

# Módulo 3. Aprendizaje basado en un problema real (ABP) en el fenómeno de *Sports Analytics* (II)

## 3.1. Introducción del módulo

Siguiendo el CASO mediante APB realizado anteriormente, seguiremos aprendiendo diferentes conceptos que nos pueden ser útiles para tener un mejor enfoque con relación al fenómeno de *Sports Analytics*. En este módulo se tratará de aprender qué es esto de *Big Data*, su correcta definición aplicada en un deporte como el fútbol y qué relación tienen la investigación y la academia. Además, informaremos sobre las revistas científicas más populares que tratan aspectos relacionados con *Sports Analytics*, la importancia de preguntarse de dónde vienen los datos que analizamos, para pensar sobre aspectos tan relevantes como la validación, la exactitud y los sesgos en esos datos antes de analizarlos. Recordaremos las diferentes partes de la estadística (descriptiva e inferencial), que muchos de los académicos enseñan en la academia y las tendremos siempre en cuenta; finalmente se mencionará la importancia de los modelos estadísticos y también la bioestadística deportiva. Nos daremos cuenta de la importancia que recae también en tener la habilidad crítica para saber leer y entender estos conceptos, además de trabajar en un ambiente multidisciplinar con otros especialistas del mundo cuantitativo para realizar correctas interpretaciones y avanzar en el mundo de *Sports Analytics*.

## 3.2. Análisis del 3º párrafo del CASO

Recordando el texto de la anterior actividad ABP del CASO presentado, seguiremos en este módulo trabajando algunos conceptos con los próximos párrafos pendientes.

Después de leer detenidamente el siguiente párrafo, voy a poner en negrita las palabras que me puedan ayudar a hacer preguntas para consolidar algunos conceptos relacionados indirectamente con *Sports Analytics*.

“Minutos más tarde, **el científico de datos** me habla de la importancia de estar familiarizado con la **investigación**, y cómo esta se adapta a los fenómenos de **Big Data** y *Data Science*. También me explica que su grupo se reúne cada martes para comentar algunos **artículos científicos** de interés relacionados con el fútbol. Me comenta que en el inicio de liga le ha sorprendido el número de tarjetas amarillas y por ello le gustaría saber cómo podrían influir estas en el fútbol. Rápidamente le comento que me suena un

estudio que se publicó hace unos años en la **revista Journal of Quantitative Analysis in Sports** (Anders, A.y Rotthoff, K. W. (2011). Yellow Cards: Do They Matter? *Journal of Quantitative Analysis in Sports*, 7(1)).

1. **¿Qué relación hay entre Investigación y Big Data? ¿Y entre Big Data y el fútbol?**
2. **¿Qué es un artículo científico? ¿Cómo se conforman? ¿Qué tipología hay según el ámbito de especialización?**
3. **¿Qué revistas de Sports Analytics puedo encontrar?**

¿Había formulado preguntas similares? Puede ser que no, pero probablemente ha pensado en preguntas relacionadas que lo hacen pensar en cómo responder y adquirir más información sobre ello. Vamos a intentar responder a las preguntas.

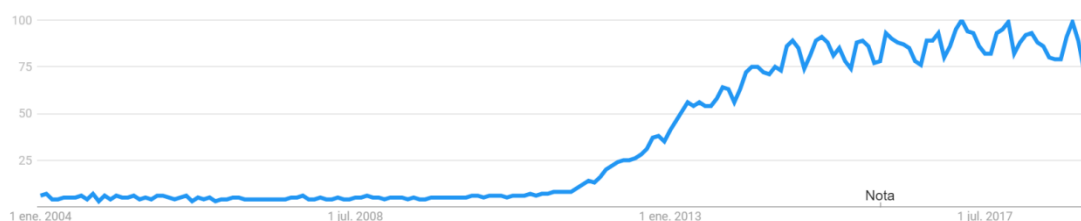
### 1. **¿Qué relación hay entre Investigación y Big Data? ¿Y entre Big Data y el fútbol?**

Como muchos sabemos, la palabra *Big Data* ha sido muy utilizada recientemente. Uno de los puntos para tener en cuenta es que el *Big Data* justamente no es solo recopilar grandes volúmenes de datos, sino tener la habilidad de saber qué tipos de datos trabajo o no trabajo, y cuáles son útiles y cuáles no son útiles.

Hay que preguntarse antes, por ejemplo, por la validación de los datos y los diferentes tipos de sesgo que puede haber. En nuestro campo de análisis del deporte, si no se tiene en cuenta la validación interna o la validación externa, básicamente no podemos avanzar. Antes de analizar y hacer un gráfico, y perder tiempo malinterpretando desde el inicio del estudio o problema, hay que utilizar el pensamiento crítico delante del fenómeno del *Big Data*. Por ello, aparte de formular buenas preguntas ante un problema planteado, será también importante preguntarse por la calidad de los datos y el sentido que tiene la validación interna y externa.

