

# Módulo 1. Fundamentos de estadística

## Introducción

Bienvenido al primer módulo del curso 'Fundamentos de *data science* y *storytelling*', donde explorarás los fundamentos esenciales de la estadística aplicada al ámbito deportivo. Este módulo está diseñado para proporcionar a los profesionales y entusiastas del deporte las herramientas necesarias para entender y aplicar estadísticas en sus áreas de interés, mejorando la toma de decisiones y optimizando estrategias basadas en datos.

La estadística es una poderosa herramienta que, correctamente utilizada, puede ofrecer *insights* significativos sobre el rendimiento, las tácticas y la planificación en diversos deportes. A lo largo de este módulo, aprenderás no solo los conceptos básicos de la estadística, sino también cómo aplicar estos conocimientos para analizar y mejorar el rendimiento en el fútbol, baloncesto, fútbol sala y balonmano. Con un enfoque práctico, cada lección está diseñada para ser aplicable y relevante en el contexto deportivo real.

## Objetivos del módulo

Los objetivos principales de este módulo son los siguientes.

1. Introducir los principios básicos y conceptos de la estadística, facilitando una comprensión clara de su importancia y aplicación en el deporte.
2. Mostrar cómo las técnicas estadísticas pueden ayudar a tomar decisiones más informadas y efectivas en el ámbito deportivo.
3. Desarrollar habilidades prácticas para el análisis de datos deportivos, utilizando ejemplos y datos reales de fútbol y otros deportes mencionados.

Además de los objetivos generales, este módulo está diseñado para que puedas aplicar los conocimientos adquiridos en situaciones reales y simulaciones basadas en datos históricos y actuales de deportes populares. Mediante la realización de ejercicios prácticos y el análisis de casos de estudio, podrás ver el impacto directo de la estadística en la estrategia y el rendimiento deportivo.

Para asegurar una comprensión profunda y aplicable, cada concepto estadístico introducido será acompañado de ejemplos detallados relacionados con el fútbol, el baloncesto, el fútbol sala y el balonmano. Estos ejemplos serán discutidos tanto desde un punto de vista teórico como práctico, facilitando así la conexión entre la teoría estadística y su aplicación en el campo deportivo.



En este módulo, también se enfatizará la importancia de la interpretación de los datos, no solo su cálculo. Aprenderás a leer y analizar tablas de resultados deportivos, interpretar gráficos y entender cómo las medidas de centralidad y dispersión pueden prever tendencias y comportamientos en los deportes. Este enfoque no solo mejora tu capacidad analítica, sino que también te prepara para presentar y discutir hallazgos estadísticos con otros miembros de tu equipo y *stakeholders*.

Finalmente, revisaremos varios estudios de caso que demuestran la aplicación efectiva de la estadística en decisiones críticas dentro del deporte. Estos casos ayudarán a solidificar la teoría aprendida y mostrarán cómo los profesionales pueden utilizar la estadística para mejorar sus decisiones y estrategias. Los estudios de caso incluirán análisis de partidos, evaluación de la efectividad de diferentes formaciones y tácticas, y el seguimiento de la progresión y el desarrollo de los jugadores a lo largo del tiempo.

A través de este módulo integral, no solo adquirirás conocimientos fundamentales en estadística, sino que también desarrollarás la capacidad de aplicar estos conocimientos de manera práctica y efectiva en tu entorno profesional. Con una combinación de teoría robusta y ejemplos prácticos, este módulo servirá como una base sólida para los siguientes módulos del curso y para las aplicaciones profesionales futuras de cada participante.

¿Qué es la estadística y su importancia en el deporte?

La estadística es una rama de las matemáticas que se ocupa de la recolección, análisis, interpretación y presentación de datos numéricos. Su aplicación en el deporte se ha vuelto esencial para mejorar la competitividad y efectividad de los equipos y atletas. A través de métodos estadísticos, los entrenadores y analistas deportivos pueden extraer información valiosa de grandes cantidades de datos, lo que permite una mejor toma de decisiones basada en evidencia real y no solo en intuiciones.

En el contexto deportivo, la estadística ayuda a analizar el rendimiento de los jugadores y equipos, predecir resultados de partidos, y optimizar estrategias y tácticas. Por ejemplo, en el fútbol, análisis estadísticos pueden indicar cuál formación es más efectiva contra un determinado rival, basándose en datos históricos de enfrentamientos previos y rendimientos actuales. Esta capacidad para analizar y utilizar datos objetivos permite a los equipos adaptar sus entrenamientos y estrategias de juego de manera más efectiva.

La estadística también juega un papel crucial en la planificación de entrenamientos y la gestión de la carga física de los atletas, ayudando a prevenir lesiones y maximizar el rendimiento. Además, permite evaluar la efectividad de diferentes



técnicas de entrenamiento y regímenes nutricionales, proporcionando una base sólida para decisiones sobre la preparación física y técnica de los deportistas.

La importancia de la estadística en el deporte se extiende a la gestión deportiva, donde se utiliza para optimizar la asignación de recursos, mejorar la interacción con los aficionados y aumentar los ingresos mediante la venta de entradas y productos. En resumen, la estadística es indispensable para la modernización y profesionalización del deporte, proporcionando herramientas que permiten a los equipos y atletas alcanzar su máximo potencial.

