



UNIVERSIDADE DA CORUÑA

Facultad de Economía y Empresa

Trabajo de
fin de grado

APLICACIÓN DE LA ESTADÍSTICA EN EL ÁMBITO DEPORTIVO

Autor: David Pousa Martínez

Tutora: Pilar Uriz Tomé

Doble Grado en Ciencias Empresariales y Turismo

Año 2021

Resumen

A pesar de que ciencia y deporte conviven desde hace siglos, en las últimas décadas han constituido un importante binomio que se retroalimenta y progresa día a día. En el presente trabajo se pretende analizar en qué medida y de qué manera, el empleo de la de alguna técnica estadística, puede influir en el ámbito deportivo y, más concretamente, en el fútbol.

Vivimos en una sociedad en la que estamos rodeados de infinidad de datos y en la que la simple contabilización de los mismos no se concibe como posible solución. Sin embargo, las avanzadas tecnologías con las que convivimos nos ofrecen la posibilidad de procesarlos y gestionarlos para poder sacarles provecho. Este hecho ha provocado que el análisis de datos haya cobrado una gran importancia para cualquier equipo o deportista, hasta el punto de ser considerado uno de los aspectos más influyentes a la hora de obtener el éxito deportivo.

En este contexto, la estadística se erige como una de las herramientas fundamentales empleadas para el tratamiento de datos, de manera que, a lo largo de esta investigación, no solo conoceremos las diversas aplicaciones que esta puede tener en el ámbito deportivo, sino que también trataremos de demostrar a través de un análisis empírico la relación existente entre estos dos fenómenos: estadística y deporte.

Abstract

Although science and sport have coexisted for centuries, in recent decades they have built an important binomial that feeds back and progresses day by day. This project aims to analyse to what extent and in what way the use of any statistical technique have influence on sport field and, more specifically, on football.

We live in a society in which we are surrounded by an infinity of data and in which the simple accounting of data is not considered as a possible solution. However, the advanced technologies offer us the possibility of processing and managing them in order to make the most of them. This fact has meant that data analysis has taken on great importance for any team or athlete, to the point of being considered one of the most influential aspects to achieving sporting success.

In this context, statistics is one of the fundamental tools used for data processing, so that, throughout this research, we will not only learn about the various applications that it can have in the field of sport, but we will also try to demonstrate through an empirical analysis the relationship between these two subjects: statistics and sport.

Palabras clave:

Deporte, fútbol, estadística, análisis, datos, big data.

Índice

INTRODUCCIÓN	7
1. LA ESTADÍSTICA EN EL ÁMBITO DEPORTIVO	9
1.1. Antecedentes	9
1.2. Función que desempeña la estadística en el ámbito deportivo	11
2. BIG DATA.....	15
2.1. Qué es el big data y su incursión en el deporte	15
2.2. Big data en el fútbol	16
2.2.1. Eventing.....	16
2.2.2. Tracking	18
3. HERRAMIENTAS EMPLEADAS POR LOS PROFESIONALES PARA EL ANÁLISIS.....	20
3.1. Videoanálisis y Media Coach	20
3.2. Sistema de posicionamiento global (GPS).....	21
3.3. La estadística como herramienta en sí misma	23
4. LA IMPORTANCIA DE JUGAR EN CASA EN LOS PARTIDOS DE FUTBOL DE PRIMERA Y SEGUNDA DIVISIÓN.	25
4.1. La importancia de jugar como local	25
4.2. Metodología	28
4.2.1. Técnicas empleadas para el análisis.....	28
4.2.2. Muestra, fuentes y variables.....	32
4.3. Resultados	33
5. EJEMPLOS CONCRETOS	40
5.1. “Método Monchi”	40
5.2. Liverpool de Jürgen Klopp	43
CONCLUSIONES	45
BIBLIOGRAFÍA.....	48

Índice de figuras

<i>Figura 1. Infografía del rol del bioestadístico deportivo</i>	13
<i>Figura 2. Cortador recogiendo los “eventos” de un partido.</i>	17
<i>Figura 3. Jugadores de LaLiga Santander con mejores registros de goles, disparos y pases.</i>	17
<i>Figura 4. Estadísticas empleadas mediante la retransmisión de BeIN Sports</i>	18
<i>Figura 5. Instalación de una de las cámaras encargadas del “tracking</i>	19
<i>Figura 6. Jugadores del FC Barcelona entrenando con sus chalecos que integran el Sistema de Posicionamiento Global (GPS</i>	22
<i>Figura 7. Mapa de calor (a la izquierda) y de toques (derecha) del jugador del FC Barcelona, Lionel Messi, durante el encuentro disputado el 24/10/20 entre el FC Barcelona y el Real Madrid CF.</i>	23
<i>Figura 8. Porcentaje de victorias locales de 5 grandes ligas de distintos deportes.</i>	26
<i>Figura 9. Correspondencias relativas al análisis de la variable “goles</i>	33
<i>Figura 10. Operaciones económicas más exitosas del Sevilla FC en la era Monchi</i>	41
<i>Figura 11. Ingresos y gastos de los equipos de la liga española en la era Monchi (en millones).</i>	42

Índice de tablas

<i>Tabla 1. Descriptivos básicos</i>	34
<i>Tabla 2. Posición final del equipo local en 1ª división</i>	34
<i>Tabla 3. Posición final del equipo local en 2ª división</i>	35
<i>Tabla 4.: Posición final del equipo visitante en 1ª división</i>	36
<i>Tabla 5. Posición final del equipo visitante en 2ª división</i>	36
<i>Tabla 6. ANOVA de un factor para la ventaja de jugar como local</i>	37

Introducción

Debido a la gran importancia que puede llegar a tener en las diversas actividades que realiza el ser humano, entre ellas el deporte, la ciencia de las matemáticas ha sido objeto de estudio prácticamente desde sus orígenes.

El deporte siempre ha generado una gran pasión en el ser humano, siendo este entendido como una parte fundamental de nuestro ADN. Esto no ha cambiado, sin embargo, debido sobre todo al constante avance tecnológico, el deporte ha sufrido una auténtica revolución en todos sus ámbitos y, con ello, también nuestra manera de consumirlo y disfrutarlo.

La sociedad tan tecnológicamente avanzada en la que vivimos provoca que nos encontremos ante una cultura orientada a los datos, no obstante, lo importante no es el incalculable volumen de datos que se genera cada segundo, sino que lo realmente valioso es el arte de saber interpretarlos y obtener así información útil de los mismos. Así, el análisis de datos se ha consolidado como un verdadero hallazgo en el sector deportivo.

Dentro del análisis de datos, la estadística es una de las disciplinas más utilizadas, despertando el interés de un amplio abanico de profesionales de diferentes ámbitos como pueden ser: entrenadores, videoanalistas, cazatalentos, directores deportivos, médicos, preparadores físicos, nutricionistas o incluso los propios futbolistas. Al sustentar sus decisiones en criterios fundamentalmente numéricos frente a otros más subjetivos, este grupo de expertos son capaces de alcanzar una mayor probabilidad de éxito. No debemos olvidarnos de otro de los grandes protagonistas de esta ecuación, el aficionado, que cada vez muestra un mayor interés e interactúa con todo tipo de estadísticas generadas durante cualquier evento deportivo, haciendo de esta manera su experiencia mucho más plena.

De esta manera, a través del presente trabajo trataré de reflejar la importancia que la estadística supone para la analítica, así como su aplicación en el ámbito deportivo.

Asimismo, profundizaremos en conceptos como el de Big Data y seremos conscientes de cómo se lleva a cabo todo el proceso de recopilación, almacenaje y extracción de información útil a partir de la ingente cantidad de datos que inunda nuestra sociedad.

Para ello, comenzaremos analizando como ha ido evolucionando el uso de la estadística en el ámbito deportivo desde sus orígenes y nos adentraremos en el ámbito de la bioestadística.

Posteriormente, abordaremos el fenómeno del Big Data centrándonos en el deporte objeto de este estudio, el fútbol y más concretamente en el modelo empleado por la Liga Profesional de Fútbol Española para su gestión.

Con el objeto de comprender el verdadero efecto que puede llegar a tener sobre el resultado y/o el rendimiento de un futbolista el condicionante de jugar como local o visitante, en el cuarto capítulo se propone y resuelve un análisis empírico obteniendo así conclusiones clarificadoras.

Finalmente, tratando de ejemplificar de qué manera lo visto durante este estudio puede ser implantado en un modelo real, se recogerán una serie de casos reales en los que el uso de la estadística condujo al éxito deportivo.

1. LA ESTADÍSTICA EN EL ÁMBITO DEPORTIVO

1.1. Antecedentes

En el mundo globalizado en el que vivimos hoy en día, las tendencias han cambiado o evolucionado de forma acelerada y, como no podía ser de otra manera, esto se ha hecho notar también en el ámbito deportivo.

Para poder comprender como el análisis estadístico ha ido extendiéndose a lo largo de los años a diversos deportes, es necesario comenzar hablando de la **sabermetría**. La **sabermetría** ha sido definida como el análisis empírico del beisbol a través de la evidencia objetiva, obtenida mediante la aplicación de métodos estadísticos (James & Wirth, 1986). Aunque este término deriva, a su vez, del acrónimo SABR (*Society for American Baseball Research*) (Davids, 1971), sin embargo no sería hasta los años 80 cuando fue acuñado en uno de los afamados artículos de beisbol de Bill James.

El primer libro ligado a la sabermetría, anterior incluso a la popularización del término, fue "*Baseball Percentages*" (Cook, 1964), que supuso una verdadera innovación literaria en estadística deportiva y que serviría de referencia para que posteriormente Erick Walker publicara "*Sinister First Baseman*" (Walker, 1982), obra basada en la construcción de una modelo estadístico a través de la habilidad del bateador y su capacidad de embasarse¹.

Cuando hablamos de sabermétricos, sin duda, el más famoso es George William James más conocido como Bill James. En la década de 1970, este escritor, historiador y estadístico estadounidense, considerado por muchos como el patriarca del análisis matemático del beisbol, comenzó a escribir artículos de este deporte desde un enfoque diferente al que lo hacían los otros estadísticos tradicionales. Mientras que la mayoría

¹ Según la RAE (Real Academia Española); dicho de un jugador de béisbol: alcanzar la base.

de los informes tradicionales eran meras recopilaciones de los eventos del partido y las posteriores opiniones de los jugadores, James, trataba de responder preguntas específicas sobre el comportamiento de los jugadores, a través de una metodología y análisis de datos que distaba de lo común de la época.

El primero de estos artículos fue publicado en 1977 bajo el nombre de “*Baseball Abstracts*”, en los que comentaba los resultados estadísticos de la temporada anterior. Las posteriores ediciones de los también denominados <<abstractos de James>>, obtuvieron un gran aumento en la cuota de lectores, convirtiéndose así en imprescindibles para los aficionados a este deporte (James & Wirth, 1986). La popularidad conseguida por estos artículos, inspiró a muchos sabermétricos a basar parte de su investigación en las estadísticas y parámetros desarrollados en los escritos de Bill James.

Sin embargo, el libro que realmente dio a conocer la filosofía sabermétrica fue, sin duda, “*Moneyball*” (Lewis, 2004). Este libro, llevado posteriormente al cine en 2011 de la mano de Billy Beane y con Brad Pitt como actor principal, cuenta como la filosofía de Bill James encontró un escenario idóneo para reproducirse en los Atléticos de Oakland, equipo profesional de béisbol de la primera división de los Estados Unidos.