*Big Data* ha sido una palabra y un fenómeno de gran interés, pero últimamente parece que este interés ha ido disminuyendo (al menos en Estados Unidos) paulatinamente, como vemos en la siguiente figura de *googletrends*.

**Figura 1: Consulta de la evolución del interés a Google Trends en general (todos los países) para el fenómeno Big data, Todo el mundo, 2004-actualidad**



Fuente: elaboración propia con base en googletrends

**Figura 2: *Big data* solamente en los Estados Unidos, 2004-actualidad**



Fuente: elaboración propia con base en googletrends

Observando los dos cuadros, podemos decir que el mote *Big data* está bajando.

Según alguna bibliografía especializada en el mundo de la investigación, autores como Zhang o Malhouse nos reafirman y recuerdan que “The *big data era* doesn’t just mean large amounts of data; it also means increased ease and ability to collect data of all types, in all walks of life”.

### ***Big Data* y fútbol**

Subsección escrita por José María Fernández Ponce, profesor titular de la Universidad de Sevilla, miembro del Departamento de Estadística e Investigación Operativa e investigador del grupo “Métodos Cuantitativos en Investigación, FQM-328” de la Junta de Andalucía. También posee el título de entrenador de fútbol (Nivel I) de la Federación Andaluza de Fútbol. En la actualidad, es asesor externo para la academia de fútbol del Cádiz Club de Fútbol, donde desarrolla un proyecto de elaboración de una base de datos para la selección, detección y desarrollo de talento deportivo.

*Big Data* es un término muy utilizado en cualquier ámbito de nuestra sociedad, aunque en algunos casos lleva a confusión y a interpretaciones erróneas, debido a la complejidad intrínseca que el concepto implica. En el fútbol profesional, y en el deporte de alto rendimiento en general, la tecnología *Big Data* no se puede obviar como herramienta de ayuda en la toma de decisiones, ya que los presupuestos que manejan los clubes son de varios millones de euros, es decir, empresas cuyo manejo económico es alto o incluso astronómico, si nos referimos a los equipos tops de La Liga. La cuestión es si el *Big Data* en el fútbol es simplemente una moda que tarde o temprano quedará en nada o una necesidad imperiosa para conseguir la modernización y, a la vez, optimizar el rendimiento deportivo de los equipos. Intentaré en este apartado explicar la idea de lo que puede aportar el *Big Data* al fútbol en su dimensión deportiva, para justificar su implementación en los equipos profesionales.

¿Qué es *Big Data*? La traducción del término, o al menos como se traduce al castellano, es *Datos Masivos*. Si bien es cierto que la definición que aporta la Real Academia Española es bastante clara y concisa, puede que no lo sea tanto para alguien no experto en computación o matemáticas. Lo primero que se viene a la mente es: “datos masivos significa muchos datos”, pero ¿cuántos datos son muchos datos? Pensemos en un ordenador de mesa. Una de las características en la que nos fijamos a la hora de comprarlo es su memoria en disco duro, es decir, la capacidad de almacenamiento de datos del ordenador. Lo normal es que esta memoria oscile entre 500 Gigabytes (GB) y 2 Terabytes (TB). Para entender lo que significan estas unidades, veamos el siguiente caso. En 1GB se pueden almacenar aproximadamente unas 300 canciones en mp3 de tamaño medio (de unos tres minutos y medio de duración y con una calidad media). Luego, en una PC de 500GB se podrían almacenar unas 150 000 canciones en mp3. Teniendo en cuenta que 1TB son 1000GB, por una regla de tres sencilla se obtiene que en un ordenador cuya memoria de disco duro es de 2TB se podrían almacenar, aproximadamente, 600 000 canciones en mp3. Si cada una de ellas tiene un tiempo medio de duración de tres minutos y medio, tardaríamos casi cuatro años en escucharlas sin descansar. Otro ejemplo muy clarificador es el juego FIFA19 para PC, que ocupa 41 GB en la memoria; por lo tanto, podríamos almacenar en un ordenador de 2TB unos 48 juegos de la misma capacidad.

