

Proposta d'un mètode per quantificar la conducta tàctica dels equips de futbol

■ JOSÉ MARÍA FERNÁNDEZ PONCE

Professor Titular de Càlcul de Probabilitats i Estadística Matemàtica.
Universidad de Sevilla.
Entrenador de Futbol, Nivell I

■ JOSÉ PINO ORTEGA

Doctor en Ciències de l'Activitat Física i Esport.
Professor de la Facultad de Ciencias del Deporte.
Universidad de Extremadura.
Entrenador de Futbol, Nivell II

■ Paraules clau

*Anàlisi de regressió logística,
Llei de Poisson, Tàctica i estratègia*

Resum

En aquest treball es presenten alguns mètodes estadístics per quantificar la conducta tàctica dels equips de futbol. Primer, es proposa una anàlisi de l'efectivitat d'un equip, entenent aquesta com les possibilitats d'aconseguir gol en funció d'altres variables que puguin ser influents. També es proposa un mètode per mesurar la qualitat ofensiva i defensiva dels equips, el qual ens permet de definir un criteri per a ordenar-les.

■ 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

Introducció

En l'actualitat, és molt comuna, per part de tots els mitjans de comunicació, la presentació d'algunes estadístiques per explicar la conducta d'un equip en un determinat partit. Les estadístiques esmentades acostumen a tenir algunes variables poc informatives, com per exemple el tant per cent del temps total de possessió de pilota d'un equip. És obvi que no seria assenyat anar gaire més enllà, perquè seria incomprendible per a una gran majoria de l'audiència. No és menys cert, però, que l'aplicació de les tècniques estadístiques, és a dir, l'anàlisi de dades en el futbol és molt interessant i relativament moderna.

Reep and Benjamin (1968) plantegen un model estadístic per explicar el nombre de passades en un partit. Però no és fins a final dels anys vuitanta quan es torna a reprendre l'anàlisi de dades en el futbol amb els treballs d'Ali (1988) i Franks (1988), encara que des de 1970 el professor Dr. Ian Franks es va dedicar a l'anàlisi quantitativa en el futbol, en concret, en processos i tècniques d'entrenaments en l'equip olímpic canadenc, des del 1980 fins al 1983. A partir d'aquests anys apareixen en la literatura estadística nombrosos estudis sobre les diferents aplicacions d'anàlisi de dades en futbol. En aquest treball presentem algunes de les aplicacions més interessants en tàctica.

Un dels aspectes del futbol que més passió aixeca entre els aficionats i especia-

listes és la discussió sobre els sistemes de jocs utilitzats pels entrenadors i les diverses tàctiques desenvolupades al llarg d'un partit. També es presten molt a la discussió les alineacions presentades en un partit, el quan i qui ha de ser substituït. En definitiva, qualsevol procés en la presa de decisió d'un entrenador de futbol és fàcilment discutible per qualsevol persona que presenciï un partit, encara que ignori els fonaments bàsics de la tàctica en el futbol. Però la presa de decisió esmentada, es basa en alguna conclusió lògica i coherent? O solament es basa en un procés intuïtiu i fonamentat en l'experiència de l'entrenador o del mateix/a aficionat/da que presencia el partit? És evident que el nostre esport es fa cada any més científic. Des dels preparadors físics, passant per un equip mèdic fins a arribar a un grup de psicòlegs. Què entenem per una anàlisi de dades que pugui ser de gran ajut a l'entrenador a l'hora de la presa de decisions? És evident que no n'hi ha prou a presentar el tant per cent de possessió de pilota, perquè aquesta dada no és informativa sobre les tàctiques o estratègies que fan servir els equips. Tampoc no és una dada interessant el nombre de xuts a porta si no va acompanyada per una altra informació, com ara el nombre de passades realitzades fins arribar al xut esmentat o des d'on i com es va començar la jugada, si mitjançant una recuperació de la pilota a



l'equip contrari o mitjançant una recuperació per jugada d'estratègia. Les targetes grogues o vermelles tampoc no són una dada rellevant del joc utilitzat per un equip, encara que sí que és cert que el nombre de targetes tretes a un equip pot influir-hi negativament en el resultat final del partit. Un treball interessant en aquest sentit va ser el desenvolupat per Ali (1988). En definitiva, quines serien les variables que s'han d'observar durant un partit per treure la major informació possible de l'estil de joc d'un equip. En primer lloc, hem de plantejar-nos una variable objectiu, és a dir, una variable els valors de la qual ens indiquin si un equip és més o menys eficaç. Per fer-ho, si analitzem el joc del futbol, podem concloure que l'objectiu final i més important en aquest esport és marcar gol i, per tant, qualsevol conjunt d'estratègies o tàctiques que desenvolupi un equip van destinades a aquesta finalitat, o bé a impedir que et marquin un gol. A partir d'aquesta premissa podem establir que el nostre principal objectiu serà quantificar l'efectivitat de les tàctiques d'un equip a efectes d'aconseguir gol. El problema següent a resoldre seria com podem quantificar aquesta efectivitat.