De la mano del joven economista, Peter Brand, y debido a los pocos recursos de los que disponían, comienzan a diseñar una estrategia diferente a la que llevaban a cabo los otros equipos de la liga. La gran diferencia es que el sistema era lo suficientemente objetivo y sencillo de interpretar como para encontrar a jugadores que, teniendo un buen rendimiento, habían sido infravalorados debido a que, por ejemplo, lanzaba la pelota de forma poco convencional o no eran tan mediáticos como otros. De esta manera, la estadística permitía encontrar a buenos jugadores por un precio casi ridículo.

Aunque al inicio los resultados fueron negativos, Brand mantuvo firme propuesta con el fin de que sus fichajes demostrarán lo que las estadísticas decían. De esta manera, el trabajo comenzó a dar sus frutos y los Athletics ganaron 20 partidos seguidos, alcanzando un nuevo récord en la Liga Americana y dejando una huella imborrable en la historia del beisbol estadounidense, hasta el punto de que muchos equipos de las grandes ligas copiaron su estrategia deportiva y llegando a influir en otros deportes colectivos.

La sabermetría cambió la forma de concebir el béisbol y constituye un ejemplo perfecto de cómo el uso de la estadística aplicada en diferentes campos o disciplinas es de gran

utilidad. Cabe destacar que, si bien es cierto que fue en el béisbol donde nació y donde más se utiliza, otros deportes como el fútbol americano, el hockey, el baloncesto e incluso el propio fútbol, han incorporado las herramientas de la sabermetría para confeccionar sus plantillas y valorar a los jugadores.

1.2. Función que desempeña la estadística en el ámbito deportivo

La estadística es una disciplina cada vez más esencial en la vida. Todo lo que nos rodea es susceptible de ser medido y cuantificado con el objetivo de poder interpretar nuestro entorno a fin de tomar las decisiones más adecuadas en cada caso y actuar en consecuencia.

El deporte es mucho más que estadísticas, pero éstas son una forma innegable de intentar comprender en profundidad ciertos fenómenos. En el caso de que solo se considerase el factor humano y no lo que aporta la utilización de modelos estadísticos a la solución de muchos problemas en el deporte, los resultados en la preparación deportiva estarían más sujetos a la casualidad y no a la causalidad. Esta es una metodología de análisis cada vez más presente en nuestra sociedad y que nos ayudará no sólo a comprender el origen de determinados hechos sino llevar a procesar y analizar la información para poder emplearla en nuestro beneficio.

El estadounidense Tim Chartier, manifestó en una conferencia magistral que tuvo lugar en 2018 en Panamá, con motivo de la primera Semana de la Ciencia de Panamá organizada por la Secretaría Nacional de Ciencia, Tecnología e Innovación (SENACYT), que las estadísticas y el deporte son "una pareja perfecta". Para hacer más evidente su afirmación puso varios ejemplos de casos reales de éxitos deportivos, en los que los análisis de los datos a través de las estadísticas tuvieron un gran protagonismo.

Por ejemplo, apuntó que, "En baloncesto, los Golden State Warriors saben – a través de los cálculos – que el tercer tiempo es en el que mejor rendimiento tienen y anotan más puntos. Además, a través de la analítica y de los registros de su jugador franquicia, Stephen Curry, mediante una proyección a final de temporada y en base a los playoffs, descubrieron que el cuarto período es el idóneo para que este jugador pueda descansar".

Chartier también señaló que, en el Mundial de fútbol de 2014, Alemania contrató a un grupo de analistas expertos para averiguar que deberían hacer para ganar el campeonato. "En la Copa Mundial de la FIFA de 2014, todo el mundo daba a Brasil como ganador, pero no fue así, los alemanes sabían, a través de la analítica, como alcanzar la gloria", apuntó.

Asimismo, Tim Chartier quiso enfatizar que la analítica no solo sirve de gran ayuda en los deportes más afamados, sino que también es una gran herramienta en deportes menos convencionales.

La estadística deportiva o, a la cuál haremos referencia también, a lo largo del trabajo, con el término de analítica del deporte ("*Sports Analytics*"), es la encargada, a través de la aplicación de modelos estadísticos, de:

- Brindar información objetiva a los deportistas o a su cuerpo técnico, de manera que estos datos sirvan para redireccionar adecuadamente su preparación.
- Ofrece la posibilidad de obtener información objetiva sobre el desempeño de los atletas y para examinar también a los rivales.
- Mejorar la eficiencia en la detección de talentos deportivos.
- Obtener una mayor rigurosidad en el establecimiento de características modelo.
- Conseguir que los entrenamientos se conviertan en verdaderas herramientas de recogida de información confiable para perfeccionar el rendimiento de los atletas.
- Tras verificar su efectividad, diseñar o utilizar un nuevo sistema de método (Vidaurreta Bueno & Mesa Anoceto, 2011).

No obstante, la ventaja más valiosa que aporta el uso de la estadística en el deporte es que permite realizar un análisis más profundo y detallado de los sucesos; proporciona la posibilidad de ir más allá de lo superfluo de los datos, llegando así a conocer el verdadero origen de los resultados, y obteniendo conclusiones más adecuadas para la toma de decisiones futuras.

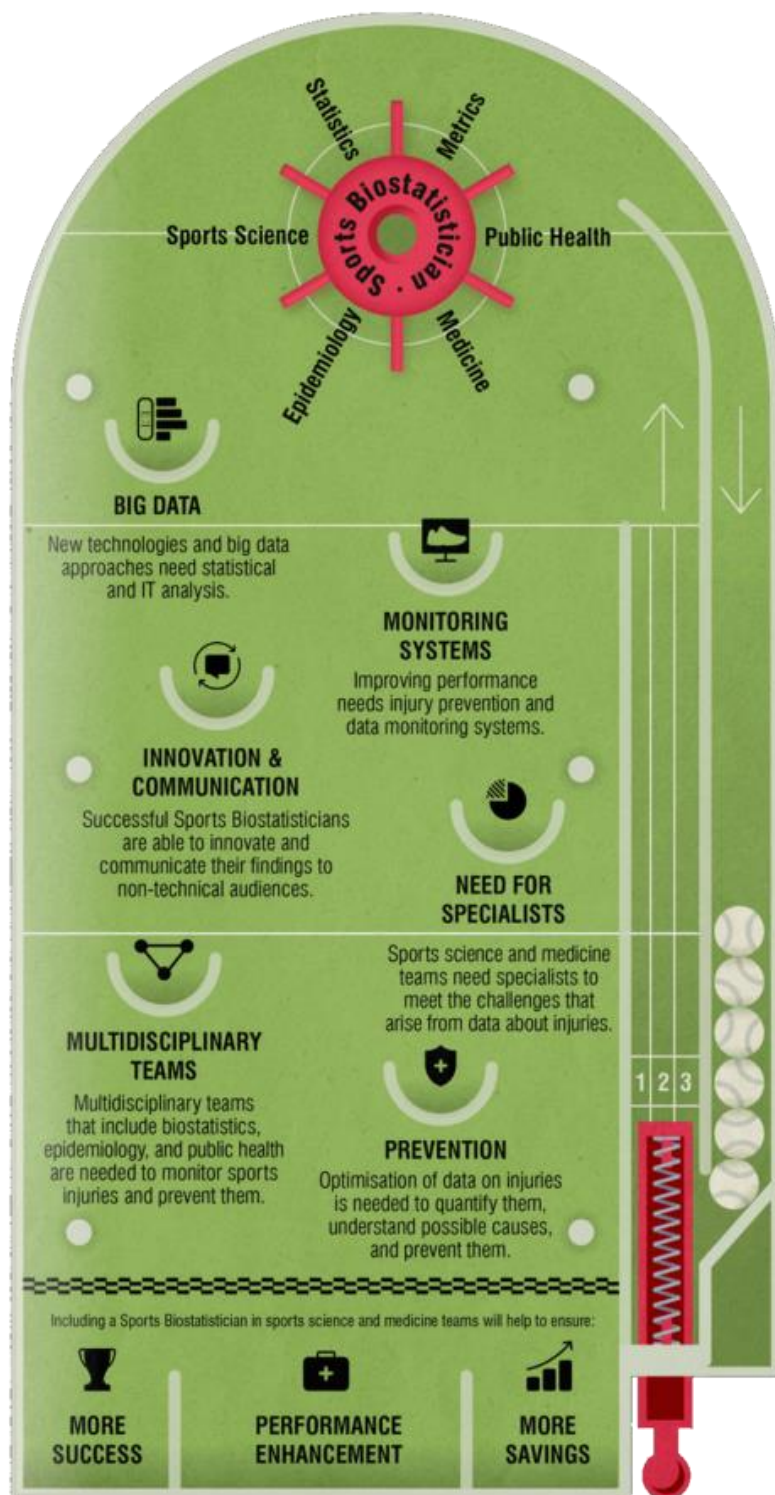
Una de las principales amenazas con las que cuenta un deportista con respecto a la no consecución de sus metas, son las lesiones deportivas, que acarrear graves problemas tanto para el rendimiento como para la salud, por lo que surge aquí la posibilidad de una nueva aplicación de la estadística para minimizar la amenaza de lesiones.

Como ya sabemos, el calentamiento, los ejercicios de estiramiento y un equipo de protección adecuado, son imprescindibles para evitar las lesiones, pero ahora se suma otra variable: la estadística. A finales del siglo XIX el médico francés Pierre Charles Alexandre Louis (1787-1872) introdujo lo que denominó como *méthode numérique*. Con este sistema comenzaría a definirse en medicina el valor numérico de los síntomas, la evolución, la duración de las dolencias y enfermedades, o la eficacia real de los tratamientos (Carreón García, Moreno Altamirano, & García de la Torre, 2018).

Se abrió así el camino que conduciría a la **bioestadística**. La bioestadística se definiría como la “aplicación de métodos estadísticos a la solución de problemas biológicos” (Rolfh & Sokal, 1986).

La evolución de esta rama de la estadística ha derivado en la creación de la figura del bioestadístico deportivo. Los investigadores Caroline Finch y Martí Casals propusieron esta nueva profesión en sus últimos trabajos publicados en las revistas *Injury Prevention* y *British Journal of Sports Medicine (BJSM)*. Según indican en los mismos, puede ayudar a optimizar los datos sobre las lesiones para cuantificarlas, comprender sus posibles causas y poder así prevenirlas (Casals & Finch, 2016).

Figura 1. Infografía del rol del bioestadístico deportivo



Fuente: revista BJSM.

Como podemos apreciar en la *Figura 1*, el bioestadístico deportivo combina habilidades de salud pública, medicina, epidemiología², ciencias del deporte y estadística avanzada.

Los autores de esta publicación (Casals & Finch, 2016) destacan también el apoyo que supone para la prevención de lesiones, el fenómeno del big data o la utilización de sistemas de monitorización, como, por ejemplo, sistemas de monitorización de la frecuencia cardíaca, sistemas de vídeo, de control a través de GPS...

Avanzar en la prevención de lesiones – como vemos en esta figura – exige un mayor número de equipos multidisciplinarios que combinen todas estas ciencias, ya que, supone un problema de salud pública y se requiere una mayor disponibilidad informativa de calidad sobre lesiones.

Estos autores también consideran que los bioestadísticos deportivos deben tratar de explicar sus ideas, esencialmente estadísticas, de manera que todos los involucrados en la comunidad deportiva, como entrenadores, deportistas, preparadores físicos, fisioterapeutas... puedan ser conscientes de su utilidad para la prevención y la predicción de lesiones.

Estados Unidos y Australia son considerados los países a la cabeza de la expansión de la bioestadística y llevan ya muchos años apostando por este tipo de profesionales. El fútbol australiano y el cricket son dos de los deportes en los que más se han aplicado estos nuevos conocimientos, pero Estados Unidos se ha sumado rápidamente, de manera que el baloncesto, el beisbol y el fútbol americano ya incorporan la bioestadística.