#### Importancia de la interpretación de los datos

La correcta interpretación de los datos estadísticos es vital en el ámbito deportivo, ya que permite transformar la información cruda en conocimientos aplicables que impactan directamente en la toma de decisiones y en la estrategia general de los equipos y organizaciones deportivas. Interpretar datos con precisión significa entender no solo lo que los números muestran, sino también el contexto en el que se generan y cómo pueden utilizarse para predecir tendencias y resultados futuros.

En el deporte, una interpretación efectiva de los datos puede revelar patrones ocultos en el rendimiento de los jugadores, la efectividad de las tácticas y las dinámicas del juego, lo que es esencial para la preparación antes de los partidos y las modificaciones tácticas durante los mismos. Por ejemplo, el análisis de las estadísticas de tiro de un equipo de baloncesto puede ayudar a identificar qué jugadores son más efectivos en momentos críticos, lo que a su vez puede influir en las decisiones del entrenador durante el juego.

Además, la interpretación de los datos ayuda a mejorar el *scouting* de jugadores y la selección de talentos, identificando las características que mejor se adaptan al estilo de juego del equipo. También es crucial para evaluar la salud y el rendimiento físico de los atletas, permitiendo ajustes personalizados en los programas de entrenamiento para optimizar el rendimiento y prevenir lesiones.

#### Uso práctico de la estadística en deportes específicos

La aplicación práctica de la estadística en deportes específicos se manifiesta a través del análisis detallado de cada aspecto del juego, lo que permite una comprensión más profunda de las estrategias y tácticas efectivas. En deportes de equipo como el fútbol y el baloncesto, la estadística se utiliza para evaluar el rendimiento de los jugadores, analizar las formaciones y entender mejor las interacciones entre diferentes jugadores y equipos.



Por ejemplo, en el fútbol, la estadística se emplea para medir la efectividad de diferentes formaciones, la precisión en los pases, la efectividad defensiva y la capacidad ofensiva. Esta información es crucial para los entrenadores que buscan maximizar la efectividad de su equipo ajustando la alineación y la táctica según el oponente y las condiciones del juego. En el baloncesto, el análisis estadístico puede mostrar las tendencias de los jugadores en tiros de campo y tiros libres, lo cual es fundamental para diseñar jugadas que maximicen las posibilidades de anotación.

En deportes individuales, como el atletismo y la natación, la estadística ayuda a monitorizar el rendimiento y el progreso de los atletas a lo largo del tiempo, proporcionando datos claves que pueden influir en la modificación de técnicas y en la planificación del entrenamiento. Esto es esencial para preparar a los atletas para competiciones importantes y para realizar ajustes tácticos que pueden ser decisivos durante los eventos.

Estas aplicaciones de la estadística en diferentes deportes demuestran su valor como herramienta para mejorar el rendimiento y la estrategia, tanto en deportes de equipo como individuales. La capacidad de utilizar datos para tomar decisiones informadas es una ventaja competitiva que puede diferenciar a los equipos y atletas de élite de sus competidores.

## **Clasificación de variables: cualitativas y cuantitativas**

En estadística, las variables son elementos básicos en la recolección y análisis de datos. Se clasifican en dos categorías principales: cualitativas y cuantitativas. Esta distinción es crucial para elegir el método estadístico adecuado para analizar los datos y para interpretar correctamente los resultados obtenidos.

### **Variables cualitativas (o categóricas)**

Las variables cualitativas, también conocidas como categóricas, son fundamentales en la estadística deportiva porque permiten clasificar elementos en categorías basadas en atributos o cualidades distintivas. Estas variables sirven para describir características que no requieren mediciones numéricas para su análisis, pero son vitales para entender los contextos deportivos y tomar decisiones estratégicas. Se dividen, principalmente, en dos tipos: nominales y ordinales, cada una con un papel específico en el análisis de datos.



## Variables nominales

Las variables nominales clasifican datos en categorías que no tienen un orden específico o jerárquico. Estas variables son útiles para segmentar datos en grupos distintos sin ninguna implicación de orden. En el contexto deportivo, pueden notarse los siguientes ejemplos típicos.

**Nacionalidad de los atletas:** permite agrupar a los atletas por su país de origen, útil en análisis de torneos internacionales o competiciones donde la diversidad de participación es un factor de interés.

**Tipo de disciplina deportiva:** en eventos multideportivos como los Juegos Olímpicos, clasificar a los atletas por el deporte que practican ayuda a organizar y analizar los resultados de manera más efectiva.

**Color de equipamiento:** en deportes de equipo, el color de la camiseta puede ser crucial para identificar rápidamente a los equipos en estudios de video para análisis táctico o en investigaciones de mercadotecnia y *branding* deportivo.

El análisis de variables nominales a menudo involucra el uso de modas o frecuencias para identificar la categoría más común o para explorar la distribución de casos entre las categorías. Este tipo de variable es esencial para cualquier análisis descriptivo que busque clasificar o identificar elementos dentro de un conjunto de datos, sin necesidad de inferir orden o magnitud.