La forma més senzilla seria mitjançant percentatges, és a dir, quin tant per cent de xuts a porta realitzats per un equip acaben en gol. Però, de fet, el problema és més complicat, perquè hem de sotmetre a una anàlisi si existeixen d'altres tipus de variables que influeixen significativament en la nostra variable objectiu o dependent. De moment cal dir que la nostra variable objectiu serà dicotòmica, és a dir, que només pren dos valors. Quan un equip posseeix la pilota només es poden produir dos esdeveniments, que aconseguixin el gol o que no. L'efectivitat la quantificarem fent servir la probabilitat d'aquests dos esdeveniments. Recordem que una probabilitat es pot interpretar com un valor entre zero i un que multiplicat per cent ens dona un tant per cent, i que en definitiva ens dona el grau de possibilitat d'ocurrència d'un esdeveniment del qual tenim incertesa a causa del seu caràcter aleatori. Tan bon punt tenim cla-

rament identificada la nostra variable objectiu necessitem determinar quines seran les variables explicatives, és a dir, necessitem definir un conjunt de variables que a priori pensem que poden influir en la probabilitat d'aconseguir gol i modificar-la. Pollard (1997) demostra que un conjunt possible de variables explicatives i per tant influents en el càlcul d'aquesta probabilitat són:

- La distància des del punt on es troba la pilota al centre de la porteria, just abans de fer-se el xut a porta.
- La zona del camp on s'origina la recuperació de la pilota, (Pollard (1997) divideix el terreny de joc en sis zones de la mateixa àrea tot al llarg del camp).
- Distància de l'oponent més proper al posseïdor de la pilota.
- L'angle que forma amb la línia de meta la recta que uneix la pilota amb el pal més proper abans d'ésser llançada.
- El nombre de tocs que dona a la pilota el llançador abans de xutar a la porteria.

D'una anàlisi realitzada sobre 489 possessions de pilota tan sols en 47 d'aquestes es va aconseguir el gol, és a dir, un percentatge del 10% i es va comprovar que aquesta efectivitat depenia de la zona des d'on es llançava el xut a porta. Per això cal distingir la probabilitat d'aconseguir gol, segons la zona on es trobi la pilota quan és recuperada per un equip que comença a realitzar les seves tàctiques ofensives.

Tan bon punt es posseeix una mostra amb tot aquest conjunt d'informació, es realitza una anàlisi de regressió logística, disponible en el paquet informàtic estadístic SPSS per a Windows. Amb el conjunt de les probabilitats d'aconseguir gol d'un equip, segons la zona on es recuperi la pilota, un entrenador podrà utilitzar determinades tàctiques i estratègies defensives per contrarestar les probabilitats esmentades. També pot servir per comprovar quina banda és més forta o més feble d'un equip en funció de l'efectivitat de la zona en qüestió. Fins i tot pot succeir que les variables explicatives exposades abans siguin significatives i importants per a un

equip i per a un altre no ho siguin, o també que un equip tingui variables explicatives diferents de les d'un altre equip a causa de les diferències existents entre les tàctiques i estratègies utilitzades pels entrenadors. Tot això serà tractat a la secció "Com mesurar l'efectivitat de la tàctica".

A la secció "Paràmetres ofensius i defensius" es presenta un model estadístic per quantificar la qualitat ofensiva i defensiva dels equips. En funció d'aquests paràmetres es proposa un criteri per ordenar els equips al final de lliga.

Finalment, a la secció "Altres aplicacions estadístiques" s'expliquen breument d'altres aplicacions interessants de l'estadística.

Com mesurar l'efectivitat de la tàctica

No hi ha dubte que el futbol és l'esport més famós del món. El futbol professional ha esdevingut en un gran negoci on es mouen grans quantitats de diners i, tanmateix, resulta difícil de creure i sorprendent que encara no s'hagin tingut gaire en compte els estudis científics existents sobre l'efectivitat de les estratègies i l'estil de joc, des d'un punt de vista matemàtic i d'anàlisi de dades provinents del mateix joc. Han estat molts els intents d'obtenir conclusions fiables de les dades procedents d'un partit de futbol, però la majoria es basaven en la visió directa del partit (la majoria de la presa de decisions realitzades per un entrenador durant un partit són d'aquest tipus). Les conclusions esmentades s'han presentat en multitud de congressos i conferències, tot intentant d'aportar solucions als entrenadors dels equips de futbol. Ali (1988) proposa la solució més rigorosa des del punt de vista matemàtic i que continua amb el treball de Pollard i Reep (1997).