En Europa, especialmente en los países nórdicos y Reino Unido, comienzan a interesarse tanto por la tecnología como por la figura del bioestadístico deportivo y su inclusión en el deporte profesional.

² El Diccionario de Epidemiología (Porta, 2014) define la epidemiología como «el estudio de la aparición y distribución de sucesos, estados y procesos relacionados con la salud en poblaciones específicas, incluyendo el estudio de los determinantes que influyen en esos procesos y la aplicación de este conocimiento para controlar los problemas de salud relevantes». Por lo tanto y más concretamente, la epidemiología de las lesiones deportivas es el estudio de la distribución y las variables que intervienen en la aparición de lesiones en grupos de población, con el propósito de establecer medidas de prevención.

2. BIG DATA

2.1. Qué es el Big Data y su incursión en el deporte

Previamente a explicar la llegada de este fenómeno y el papel que desempeña en el mundo del deporte y, más concretamente, en el fútbol, debemos entender a qué se refiere este término.

El **big data** hace referencia a una gran cantidad o volumen de datos que se intentan procesar para extraer información. La utilidad del mismo suele definirse de acuerdo con sus “cinco uves” puesto que permite analizar una gran **variedad** y un gran **volumen** de datos, a gran **velocidad**, de forma **veraz** y otorgando **valor**.

Como mencionábamos al principio de este trabajo, el origen de obtener datos estadísticos en el deporte se remonta a la época en la que Bill James comenzó a analizar a los jugadores de béisbol, haciendo anotaciones todavía a papel y lápiz. El paso de los años, la constante evolución tecnológica y el comienzo de la informatización, supuso la posibilidad de conseguir y tratar un gran volumen de datos. El big data otorga la posibilidad de gestionar esta ingesta cantidad de datos, una cuantía que no podría ser procesados por un software convencional o un ordenador personal (Gamboa, 2019).

Sin duda, el análisis del deporte a través del big data ha transformado este ámbito, sobre todo a niveles profesionales, donde su uso es más significativo. No obstante, es cierto que hemos empezado a prestar atención a esta forma de entender el deporte desde hace relativamente poco, puesto que hace 20 años esta concepción era impensable. Por tanto, nos encontramos todavía en la primera etapa de una nueva y revolucionaria era en el deporte: **la era del Big Data** (Silva, 2015)

Cualquier dato que se genere, y durante el desarrollo de un evento deportivo se producen una innumerable cantidad de ellos, es susceptible de ser analizado: un golpeo de balón, la longitud de ese lanzamiento, la velocidad en una carrera o el porcentaje de acierto en lanzamiento de penaltis.

Es importante enfatizar la importancia del valor humano en el análisis, ya que los datos sin un contexto en los que enmarcarlos no tendrían tanta relevancia. Así, los analistas deportivos relacionan una serie de variables del entorno, un contexto y una situación, de manera que, aplicando su capacidad de interpretación, podrán mejorar o controlar distintos aspectos del juego.

Una de las posibilidades más útiles que nos brinda el análisis de los macrodatos es la de la predicción, ya que, si se da la posibilidad de encontrar patrones, es muy probable que estos vuelvan a ocurrir con cierta frecuencia (Román, 2018).

2.2. Big data en el fútbol

El big data se ha ido introduciendo paulatinamente en los diferentes deportes como en la NBA, la NFL o el beisbol y, aunque no lo ha hecho de forma simultánea, finalmente también ha llegado a un mundo tan impredecible como es el fútbol.

Puesto que, como todos los deportes, el fútbol también es un negocio en el que los errores pueden ocasionar una pérdida de millones de euros, los profesionales del 'big data' han conseguido un especial protagonismo, con el objetivo de intentar reducir lo máximo posible ese margen de error. Este es el nuevo fútbol, el de los datos, en el que todavía existe un gran potencial por explotar.

Pero en esta nueva era del fútbol, ¿cómo se lleva a cabo la recogida de todos estos datos? ¿Cómo se llegan a, por ejemplo, generar estadísticas específicas con respecto a determinados jugadores? Para poder contestar a estas preguntas, debemos conocer las diferencias entre "eventing" y "tracking".

2.2.1. Eventing

El término "**eventing**" hace referencia a los propios eventos que suceden durante el partido ya sean pases, remates, goles, etc. Estos datos se recogen manualmente a través de los llamados *taquígrafos* del fútbol o *cortadores*, de manera que cada partido es analizado por tres personas. Dos de estas personas se encargan de registrar los

datos del equipo que les fue asignado, mientras que el tercer analista se encarga de revisar que todos estos datos se hayan recogido correctamente. Cada analista utiliza un teclado con el que registran los “eventos”, basados en cuatro variables: de qué acción se trata (pase, gol, asistencia, regate...), en qué parte del campo ha sucedido, en qué momento y qué jugador (Hidalgo, 2018). Gracias a este proceso podemos saber que, por ejemplo, en la temporada 2019/2020, de media, se marcaron 2,4 goles y se produjeron 21 disparos por partido (Torres, 2020).

Figura 2. Cortador recogiendo los “eventos” de un partido.

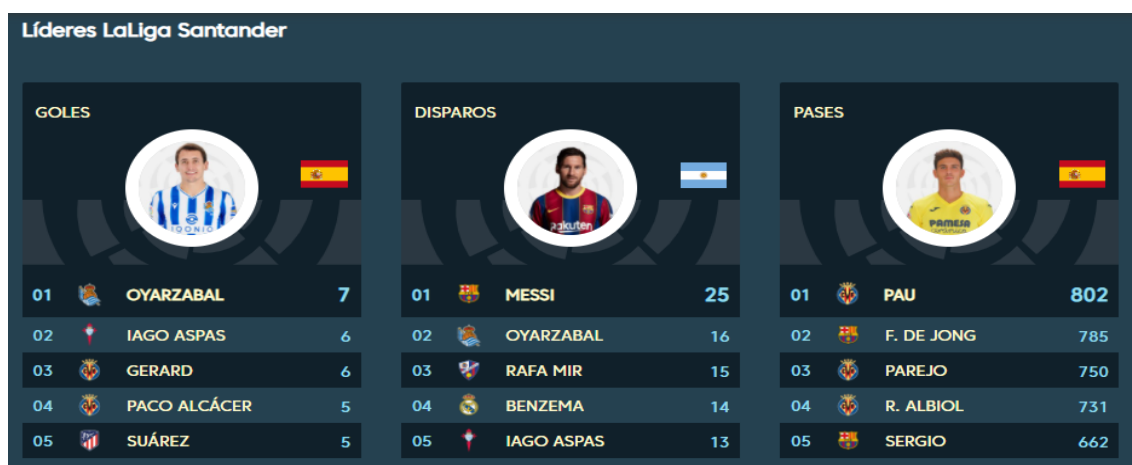


Fuente: Opta Sports

Debido al auge que experimentan hoy en día a nivel mundial tanto el fútbol como el Big Data, han surgido una gran cantidad de empresas dedicadas a ofrecer este tipo de estadísticas.

Una de las primeras fue **Opta Sports**; empresa nacida en 1996 en Inglaterra con el objetivo de proporcionar información a todos aquellos aficionados al fútbol con curiosidad por conocer los datos que generaban sus jugadores y equipos favoritos. Tal fue su éxito, que solamente un año más tarde comenzó a analizar partidos a nivel profesional de la Premier League (liga de fútbol inglesa) y fue contratada por SkySports, principal cadena de televisión deportiva en Reino Unido, para sus retransmisiones televisivas (Carranza, 2017).

Figura 3. Jugadores de LaLiga Santander con mejores registros de goles, disparos y pases.



Fuente: datos obtenidos de la web oficial de LaLiga y recopilados por Opta, proveedor oficial de datos de la competición.

Gracias a su vocación internacional, a los servicios que ofrecen y a la gran rapidez con la que trabajan, han ido traspasando fronteras y aumentando su cobertura hasta diferentes países y deportes. Esto les ha llevado a alcanzar un acuerdo con LaLiga en la temporada 2017-2018 para ser el **proveedor oficial de datos de la competición española**, tanto en primera y segunda división, como en la primera división femenina.

Cabe destacar también que Opta Sports, además de ser el proveedor oficial de datos tanto de distintas competiciones como de clubes, trabaja estrechamente con multitud de medios de comunicación punteros como pueden ser *France Football*, *L'Equipe* o la *BBC*. En España podríamos destacar su colaboración con los medios de prensa *Marca* y *AS* o con la cadena de televisión de eventos deportivos *BeIN Sports* (Perform group, s.f.).

Al igual que Opta Sports, existen otros proveedores de datos deportivos entre los que sobresalen Instat, Mediacoach, Stats Prozone, Tracab o Beemray.

Figura 4. Estadísticas empleadas mediante la retransmisión de BeIN Sports



Fuente: Opta Sports

2.2.2. Tracking

Por otro lado, el **“tracking”** es un sistema de seguimiento óptico que se lleva a cabo de forma automática y cuya función es obtener la posición de los jugadores y la pelota durante el transcurso del partido.

En cada liga de fútbol, el método de “trackeo” se realiza de una forma diferente. Centrándonos en la liga española, el sistema de “tracking” utilizado es proporcionado por el **sistema Tracab** desarrollado por la empresa ChyronHego.

El seguimiento se realiza a través de 4 cámaras que han sido instaladas en todos los estadios de primera y segunda división. Estos dispositivos proporcionan información a 25 fotogramas por segundo, por lo que en cada uno de esos fotogramas recopila información sobre los 23 protagonistas del terreno de juego (22 futbolistas y el balón). Dispondremos así de las coordenadas X (jugadores equipo local), Y (jugadores equipo visitante) y Z (pelota). De manera que, considerando que un partido de fútbol tiene una duración de unos 5.400 segundos, obtendremos en torno a 135.000 fotogramas por partido (25 fotogramas/segundo x 5400 segundos), lo que multiplicado por cada una de las coordenadas existentes (22+1), provocará que se recopilen más de 3 millones de puntos aproximadamente por partido.

Esto permite crear una especie de mapa 3D en tiempo real, a partir del cual se pueden obtener medidas como las zonas en las que se produjo la posesión del balón, distancia recorrida, velocidad máxima alcanzada o número de pases. Gracias a estas mediciones, después también es posible realizar estudios más globales de los equipos, como la posesión de balón, los tiros a puerta totales, los mapas de calor...(Gamboa, 2019).

Figura 5. Instalación de una de las cámaras encargadas del “tracking”



Fuente: imagen obtenida de la página oficial de ChyronHego (chyronhego.com)

3. HERRAMIENTAS EMPLEADAS POR LOS PROFESIONALES PARA EL ANÁLISIS

3.1. Videoanálisis y Media Coach

En un mundo tan tecnológicamente avanzado como en el que vivimos, a un analista o scouter futbolístico de primer nivel se le exigen altas competencias informáticas en el empleo de herramientas como las de videoanálisis.

El principal objetivo del videoanálisis es el estudio, mediante la visualización de videos, de las acciones que tienen lugar en un partido o entrenamiento de fútbol, ya sea un análisis individual o grupal. Se trata de obtener una radiografía tanto de tu equipo como del rival, puesto que permite analizar, por ejemplo, sus tácticas, las fases del partido en las que se muestran más vulnerables, sus puntos fuertes y débiles o los patrones de comportamiento de sus estrellas dentro del campo (Pérez, 2018).

