados demostraron que la ventaja de ser local varía de una temporada a otra y que en algunas temporadas existen equipos con ventajas de ser local negativas. Se dedujo de este trabajo que no hay diferencias significativas para la ventaja de ser local entre las diferentes categorías, que el año influye de manera importante en dicho coeficiente y el tipo de club también influye en determinar la ventaja de ser local pero no de manera tan determinante como la temporada.

Conclusiones

Nuestro objetivo en este trabajo ha sido presentar diversos métodos cuantitativos para explicar, entender mejor y contrarrestar el comportamiento táctico de los equipos de fútbol. De nuestras charlas con algunos entrenadores y preparadores físicos de equipos de primera línea a nivel nacional llegamos a la conclusión que la mayoría de ellos eran escépticos acerca de la utilidad del análisis de datos avanzado para la mejora de la calidad de sus equipos basándose en los aspectos cualitativos de este deporte. Es evidente que uno de los factores más importantes

en el fútbol de hoy en día es la habilidad de sorprender al contrario con un gesto técnico o táctico que no pueda contrarrestarse pero no es menos cierto que a medida que avanza un partido o una temporada dicha habilidad va disminuyendo por diversas razones y el comportamiento de los equipos se hace menos sorpresivo y por tanto más constante y fácil de cuantificar. Si poseemos las herramientas necesarias para conocer mejor dicho comportamiento habremos empezado a ganar el partido antes que el balón esté rodando por el terreno de juego. En este trabajo hemos desarrollado algunas de las aplicaciones más importante del análisis de datos avanzado aplicado al fútbol como puede ser el estudio de la efectividad de un equipo en función de variables explicativas que pueden influir significativamente en la probabilidad de conseguir gol. También se ha desarrollado un método para explicar mediante parámetros el comportamiento ofensivo y defensivo de los equipos.

Bibliografía

- Ali, A. H.: "A Statistical Analysis of tactical movement patterns in soccer", *Science and Football*, pp. 302-308, London: Spon, 1988.
- Clarke, S. R. y Norman, J. M.: "Home ground Advantage of individual clubs in English soccer", *The Statistician*, 44 (1995), pp. 509-521.
- Dixon, M. J. y Coles, S. G.: "Modelling association football scores and inefficiencies in the football betting market", *Applied Statistics*, 46, (2) (1997), pp. 265-280.
- Dixon, M. J. y Robinson, M. E.: "A Birth process model for association football matches", *The Statistician*, 47 (3) (1998), pp. 523-538.
- Franks, I. M.: "Analysis of association Football", *Soccer Journal* 33 (5) (1988), pp. 35-43.
- Pollard, R. y Reep, C.: "Measuring the effectiveness of playing strategies at soccer", *The Statistician*, 46 (4) (1997), pp. 541-550.
- Reep, C. y Benjamin, B.: "Skill and Chance in Association Football", *Journal of Royal Statistical Society A*, 131, (1968), pp. 581- 585.
- Ridder, G.; Cramer, J. S. y Hopstaken, P.: "Down to ten: Estimating the effect of a red card in soccer", *Journal of the American Statistical Association*, 89, 427, (1994), pp. 1124-1127.