Anàlisi de la possessió de pilota

La unitat bàsica que utilitzarem per realitzar l'anàlisi d'un equip és la possessió de pilota. Una possessió comença quan un jugador recupera la pilota, indepen-

dentment de la forma, i l'hi passa a un altre jugador del mateix equip. El jugador ha de tenir prou control sobre la pilota perquè li permeti d'enviar-la en qualsevol direcció. La possessió de l'equip continuarà amb un seguit de passades entre jugadors del mateix equip i finalitzarà quan es produeixi un dels esdeveniments següents:

- La pilota està fora de joc.
- Un jugador de l'equip contrari toca la pilota (per una entrada, intercepció o una parada). S'exclou el cas que la direcció de la pilota sigui canviada pel toc momentani d'un jugador contrari.
- Si s'incompleix alguna de les regles del joc.

Cada possessió d'un equip està formada per diversos components. Per exemple, pot ser-ne la causa una passada llarga endavant des de mig camp. Per mesurar l'efectivitat d'aquestes jugades cal quantificar-les. En cada possessió de pilota poden donar-se diferents circumstàncies, vegem-ne algunes.

Els gols

L'objectiu primordial del futbol i, per tant, de la possessió de la pilota és aconseguir gol. Pollard i Reep (1997) van comprovar que de 6.000 possessions d'equip solament 47 van acabar en gol. Per tant, si els gols es codifiquen com una variable binària, prop del 99% de les possessions de pilota no acaben en gol.

Els xuts a porta

Normalment, un gol és precedit per un xut a porta, entenent com a xut a porta un intent directe de marcar un gol per un jugador que copeja el baló cap a la porteria del contrari. De la mateixa forma, només el 8% de les possessions de pilota acaben en xut a porta, per tant, tampoc no seria una variable apta per quantificar-ne l'efectivitat. També passa que la probabilitat d'aconseguir gol varia considerablement depenent del punt des d'on es faci el xut a porta. Per exemple, si el xut a porta es realitza des de dins de l'àrea de penal és 15 vegades més probable que s'obtingui gol que no pas si el xut es fa des de fora. Així doncs, sembla lògic ponderar

cada xut a porta segons la probabilitat d'aconseguir gol des de la posició esmentada.

Xuts a porta ponderats

En aquest cas codificarem els xuts a porta mitjançant dos valors: 0 en el cas que la possessió de pilota no produeixi un xut a porta i valdrà p si la possessió de pilota acaba en xut a porta. Evidentment, p seria un valor aproximat de la probabilitat d'aconseguir gol. Una estimació per al valor de p es pot obtenir mitjançant una anàlisi de regressió logística basada en les possessions d'equip que acaben en xuts a porta. Fins i tot amb aquesta codificació, tindríem que, de les possessions de pilota, en el 92% dels casos tindríem un valor de 0. Per tant, sembla lògic codificar les possessions de pilota que no acaben en xuts a porta. Per exemple, podem distingir entre les possessions de pilota que són relativament reeixides (les que acaben en un córner) i les que són menys reeixides (com pot ser la pèrdua de la possessió sense haver sobrepassat la línia de mig camp).

L'efectivitat de la possessió de pilota

Per quantificar la variable efectivitat, primer es classifica la possessió de cada equip en dues variables, la zona on s'origina i el tipus de possessió. El terreny de joc es divideix en sis zones de la mateixa amplada. Una anàlisi prèvia dels xuts a porta suggereix que la probabilitat de marcar gol depèn de si la possessió de pilota s'ha originat com a jugada d'estratègia (per exemple, un lliure directe) o de joc directe; aquesta informació binària es representa com a "tipus de possessió". Per a una possessió de tipus j que comença a la zona i , la probabilitat de marcar un gol p_{ij} es pot aproximar per

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

On $i = 1, \dots, 6$ segons la zona (vegeu *figura 1*), $j = 1$ (si és joc directe) o $j = 2$ (jugada d'estratègia), p_{ijk} denota la k -èsima possessió de pilota de tipus j originada

a la zona i i és igual a la probabilitat aproximada de marcar gol p de què la possessió de pilota si finalitza en un xut o 0 en cas contrari. n_{ij} és el nombre total de possessions de pilota que s'originen a la zona i de tipus j .

Per tant, $p_{51} = 0,014$ significaria que de 1.000 possessions de pilota que s'originen a la zona 5 en joc directe l'equip esperaria d'obtenir 14 gols.

Si es tenen calculats els valors per a cada p_{ij} seria possible assignar cadascun d'aquests a cada possessió de pilota observada, segons el tipus de possessió i de la zona on s'origini la possessió de pilota següent. Per exemple, si al final d'una possessió de pilota, immediatament, la continua tenint el mateix equip, que recupera la pilota en joc directe en la seva zona 4, aleshores el valor de l'esdeveniment per a la primera possessió de pilota hauria de ser p_{42} . Tanmateix, si la possessió no es recupera, és a dir, si la pilota passa a l'equip contrari en joc directe, aleshores el valor de l'esdeveniment per a la primera possessió hauria de ser $-p_{31}$, atès que la possessió següent seria amb l'equip contrari a la seva pròpia zona 3. El signe negatiu indica, doncs, que la possessió inicial de la pilota finalitza amb la possessió dels contraris, el valor p_{ij} seria la probabilitat esperada que l'equip contrari marqui des de la situació en què van recuperar la pilota.