## Variables ordinales

Las variables ordinales, a diferencia de las nominales, implican un orden o jerarquía entre las categorías, aunque la magnitud exacta entre los niveles no sea conocida. Esto es particularmente útil en el deporte para clasificar o calificar atributos que tienen un grado inherente de intensidad o preferencia, pero donde las diferencias exactas entre esos grados no son cuantificables. A continuación, se mencionan algunos ejemplos.

**Nivel de competencia:** atletas o equipos pueden ser clasificados como principiantes, intermedios o avanzados, lo que ayuda en la organización de ligas y competiciones para asegurar que las competencias sean justas y equilibradas.

**Categorías de peso en deportes de combate:** en boxeo, judo o lucha, las categorías de peso ordenan a los competidores en clases que suponen un equilibrio de fuerza y habilidad, facilitando enfrentamientos más seguros y competitivos.



**Grado de lesión:** en medicina deportiva, las lesiones pueden clasificarse en grados como menor, moderado y severo, lo que es crucial para diagnosticar tratamientos y estimar tiempos de recuperación.

El análisis de variables ordinales puede incluir técnicas como la mediana o percentiles que reflejan la centralidad o dispersión de los datos dentro de un orden definido. Estas variables son particularmente útiles para evaluar tendencias o para realizar análisis comparativos en situaciones donde el orden es importante, pero las diferencias exactas entre las categorías no son necesarias para el propósito del estudio.

### **Variables cuantitativas (o numéricas)**

Las variables cuantitativas, también conocidas como numéricas, son importantes en el análisis estadístico dentro del deporte y en muchos otros campos, debido a que proporcionan una base medible y objetiva para la recopilación de datos y el análisis posterior. Estas variables se clasifican en dos tipos principales: variables discretas y variables continuas, cada una con características y aplicaciones específicas en el análisis de datos deportivos.

#### Variables discretas

Las variables discretas son aquellas que toman valores contables y finitos. En el contexto deportivo, estas variables son particularmente útiles porque permiten el recuento de ocurrencias de eventos específicos dentro de un juego o temporada. A continuación, se nombran algunos ejemplos típicos de variables discretas en deportes.

**Número de goles:** en deportes como el fútbol o el hockey, contar los goles durante un partido proporciona una medida directa del rendimiento ofensivo de un equipo.

**Número de faltas:** este indicador puede ayudar a analizar la disciplina de un equipo o jugador durante un partido.

**Cantidad de cambios:** en juegos donde los cambios son limitados, el número de cambios realizados puede impactar en la estrategia del juego.

El análisis de variables discretas a menudo implica técnicas estadísticas que consideran la naturaleza contable de los datos, como las pruebas de chi-cuadrado para independencia, que pueden examinar si dos variables discretas están relacionadas entre sí. Además, los modelos de regresión logística pueden ser



adecuados cuando la variable dependiente es discreta, como predecir si un equipo gana o pierde basado en diversas métricas de juego.

## Variables continuas

Por otro lado, las variables continuas son aquellas que pueden tomar cualquier valor dentro de un rango específico y son esenciales para medir magnitudes. En el deporte, estas variables ofrecen una rica fuente de información sobre el rendimiento y las capacidades físicas de los atletas.

**Velocidad de un lanzamiento:** en el balonmano, la velocidad de lanzamiento es crucial y se mide en millas por hora o kilómetros por hora.

**Distancia recorrida:** en deportes de equipo como el fútbol o el rugby, se mide la distancia total que un jugador recorre durante un partido, proporcionando *insights* sobre su condición física y esfuerzo.

**Tiempo de posesión:** en baloncesto o balonmano, se contabiliza el tiempo que tarda el equipo en realizar el ataque/defensa, dato que puede ser útil a la hora de plantear tácticas.

El análisis de variables continuas frecuentemente utiliza técnicas como la estadística descriptiva para resumir los datos (medias, medianas, modas), y la inferencia estadística, como las pruebas t y ANOVA, para comparar medias entre diferentes grupos o condiciones. La correlación y la regresión lineal son también métodos comunes para investigar relaciones entre variables continuas, como la relación entre la velocidad de un atleta y su tiempo de reacción.

La habilidad para identificar y clasificar correctamente las variables cuantitativas sirve no solo para seleccionar los métodos estadísticos apropiados, sino también para la interpretación correcta de los resultados. Por ejemplo, entender si una variable es discreta o continua puede determinar el tipo de gráfico más adecuado para su representación visual y el análisis estadístico más efectivo para extraer conclusiones válidas. En el deporte, donde las decisiones basadas en datos se están volviendo cada vez más críticas, tener un sólido entendimiento de estas variables y cómo analizarlas puede proporcionar ventajas competitivas significativas, permitiendo a entrenadores y analistas maximizar el rendimiento y mejorar las estrategias de juego.