Con el uso de internet, el flujo de datos se ha disparado no solo por su volumen (la memoria que se necesitaría para almacenarlos y tratarlos) sino también por la variedad. Por ejemplo, hoy en día usamos muchos tipos de redes sociales: Facebook, Twitter, WhatsApp, entre otras. En ellas se suele intercambiar todo tipo de información: mensajes, fotos, videos, gifs y demás. Esta variedad, unida al volumen de datos, puede originar un problema de *Big Data*. Podríamos tener tantos datos (masivos en volumen y/o variedad) que resultara imposible tratarlos en una PC convencional, sobre todo porque excederían en capacidad los TB del ordenador de mesa. Un caso real de *Big Data* es Amazon. A esta empresa le interesa tener satisfechos a todos sus clientes, con vistas a obtener mayores beneficios. Para ello los clasifica según sus gustos y su historial de compras, para mandarles publicidad de manera individualizada a los dispositivos electrónicos desde donde compraron o, simplemente, visualizaron los productos. ¿Cuántos clientes tiene Amazon a nivel mundial? En Estados Unidos son cerca de 93 millones de clientes *prime*, por tanto, el volumen de todos ellos con sus datos a nivel mundial no pueden ser gestionados desde una PC normal. Toda esta problemática aquí expuesta no es muy exhaustiva, ya que la realidad es algo más compleja, pero sí nos permite entender el principio del concepto de la tecnología *Big Data* como el conjunto de *hardware* (es decir, componentes electrónicos y ordenadores) y de *software* (programas informáticos) que permiten tratar y analizar datos masivos. No es suficiente con una sola máquina, sino que se necesitan varias (ordenadores conectados en red) que contengan toda la información y, para poder manipularlos, además, se requiere de

un software que facilite el reparto de esa información y la obtención de los resultados (lo que se llama, en el argot informático, la computación en paralelo y/o distribuida). Pero el asunto es aún más complicado. La aparición de esta tecnología informática para el tratamiento de datos masivos ha originado la necesidad de un nuevo enfoque de la Estadística. Parafraseando al profesor Daniel Peña, de la Universidad Carlos III de Madrid, la aparición del *Big Data* ha supuesto, para la estadística clásica (aquella que trata con *muestras pequeñas*), algo equivalente a lo que supuso la Teoría de la Relatividad de Einstein a la física clásica de Newton. En definitiva, el *Big Data* ha obligado a los matemáticos/estadísticos plantearse problemas desconocidos hasta el momento y aportar soluciones a este nuevo paradigma.

**Figura 3: WimU en jugadores profesionales de fútbol**



Fuente: [https://www.culemania.com/palco/real-track-systems-primer-ejemplo-exito-barca-innovation-hub-231195\\_102.html](https://www.culemania.com/palco/real-track-systems-primer-ejemplo-exito-barca-innovation-hub-231195_102.html)

¿Y en el fútbol, existen tantos datos tal que resulte imposible su manipulación en una PC? El avance en las Tecnologías de la Información ha provocado que, hoy en día, estemos viviendo una revolución tecnológica del fútbol. Por ejemplo, desde la temporada 2015-16, la *International Football Association Board* permite a los equipos utilizar GPS en partidos oficiales, y no ha sido hasta esta última temporada cuando se ha permitido que estos datos lleguen en tiempo real a los banquillos de los equipos a través de dispositivos electrónicos (tabletas, móviles u otros). Los GPS que llevan los jugadores se pueden observar claramente, por debajo de las camisetas, en la espalda justo entre los omóplatos, a la altura de los hombros (ver figura 1). Hay varias empresas que comercializan este producto, pero entre todas ellas podemos destacar a la andaluza *Real Track Systems*, que trabaja con su sistema WIMU PRO. Esta empresa de aplicación de los GPS a las Ciencias del Deporte está bastante extendida a nivel internacional y, por supuesto, entre los equipos de la LFP. Este dispositivo origina datos cada segundo por jugador, en más de 250 variables registradas (frecuencia cardíaca, saturación de oxígeno, distancia recorrida, velocidad máxima, aceleraciones/minuto, desaceleraciones/minutos, *esprints*, información táctica, entre otras muchas más).

También cabe resaltar cómo La Liga ha desarrollado su propio software (Mediacoach) para toma de datos durante los partidos, que se pueden complementar con los de Wimupro. El análisis de estos datos puede ser muy útil según la información que se quiera extraer. Una de esas utilidades podría ser detectar posibles lesiones de los jugadores, mediante la programación de algoritmos que avisen con alertas de dichos riesgos y, por tanto, sea aconsejable la sustitución de algún jugador o la disminución de su carga en los entrenamientos. También podría ser bastante útil el análisis de reconocimiento de patrones tácticos a partir de los datos aportados por estas tecnologías. Pero ¿cuál es el problema en la realidad? El problema en la realidad es que con esta tecnología, suponiendo que un equipo tiene 23 jugadores y realizan 200 sesiones de entrenamiento de 90 minutos en la temporada, y registrando únicamente 100 variables por segundo para cada jugador (en la realidad podrían ser más), y suponiendo que cada dato ocupa el menor valor posible en memoria (6 bytes), se tendría, en total, un volumen de datos de 149 TB, es decir estaríamos tratando con un volumen de datos muy superior al de una PC normal. Si además consideramos que durante una temporada hay unos 40 partidos, se ve claramente cómo en el fútbol profesional existe la necesidad de trabajar con tecnología *Big Data* y con analistas de datos profesionales procedentes del ámbito de las matemáticas/estadística. Como curiosidad, podemos comentar que el Liverpool FC tiene tres analistas de datos con este perfil científico en su Departamento de Análisis de Rendimiento. Es más, todos los equipos de la Premier League Inglesa cuentan, al menos, con un especialista en este campo, por no enumerar los correspondientes en el deporte profesional americano (NBA, MLB y NFL).