En aquest moment, per a cada possessió de pilota es podria assignar el valor p o qualsevol dels anteriorment calculats, als quals ens referirem com a l'"efectivitat preliminar". Fent servir aquests nous valors, el valor mitjà de l'esdeveniment per a les possessions de pilota originades en cada zona i segons el tipus, es podrien calcular fent servir les efectivitats preliminars en lloc dels xuts a porta ponderats. Els nous valors mitjans s'anomenaran efectivitats y_{ij} i seran la nova estimació del valor corresponent a una possessió de pilota de tipus j originada a la zona i .

Cada nova possessió de pilota es podrà calcular segons un valor determinat com abans, mitjançant 'on' i 'com' comença la següent possessió de l'equip. Aquest valor serà l'efectivitat, definit anteriorment, i



substituirà l'efectivitat preliminar. Aquest procés iteratiu es continua en cada estat i la nova efectivitat mitjana es calcula prenent com a base les efectivitats des de la iteració anterior. Quan cap dels valors y_{ij} varia més de 0,001 en una iteració, aleshores el procés acaba i l'efectivitat roman constant.

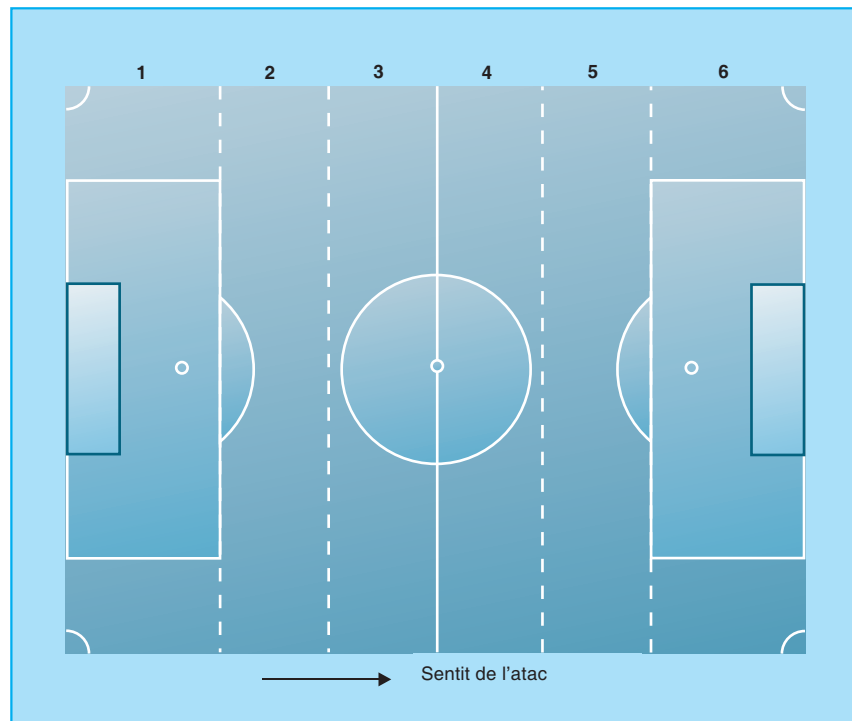
L'efectivitat serà el nostre valor final. Especificarem els dos usos distints que posseïx aquesta variable. En primer lloc, l'efectivitat es pot utilitzar per quantificar l'esdeveniment esperat d'una possessió de pilota de tipus j originada a la zona i . I en segon lloc, es pot usar per mesurar l'esdeveniment actual de la possessió de pilota, segons l'equip, la zona i el tipus de possessió següent. Per exemple, si l'efectivitat $y_{51} = 0,025$, aleshores per cada 1.000 possessions de pilota que origins en joc directe a la zona 5 s'espera d'aconseguir 25 gols més que els concedits. Alternativament, una possessió de pilota que finalitza i és seguida per una recuperació de pilota en joc directe a la zona 5 hom li assignaria un pes de $y_{51} = 0,025$.

En termes de probabilitat, l'efectivitat de la possessió de pilota seria la probabilitat estimada de marcar gol menys la probabilitat estimada que ens marquin un gol, basada en l'esdeveniment de la possessió. Encara que l'efectivitat és menys fàcil d'interpretar que un gol o un xut, té l'avantatge de quantificar i distingir entre la gran varietat de possessions frustrades. De fet, el desenvolupament de l'efectivitat té una connexió directa amb marcar gol. L'efectivitat es basa en l'estimació de la probabilitat de marcar gol, la qual es basa en el fet de marcar un gol, mitjançant una anàlisi de regressió logística.