El estudio de las distribuciones de probabilidad constituye una base esencial en el análisis estadístico, siendo aplicable en una variedad amplia de disciplinas que van desde la economía y la ingeniería hasta áreas menos convencionales como la psicología y, particularmente, el deporte. La relevancia de estas distribuciones en el



deporte radica en su capacidad para modelar y predecir una variedad de situaciones y resultados que dependen de eventos aleatorios y variables inciertas. Este texto no solo explora los conceptos básicos de probabilidad, sino que también introduce algunas de las distribuciones más comunes: la normal, la binomial y la Poisson. Cada una de estas distribuciones proporciona herramientas distintas para el análisis y la interpretación de datos en situaciones deportivas, desde el rendimiento de los atletas hasta las estrategias de juego y la toma de decisiones tácticas. La distribución normal, por ejemplo, es ideal para modelar características que siguen una simetría alrededor de un promedio, como el tiempo promedio en carreras o la altura y peso de los jugadores. Por su parte, las distribuciones binomial y de Poisson son excepcionales para eventos discretos, como el número de goles o puntos anotados en un partido. Al discutir estas distribuciones y sus aplicaciones en el deporte, este texto busca proporcionar una comprensión clara de cómo la estadística puede ser una herramienta poderosa para mejorar el análisis y la planificación en el ámbito deportivo.

### Conceptos básicos de probabilidad

La probabilidad es una disciplina matemática que cuantifica la incertidumbre asociada a diferentes tipos de eventos, desde fenómenos naturales hasta decisiones humanas y resultados deportivos. Es una medida cuantitativa que indica qué tan probable es que ocurra un evento, expresándose como un número entre 0 y 1. Un valor de 0 significa que el evento es imposible, mientras que un valor de 1 indica que el evento es seguro y ocurrirá con certeza. Esta escala proporciona una manera de medir la certeza de los eventos de una manera matemáticamente rigurosa y objetiva.

En la teoría de la probabilidad, los investigadores y analistas utilizan modelos matemáticos para simular y entender comportamientos en contextos de incertidumbre, permitiendo realizar inferencias y predicciones basadas en la estructura de estos modelos. Por ejemplo, en el contexto de un partido de fútbol, la probabilidad puede ayudar a predecir el resultado del partido basándose en estadísticas pasadas y la *performance* actual de los equipos, entre otros factores.

La probabilidad se divide en dos categorías principales: empírica y teórica. La probabilidad empírica se deriva de la experiencia directa y la observación de eventos. Por ejemplo, si observamos que un equipo de fútbol ha ganado 30 de sus últimos 50 partidos en casa, podríamos estimar que la probabilidad de que el equipo gane el próximo partido en casa es de 0.6. Este tipo de probabilidad es útil porque se basa en datos reales y puede ajustarse a medida que se disponga de más información.



Por otro lado, la probabilidad teórica se basa en modelos matemáticos y suposiciones lógicas. No depende directamente de la observación de eventos reales, sino de la definición teórica de cómo deberían comportarse los eventos bajo condiciones ideales. Un ejemplo clásico es el lanzamiento de una moneda equilibrada, donde la probabilidad de obtener cara o cruz es siempre 0.5, independientemente de cualquier otro factor.

Las variables aleatorias son conceptos fundamentales en la teoría de la probabilidad. Una variable aleatoria puede tomar diferentes valores dependiendo del resultado de un proceso aleatorio. Estas variables se clasifican en dos tipos: discretas y continuas. Las variables discretas son aquellas que pueden tomar un conjunto finito o contable de valores. Por ejemplo, el número de goles marcados en un partido de fútbol es una variable discreta porque el número puede ser contado y cada gol se considera un incremento unitario.

En contraste, las variables continuas se refieren a resultados que pueden tomar cualquier valor numérico dentro de un rango o intervalo, a menudo asociados con mediciones. Un ejemplo sería el tiempo que tarda un atleta en completar una carrera. Este tipo de variables es crucial en experimentos donde la precisión y la variabilidad continua son inherentes a la naturaleza del fenómeno observado.

El entendimiento profundo de estas variables es esencial para aplicar correctamente métodos estadísticos y realizar análisis adecuados. Por ejemplo, saber si una variable debe tratarse como discreta o continua, afecta el tipo de pruebas estadísticas que se pueden utilizar, como pruebas de chi-cuadrado para variables discretas o pruebas t para variables continuas.

Además, la probabilidad juega un papel crucial en el desarrollo de modelos predictivos en deportes y otras áreas. Permite a los entrenadores y gestores deportivos hacer predicciones más precisas sobre los resultados de los partidos, evaluar estrategias y mejorar el rendimiento de los atletas basándose en datos históricos y tendencias calculadas.

### Distribuciones comunes

La aplicación de las distribuciones de probabilidad en estadísticas sirve para comprender y modelar diversos fenómenos, tanto en ciencias naturales como en disciplinas humanas, incluido el deporte. Este concepto se utiliza para describir y predecir comportamientos y resultados basados en la incertidumbre inherente de los eventos deportivos. Las distribuciones de probabilidad más comúnmente utilizadas en el ámbito deportivo son la normal, la binomial y la de Poisson, cada una adecuada para diferentes tipos de datos y situaciones.



### Distribución normal

La distribución normal, o gaussiana, es notable por su relevancia y omnipresencia en muchos campos del saber. Se caracteriza por su curva simétrica en forma de campana centrada en la media,  $\mu$ , y su dispersión determinada por la desviación estándar,  $\sigma$ . Esta distribución es fundamental porque muchos fenómenos tienden a comportarse de manera normal bajo condiciones naturales.