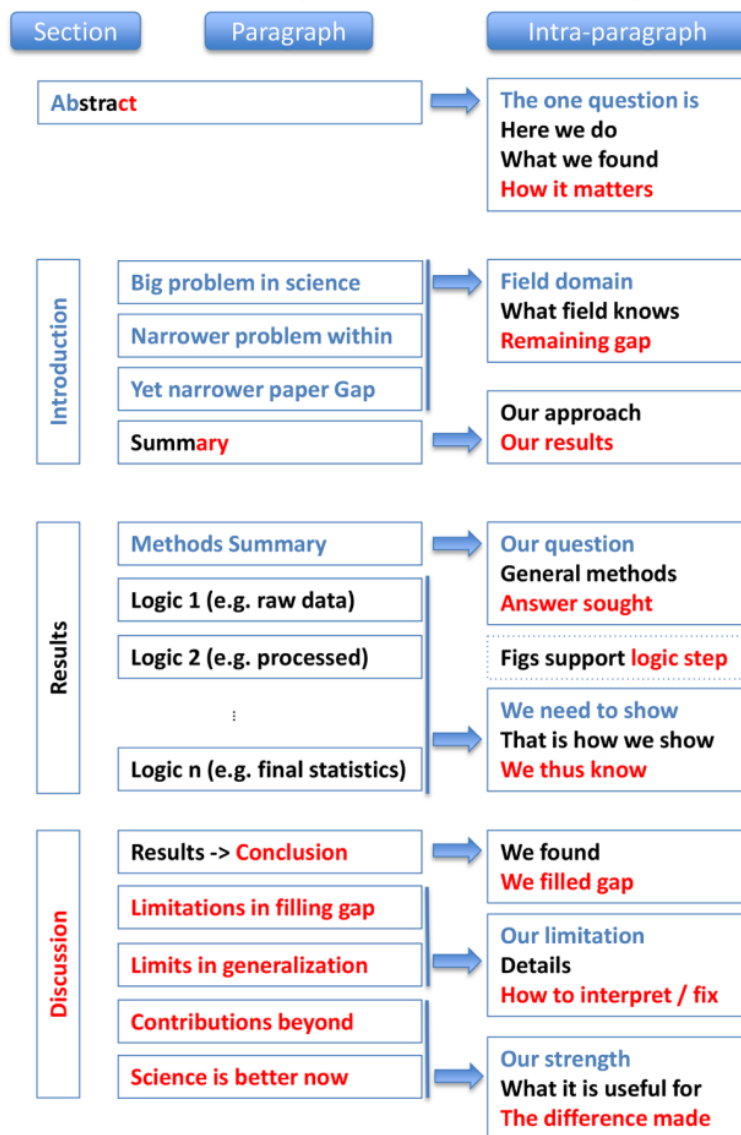
En definitiva, podemos afirmar que *Big Data* en el fútbol no es solamente una máquina fenomenal, sino más bien, que son muchas máquinas haciendo un trabajo que el humano no podría hacer. Evidentemente, el *Big Data* no es un sistema infalible, y menos en el fútbol, que es, probablemente, el deporte más azaroso de los que existen, aunque parezca todo lo contrario. No obstante, es una herramienta necesaria en los equipos profesionales de fútbol para la toma de decisiones, donde ahí sí debe intervenir el factor humano basado en la experiencia y el conocimiento. El *Big Data* ya está presente en el fútbol español y, tarde o temprano, los equipos profesionales tenderán a contar con especialistas en matemáticas, estadística y computación (algunos ya los tienen) en sus secretarías técnicas, porque, por mucha privacidad de datos que se garantice ¿contrataríamos para nuestro club de fútbol a una empresa de analítica del deporte que también trabaje para nuestro rival deportivo?

## **2. ¿Qué es un artículo científico? ¿Qué partes tiene normalmente? ¿Qué tipología hay según el ámbito de especialización?**

Cuando entramos en el mundo científico, parece que muchos conceptos o habilidades son obvios, pero no es así. Conocer la información y qué tipo de secciones puede

encontrarse (ejemplo: introducción, métodos, resultados, discusión y conclusiones) es realmente importante para poder interpretar mejor el contenido del artículo en cuestión.