La *taula 1* proporciona l'efectivitat en gols marcats per cada 1.000 possessions, segons la zona on s'origini i el tipus de joc basats en 5.844 possessions de pilota del Mundial de Futbol celebrat a Mèxic el 1986.

Per a qualsevol zona, les possessions originades en joc directe tenen més efectivitat que no pas les jugades en estratègies. Això és presumiblement un reflex del temps extra que una jugada d'estratègia proporciona a l'equip contrari per a situar-se en posicions defensives.

■ FIGURA 1.
Divisió del terreny de joc en sis parcel·les.



■ Taula 1.
Possessions i efectivitat de possessions.

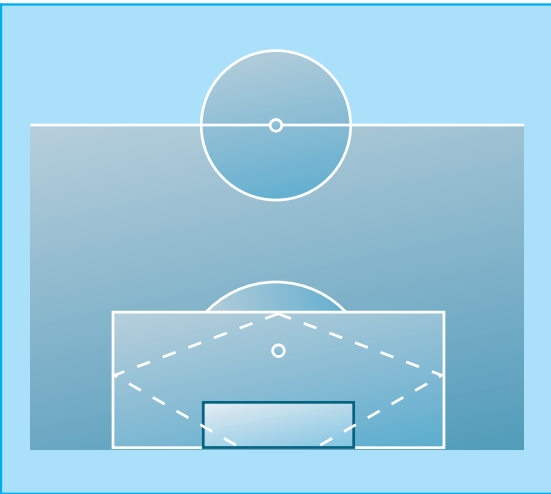
ZONA D'ORIGEN	NOMBRE POSSESSIONS JOC OBERT	NOMBRE POSSESSIONS EN ESTRATÈGIES	EFFECTIVITAT EN JOC OBERT	EFFECTIVITAT EN ESTRATÈGIES
1	865	651	5,9	2,2
2	822	244	8,5	0,5
3	837	321	6,2	2,2
4	473	450	10,9	8,5
5	318	336	24,8	12,6
6	111	416	78,3	18,0

L'alta efectivitat de la recuperació de la pilota a la zona 6 (1 gol per cada 13 possessions) indica la importància que els equips haurien de donar a la recerca de la possessió en aquesta zona, mitjançant la pressió en l'àrea de penal esmentada. Aquesta és una de les estratègies fonamentals d'un estil de joc directe i agressiu. Potser sigui sorprenent que l'efectivitat, la variable de l'esdeveniment de la possessió

de pilota d'un equip, hagi de ser tan dependent del tipus de possessió i de la zona d'origen, els quals són components inicials de la possessió.

És molt possible que per a una possessió s'obtingui que l'efecte d'aquests components inicials sigui poc important. La majoria de les possessions són de molt curta durada. De 23.000 possessions analitzades el 1958 a la primera divisió

■ FIGURA 2.
Zona de l'àrea amb més probabilitat de gol.



anglesa tan sols en el 5% es van realitzar més de 4 passades completes. En el mundial de 1986 fins i tot es va arribar a un 15%, car aquest es va caracteritzar per la conservació de pilota per part dels equips.

L'efectivitat dels xuts a porta

La variable 'xuts ponderats' requereix valors per a la probabilitat de marcar un gol segons les circumstàncies. Per estimar aquestes probabilitats es va realitzar una anàlisi a partir de 489 possessions de pilota que van acabar en xut a porta. Es va aconseguir gol en tan sols 47 ocasions, cosa que dona un percentatge d'encert del 9,6%, el qual és consistent amb una raó d'1 sobre 10 ja establert per diversos autors anteriorment, on es relacionava la probabilitat d'aconseguir gol d'acord amb la localització del xut a porta.

La figura 2 disposa una zona marcada dintre de l'àrea de penal on són marcats la majoria dels gols. Aquesta zona està compresa entre dues línies imaginàries que formen un angle de 45° amb la línia de meta i tenen com a origen cada pal i una longitud de 16,38 metres. La zona queda tancada unint els punts finals d'aquestes línies al límit de l'àrea de meta. La probabilitat de marcar d'un xut en aquesta zona és de 0,189 ($n = 206$), comparada amb 0,14 des fora ($n = 278$). Se n'exclouen els penals.

És evident que la localització del xut té molta influència en la probabilitat d'aconseguir gol. Es va observar que existien d'altres variables que podrien influir en la probabilitat d'aconseguir gol. Així, es planteja el model següent: la variable dependent pren dos valors, 1 si s'aconsegueix gol i 0 si no s'aconsegueix. Les variables explicatives són:

- X_1 : seria la distància entre el punt mitjà de la porteria situat sobre la línia de gol i el punt del xut a porta;
- X_2 : seria l'angle en radians que formen la recta que uneix el punt des d'on es realitza el xut i el pal més proper a la pilota i la línia de gol;
- X_3 : val 0 si el jugador realitza el xut quan només ha tocat una vegada la pilota i val 1 si l'ha tocat més d'una vegada abans de realitzar el xut a porta;
- X_4 : val 0 si realitza el xut a una distància inferior a una iarda del defensor més pròxim i zero en cas contrari;
- X_5 : val 0 si la possessió de pilota s'ha originat en joc directe i valdrà 1 si és mitjançant una jugada d'estratègia.