En el contexto deportivo, la distribución normal se utiliza extensamente para modelar y analizar variables que naturalmente tienden a agruparse alrededor de un promedio. Por ejemplo, características físicas como la altura y el peso de los atletas suelen seguir una distribución normal, permitiendo a los entrenadores y científicos del deporte evaluar cómo se comparan individualmente los atletas dentro de un grupo. Además, rendimientos como tiempos de carrera o puntuaciones en eventos deportivos frecuentemente se distribuyen de manera normal, facilitando la identificación de rendimientos estadísticamente significativos o anómalos.

### Distribución binomial

La distribución binomial es esencial para modelar el número de éxitos en una serie de pruebas independientes, donde cada prueba tiene dos posibles resultados: éxito o fracaso. Los parámetros que define esta distribución son  $n$ , el número de ensayos, y  $p$ , la probabilidad de éxito en cada ensayo.

En deportes, esta distribución tiene aplicaciones prácticas significativas. Por ejemplo, en el baloncesto, puede utilizarse para calcular la probabilidad de que un jugador realice un determinado número de tiros libres exitosos sobre un total de intentos, basándose en su rendimiento histórico. Esto es útil no solo para estadísticas de juego, sino también para desarrollar estrategias específicas de entrenamiento y para la toma de decisiones tácticas en situaciones críticas del juego.

### Distribución de Poisson

La distribución de Poisson se emplea para modelar el número de veces que un evento ocurre en un intervalo de tiempo o espacio dado, asumiendo que estos eventos suceden con una tasa constante y de manera independiente entre sí. Esta distribución es particularmente útil en deportes para predecir eventos que son aleatorios y espaciados.

Por ejemplo, en el fútbol, la distribución de Poisson puede ser usada para modelar el número de goles que se espera que un equipo marque en un partido, dado su promedio de goles por partido. Este tipo de modelado es invaluable para los analistas deportivos que buscan predecir resultados de partidos o desarrollar



modelos de apuestas deportivas. Además, puede aplicarse para estimar la ocurrencia de lesiones o faltas en un partido, lo que ayuda en la preparación y estrategia de los equipos.

#### Aplicaciones en el deporte

Las distribuciones de probabilidad tienen un amplio rango de aplicaciones en el mundo deportivo, fundamentalmente para optimizar estrategias, perfeccionar programas de entrenamiento y realizar predicciones basadas en análisis estadísticos. Cada tipo de distribución ofrece herramientas específicas que se adaptan a diferentes necesidades y situaciones dentro de los deportes, desde el análisis del rendimiento individual de los atletas hasta la evaluación de tácticas de equipo y la predicción de resultados en competiciones.

La distribución normal, por su naturaleza de modelar variables que tienden a agruparse alrededor de un promedio, es excelente para identificar atletas que destacan del grupo. En atletismo, por ejemplo, el tiempo de carrera en eventos como los 100 metros planos puede seguir una distribución normal. Los entrenadores pueden usar esta distribución para identificar corredores cuyos tiempos se desvían significativamente del promedio, señalando potencial para competencias de élite o la necesidad de intervenciones en su entrenamiento. Del mismo modo, en deportes como la natación o el ciclismo, donde los tiempos y las medidas de rendimiento son críticos, la distribución normal ayuda a clasificar a los atletas según su rendimiento en relación con el grupo, facilitando decisiones estratégicas sobre formación y selección.

La distribución binomial es particularmente útil para analizar situaciones donde los resultados se dividen claramente entre éxito y fracaso. En el baloncesto, esta distribución puede aplicarse para estudiar la efectividad de los tiros libres de un jugador. Si un jugador tiene un 75 % de probabilidad de encestar cada tiro libre, la distribución binomial puede calcular la probabilidad de que el jugador acierte una cantidad específica de tiros en una serie dada, lo cual es crucial para diseñar estrategias en los últimos minutos de un juego. Otro ejemplo podría ser en el fútbol, donde un delantero podría tener ciertas probabilidades de convertir un penal; la distribución binomial permite al entrenador entender mejor las posibilidades de éxito en estos momentos críticos.

Por último, la distribución de Poisson es excelente para modelar el número de veces que un evento ocurre en un intervalo de tiempo determinado, bajo la suposición de que estos eventos suceden de manera independiente y a una tasa constante. En el fútbol, por ejemplo, se puede utilizar para predecir el número de goles que un equipo podría marcar en un partido, basado en la tasa promedio de goles por partido. Esta información es invaluable para los entrenadores al planificar tácticas ofensivas y



defensivas. En deportes como el hockey sobre hielo o el rugby, la distribución de Poisson también puede ser útil para estimar la frecuencia de penalizaciones o *tries* durante un partido, lo que ayuda a los entrenadores a ajustar sus estrategias y preparaciones de juego.

En resumen, las distribuciones de probabilidad permiten a los profesionales del deporte tomar decisiones informadas y científicamente fundamentadas. Al entender y aplicar estas distribuciones, los entrenadores y analistas pueden mejorar significativamente sus estrategias, optimizar el entrenamiento, y aumentar las probabilidades de éxito en competiciones. Estos métodos estadísticos no solo mejoran la comprensión de los fenómenos deportivos, sino que también contribuyen a elevar el nivel de rendimiento y competitividad en el deporte.