Existen una gran cantidad de softwares empleados a diario por los cuerpos técnicos de los equipos de mayor nivel mundial como podrían ser Er1c, Nac Sport y Sportcode. Sin embargo, en este apartado nos centraremos en el análisis de Media Coach.

Media Coach es una herramienta de videoanálisis creada en 2011 para los clubes de 1ª y 2ª división de España a través de un importante acuerdo entre La Liga de Fútbol Profesional Española y el grupo líder en el sector audiovisual europeo, Media Pro.

El punto fuerte de este sistema, que lo sitúa como la primera opción a la hora de elegir dentro del mercado del análisis de los partidos, es que se caracteriza por generar postproducción de vídeos, informes y estadísticas de manera muy sencilla e intuitiva ofreciendo una cobertura total en un mismo software. Esto permite ahorrar tiempo a los

analistas y miembros de los cuerpos técnicos a la hora de editar los vídeos, recibiendo información pre-partido (Mediacoach Desktop), en vivo (Mediacoach Live) y post-partido (Mediacoach Portal) (Sanchez, 2015).

Para elaborar sus contenidos se nutre de **dos proveedores**. Por un lado, como ya comentamos anteriormente, las cámaras instaladas en los estadios serán las encargadas de proporcionar información a través del proceso de “tracking”, mientras que, por otra parte, Opta, la empresa proveedora de datos deportivos oficial de LaLiga, se encarga del proceso de “eventing” y de recopilar todos aquellos datos que son más objetivables (disparos, pases, faltas, corners...) (Olmeda, 2020).

Es importante enfatizar la democratización que supuso su llegada en el mundo del análisis, puesto que, como explica su director general, Nacho Roca (2011), este tipo de herramientas de análisis se comercializaban por un precio en torno a 140.000€ por temporada, asumible únicamente por los equipos con mayor capacidad económica.

Con este acuerdo, esa cifra se rebajó a una media de en torno a 30.000 euros por equipo, de manera que los equipos más modestos de primera división disponen de la misma herramienta y, por tanto, mismos datos en calidad y cantidad, que los grandes equipos (Martínez, 2019).

Sin embargo, la recogida de datos sin un análisis de los mismos se convierte en una función superflua. Es en ese punto donde estriba la gran diferencia entre unos equipos y otros, de manera que la cantidad de recursos humanos destinados a sacarle el máximo partido a este servicio se consolida como la principal ventaja competitiva ya que, mientras que algunos clubes cuentan con una plantilla de hasta quince analistas, otros apenas poseen una única persona trabajando en ese análisis de los datos

3.2. Sistema de posicionamiento global (GPS)

Por otro lado, los equipos cuentan también con sus propios métodos de geolocalización como son los **Sistemas de Posicionamiento Global (GPS)**. Este tipo de dispositivos *wearables* o tecnología “vestible” adoptan, generalmente, la forma de un chaleco inteligente que es colocado en el torso del deportista.

A pesar de que en su origen fueron creados para deportistas que recorrían largas distancias durante el entrenamiento, posteriormente, gracias al avance que ha experimentado la tecnología y el desarrollo de softwares específicos, su uso se ha ido expandiendo a otras disciplinas, destacándose su introducción en los deportes colectivos (Dobson & W. L. Keogh, 2007).

Dentro de estos últimos estaría enmarcado el fútbol, deporte en el que hasta 2015 estos dispositivos solo podían ser empleados en los entrenamientos. Sin embargo, ese mismo año, el uso de sistemas portátiles de monitorización del rendimiento durante los encuentros, fue aprobado por la Federación Internacional de Fútbol Asociado (FIFA). Este fue un hecho muy significativo puesto que otorgaba a los analistas técnicos la posibilidad de transmitir datos de gran valor al entrenador durante el transcurso de los partidos. Seis años más tarde, estos dispositivos se han convertido en herramientas indispensables para la gran mayoría de los equipos profesionales (Pulgarín, 2018).

Figura 6. Jugadores del FC Barcelona entrenando con sus chalecos que integran el Sistema de Posicionamiento



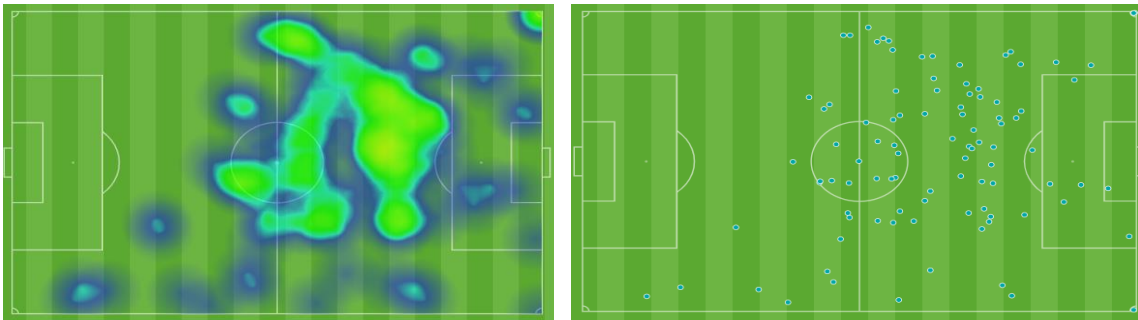
Fuente: foto de Miguel Ruiz (fotógrafo oficial del FCB)

Los GPS pueden ser empleados para cuantificar los niveles de estrés y esfuerzo físico a los que son sometidos los deportistas, evaluar la carga de trabajo y planificar las pertinentes sesiones de entrenamiento a partir de estos resultados, prevenir o recuperarse de lesiones o incluso para establecer un plan nutricional para un determinado jugador en función de la fatiga y desgaste físico que presenta. Estas son algunas de las múltiples aplicaciones en el deporte de los Sistemas de Posicionamiento Global, cuyo fin primordial es el de optimizar el rendimiento deportivo (Aughey, 2011).

La gran cantidad de información que puede ser recopilada por estos dispositivos se divide en tres segmentos:

- **Fisiológico:** hace referencia a parámetros como frecuencia cardíaca, niveles de oxígeno o fatiga.
- **Cinemático:** recopila información relativa a la velocidad, aceleración, desaceleración o distancia recorrida.
- **Parámetros de singularidad táctica:** podemos destacar los usualmente empleados mapas de calor, que permiten ver el espacio en el campo de juego en el que más se movió el jugador

Figura 7. Mapa de calor (a la izquierda) y de toques (derecha) del jugador del FC Barcelona, Lionel Messi, durante el encuentro disputado el 24/10/20 entre el FC Barcelona y el Real Madrid CF.



Fuente: Opta Sports.

3.3. La estadística como herramienta en sí misma

El avance tecnológico permite capturar un amplio volumen de datos y procesarlos incluso para su análisis en tiempo real, lo que provoca que, hoy en día, un gran número de las decisiones que se toman en el mundo del fútbol se amparen bajo los datos y dejen a un lado las “corazonadas”.

Como acabamos de ver en los apartados anteriores, las acciones de los jugadores en un partido o un entrenamiento pueden ser capturados por multitud de fuentes y dispositivos. Sin embargo, lo que realmente convierte en útil esta información es el procesamiento de la misma.

De esta manera, la estadística se constituye ya como esencial en el mundo del fútbol y como una herramienta en sí misma para elaborar tendencias, patrones o identificar correlaciones que, bajo la simple observación, parecían ocultas pero permiten valorar o

comprender mejor el rendimiento del deportista o de un equipo a lo largo de una temporada (Vidaurreta Bueno & Mesa Anoceto, 2011).

Los métodos estadísticos además de facilitar la recogida y organización de los datos, posibilitan conocer si se han cumplido o no los objetivos trazados, aportan información complementaria al entrenador para el proceso de toma de decisiones y permitan conocer mejor este deporte y los distintos factores que lo rodean alcanzando así rendimientos deportivos muy elevados. Estas técnicas, como veremos a lo largo del apartado 5, incluso favorecen la revolución de áreas tan importantes en el fútbol como son los fichajes y la captación de talento.

Finalmente, señalar que es recomendable que todo este proceso de extracción de información se lleve a cabo bajo un método que asegure la fiabilidad y evite distorsiones, incluyendo soportes informáticos o softwares que ayuden a completar esta metodología (Alvarez Medina, Parra Artal, Murillo Llorente, & García Felipe, 2018).

4. LA IMPORTANCIA DE JUGAR EN CASA EN LOS PARTIDOS DE FUTBOL DE PRIMERA Y SEGUNDA DIVISIÓN.

4.1. La importancia de jugar como local

Como se viene indicando a lo largo del trabajo, en los últimos años hemos sido testigos de una tendencia hacia la profesionalización y especialización en el ámbito deportivo. Con el objetivo de alcanzar el éxito se han llevado a cabo una gran cantidad de estudios y trabajos a fin de identificar los distintos parámetros y condicionantes que pueden afectar a los deportistas.

La ventaja de jugar en casa, también denominada **“home advantage” (HA)**, es un importante factor a tener en cuenta a la hora de analizar las distintas competiciones y modalidades deportivas. Precisamente, este hecho ha provocado que sean muy numerosas las investigaciones realizadas sobre esta materia hasta la fecha.

El efecto provocado en el rendimiento al competir como local se empezó a investigar en la década de los 70 (F.Barsky & Schwartz, 1977) cuando se demostró que equipos de baloncesto, hockey, béisbol y fútbol americano obtenían unos mejores resultados cuando jugaban en casa. Hoy en día son muchos los estudiosos que siguen estos cauces y siguen investigando a cerca de la influencia del factor campo en los distintos deportes. De esta manera, se ha llegado a comparar incluso la incidencia de este factor en las distintas modalidades deportivas.

Figura 8. Porcentaje de victorias locales de 5 grandes ligas de distintos deportes.

Liga	Partidos ganados como local
MLB (béisbol)	53.9%
NHL (hockey sobre hielo)	55.7%
NFL (fútbol americano)	57.3%
NBA (baloncesto)	60.5%
MLS (fútbol)	69.1%

Fuente: tabla de elaboración propia a partir del artículo publicado por (J.Dubner, 2011)

Stephen J. Dubner analizó en 2011 la ventaja que obtenían los equipos locales en las ligas más importantes de los deportes más practicados y con más seguidores en USA. Atendiendo a la definición de (V.Carron & Courneya, 1992) la ventaja de jugar en casa se produce cuando los equipos locales ganan más del 50% de los partidos en las competiciones de ida y vuelta. Así, como podemos apreciar en la tabla anterior, la presencia de valores superiores al 50%, pone de manifiesto el hecho de que jugar en casa supone una ventaja a favor de los locales. Entre los distintos deportes se observan oscilaciones en los porcentajes de victorias como local, que pueden ser debidas a las diversas características de cada uno de ellos, así como a la distinta manera de organizar el sistema de competición en cada caso ya que, por ejemplo, mientras que en la NBA las eliminatorias finales se deciden a favor del primero que gana 4 encuentros de 7, en la NFL el ganador es el que sale victorioso de un único enfrentamiento.

Es importante destacar que, a la hora de analizar el factor campo como ventaja para el equipo local no debemos centrarnos única y exclusivamente en la propia localización en la que tiene lugar el evento deportivo, es decir, el estadio, pues son muchos los **condicionantes** que pueden también influir en el momento de competir bien como visitante o bien como local. Dentro de los numerosos estudios que se han realizado sobre los factores que pueden tener incidencia en el rendimiento (V.Carron & Courneya, 1992) se proponen 4 posibles variables que afectan al beneficio de jugar en casa:

- **Familiaridad:** asociado a las características del césped, las dimensiones del terreno de juego (pueden variar dentro de unos límites pautados), las condiciones meteorológicas predominantes en la zona y a las que están acostumbrados los jugadores, el sentimiento de encontrarse en un ambiente familiar y amigable contra la hostilidad percibida por los visitantes...