Amb aquest model és apropiat utilitzar un model de regressió logístic. Es va realitzar separatament per a rematades de cap i xuts a porta amb el peu.

Xuts a porta amb el peu

Analitzant 410 xuts a porta s'obté que la variable x_3 no és significativa, després el model final queda com:

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

Per la qual cosa la probabilitat de marcar es pot estimar mitjançant la fórmula:

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

Per tant, per a cada possessió de pilota d'un equip amb valors particulars per a les variables x 's es pot estimar la probabilitat d'aconseguir gol. Per exemple, suposem un xut des de 14,56 metres, frontal a porteria amb un oponent a menys d'una iarda i la possessió del qual

es va originar en joc directe, aleshores el valor de $i = -3.328$ i l'estimació de la probabilitat seria $p = 0,35$.

Aquest model també permet d'altres interpretacions. Per exemple, observant el coeficient de la primera variable, $exp(0,219) = 1,24$, indica que per cada iarda que ens acostem a porta la probabilitat de marcar gol augmenta en un 24%. De forma similar, $exp(0,947) = 2,58$, significa que un jugador que té el seu oponent a més d'una iarda augmentarà més del doble la probabilitat de fer gol.

Anàlisis similars es poden realitzar per a les rematades i els llançaments des del punt de la pena màxima.

Aplicació en tàctica

L'efectivitat també es pot utilitzar per quantificar l'actual esdeveniment de possessió de pilota d'un equip, com s'ha descrit a la secció "Anàlisi de la possessió de la pilota". Per tant, per assignar l'efectivitat d'una estratègia particular, es pot calcular la mitjana de les efectivitats de totes les possessions basades en les estratègies esmentades. Així es poden comparar diferents estratègies. Com a exemple simple, suposem un equip que està en la zona 6. Hi ha dues estratègies bàsiques: xutar a porta o realitzar una passada en curt. Utilitzant les dades del mundial de 1986, els xuts a porta van tenir una efectivitat de 21,7 i les passades en curt 3,5. Fent servir un test no paramètric (ja que la distribució dels valors d'efectivitat no segueix una normal) aquestes diferències corresponents a la zona 6 eren clarament significatives ($p < 0,01$).

Paràmetres ofensius i defensius

Característiques del model

Una de les mesures acceptada per la RFEF i la LFP per classificar els equips empatats a punts és la diferència entre els gols aconseguits i els encaixats, tot afavorint, en cas d'igualtat, el qui més gols hagi aconseguit. La justificació d'aquest criteri de presa de decisió és el fet de premiar el



joc ofensiu a efectes de millorar la qualitat de l'espectacle. Però, realment, aquest criteri està fonamentat en una certa lògica? Des del punt de vista matemàtic es pot realitzar una ordenació d'elements si ens basem en un nombre real per a la comparació. Per exemple, els components d'un equip de futbol poden ser ordenats per la seva alçada, des del més baix fins al més alt. La dificultat radica quan volem ordenar elements d'un conjunt comparant al mateix temps dues o més característiques. L'ordenació actual, establerta pel reglament oficial del futbol, s'anomena Lexicogràfica, és a dir, primer ordenem per la característica que considerem més important, (en el nostre cas, el nombre de punts aconseguits) i en cas d'empats s'ordenen els equips segons la diferència de gols a favor, i si novament hi ha empat el que més gols hagi aconseguit. No seria estrany trobar equips que no puguem ordenar, és a dir, que no puguem decidir quin és millor, perquè empaten en tots aquests criteris. Aquest criteri pot presentar fins i tot alguns dubtes sobre la qualitat dels equips pel que fa a les seves referències ofensives i defensives. No oblidem que el futbol es basa en dos blocs tàctics, principis ofensius i principis defensius, i que realitzar-los a la perfecció redunda en la millora de la qualitat de l'espectacle. Per tant, seria interessant crear un altre criteri d'ordenació on els empats siguin "gairebé impossibles" i que, a més a més, sigui un criteri que tingui en compte i valori en justa mesura, tant el joc ofensiu com el joc defensiu.