**Figura 4: Summary of a paper's structural elements at three spatial scales: Across sections, across paragraphs, and within paragraphs**



Fuente: Mensh, B. y Kording, K. (2017).

### 3. ¿Qué revistas puedo encontrar de *Sports Analytics*?

Con la creciente fascinación de la sociedad por el deporte y la progresiva disponibilidad de datos relacionados con el deporte, hay grandes oportunidades para realizar una investigación de análisis deportivo. En el artículo publicado por Swartz (2018), se explican algunos problemas relevantes que pueden encontrar algunos científicos cuantitativos para publicar en el campo de analítica deportiva. Además, se proporcionan algunas experiencias y opiniones que nos hacen conocer las revistas que hacen referencia a estadística deportiva, tecnología y computación deportiva, análisis del

deporte, y ciencias del deporte y medicina del deporte. Las dos revistas académicas más populares del análisis del deporte provienen de Estados Unidos: *Journal of Quantitative Analysis in Sports (JQAS)* (<https://www.degruyter.com/view/j/jqas?lang=en>) y *Journal of Sports Analytics (JSA)* (<https://www.journalofsportsanalytics.com/>).

### 3.3 Análisis del 4º párrafo del CASO

Si nos fijamos en el siguiente párrafo:

“El científico de datos tiene un poco de prisa y me presenta a uno de los analistas tácticos o de juego del FCB del primer equipo. Este me facilita una **base de datos (validada por el equipo de IT)** de los últimos 15 partidos de liga de todos los equipos (**no han podido obtener los 10 primeros partidos**), donde hay diferentes **características** de los equipos (ejemplo: Estadio Home/Away), formación inicial (ejemplo: 4-3-3, 4-4-2), porcentaje posesión, number shot left, number shot middle, number shot right, shot conversion rate, número de corners, número de faltas, resultado final) para que pueda **ver algunos patrones y describir comportamientos** de alguna de ellas. El analista me comenta que **no tiene una pregunta concreta ahora mismo** para hacerme, pero hace tiempo que le gustaría **saber si hay una asociación** entre el hecho de marcar antes que el contrario y el resultado final del partido”.

Nos podríamos hacer preguntas del siguiente tipo:

4. **¿Qué es validación de los datos? ¿Qué son realmente los sesgos y qué tipos hay?**
5. **¿Qué diferencia hay entre muestra de datos y población? ¿Qué tipo de variables nos podemos encontrar?**
6. **¿Qué nos aporta la estadística descriptiva?**
7. **¿Y la estadística inferencial?**

Para contestar la mayoría de las preguntas (de la 4 a la 7) es importante y útil refrescar y aprender conceptos del mundo de la educación estadística. Si quien está realizando el curso viene del mundo del deporte, hay diferentes libros que se han explicado en la academia en donde se ejemplifican algunos de estos conceptos mediante deporte y estadística (O'Donoghue, P. y Holmes, L., 2014). Otros conceptos relacionados pueden ser también explorados en detalle en bibliografía más aplicada y moderna o más focalizada en estadística avanzada y deporte (O'Donoghue, P., 2013).

Aunque no es posible abarcar ni resumir el conocimiento de la estadística y de la estadística deportiva más avanzada en un minicurso, se mencionarán, a continuación, algunos aspectos académicos que no pueden pasarse por alto en este ámbito antes de

empezar a analizar delante de un ordenador y de adentrarse más en el histórico e inicial *statistical thinking* (Albert, J. y Koning, R. H. Eds.,2007).

## **Un viaje por algunos conceptos básicos del mundo de la estadística**

Muchos de nosotros hemos estudiado una asignatura de estadística en nuestras carreras universitarias, pero ya no recordamos de qué trataba ni qué utilidad tenía. Algunos han tenido la suerte de tener un profesor más motivador, que puede haberles mostrado algún video como este: [https://www.youtube.com/watch?v=cQx\\_yPOpPpk](https://www.youtube.com/watch?v=cQx_yPOpPpk) para empezar esta asignatura llena de interesante conocimiento. La estadística ya no es definida como una de las ramas de las matemáticas, como ya hemos visto, sino que es la ciencia de aprender desde los datos, donde se mide, se controla y se comunica la incertidumbre. Se trata de una ciencia joven, que alguna parte de la sociedad todavía no conoce. Una prueba de ello es que normalmente se confunde al estadístico con el estadista (profesión del Estado) o se piensa que el estadístico es una especie de informático que calcula medias, medianas y que hace algunos cálculos numéricos. La estadística cuantifica la incertidumbre y asesora sobre cómo recoger los datos para que estos aporten la máxima información posible. En la medicina se ha estudiado la bioestadística, que es la ciencia que hace referencia a la aplicación y al desarrollo de la estadística en las ciencias de la vida y, más particularmente, en las ciencias de la salud. La gran diferencia entre la estadística y la bioestadística es que, en esta última, usted tiene que estar familiarizado con otras disciplinas, como epidemiología, genética, demografía, salud pública, etcétera. Lo mismo pasa en *Sports Analytics*. La gran diferencia entre esta y estadística es que hay que conocer también otras disciplinas para conocer más sobre este fenómeno. Cuando leemos artículos científicos o queremos realizar proyectos de investigación, es importante entender diferentes términos estadísticos básicos:

- Cómo diferenciar entre población (objetivo de nuestro estudio y todos los datos que queremos conocer) y muestra (subconjunto de la población y los datos con los cuales podemos trabajar). A partir de aquí, es importante conocer cómo se selecciona la muestra y, para ello, diferentes técnicas de muestreo (por ejemplo: aleatorias, estratificadas) y sus posibles sesgos de selección o información (Elfil, M. y Negida, A., 2017).
- Saber diferenciar la naturaleza, el nivel de medición o escala y el tipo de variables (cualitativas o cuantitativas) o características sobre las cuales trabajamos es clave para realizar análisis adecuados sobre estas. Un punto importante del correcto uso de la estadística es conocer bien qué es lo que realmente queremos contestar. Preguntas como qué ocurre (estadística descriptiva), qué ocurrió (diagnóstico), qué ocurrirá (predicción) o qué deberíamos hacer (prescripción) nos señalan posibles herramientas

estadísticas que se pueden utilizar. Otra distinción clave para tener en cuenta es que los estudios pueden explorar datos observados (estadísticas descriptivas) o utilizar datos observados de una muestra para inferir la población de estudio (estadísticas inferenciales). De hecho, actualmente, según uno de los científicos y estadísticos más populares del mundo de la salud, hay tres campos bien diferenciados a estudiar que responden a tres preguntas diferentes y nos pueden acercar al objetivo del análisis rápidamente: Estadística descriptiva, predicción e inferencia causal (Hernán, M. A., Hsu, J. y Healy, B., 2019).

La estadística descriptiva nos muestra los datos que tenemos en ese momento mediante, por ejemplo, el uso de tablas de frecuencia y gráficos adecuados, según las variables con las que estemos trabajando. También se utilizan diferentes medidas de centralización, dispersión, posición y forma para las variables cuantitativas. Por ejemplo, en la Universidad de Carolina del Norte, en Chapell Hill, solían presumir que sus estudiantes de Geografía eran los que lograban un salario medio más alto al licenciarse, a diferencia de lo que ocurría con los estudiantes de otras facultades. Lo que no sabían es que allí había un alumno que se licenció en Geografía vía una beca deportiva. Este alumno era Michael Jordan, cuyo salario, como usted puede suponer, era diferente al de los otros que se licenciaron. En este caso, no tendríamos que calcular la media del salario, sino otra medida de centralización, como la mediana. Tal como nos decía el infografista Alberto Cairo: “la estadística no miente, miente la persona que la manipula”. También sería importante acompañar esta medida de centralización con alguna medida de dispersión, para conocer siempre la variabilidad de sus datos. Cuando describimos variables cuantitativas, es útil utilizar gráficos de cajas, también llamados *box-plot* u otras alternativas que hay en la bibliografía, donde se muestran varias medidas, como el mínimo, el máximo, la mediana y el primer y tercer cuartil, que permiten descubrir posibles *outliers*. Imaginen que el analista dice que el percentil 3 de goles de córner recibidos de los equipos de la liga es 2. Esto querrá decir que el 3 % de los equipos reciben 2 o menos goles de córner. Por ello, será siempre de interés saber comunicar e interpretar este tipo de medidas descriptivas básicas.

La inferencia estadística pretende obtener conclusiones sobre una población a partir de una muestra. Es importante conocer dos conceptos, como lo son las diferencias estadísticamente significativas o las diferencias clínicamente, tácticamente, técnicamente. económicamente (u otras) relevantes. ¿Cuál de los dos conceptos es más importante? Probablemente el segundo, pero, para contestar, necesitamos el primero. Estamos en un mundo donde solemos compararnos constantemente. Uno de los puntos que nos permite comparar, relacionar, testear o estimar es el que se conoce como contraste o test de hipótesis. También está el concepto de modelización. El fundamento de muchas conclusiones científicas publicadas es el concepto de significación estadística, normalmente evaluada mediante un índice denominado *p-valor*. Un *p-valor* es la probabilidad, según un modelo estadístico especificado, de que un estadístico que