- **El desplazamiento:** en el caso de que el equipo visitante tenga que recorrer largas distancias hasta el lugar en el que se producirá el evento, sus jugadores podrían acusar fatiga debido al viaje realizado.
- **Territorialidad:** definida por (Neave & Wolfson, 2003) como el “sentimiento de proteger un determinado lugar o ejercer un sentido de territorialidad como respuesta a la percepción de invasión”, este sentimiento de territorialidad se suele relacionar con una mayor producción de testosterona por parte de los deportistas cuando compiten en casa y con conductas y acciones más agresivas para intimidar al rival.
- **Apoyo de la propia afición:** podemos asociar este factor con dos fenómenos que se complementan: la *facilitación social* y la *teoría del impulso*.

El concepto de la *facilitación social* hace referencia a que la asistencia de público afecta al rendimiento de los deportistas pues ejerce influencia sobre el mismo, así como sobre su actuación en el juego (Cox, 2008). Por otra parte, la *teoría del impulso* manifiesta que la presencia de un determinado impulso (en este caso la activación por parte del público) provoca una respuesta dominante³, en este caso, en el deportista.

Algunos autores, (Zajonc, 1965) fueron más allá fusionando ambas teorías, demostrando que si se produjese un mayor impulso por parte del público (a través de silbidos, aplausos, abucheos, ovaciones...), esto conllevaría un aumento de las respuestas dominantes tanto positivas como negativas. La diferencia radica en que aquellos deportistas que cuenten con una mayor experiencia, mejores cualidades técnicas y tengan un mayor control en este tipo de situaciones experimentarán una mejora de su rendimiento, mientras que, aquellos que no tengan tanta experiencia, no tengan tan desarrolladas sus capacidades o no cuenten con tanto dominio sobre estas situaciones, se verán perjudicados por el aumento de esta activación. Aquí es donde puede comenzar

³ Respuesta que se produce con mayor probabilidad ante una situación específica. La misma puede ser debida a entrenamiento, costumbre, preferencias personales o factores innatos. De este modo, en tareas bien ejercitadas o fáciles la respuesta dominante será la correcta, mientras que en tareas de mayor complicación o mal aprendidas la respuesta dominante tenderá a ser la incorrecta (Zajonc, 1965).

a actuar también otro condicionante como es el conocido como “miedo escénico”.

El “miedo escénico” fue relacionado con el fútbol por primera vez en 1986 cuando Jorge Valdano lo empleó en un artículo de la revista *occidente*. El exjugador y exentrenador argentino, valiéndose de la expresión acuñada en su origen por Gabriel García Márquez, trataba de hacer referencia con ella al temor que puede sufrir un futbolista al saltar al terreno de juego ante la incertidumbre que le provoca no conocer el resultado de su futura actuación (El miedo escénico, 1994). El hecho de que la decisión tomada por un deportista pueda ser correcta o errónea, provoca que en muchas ocasiones aparezca el miedo a fallar. Este sentimiento puede aflorar tanto internamente, a partir de las propias sensaciones que condicionan al sujeto, así como de manera externa, es decir, a partir de la influencia ejercida por el entorno social (hinchada, influencia del entorno familiar, opiniones externas...) que componen esa atmósfera particular que rodea una competición (Vigara, 2017).

Hoy en día, debido a la pandemia a la que nos enfrentamos, las restricciones han obligado a la no asistencia de público en la mayor parte de los eventos deportivos. Esto ha provocado que el “miedo escénico” se haya visto reducido enormemente y que la influencia del entorno local en el equipo visitante cada vez sea menor. Debido a la no exigencia del público local sobre el equipo, así como la falta de intimidación que sufría el equipo rival, los partidos a puerta cerrada han cambiado totalmente la competición. Más adelante profundizaremos sobre en qué medida esta situación ha modificado los resultados obtenidos tanto por los equipos locales como por los visitantes.

4.2. Metodología

4.2.1. Técnicas empleadas para el análisis

A través del siguiente análisis empírico trataremos de investigar hasta qué punto la condición de jugar como local tiene un efecto significativo tanto en los resultados obtenidos como en el rendimiento de los jugadores.

Antes de pasar a comentar los métodos utilizados en este estudio para el análisis y representación de los datos es importante realizar una distinción entre las distintas **variables** que nos podemos encontrar. En primer lugar, una variable estadística es una característica que puede presentar diferentes valores (variables cuantitativas) o modalidades (variables cualitativas) y cuyas oscilaciones pueden medirse u observarse. Esta distinción es de gran importancia debido a que las técnicas estadísticas que es adecuado utilizar difieren según sean de uno u otro tipo.

Una de las técnicas utilizadas en esta investigación es la que analiza la independencia/relación de las dos variables que forman las tablas de contingencia. Estas tablas de doble entrada están compuestas por filas (representan una variable) y columnas (representan la otra variable), cada una de ellas con sus diferentes modalidades o niveles, dando así lugar a celdas o casillas que contienen la información necesaria para interpretar los resultados.

A través de estas tablas trataremos las variables cualitativas estudiándolas a partir de su único aspecto cuantificable: la frecuencia. Las tablas de contingencia representan la distribución bidimensional de frecuencias de manera que se pretende estudiar la posible relación entre dos variables cualitativas y, en el caso de que esta exista, medir o cuantificar el grado de asociación y, en algunos casos, la dirección de la misma.

Cuando creamos tablas de contingencia, la información contenida en las casillas puede venir expresada de diversas formas; en nuestro caso trabajaremos con las siguientes medidas:

- **Frecuencias conjuntas:**
 - **Observadas.** Número de casos reales u observados conjuntamente para cada par de niveles de cada una de las variables (n_{ij})
 - **Esperadas.** Valor teórico o esperado; es decir el número de casos que debería tener cada casilla suponiendo que la hipótesis de independencia de las variables fuese cierta (\hat{E}_{ij}); es decir suponiendo que

$$\hat{E}_{ij} = \frac{n_{i.} * n_{.j}}{n} ; \quad \forall i, j$$

Siendo n el tamaño de la muestra.

- **Frecuencias (totales) marginales**
 - **Por filas.** Sumas de las frecuencias conjuntas de toda la fila ($n_{i.}$)

- **Por columnas.** Suma de las frecuencias conjuntas de toda la columna (n.j)
- **Porcentajes:**
 - **Por filas.** Valor de cada casilla con respecto al total marginal de su fila.
 - **Total.** Valor que la frecuencia observada de una casilla representa respecto al número total de casos.
- **Residuos:** El análisis de los residuos resulta de interés ya que el patrón de os mismos nos indicará cuáles son las combinaciones de niveles de la tabla que, en su caso, causan la asociación existente entre las variables. Dado que dicho patrón no siempre es sencillo de detectar se suele recurrir a su comparación con una distribución de probabilidad conocida; por ello es preciso definir:
 - **Residuos.** Representan las diferencias entre las frecuencias observadas y esperadas ($n_{ij} - \hat{E}_{ij}$)
 - **Residuos estandarizados o tipificados (eij).** Tienen una distribución normal con esperanza 0 y con varianza V_{ij}

$$e_{ij} = \frac{n_{ij} - \hat{E}_{ij}}{\sqrt{\hat{E}_{ij}}}$$

La estimación de la varianza viene dada por

$$\widehat{V}(e_{ij}) = (\widehat{V}_{ij}) = \left(1 - \frac{n_{i.}}{n}\right) \left(1 - \frac{n_{.j}}{n}\right)$$

- **Residuos corregidos o ajustados (dij).** Se distribuyen asintóticamente como normal con media cero y desviación típica 1, por lo que, por ejemplo un dij mayor que 1,96 indica que la desviación entre n_{ij} y \hat{E}_{ij} es significativamente distinta de cero con un nivel de significación del 5%.

$$d_{ij} = \frac{e_{ij}}{\sqrt{\widehat{V}_{ij}}}$$

Al distribuirse normalmente con media 0 y desviación típica 1 su interpretación es muy sencilla pues, empleando un nivel de confianza del 95%, los residuos mayores de 1,96 nos mostrarán celdas con más casos de los que debería de haber si existiese independencia entre las variables, mientras que, por el contrario, los residuos menores de -1,96 señalarán casillas con menos casos de los que deberían existir bajo condición de independencia.

En este TFG, las **variables que formarán las tablas de contingencia** son: posición en la clasificación final (con 4 niveles tanto en primera como en segunda división: posición 1 – 5, posición 6 – 10, posición 11 -15 y posición 16 – 20 si es primera división y posición 16 – 22 si es segunda división) y resultado (con tres niveles: victoria local, victoria visitante y empate).

Posteriormente, llevaremos también a cabo un análisis de la varianza (ANOVA). El **ANOVA** es una técnica de análisis multivariante que se agrupa dentro de las de análisis de la dependencia o explicativas, es decir, entre aquellas en las que es posible distinguir entre variables explicativas (independientes) y explicadas (dependientes). El objetivo del ANOVA es estudiar la relación entre una o varias variables dependientes cuantitativas y una o varias variables independientes cualitativas (factores). En otras palabras, estudia en qué medida una o varias variables independientes o factores no métricos influyen en el comportamiento de una o varias variables dependientes métricas.

En este caso, el modelo empleado será el de **análisis de la varianza de un solo factor**, división, con dos niveles (1ª y 2ª división) y con, en cada caso, una variable dependiente cuantitativa que será: puntos, goles total, goles locales, goles visitantes y goles, lo que dará lugar a cinco ANOVA diferentes.

El modelo de análisis de la varianza de un factor supone determinar si el valor esperado de la variable dependiente es el mismo (igual población) o es diferente (distinta población) en cada uno de los niveles del factor. Es posible observar que las medias muestrales de los distintos grupos son diferentes; la cuestión es determinar si dichas diferencias son significativas, de tal manera que supongan que las muestras no proceden de una única población. En este TFG lo que se pretende es determinar si el valor esperado o medio de cada una de las variables dependientes cuantitativas son iguales o distintas en cada uno de los niveles (1ª y 2ª división); es decir, evaluar si el factor división puede explicar en cierto modo los cambios que se producen en: puntos, goles locales, goles visitantes...

La homogeneidad de los datos suele medirse a través de su varianza o de su suma de cuadrados; se trata de ver cuánta dispersión tienen los valores de la variable respecto a su media: cuanto menor dispersión respecto a la media global mayor será la homogeneidad. Por ello se trata de descomponer la suma de cuadrados total (SCT) en dos partes: la provocada por los diferentes niveles del factor (SCE) y la provocada por la propia variabilidad muestral (SCD); obviamente cuanto mayor sea SCE mayor sea la influencia del factor.

Teniendo esto en cuenta, se calcula el estadístico del contraste (F), que viene dado por

$$F = \frac{SCE/k - 1}{SCD/n - k}$$

Donde k es el número de niveles del factor, en nuestro caso 2.

Un valor de F estadísticamente significativo denotará una relación entre las variables, lo cual significa que las medias de las variables dependientes difieren entre los grupos o niveles de la variable independiente.

4.2.2. Muestra, fuentes y variables

Muestra

La muestra está compuesta por 46 equipos pertenecientes a la primera y segunda división de la liga española de fútbol profesional. El estudio se ha realizado para dos temporadas, en concreto, las temporadas 2018-2019 y 2019-2020.

Fuentes de información

Para la recogida de datos se ha empleado la página oficial de La Liga (<http://www.lfp.es>). El análisis y tratamiento de los datos se llevó a cabo a través de los programas informáticos Excel y SPSS.