El model matemàtic

El criteri que es proposa a continuació està basat en el treball de Dixon i Coles, 1997. La hipòtesi de partida és que el nombre de gols marcats per un equip com a local i com a visitant segueix lleis de probabilitat de Poisson, les mitjanes de les quals es determinen per qualitats ofensives i defensives. La *taula 2* presenta una aproximació de la probabilitat del nombre de gols aconseguits a casa pels equips de la Premier League en la temporada 1994-1995. Realitzant un contrast no paramètric de bondat d'ajust no podem rebutjar la hipò-

tesi que els percentatges esmentats són aproximats als corresponents d'una llei de Poisson. Els tants per cent esperats de la llei de Poisson són calculats utilitzant una aproximació a la mitjana amb les dades que tenim, en el nostre cas la mitjana val:

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

De la mateixa forma podem fer amb els gols dels equips visitants, representats a la *taula 3*.

En aquest segon cas, el nombre mitjà de gols per partit aconseguits pels equips visitants és de:

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

A la *taula 4* es presenten els percentatges dels diferents resultats. Així podem estudiar estadísticament si els nombres de gols marcats pels equips locals són independents dels nombres de gols marcats pels equips visitants. Al final s'arriba a la conclusió que no tenim prou evidència estadística com per rebutjar la independència entre aquestes dues variables.

Així, per exemple, en el 8,2% dels partits el resultat final va ser d'empat a zero i en el 4,5% el resultat final va ser de 3 a 1. En definitiva, i després d'aquesta anàlisi prèvia, es pot realitzar el model estadístic

■ **TAULA 4.**
% de resultats en la temporada 1994-1995.

LOCAL	VISITANT				
	0	1	2	3	Més 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
Més de 4	1,6	1,8	1,1	0,6	0,1

■ **TAULA 2.**
Percentatges de gols d'equips locals.

NOMBRE DE GOLS	%	% ESPERATS DE POISSON
0	22,6	24,41
1	33,5	34,42
2	25	24,26
3	13,1	11,40
Més de 4	5,8	5,51

■ **TAULA 3.**
Nombre de gols aconseguits pels visitants.

NOMBRE DE GOLS	%	% ESPERATS DE POISSON
0	33,4	34,10
1	36,4	36,68
2	19,5	19,70
3	7,9	7,00
Més de 4	2,8	2,52

■ **TAULA 5.**
Puntuacions de la temporada 1994-1995 de la Premier League.

EQUIP	ATAC	DEFENSA	POSICIÓ FINAL	POSICIÓ 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

següent. En un partit entre els equips i i j , siguin:

- X_{ij} = Nombre de gols marcats pel local i quan s'enfronta al j .
- Y_{ij} = Nombre de gols marcats pel visitant j quan s'enfronta a l' i .

Els percentatges d'aquestes variables poden ser aproximades per lleis de probabilitat de Poisson, la mitjana de les quals, que en definitiva seria el nombre mitjà de gols per partit, dependrà dels paràmetres ofensius i defensius de cada equip. Així, l'equip i tindrà dos paràmetres, un d'ofensiu (a_i) i un altre de defensiu (b_i). De tal

manera, que el paràmetre ofensiu mesura la raó "d'atac" i el paràmetre defensiu la raó de "defensa". El nombre mitjà de gols marcats per l'equip i com a local quan s'enfronta a l'equip j (m_{ij}) dependrà dels paràmetres anteriors segons el model

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

On $g > 0$ és un paràmetre que ens permet de tenir en compte l'avantatge de jugar a casa. Fent servir el mateix raonament, podem modelar el nombre mitjà de gols marcats per l'equip j com a visitant quan s'enfronta a l'equip i (d_{ij}) de la forma següent:

$$d_{ij} = a_j b_i.$$

Aquests paràmetres, a priori, són desconeguts i cal calcular-los segons que es va desenvolupant la competició. Els mètodes estadístics per obtenir-los són força complexos com per desenvolupar-los aquí, així que fem esment dels treballs de Dixon i Coles (1997) i Dixon i Robinson (1998) per veure'ls de forma més àmplia.

Ordenació paramètrica

Basant-nos en els paràmetres ofensius i defensius calculats anteriorment podem construir un criteri d'ordenació entre equips. Fixem-nos que per a un equip qualsevol l'ideal és que tingui el seu paràmetre ofensiu com més alt millor i el seu paràmetre defensiu com més baix millor. Així podem definir la raó de qualitat de l'equip i com el quocient entre el paràmetre ofensiu i el defensiu:

$$C_i = a_i / b_i.$$

Per tant, un equip serà millor que un altre, com més gran sigui la seva raó de qualitat. Resulta curiós que si ordenem la classificació final d'una lliga, per exemple la classificació final de la temporada 1994-1995 de la Premier League, se n'obté una altra de diferent de la que es va obtenir (vegeu *taula 5*). Encara que en aquella temporada no modificava d'una manera important la classificació (en el nostre exemple solament es veuen afectades les posicions desena i onzena), pot donar-se el cas que en algunes temporades hi hagi empats a punts entre el quart classificat i el cinquè i decidir quin dels dos és millor és una decisió en què hi ha en joc molts milions de dòlars, atès que el quart lloc té dret a jugar la Champions League mentre que el cinquè només juga la Copa de la UEFA; de la mateixa manera pot passar per als llocs de descens.