### Estadística descriptiva

La estadística descriptiva desempeña un papel crucial en el análisis de datos en una amplia variedad de campos, incluyendo, de manera significativa, el deporte. Esta rama de la estadística ofrece herramientas esenciales para resumir o describir las características fundamentales de un conjunto de datos. Estas herramientas son esenciales a menudo con el objetivo de realizar comparaciones sistemáticas o de informar decisiones críticas basadas en datos objetivos. En el ámbito deportivo, la estadística descriptiva es invaluable para entrenadores, jugadores y analistas, ya que les permite comprender de manera más efectiva el rendimiento y las capacidades tanto individuales como de equipo, facilitando así la optimización de estrategias y fomentando una mejora continua en todas las áreas de práctica y competición.

Este texto se centra específicamente en las medidas de centralidad y dispersión, que son dos de los componentes más fundamentales de la estadística descriptiva. Además, se proporcionan ejemplos concretos aplicados a deportes como el fútbol, el baloncesto y el balonmano, mostrando cómo estos conceptos se utilizan para evaluar y mejorar el rendimiento en diferentes contextos deportivos. Con estos fundamentos, se puede profundizar en cómo la recolección y el análisis detallado de datos desempeñan un rol esencial en el desarrollo y la ejecución de estrategias deportivas efectivas.

### Medidas de centralidad: media, mediana y moda

Las medidas de centralidad, también conocidas como medidas de tendencia central, son herramientas estadísticas esenciales que proporcionan un resumen conciso que representa el punto central de un conjunto de datos. Estas medidas ofrecen una



manera sencilla, pero poderosa de comprender la distribución de los datos. Las medidas de centralidad más comúnmente utilizadas son la media, la mediana y el modo. Cada una de estas tiene aplicaciones específicas y es útil en diferentes contextos, incluidos varios deportes como el fútbol, baloncesto y balonmano.

### Media (promedio)

La media, comúnmente conocida como el promedio, es quizás la medida de centralidad más conocida y utilizada. Se calcula sumando todos los valores numéricos de un conjunto de datos y luego dividiendo esta suma total por el número de valores en el conjunto. Matemáticamente, si tenemos un conjunto de valores  $x_1, x_2, \dots, x_n$ , la media  $\bar{x}$  se calcula como:

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n}$$

Donde  $\sum_{i=1}^n x_i$  denota la suma de todos los valores, y  $n$  es el número de valores.

En el ámbito deportivo, la media se utiliza para calcular el promedio de varios tipos de datos cuantitativos, como el número de puntos, goles o cualquier otra estadística relevante. Por ejemplo, en baloncesto, la media de puntos anotados por un jugador a lo largo de una temporada puede ser un indicador crucial de su rendimiento ofensivo. Si un jugador anota 25, 30 y 20 puntos en tres juegos consecutivos, su promedio de puntos por juego se calcularía sumando los puntos anotados en cada juego (25 + 30 + 20) y dividiendo el resultado por el número de juegos jugados, lo que resultaría en:

$$\text{Promedio de puntos por juego} = \frac{25 + 30 + 20}{3} = 25$$

### Mediana

La mediana es el valor que se encuentra en el medio de un conjunto de datos cuando estos están ordenados en forma ascendente o descendente. Si el número de observaciones en el conjunto de datos es impar, la mediana es simplemente el valor medio. Si el número de observaciones es par, la mediana se calcula como el promedio de los dos valores medios. Este valor es especialmente valioso para entender el rendimiento típico en situaciones donde los datos pueden estar sesgados por valores atípicos.

Por ejemplo, en el deporte del balonmano, consideremos un jugador que ha marcado 2, 4, 5, 7 y 8 goles en cinco partidos distintos. Al ordenar los datos (2, 4, 5, 7,



8), es claro que la mediana, o el número medio de goles por partido, es 5. Este valor mediano proporciona una medida robusta del rendimiento del jugador, minimizando el efecto de partidos anómalamente buenos o malos.

### Moda

El modo es la medida de centralidad que identifica el valor o valores que aparecen con mayor frecuencia en un conjunto de datos. Es particularmente útil en el análisis de datos categóricos o para identificar el resultado más común en un conjunto de datos. En el contexto deportivo, conocer el modo puede ayudar a identificar tendencias o patrones comunes en el rendimiento de un equipo o jugador.

Por ejemplo, en el fútbol, si se registra que un equipo ha marcado 1 gol en cinco partidos, 2 goles en diez partidos y 3 goles en cinco partidos durante una temporada, el número de goles que más frecuentemente ha marcado el equipo, el modo, sería 2 goles por partido. Este dato es útil para entender la capacidad ofensiva más común del equipo y puede influir en las decisiones tácticas para mejorar el rendimiento del equipo en futuros partidos.

### Medidas de dispersión: rango, varianza y desviación estándar

Las medidas de dispersión son estadísticas vitales que describen cuán esparcidos están los datos alrededor de un centro, que generalmente es la media. Estas medidas son cruciales para comprender la variabilidad en un conjunto de datos, lo que ayuda a evaluar la consistencia, la variabilidad y los posibles valores atípicos dentro de los datos. Entre las medidas de dispersión más utilizadas se encuentran el rango, la varianza y la desviación estándar, cada una proporcionando una perspectiva única sobre la distribución de los datos.