Definición de variables

➤ **Puntos**

Recoge la puntuación obtenida por los equipos tanto locales como visitantes. Para la victoria visitante el calor empleado es el 0, el valor 1 equivale a un empate y el 2 se identifica con una victoria del equipo local.

➤ **Goles locales**

Número de veces que el equipo que juega en casa consigue marcar en la portería contraria.

➤ **Goles visitantes**

Número de veces que el equipo visitante consigue marcar en la portería contraria.

➤ **Goles**

Indica la cantidad de goles marcados por los equipos una vez finalizado ya el encuentro. Para esta variable se emplearán las siguientes correspondencias:

Figura 9. Correspondencias relativas al análisis de la variable "goles"

VALOR	CORRESPONDENCIA
0	3 o más goles a favor del equipo visitante
1	2 goles a favor del equipo visitante
2	1 gol a favor del equipo visitante
3	Empate
4	1 gol a favor del equipo local
5	2 goles a favor del equipo local
6	3 o más goles a favor del equipo local

Fuente: elaboración propia

➤ **Goles total**

Determina la diferencia de goles existente al finalizar los partidos a favor o en contra del equipo local: +3; -2; 0; +1...

4.3. Resultados

En la siguiente tabla figuran los descriptivos básicos de las variables que hemos seleccionado para analizar la ventaja de jugar como local en la primera y segunda división española para las temporadas 2018-2019 y 2019-2020. Una de las conclusiones más destacables que podemos extraer es la mayor media que los goles locales ($1,35 \pm 0,81$) tienen frente a los goles visitantes ($1,02 \pm 1,11$) en el cómputo global de las dos divisiones y temporadas, mientras que su dispersión es ligeramente menor (0,81 frente a 1,11); también debemos señalar que ambas variables presentan una distribución ligeramente asimétrica positiva.

Tabla 1. Descriptivos básicos

Variables	Media	Desviación típica	Coefficiente de asimetría	Sumatorio
Puntos	1,18	1,03	-0,34	1985
Goles locales	1,35	0,81	0,86	2268
Goles visitantes	1,02	1,11	1,04	1712
Goles	4,14	1,03	0,27	6959
Total goles	0,33	1,86	-0,03	556

Fuente: elaboración propia

A continuación, se analizarán las tablas de contingencia elaboradas a partir de los datos de las jornadas del conjunto de las dos temporadas. Las tablas 1 y 2 muestran que aquellos equipos de primera división que finalizan la temporada ocupando los primeros puestos (del primer puesto al quinto) ganan un mayor número de partidos como locales (65,08%) que los equipos de segunda división que terminan en las mismas posiciones (60,10%). El análisis de los residuos corregidos (RC) nos muestra que ambos valores (6,45 y 5,71, respectivamente) son positivos y claramente significativos, indicando un número de victorias realmente obtenidas claramente superior a las esperadas; lo que indica una clara influencia del factor “jugar en casa” en el número de victorias obtienen tanto los equipos de segunda como, sobre todo, los de primera división.

De la misma manera, es importante subrayar que tanto en primera (27,51%) como en segunda división (34,25%) los equipos más alejados de la cabeza en la tabla clasificatoria (puestos del 16 a 20 en primera división y del 16 a 22 en segunda división) obtienen unos porcentajes de victorias en campo propio mucho menores a los obtenidos por los primeros clasificados. Los valores (-4,11 y -2,72, respectivamente) de los residuos corregidos – ambos negativos y significativos – indican que el número de victorias realmente obtenidas claramente inferior a las esperadas.

Por otra parte, en cómputo global, los porcentajes de victorial local son realmente parecidos entre la primera (44,87%) y la segunda división (42,86%).

Tabla 2. Posición final del equipo local en 1ª división

	Victoria Local	Victoria Visitante	Empate	Total
Posición 1 -5	123	23	43	189
F. Esperada	84,80	50,48	53,72	189,00
%	65,08	12,17	22,75	100,00
Residuo Corregido	6,45	-5,21	-1,99	
Posición 6 -10	92	47	53	192
F. Esperada	86,15	51,28	54,57	192,00
%	47,92	24,48	27,60	100,00

Residuo Corregido	0,73	-0,69	-0,25	
Posición 11 -15	74	53	63	190
F. Esperada	85,3	50,8	54,0	190,0
%	38,95	27,89	33,16	100,00
Residuo Corregido	-1,41	0,36	1,41	
Posición 16 -20	52	80	57	189
F. Esperada	84,80	50,48	53,72	189,00
%	27,51	42,33	30,16	100,00
Residuo Corregido	-4,11	4,79	0,52	
Total	341	203	216	760
F. Esperada	341,00	203,00	216,00	760,00
%	44,87	26,71	28,42	100,00

Fuente: elaboración propia

Tabla 3. Posición final del equipo local en 2ª división

	Victoria Local	Victoria Visitante	Empate	Total
Posición 1 -5	125	36	47	208
F. Esperada	89,14	51,32	67,53	208
%	60,10	17,31	22,60	100,00
Residuo Corregido	5,71	-2,80	-3,45	
Posición 6 -10	89	43	80	212
F. Esperada	90,86	52,31	68,83	212
%	41,98	20,28	37,74	100,00
Residuo Corregido	-0,22	-1,47	1,53	
Posición 11 -15	82	50	80	212
F. Esperada	90,9	52,3	68,8	212
%	38,68	23,58	37,74	100,00
Residuo Corregido	-1,06	-0,36	1,53	
Posición 16 -22	100	99	93	292
F. Esperada	125,14	72,05	94,81	292
%	34,25	33,90	31,85	100,00
Residuo Corregido	-2,72	3,84	-0,22	
Total	396	228	300	924
F. Esperada	396,00	228,00	300,00	924
%	42,86	24,68	32,47	100,00

Fuente: elaboración propia

En las tablas 4 y 5 se pone de manifiesto la posición obtenida por los equipos visitantes tanto para 1ª como para 2ª división. Mientras que en 1ª división los equipos peor clasificados obtienen un porcentaje de derrota por encima del 50% (57,89%), en 2ª división el porcentaje es inferior (53,22), con residuos corregidos significativos (3,1 y 3,3, respectivamente)

Por el contrario, algo que realmente destaca es que el porcentaje de victorias visitantes en primera (15,79%; RC=-3,36) y segunda (15,25%; RC=-3,95) división para estos mismos equipos son prácticamente iguales. Por último, podemos observar que los equipos mejor clasificados de primera división obtienen un menor porcentaje de derrotas como visitantes (24,87%; RC=-6,38) en comparación con aquellos que ocupan los primeros puestos en segunda división (33,65; RC=-3,08).

Tabla 4.: Posición final del equipo visitante en 1ª división

	Victoria Local	Victoria Visitante	Empate	Total
Posición 1 -5	47	80	62	189
F. Esperada	84,80	50,48	53,72	189
%	24,87	42,33	32,80	100,00
Residuo Corregido	-6,38	5,60	1,54	
Posición 6 -10	87	53	52	192
F. Esperada	86,15	51,28	54,57	192
%	45,31	27,60	27,08	100,00
Residuo Corregido	0,11	0,28	-0,40	
Posición 11 -15	97	40	52	189
F. Esperada	84,8	50,5	53,7	189
%	51,32	21,16	27,51	100,00
Residuo Corregido	1,53	-1,70	-0,27	
Posición 16 -20	110	30	50	190
F. Esperada	85,25	50,75	54,00	190
%	57,89	15,79	26,32	100,00
Residuo Corregido	3,10	-3,36	-0,63	
Total	341	203	216	760
F. Esperada	341,00	203,00	216,00	760
%	44,87	26,71	28,42	100,00

Fuente: elaboración propia

Tabla 5. Posición final del equipo visitante en 2ª división

	Victoria Local	Victoria Visitante	Empate	Total
Posición 1 -5	71	72	68	211
F. Esperada	90,43	52,06	68,51	211
%	33,65	34,12	32,23	100,00
Residuo Corregido	-3,08	3,62	-0,08	
Posición 6 -10	71	63	75	209
F. Esperada	89,57	51,57	67,86	209
%	33,97	30,14	35,89	100,00
Residuo Corregido	-2,23	1,81	0,99	
Posición 11 -15	97	48	64	209

F. Esperada	89,6	51,6	67,9	209
%	46,41	22,97	30,62	100,00
Residuo tipificado	0,78	-0,50	-0,47	
Residuo Corregido	0,89	-0,57	-0,53	
Posición 16 -22	157	45	93	295
F. Esperada	126,43	72,79	95,78	295
%	53,22	15,25	31,53	100,00
Residuo Corregido	3,30	-3,95	-0,34	
Total	396	228	300	924
F. Esperada	396,00	228,00	300,00	924
%	42,86	24,68	32,47	100,00

Fuente: elaboración propia

Por último, hemos procedido a realizar un análisis de la varianza de un factor para la ventaja de jugar en casa cuyos resultados se recogen en la siguiente tabla:

Tabla 6. ANOVA de un factor para la ventaja de jugar como local

Variables	1ª división (media)	2ª división (media)	Media cuadrática	F	P
Puntos	1,1805	1,1818	0,0007	0,0011	0,9736
Goles total	0,3544	0,3095	0,8397	0,3860	0,5345
Goles locales	1,4427	1,2716	12,1910	9,9768	0,0016
Goles visitantes	1,0883	0,9610	6,7460	6,4151	0,0114
Goles	4,2424	4,0595	13,9399	4,0534	0,0442

Fuente: elaboración propia

Lo más destacable del ANOVA son las diferencias encontradas entre primera y segunda división en el número de goles anotados, tanto locales ($p=0,0016$) como visitantes ($p=0,0114$), con superioridad para la primera categoría. Las variables “puntos” y “goles total” son las únicas cuyo valor de p está por encima de 0,05, por lo que podemos afirmar que, bajo un nivel de confianza del 95%, no hay relación significativa entre el factor y estas dos variables dependientes, de manera que el factor división únicamente puede explicar en cierto modo los cambios que se producen en: goles locales, goles visitantes y goles.

Una vez realizado este caso práctico podemos confirmar la relación existente entre la ventaja de jugar en casa y, en este caso, el nivel del equipo, es decir, si pertenece a la primera o segunda categoría de la liga nacional de fútbol española. Con respecto a esto, con el análisis desarrollado para las temporadas 2018-2019 y 2019-2020 observamos que los equipos que finalizan en las primeras posiciones en 1º división en las primeras posiciones obtienen una mayor ventaja cuando juegan como local (65,08% de victorias

como local) tanto si lo comparamos con los equipos peor clasificados en la misma categoría (27,51%) como si lo cotejamos con los que finalizan en los primeros puestos de segunda división (60,10%). Esto se ve ratificado con el análisis de la varianza realizado en el que observamos una mayor cantidad de goles locales marcados en primera división ($p=0,0016$). Esta mayor influencia del factor campo en determinados equipos puede venir explicada por fenómenos comentados anteriormente como la facilitación social.

Todo esto se ve reafirmado también mediante el análisis de las victorias de los equipos visitantes (tablas 4 y 5), puesto que observamos que cuando un equipo local recibe en su campo a un conjunto que ha finalizado en primera división entre los puestos 16-20 consiguen un 57,89% de victorias ($RC = 3,1$). Por el contrario, cuando se enfrentan a un equipo que ha finalizado entre los 5 primeros, el porcentaje de victorias se reduce considerablemente hasta el 24,87% ($RC = -6,38$). En segunda división los porcentajes de victorias son superiores cuando se enfrentan a los primeros clasificados (33,65%; frente al 24,87% de primera división) mientras que cuando reciben a los últimos de la tabla los porcentajes son más similares (53,22%; frente a los 57,89% en primera división).