Altres aplicacions estadístiques

Dintre del futbol hi ha molts tòpics. Per exemple, acostuma a dir-se que amb deu



jugadors es juga millor que amb onze, quan un jugador és expulsat. És obvi que aquest axioma dificulta un dels principis bàsics del futbol per aconseguir la victòria en un partit, que no és altre que assolir la superioritat numèrica allà on es trobi la pilota. Està demostrat estadísticament que una expulsió afecta negativament l'equip infractor (vegeu Ridder *et al.*, 1994). En aquest treball es modela l'evolució del marcador d'un partit de futbol en termes de variables estadístiques, com ara el minut en què un jugador és expulsat, el nombre de gols aconseguits per l'equip abans de l'expulsió, el nombre de gols aconseguits després de l'expulsió esmentada. A part d'aquestes variables, se n'estudien algunes de relatives la conducta dels equips en diferents partits i diferenciant amb quins equips s'enfronten. El model es dissenya sota unes determinades suposicions, com ara:

- Els dos equips van aconseguir gols segons la llei de probabilitat de Poisson en termes de temps i de forma independent l'un de l'altre, és a dir, el nombre de gols marcats per un equip és independent del nombre de gols marcats per l'altre. Aquesta suposició, que a priori sembla irreal, està contrastada estadísticament. Però encara més, també s'ha contrastat que els intervals de temps entre gols són independents i que les intensitats de marcar un gol no són constants tot al llarg d'un partit.
- La intensitat de marcar gol d'un equip amb 11 jugadors només depèn de l'equip al qual s'enfronta, però roman constant al llarg d'un partit encara que depèn d'una funció del temps en el partit.
- Després d'una targeta vermella, la intensitat de marcar gol varia, atès que l'equip passa a tenir deu jugadors, i a partir d'aquest instant es modifiquen les condicions imposades a la hipòtesi 2.

Un altre dels tòpics usuals en el futbol és l'avantatge inicial que posseeixen els equips locals per jugar a casa. Clarke and Norman (1984) van analitzar aquesta situació. En el seu treball van modelar l'ha-

bilitat d'un equip i l'avantatge de ser local de la forma següent:

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

On w_{ij} pot prendre tres valors $\{-1, 0, 1\}$ si l'equip local perd, empata o guanya l'equip j , respectivament. La variable u_i quantifica l'habilitat de l'equip local i , la variable u_j quantifica l'habilitat de l'equip visitant j . La variable h_i quantifica l'efecte de l'avantatge de ser local per a l'equip i . Finalment, en tots els models de regressió s'introdueix una variable anomenada error aleatori de mitjana nul·la, que en aquest cas ve representada per e_{ij} . Una de les suposicions d'aquest model és que l'avantatge de ser local i l'habilitat dels equips romanen constants tot al llarg d'una temporada.

Aquest model va ser aplicat a partits de futbol de la lliga anglesa des de la temporada 1981-1982 fins a la temporada 1990-1991; en total van ser 920 equips i 20.306 partits. Els resultats van demostrar que l'avantatge de ser local varia d'una temporada a una altra i que en algunes temporades hi ha equips amb avantatges de ser local negatives. Es va deduir d'aquest treball que no hi ha diferències significatives per a l'avantatge de ser local entre les diferents categories, que l'any influeix de manera important en el coeficient esmentat i el tipus de club també influeix en determinar l'avantatge de ser local, però no de manera tan determinant com la temporada.

Conclusions

El nostre objectiu en aquest treball ha estat presentar diversos mètodes quantitius per explicar, entendre millor i contrastar la conducta tàctica dels equips de futbol. De les nostres xerrades amb alguns entrenadors i preparadors físics d'equips de primera línia a nivell nacional, arribem a la conclusió que la majoria eren escèptics sobre la utilitat de l'anàlisi de dades avançada per a la millora de la qualitat dels seus equips, basant-se en els aspectes qualitius d'aquest esport. És evident que un dels factors més importants en el futbol d'avui dia és l'habilitat

de sorprendre el contrari amb un gest tècnic o tàctic que no es pugui contraestimar, però no és menys cert que a mesura que avança un partit o una temporada, aquesta habilitat va disminuint per diverses raons i la conducta dels equips es fa menys sorprenent i per tant més constant i fàcil de quantificar. Si posseïm les eines necessàries per conèixer millor la conducta esmentada haurem començat a guanyar el partit abans que la pilota estigui rodant pel terreny de joc. En aquest treball hem desenvolupat algunes de les aplicacions més importants de l'anàlisi de dades avançada aplicada al futbol, com ara l'estudi de l'efectivitat d'un equip en funció de variables explicatives que poden influir significativament en la probabilitat d'aconseguir gol. També s'ha desenvolupat un mètode per explicar, mitjançant paràmetres, la conducta ofensiva i defensiva dels equips.

Bibliografia

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