### Rango

El rango es la diferencia entre el valor más alto y el más bajo en un conjunto de datos. Se calcula simplemente restando el valor mínimo del máximo. Matemáticamente, si tenemos un conjunto de valores  $x_1, x_2, \dots, x_n$ , donde  $x_{max}$  es el valor máximo y  $x_{min}$  es el valor mínimo, el rango ( $R$ ) se calcula como:

$$R = x_{max} - x_{min}$$

El rango proporciona una estimación rápida y fácil de la dispersión en los datos, mostrando directamente la extensión de los datos. Sin embargo, es importante notar que el rango puede ser altamente sensible a valores atípicos, ya que un solo dato extremo puede alterar significativamente el rango. En el contexto deportivo, el rango



se utiliza frecuentemente para evaluar la consistencia del rendimiento. Por ejemplo, en baloncesto, si el máximo de puntos que un equipo ha anotado en un partido es 110 y el mínimo es 70, el rango es 40. Este rango indica la variabilidad en el rendimiento de anotación del equipo de un partido a otro, lo cual puede ser útil para el entrenador al planificar estrategias que aborden esta inconsistencia.

### Varianza

La varianza es una medida estadística que describe la dispersión de los datos alrededor de su media. Indica cuán lejos tienden a estar los valores individuales del promedio. La varianza se calcula como el promedio de las diferencias al cuadrado entre cada valor y la media del conjunto de datos. Formalmente, para un conjunto de valores  $x_1, x_2, \dots, x_n$  con media  $\underline{x}$ , la varianza ( $\sigma^2$ ) se calcula como:

$$\sigma^2 = \frac{\sum_{i=1}^n (x_i - \underline{x})^2}{n}$$

En deportes, la varianza es útil para entender la consistencia en el rendimiento de los jugadores o equipos. Por ejemplo, si un portero de fútbol permite 0, 1, 2, 0 y 3 goles en cinco partidos diferentes, calcular la varianza de los goles permitidos puede ofrecer una perspectiva sobre su consistencia entre juegos. Una varianza baja indicaría un rendimiento consistentemente similar entre los partidos, mientras que una varianza alta podría indicar una gran inconsistencia.

### Desviación estándar

La desviación estándar es la raíz cuadrada de la varianza y se utiliza para medir la dispersión de los datos en la misma unidad que los datos originales. Esto la hace particularmente útil para comparar la dispersión entre diferentes conjuntos de datos. La desviación estándar ( $\sigma$ ) se calcula como:

$$\sigma = \sqrt{\sigma^2} = \sqrt{\frac{\sum_{i=1}^n (x_i - \underline{x})^2}{n}}$$

En el contexto del baloncesto, la desviación estándar de los puntos anotados por un jugador en diferentes partidos puede indicar cuánto varía su rendimiento de un juego a otro. Por ejemplo, si un jugador tiene una desviación estándar baja en los puntos anotados, esto sugiere que su rendimiento es bastante consistente de un partido a otro. Por el contrario, una desviación estándar alta podría indicar que el jugador tiene actuaciones muy variables, lo que podría ser un área para abordar durante los entrenamientos o en la preparación para los juegos.



### Ejemplos en deportes

En el ámbito del fútbol, las estadísticas descriptivas son herramientas invaluable para analizar la efectividad de diferentes formaciones y estrategias de equipo. Por ejemplo, si un equipo tiene un promedio de 2.5 goles marcados por partido y una media de 2.5 goles permitidos por partido, los datos indican que, en promedio, el equipo marca y concede un número similar de goles. Sin embargo, la mediana de goles marcados puede ser diferente, indicando si hay un sesgo en el rendimiento ofensivo del equipo. Si la mediana de goles marcados es 2, esto puede sugerir que el equipo tiende a marcar menos goles que el promedio, lo que podría influir en la decisión del entrenador para ajustar la formación o la estrategia.

Otro ejemplo en el fútbol podría ser el análisis de la varianza y la desviación estándar de las tarjetas amarillas recibidas por diferentes jugadores durante una temporada. Si el jugador A tiene una varianza de 1 y una desviación estándar de 1 en las tarjetas amarillas, y el jugador B tiene una varianza de 4 y una desviación estándar de 2, esto indica que el jugador B es más inconsistente en su comportamiento disciplinario, lo que podría ser un factor a considerar al planificar las alineaciones o las tácticas defensivas.

En el baloncesto, las medidas de dispersión como la varianza y la desviación estándar son útiles para evaluar la consistencia en el rendimiento de los jugadores en los tiros de campo. Si un jugador tiene una media de 20 puntos por partido, pero una alta desviación estándar de 10, su rendimiento es inconsistente, lo que podría ser problemático en situaciones de alta presión. Por otro lado, si otro jugador tiene una media de 15 puntos, pero una desviación estándar de 3, es más consistente, lo que podría ser preferible en escenarios cruciales. Estas estadísticas ayudan a los entrenadores a decidir cuáles jugadores son más confiables en momentos críticos del juego.