Tratando de actualizar lo máximo posible los resultados y queriendo hacer referencia a la verdadera realidad que en este momento nos rodea, es interesante comprender como la pandemia que nos asola ha influido también variando la influencia y efectividad del factor campo. Para ello analizaremos los resultados obtenidos por los equipos de la 1ª división española anteriores y posteriores al parón de la competición decretado como consecuencia de la pandemia.

Tras declarar el estado de alarma y el confinamiento domiciliario, la liga de fútbol profesional española volvía a ponerse en marcha el 11 de junio quedando todavía por disputarse 11 jornadas. Sin embargo, dado lo rocambolesco de la situación se requerían también medidas especiales para que la competición pudiese seguir su curso. Sin duda, uno de los cambios más drásticos y que más impactó tanto en la manera de ver como de practicar este deporte fue la decisión de jugar a puerta cerrada, es decir, sin público; reduciendo así uno de los factores más condicionantes en la ventaja de jugar como local, el empuje de la afición y su influencia en el rendimiento de los futbolistas.

Con respecto a esa temporada, la 2019-2020, desde el inicio de la competición hasta que se produjo el parón el 8 de marzo, se disputaron 27 jornadas. En los 270 encuentros realizados en este primer tramo; el 47,7% (129 partidos) acabó con victoria local, un

24,4% de los partidos (66) terminaron en empate y un 27,7% (75) los ganaron los equipos visitantes. Sin embargo, tras la reanudación de La Liga en las 11 jornadas (110 partidos) que quedaban por disputarse, todas ellas a puerta cerrada, se registraron un 41% de victorias locales (45 partidos), un 31,8% de victorias visitantes (35 partidos) y un 27,2% de empates (30 partidos). Por tanto, podemos observar como en el segundo tramo del campeonato liguero se produjeron un menor número de victorias a favor de los equipos de casa y un mayor número de victorias visitantes, no variando apenas el porcentaje de empates.

Llegados a este punto y acorde con lo que venimos señalando a lo largo de la investigación, podemos extraer las siguientes conclusiones:

- Los equipos de primera división obtienen una mayor superioridad al competir en su estadio que los conjuntos de segunda división exceptuando aquellos equipos que finalizan la temporada en las últimas posiciones. En este último caso, la incidencia del factor campo es superior en segunda división obteniendo un mayor número de victorias locales.
- Los equipos visitantes con mayor categoría se ven menos afectados por la ventaja que posee el rival como equipo local, por lo que podríamos concluir que en la segunda división española hay mayor igualdad entre los distintos equipos.
- Con respecto a las modificaciones introducidas en el fútbol debido al COVID´19, la ausencia de público en los estadios ha influido considerablemente a la ventaja de jugar como local. El número de victorias locales sigue siendo mayor al de las visitantes pero en un porcentaje inferior, concluyendo así que el factor campo sigue existiendo pero en menor medida debido a la desaparición del fenómeno de la facilitación social o la teoría del impulso.

5. EJEMPLOS CONCRETOS

A continuación, con el objetivo de ser conscientes de la relevancia que puede tener esta ciencia en el deporte, conoceremos dos ejemplos reales de clubes que han alcanzado el éxito gracias a la aplicación de modelos estadísticos y al trabajo diario con algunas de las herramientas que hemos mencionado anteriormente.

5.1. “Método Monchi”

En primer lugar, cabe destacar, sin lugar a duda, la figura de Ramón Rodríguez Verdejo; más conocido deportivamente como Monchi. Este exfutbolista español que comienza su andadura como director deportivo del Sevilla FC en el año 2000, labor que todavía hoy desempeña, ha sido uno de los grandes “culpables” de los éxitos cosechados por el Sevilla en los últimos años.

Con el paso de los años y de la mano de Monchi, el club hispalense se ha convertido en el **precursor en España en el empleo de la estadística** para descubrir jugadores baratos que posteriormente puedan aportar un gran beneficio al club.

El director deportivo del Sevilla FC cuenta en su equipo con “matemáticos, ingenieros, físicos, estadísticos y analistas” que llevan a cabo el desarrollo de una aplicación con la que poder ser lo más precisos posible en la búsqueda de jugadores. Se trata de encontrar el perfil de futbolista que el entrenador solicita, buscando las características clave que necesita y aplicando variables estadísticas para descartar a los que no cumplen con esos requisitos. De esta manera se ahorrará una gran cantidad de tiempo al tener que observar a un número mucho menor de jugadores (Estadio Deportivo, 2020).

Para comprenderlo de una manera más sencilla, habitualmente los ojeadores de los clubes rastrean en torno a 40 ligas de todo el mundo. En cada una de estas ligas participan 20 equipos, que a su vez cuentan con entre 20 y 25 jugadores de los que cuatro o cinco pueden ser delanteros. Es decir, en la búsqueda de un delantero, se comienza a escoger de una base de datos con entre 3.600 y 4.500 futbolistas. Poniéndonos en el hipotético caso de que el entrenador quiera alguien que fuera bien por alto y capaz de generar muchas ocasiones de gol; sabiendo que la media de duelos aéreos ganados por los delanteros de la Primera División española es actualmente del 62% y que la media de ocasiones de gol generadas por los delanteros en un partido es de 3,4; si se seleccionasen solamente a los futbolistas que superan estas medias, finalmente, el número de jugadores se reduciría drásticamente de 4.000 a 800, luego a 200 y se podría seguir minorando en función de otros criterios como podrían ser los económicos o de personalidad.

Esta forma de contratar jugadores ha impulsado al Sevilla hasta la cúspide del fútbol mundial pues, antes de la llegada de Monchi el Sevilla solamente contaba con 4 títulos en sus vitrinas y llevaba desde 1948 sin ganar nada. Por el contrario, tras el fichaje del director deportivo por el club hispalense, la entidad ha conseguido levantar 10 títulos entre los que destacan siete títulos europeos (seis Europa League y una Supercopa de Europa) (Villareal, 2019).

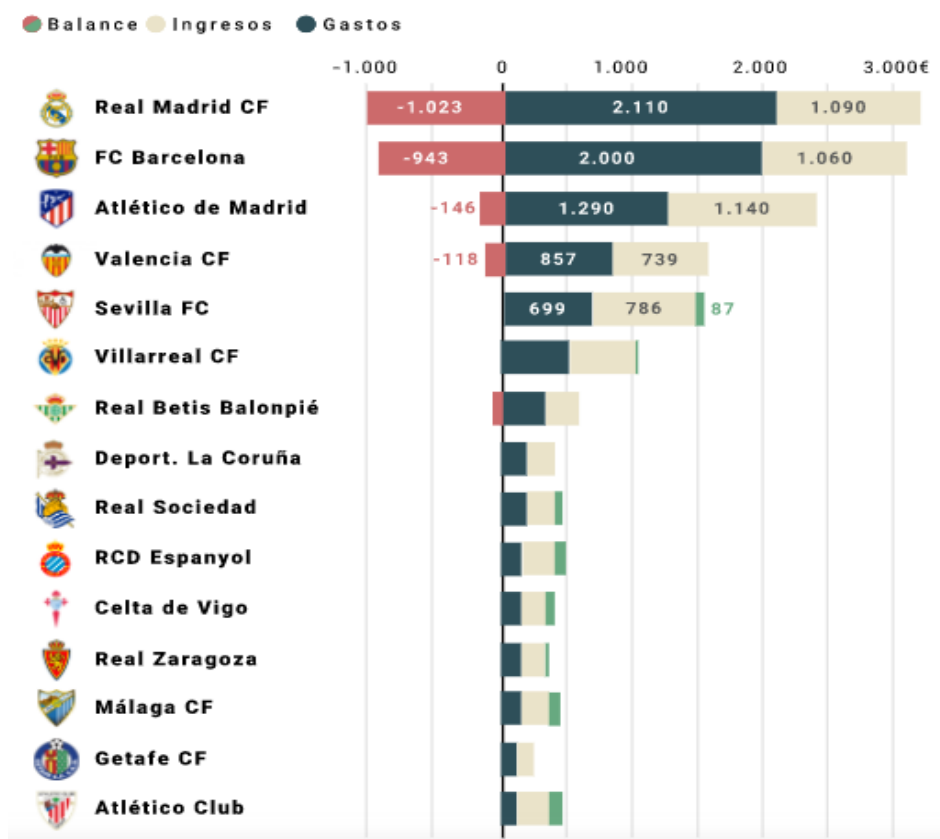
Figura 10. Operaciones económicas más exitosas del Sevilla FC en la era Monchi



Fuente: elaboración propia a partir de datos e imágenes de Transfermarkt

Los éxitos no solo se han producido a nivel deportivo, sino que el Sevilla FC ha visto como, gracias al popularmente conocido como “método Monchi”, su economía ha crecido también enormemente, llegando a generar, como podemos ver en la gráfica anterior, una plusvalía de 150,5 millones de € con la compra-venta de tan solo 6 futbolistas. Estos hechos han provocado que la entidad se haya incluso consolidado como el club español con un mejor balance entre ingresos y gastos en esta etapa (2000-actualidad).

Figura 11. Ingresos y gastos de los equipos de la liga española en la era Monchi (en millones).



Fuente: Transfermarkt

Una vez conocemos todo esto no es de extrañar que el propio director deportivo del Sevilla asegure que “no firmaré a un jugador solo por el dato, pero no firmaré a un jugador si antes el dato no me ha dado la señal. La estadística te ayuda a reducir el riesgo y a ganar tiempo”.

5.2. Liverpool de Jürgen Klopp

John W. Henry, actual dueño del Liverpool FC, es conocido también por ser uno de los protagonistas de la historia real contada en la película Moneyball; el multimillonario propietario de los Boston Red Sox, equipo profesional de la liga de béisbol de los Estados Unidos. Este empresario e inversor estadounidense, habituado a trabajar con teorías matemáticas para analizar el mercado de valores, decidió aplicar sus conocimientos en el mundo del beisbol para, posteriormente, centrarse también en el fútbol. De esta manera, en 2010, cuando el club se encontraba en una crisis económica y deportiva, John W. Henry adquiriría el Liverpool FC, equipo histórico de Inglaterra, con el objetivo de devolverlos a la gloria (Torres, 2019).

En 2015, bajo la directriz de John W. Henry, el Liverpool FC decidió confiar en la ciencia y en **Ian Graham** para llevar a cabo la elección del que sería el nuevo entrenador del equipo.

Este físico de la Universidad de Cambridge diseñó un modelo matemático que trataba de eliminar todos esos elementos aleatorios que existen en el fútbol para poder obtener conclusiones sin un ápice de distorsión. Los informes obtenidos trataban de superar la banalidad del marcador final o la clasificación con el objetivo de enfocarse en determinar el nivel real de los distintos equipos. Para ello, el propio Graham reveló que no vio ningún partido pues “las imágenes contienen un poderoso sesgo”, aseguraba (Schoenfeld, 2019).

El resultado obtenido fue muy sorprendente con respecto a uno de los entrenadores analizados; Jürgen Klopp, técnico por aquel entonces del Borussia Dortmund. El equipo alemán terminó en séptimo lugar en la temporada 2014-2015 pero, sin embargo, el modelo creado por Ian Graham determinó que por el rendimiento obtenido deberían haber terminado en segunda posición. Las estadísticas indicaban que habían sido la plantilla con peor suerte de las grandes ligas pues, el equipo de este entrenador alemán fue el que más se alejó de los resultados que le asignaban las probabilidades.