sintetiza alguna característica de los datos (por ejemplo, la diferencia de las medias al comparar dos grupos) sea igual o más extremo que el valor observado. Ahora bien, a pesar de que el p-valor puede ser una medida estadística útil, a menudo se lo emplea de forma incorrecta, se abusa de él y también se lo malinterpreta. Esto ha llevado a que algunas revistas científicas no recomienden tanto su uso. En este contexto, la American Statistical Association (ASA) facilitó una declaración formal para la comunidad científica, donde aclara algunos principios que son ampliamente aceptados y que están implícitos en la correcta utilización e interpretación del p-valor (The American Statistical Association, 2016). Un índice único no debería sustituir el razonamiento científico. También han sido propuestas algunas alternativas a este índice (como el intervalo de confianza [IC], el d de Cohen, las medidas de efecto, etc.) y vale la pena tenerlas presentes. Como consecuencia de la *Declaración de la American Statistical Association* sobre la significación estadística y los valores p, el *New England Journal of Medicine* publicó recientemente un editorial sobre *New Guidelines for Statistical Reporting in the Journal* (Harrington, D. *et al.*, 2019). El editorial comienza afirmando que tal vez algunos lectores puedan haber notado una información más parsimoniosa en relación con la información de los valores p, y explican que esto se debe a que, recientemente, se implementaron directrices de la revista que limitan el uso de los valores p para comparaciones secundarias y otros. Exponen que ahora han aclarado, ampliado y perfeccionado las directrices estadísticas para los autores, para abarcar tanto ensayos clínicos como estudios observacionales. Puede encontrar, a continuación, el enlace a las pautas estadísticas para autores: *Statistical reporting guidelines* (<https://www.nejm.org/author-center/new-manuscripts>).

- Cuando relacionamos dos variables cuantitativas que tienen una relación lineal, se especifica que estas variables están correlacionadas. Este concepto (relación lineal) se confunde muchas veces con causalidad, un concepto bien distinto para el que hay que tener en cuenta el diseño y otros aspectos más complejos.
- Es común comparar nuestra variable objetivo o dependiente con otra característica de interés o variable independiente. Aun así, por ejemplo, en medicina deportiva, como pasa en otros campos, muchas veces nuestro objetivo (por ejemplo: lesionarse [sí/no]) está asociado a más de una covariable (por ejemplo: lesión previa, edad, superficie de juego, momento de la temporada) y hay que tener en cuenta los múltiples factores, algunos latentes y, sobre todo, dinámicos que pueden influir. La especificación y la validación de los modelos estadísticos de regresión, y la toma en cuenta de variables confusoras o modificadoras, son aspectos que hay que verificar para conocer la etiología de las lesiones o los factores asociados a ellas, y también para la utilización de modelos predictivos.

- La mayoría de las preguntas del análisis del deporte para las que un staff quiere respuesta son: el diagnóstico, el pronóstico y la búsqueda de factores de riesgo de un problema en concreto. Pueden ser factores variables (la posición que ocupa en el terreno de juego) y factores no modificables (como la edad). En la medicina deportiva, también se suele hablar de factores intrínsecos (por ejemplo: edad, etnia, sexo, predisposición genética, historia previa de lesión) o extrínsecos (por ejemplo: nutrición, factores psicológicos, balance de fuerza muscular, carga, fatiga, flexibilidad). El médico trata de reconocer todo esto, mediante modelos estadísticos que identifiquen esos factores y los ayude. Junto con esto, en la última década, se viene trabajando en bioestadística aplicada a la medicina personalizada, donde se intenta responder qué dosis de tratamiento hay que suministrar a un paciente en función de sus características únicas, ya que el tratamiento no tiene por qué ser igual para todos los pacientes que sufran la misma enfermedad o lesión, ver: <https://barcainnovationhub.com/es/bioestadistica-ciencias-del-deporte/>. De todas formas, muchas veces, conocer los factores de riesgo o identificar subpoblaciones que aumentan o disminuyen el riesgo de lesión no es suficiente, y hay que intentar responder cómo y por qué se ha desarrollado esta lesión o enfermedad.