Estando ya Jürgen Klopp en el Liverpool FC y convencido del método que lo había seleccionado como el entrenador idóneo para el equipo, confió en Ian Graham para detectar jugadores talentosos y futuras promesas en el mercado de fichajes. Graham constituyó un particular equipo de trabajo compuesto por; Tim Waskett, licenciado en astrofísica Dafydd Steele, campeón de ajedrez con posgrado en matemáticas, y Will

Searman, doctorado en física. Estos analistas comenzaron a trabajar sobre una base de datos de más de 10.000 futbolistas y, mediante la aplicación de estadísticas avanzadas y algoritmos que trataban de eliminar la aleatoriedad, comenzaron a realizar fichajes tan importantes como los de Van Dijk, Robertson, Wijnaldum, Mané, Salah... futbolistas que hoy en día constituyen la columna vertebral del equipo y que son los mayores activos de la entidad en la actualidad (Schoenfeld, 2019).

Tras 5 temporadas al frente del equipo, el Liverpool de Jürgen Klopp y, por qué no, de Ian Graham, ha conseguido volver a situarse como uno de los equipos de la cúspide del fútbol. La aplicación de esta metodología de trabajo, cada vez más implementada en el deporte, vio reflejada su éxito con la consecución de 4 importantes títulos: Champions League (tras 13 temporadas sin conseguirlo), Premier League (tras 30 años sin ganarla), una Supercopa de Europa y un Mundial de Clubes. Asimismo, Klopp fue galardonado con el premio al mejor entrenador de 2019 que concede la FIFA (Federación Internacional de Fútbol Asociado) (Goal, 2020).

Conclusiones

A lo largo de un partido de fútbol siempre se han generado millones de datos, dejando así un interesante rastro estadístico del que sacar provecho. Sin embargo, es a partir de los nuevos avances tecnológicos cuando se empiezan a recopilar, analizar de manera minuciosa e interpretar, llegando a obtener información muy valiosa.

De esta manera, el analista de datos se ha convertido en una de las figuras más valiosas, incluso indispensables, dentro de los cuerpos técnicos que rodean a los deportistas. En su día a día se ven apoyados por distintos programas de análisis y empresas proveedoras de megadatos como Mediacoach, Nac Sport, Instat Scout o Wyscout.

Comencé esta investigación con casi total desconocimiento con respecto al papel de este perfil de profesional dentro de una organización deportiva. No obstante, llegado este momento y siendo conocedor de la influencia cada vez mayor que tienen en el éxito deportivo, debido a la gran diversidad de aspectos que manejan dentro de la planificación y preparación deportiva, considero que la contratación de los mejores analistas de datos llegará incluso a ser un motivo de rivalidad entre los equipos, alcanzando así importantes cotizaciones y siendo un protagonista más dentro de los mercados de fichajes.

Por otro lado, la introducción del Big Data en el mundo del deporte ha supuesto toda una revolución, proporcionándole a los responsables instrumentos para predecir y decidir de una manera más adecuada y a los deportistas potenciándoles sus capacidades y rendimiento. Por ende, ha transformado áreas de vital importancia como la captación de talento y la preparación y salud deportiva.

Al mundo del fútbol estos cambios han llegado de manera más tardía que a otros deportes, lo cual puede ser explicado por una serie de condicionantes que rodean a este deporte. El fútbol no se presta tanto a un análisis estadístico como si lo pueden hacer otros deportes, fundamentalmente debido a que la aleatoriedad y el azar poseen un rol

mucho más determinante que en otros más lineales como el béisbol. Además, la cantidad de datos totales registrados es mucho menor que en otros deportes, pues mientras que en el fútbol la media de goles por partidos es muy baja, en deportes como el baloncesto es habitual que se superen los 100 puntos. Por último, identificándose más con un condicionante sociológico, existe una tendencia más reacia hacia las innovaciones. Así, mientras que el “ojo de halcón” lleva siendo utilizado en deportes como el tenis una gran cantidad de años, su adaptación en el fútbol se producido hace relativamente poco. Otro ejemplo de esto es la reciente introducción del VAR (Video Assistant Referee) o asistencia al árbitro por video para revisar determinadas jugadas polémicas, ya que esta tecnología llevaba ya disponible desde hace años pero su aplicación se ha visto continuamente aplazada.

El cambio acaba de empezar, pero, sin lugar a duda, ha llegado para quedarse y ya no hay vuelta atrás. El aprovechamiento de la tecnología del Big data se consolidará como una herramienta más para alcanzar los objetivos, hasta el punto de que será considerada como un factor competitivo clave y generará diferencias de nivel con aquellos clubes que no tengan los suficientes recursos para su total aprovechamiento. Esto no será únicamente aplicable a nivel profesional y de las grandes ligas, si no que poco a poco comenzará a introducirse también a nivel formativo.

En cuanto a las herramientas de análisis de datos empleadas, además de la estadística, mencionada constantemente a lo largo del trabajo, merece especial mención la llegada de Mediacoach. Ha supuesto un antes y un después no solo porque se haya posicionado como líder indiscutible en nuestra liga sino también por la paridad lograda entre clubes con respecto a esta materia. Desde su nacimiento, contar con este tipo de tecnología de videoanálisis se volvió mucho más accesible económicamente para todas las entidades, disponiendo así de la misma información y no generando desigualdades.

Independientemente de las conclusiones ya extraídas en el análisis empírico llevado a cabo, es necesario resaltar la importancia que este tipo de investigaciones pueden tener para el deporte. Se ha demostrado que de la recolección masiva de datos y mediante la aplicación de técnicas estadísticas, se puede llegar a extraer información verdaderamente valiosa que condicione incluso la estrategia a seguir durante un partido, en función, por ejemplo, de si se juega como local o como visitante,

Para finalizar esta investigación, me gustaría destacar el hecho de que, desde mi humilde opinión, el empleo del big data y la estadística como recursos en el fútbol, contribuyen en todo momento su mejora, progreso y adaptación y nunca a la destrucción

de su esencia. A pesar de que se tienen en cuenta y se analizan miles de datos y aspectos del juego, una gran diversidad de variables del mismo (estado anímico, influencia de los aficionados sobre los equipos...) no pueden ser controladas. Esto forma parte de la aleatoriedad característica que rodea a este deporte y que provoca que no se pierda ese factor sorpresa que tanto fascina y cautiva a los espectadores.

Bibliografía

- Alvarez Medina, J., Parra Artal, Á., Murillo Llorente, V., & García Felipe, A. (2018). Análisis observacional de los goles de dos temporadas de la LNFS. *Revista Internacional de Medicina y Ciencias de la Actividad Física y el Deporte*, 27-42.
- Aughey, R. (2011). Applications of GPS technologies to field sports. *International Journal of Sports Physiology and Performance*, 295-310.
- Cardenas, J. (6 de Noviembre de 2015). *Networkianos*. Obtenido de <http://networkianos.com/anova-de-un-factor-que-es-como-analizar/>
- Carranza, M. A. (14 de julio de 2017). Identificación del talento en la organización : el Big Data aplicado al fútbol. Valladolid.
- Carreón García , J., Moreno Altamirano, L., & García de la Torre, G. (2018). Selección del análisis estadístico. En L. M. Altamirano, *Epidemiología clínica* (pág. 329). México: MC Graw Hill.
- Casals, M., & Finch, C. (2016). Sports Biostatistician: a critical member of all sports science and medicine teams for injury prevention. *Injury Prevention*.
- Cook, E. (1964). Percentage baseball. *Waverly Press*.
- Cox, R. H. (2008). *Psicología del deporte: conceptos y sus aplicaciones*. Madrid: Editorial Médica Panamericana.
- Davids, B. (1971). *Society for American Baseball Research*.
- Dobson, B., & W. L. Keogh, J. (2007). Methodological issues for the application of time-motion analysis research. *Strength and conditioning journal*, 48-55.
- Estadio Deportivo. (23 de septiembre de 2020). El secreto de Monchi: matemáticos y estadísticos para buscar buenos fichajes. *Estadio Deportivo*.
- F.Barsky, S., & Schwartz, B. (1977). The home advantage. *Social Forces*, 641-661.
- Fraga, A. I. (2016). Cómo el análisis de datos ha transformado el mundo del deporte. *TicBeat*.
- Gamboa, A. G. (2019). Big Data & Sports. *Moleqta*, 67.
- García, E. (28 de Junio de 2019). Expertos en Big Data: "Una máquina nunca sustituirá a un Serra". *Estadio Deportivo*.
- Goal. (8 de octubre de 2020). Obtenido de <https://www.goal.com/>
- Hidalgo, E. S. (22 de Junio de 2018). Las matemáticas del balón: así se recogen los datos durante los partidos de fútbol. *El país*.

- J. Dubner, S. (2011). *Freakonomics*.
- James, B., & Wirth, M. (1986). *The Bill James historical baseball abstract*. New York: Villard Books.
- Lewis, M. (2004). *Moneyball: The art of winning an unfair game*. WW Norton & Company.
- Martín, C. (5 de Diciembre de 1994). El miedo escénico. *El País*. Obtenido de <https://footballcitizens.com/>
- Martínez, V. C. (27 de Mayo de 2019). Estudio de mercado para fundar una empresa de Big Data. Valladolid, España.
- Neave, N., & Wolfson, S. (2003). Testosterone, territoriality, and the 'home advantage'. *Physiology & Behavior*, 269-275.
- Olmeda, M. (15 de Enero de 2020). La tecnología que ayudará a los equipos a preparar los partidos. *ABC*.
- Pérez, D. (2018). Objetivo Analista. *Los 3 Mejores Software de VideoAnálisis de Fútbol*.
- Perform group. (s.f.). *Opta Sports*. Obtenido de <https://www.optasports.com/>
- Porta, M. (2014). *Dictionary of epidemiology*. Oxford University Press & International Epidemiological Association.
- Pulgarín, J. (21 de Mayo de 2018). Toda la tecnología que se usa en el fútbol y pocos conocen.
- Rodrigo, J. A. (Enero de 2016). *Ciencia de datos*. Obtenido de ANOVA análisis de varianza para comparar múltiples medias: https://www.cienciadedatos.net/documentos/19_anova
- Rolfh, F., & Sokal, R. (1986). *Introducción a la Bioestadística*. Reverté, S.A.
- Román, V. (2018). Big data en el fútbol, cuando el deporte se ayuda de los números. *N+1, tecnología que suma*.
- Sanchez, F. S. (2015). *¿Que esconde tu rival?* Futbol de libro S.L.
- Schoenfeld, B. (29 de mayo de 2019). El arma secreta del Liverpool: el análisis de datos. *The New York Times*.
- Silva, J. P. (23 de 12 de 2015). La era del Big Data en el deporte. *Expansión*.
- Torres, D. (3 de junio de 2019). Jürgen Klopp, perdedor en la gloria. *El País*.
- Torres, D. (9 de noviembre de 2020). La Liga más seca. *El País*.
- V. Carron, A., & Courneya, K. (1992). The home advantage in sport competitions: A literature review. *Journal of Sport and Exercise Psychology*, 13-27.
- Vidaurreta Bueno, R., & Mesa Anoceto, M. (2011). Contribución de la Estadística Matemática a la preparación deportiva sostenible. *EFDeportes*.
- Vigara, D. (14 de Diciembre de 2017). *UPAD Psicología y Coaching*. Obtenido de <https://www.upadpsicologiacoaching.com/>

Villareal, A. (17 de octubre de 2019). La reinención de Monchi, el mago de los fichajes:
"Obviar el 'big data' es anacrónico. *El Confidencial*.

Walker, E. (1982). *The Sinister First Baseman and Other Observations*. Celestial Arts.

Zajonc, R. B. (1965). Social Facilitation. *Science*, 269-274.