- En el campo del análisis del deporte, es también fundamental conocer los modelos estadísticos y las nociones de probabilidad. En cuanto a los modelos estadísticos, es de gran importancia tener presente que, como algunos de los mejores entrenadores nos recuerdan, el espacio y el tiempo son claves a la hora de tomar decisiones. Se están desarrollando modelos estadísticos que tienen en cuenta estos dos conceptos (temporal y espacial) en diferentes especializaciones de *Sports Analytics*. Algunas reflexiones actuales se pueden encontrar en este sitio: <https://barcainnovationhub.com/es/espacios-futbol-perspectiva-cuantitativa/>. Aunque, como decía George Box, es también bueno recordar que “Essentially, all models are wrong, but some are useful”. Por último, respecto al concepto de probabilidad, aparte de adentrarse en conceptos como probabilidad condicionada, teorema de Bayes, es también interesante tener presente el sistema de cualificación ELO (método para calcular los niveles relativos de habilidad de los jugadores del ajedrez) y diferentes métricas aplicadas al deporte. Los modelos estadísticos y la probabilidad nos llevan a conceptos del mundo del deporte como las diferentes métricas populares descritas en baloncesto, como el PER, Usage, +/-, Winscore, PTC, entre otras. Y también en el mundo del fútbol. a los ya conocidos goles esperados (xG) entre otras, donde hay una extensa bibliografía en: [https://wikieducator.org/Sport Informatics and Analytics/Performance Monitoring/Expected Goals#Expected goals literature](https://wikieducator.org/Sport%20Informatics%20and%20Analytics/Performance%20Monitoring/Expected%20Goals#Expected%20goals%20literature) e incluso algún paquete R creado: (<https://github.com/Jelagmil>).

# Referencias

- Albert, J. y Koning, R.** (Eds.) (2007). *Statistical thinking in sports*. CRC Press.
- Albert, J., Glickman, M., Swartz, T. y Koning, R.** (Eds.) (2017). *Handbook of Statistical Methods and Analyses in Sports*. CRC Press.
- Barca Innovation Hub** (9 mayo 2018). Los espacios en el fútbol desde una perspectiva cuantitativa. Recuperado de: <https://barcainnovationhub.com/es/espacios-futbol-perspectiva-cuantitativa/>
- Barca Innovation Hub** (20 agosto 2018). La importancia de la bioestadística en las ciencias del deporte y la medicina deportiva. Recuperado de: <https://barcainnovationhub.com/es/bioestadistica-ciencias-del-deporte/>
- Elfil, M. y Negida, A.** (2017). Sampling methods in clinical research; an educational review. *Emergency*, 5(1).
- Harrington, D. et al.** (2019). *New guidelines for statistical reporting in the journal*.
- Hernán, M., Hsu, J. y Healy, B.** (2019). A second chance to get causal inference right: A classification of data science tasks. *Chance*, 32(1), 42-49.
- Horton, M.** (2018). *Algorithms for the Analysis of Spatio-Temporal Data from Team Sports*.
- Jones, I.** (2014). *Research methods for sports studies*. Routledge.
- Journal of Quantitative Analysis in Sports (JQAS)**. Recuperado de: (<https://www.degruyter.com/view/j/jqas?lang=en>)
- Journal of Sports Analytics (JSA)**.  
Recuperado de: (<https://www.journalofsportsanalytics.com/>)
- Kai Zhang** (2018). How big data has created a big crisis in science. 05/01/2019, The Conversation. Recuperado de: <https://theconversation.com/how-big-data-has-created-a-big-crisis-in-science-102835>
- Lagos, J.** Create your own github profile. Recuperado de: <https://github.com/Jelagmil>
- Malthouse, E. y Li, H.** (2017). Opportunities for and pitfalls of using big data in advertising research. *Journal of Advertising*, 46(2), 227-235.
- Mensh, B. y Kording, K.** (2017). Ten simple rules for structuring papers. *PLoS computational biology*, 13(9), e1005619.
- Miller, T.** (2016). Sports Analytics and Data Science. *Winning the Game with Methods and Models*.
- Newell, J., Aitchison, T. y Grant, S.** (2014). *Statistics for sports and exercise science: a practical approach*. Routledge.
- O'Donoghue, P.** (2013). *Statistics for sport and exercise studies: An introduction*. Routledge.
- O'Donoghue, P. y Holmes, L.** (2014). *Data analysis in sport*. Routledge.
- Osborne, J.** (17 julio 2017). Understanding sports through Statistics. Recuperado de: [https://www.youtube.com/watch?v=cQx\\_yPOpPpk](https://www.youtube.com/watch?v=cQx_yPOpPpk)

**Quintana O.** (22 marzo 2019). Real track systems. El primer ejemplo de éxito del Barca Innovation Hub. Recuperado de: [https://www.culemania.com/palco/real-track-systems-primer-ejemplo-exito-barca-innovation-hub\\_231195\\_102.html](https://www.culemania.com/palco/real-track-systems-primer-ejemplo-exito-barca-innovation-hub_231195_102.html)

**Swartz, T.** (2018). Where Should I Publish My Sports Paper? *The American Statistician*, (just-accepted), 1-18.

**Wiki Educator.** Sports Informatic and Analytics/ Performance Monitoring/ Expected Goals. Recuperado de: [https://wikieducator.org/Sport Informatics and Analytics/Performance Monitoring/Expected Goals#Expected\\_goals\\_literature](https://wikieducator.org/Sport Informatics and Analytics/Performance Monitoring/Expected Goals#Expected_goals_literature)