En el balonmano, el modo puede ser una medida particularmente útil al analizar el rendimiento defensivo de un equipo. Por ejemplo, si el modo de goles concedidos por un equipo en los últimos 10 partidos es 20, esto indica que el equipo tiende a conceder 20 goles con más frecuencia que cualquier otro número. Esto podría sugerir que el equipo tiene una estrategia defensiva efectiva para limitar los goles en la mayoría de los partidos, pero que podría ser vulnerable en partidos donde



concede significativamente más de 20 goles, lo que puede ser una señal de que se necesita mejorar la defensa en ciertas situaciones.

Estas herramientas estadísticas proporcionan una base sólida para la toma de decisiones basada en datos en el ámbito deportivo. Al utilizar medidas de centralidad y dispersión, los profesionales del deporte pueden planificar mejor sus estrategias, preparar a sus equipos para la competencia y maximizar las posibilidades de éxito. Por ejemplo, los análisis de la varianza y la desviación estándar pueden ser útiles para identificar inconsistencias en el rendimiento, mientras que el análisis de la media y la mediana puede ayudar a enfocar los esfuerzos de mejora.

### Estadística inferencial

La estadística inferencial es una rama de la estadística que se utiliza para hacer predicciones o inferencias sobre una población a partir de una muestra de datos. Es crucial en el análisis de datos en muchas áreas, incluyendo el deporte, ya que permite a los analistas deportivos, entrenadores y otros profesionales del deporte tomar decisiones informadas y fundamentadas en datos. A través de la estadística inferencial, se pueden hacer generalizaciones sobre poblaciones más grandes basándose en el análisis de muestras más pequeñas. Los dos conceptos principales en la estadística inferencial son los intervalos de confianza y las pruebas de hipótesis.

La inferencia estadística se basa en el análisis de una muestra de datos para hacer generalizaciones o predicciones sobre una población más amplia. Esto es fundamental en el deporte, donde a menudo se desea conocer la tendencia general o el rendimiento promedio de un equipo o jugador, basado en el análisis de un subconjunto de datos. La estadística inferencial utiliza técnicas como el muestreo aleatorio para obtener datos representativos de una población, y luego aplica métodos estadísticos para hacer inferencias o probar hipótesis sobre la población.

Por ejemplo, si un entrenador quiere saber si un nuevo plan de entrenamiento mejora el rendimiento de los jugadores, puede seleccionar una muestra de jugadores, implementar el plan, y luego usar la estadística inferencial para determinar si el rendimiento de estos jugadores ha mejorado significativamente. La inferencia estadística permite hacer este tipo de generalizaciones de manera científica, basada en el análisis de datos.



## Intervalos de confianza

Los intervalos de confianza son un concepto clave en la estadística inferencial. Un intervalo de confianza es un rango de valores que se utiliza para estimar un parámetro de población, como la media o la proporción. El intervalo de confianza proporciona un rango dentro del cual se espera que se encuentre el verdadero valor del parámetro, con un cierto nivel de confianza.

Para calcular un intervalo de confianza, se utiliza la siguiente fórmula general:

$$\text{Intervalo de confianza} = \underline{x} \pm z * \frac{\sigma}{\sqrt{n}}$$

Donde:

$\underline{x}$  = es la media de la muestra,

$z$  = es el valor crítico que corresponde al nivel de confianza deseado,

$\sigma$  = es la desviación estándar de la muestra,

$n$  = es el tamaño de la muestra.

Por ejemplo, si se desea calcular un intervalo de confianza del 95 % para el tiempo promedio de carrera de un atleta, se podría tomar una muestra de tiempos de carrera, calcular la media y la desviación estándar de la muestra, y luego utilizar la fórmula anterior para calcular el intervalo de confianza. Esto proporcionará un rango dentro del cual se espera que se encuentre el tiempo promedio de carrera del atleta con un 95 % de confianza.

En el deporte, los intervalos de confianza se utilizan para estimar parámetros como el rendimiento promedio de un jugador, el número de goles que un equipo podría marcar en un partido, o el tiempo promedio que un atleta puede completar un evento. Por ejemplo, si un entrenador de fútbol quiere estimar el número de goles que un delantero puede marcar en un próximo partido, podría analizar los datos de rendimiento de este delantero en partidos anteriores, calcular un intervalo de confianza para el número promedio de goles marcados, y utilizar este rango para planificar estrategias y tácticas.

### Uso práctico de los intervalos de confianza

#### **Ejemplo 1: intervalo de confianza para el rendimiento en carrera**

Supongamos que un entrenador desea estimar el tiempo promedio de carrera de un atleta. Para hacerlo, toma una muestra de tiempos de carrera de este atleta en 10



eventos. Los tiempos, en segundos, son los siguientes: 11.5, 11.6, 11.7, 11.8, 11.4, 11.6, 11.7, 11.8, 11.5, y 11.6. La media muestral  $\bar{x}$  es 11.62 y la desviación estándar  $\sigma$  es 0.14.

Para calcular un intervalo de confianza del 95 %, se utiliza un valor crítico  $z$  de 1.96 (para un nivel de confianza del 95 %):

$$\begin{aligned} \text{Intervalo de confianza} &= 11.62 \pm 1.96 * \frac{0.14}{\sqrt{10}} = 11.62 \pm 0.087 \\ &= (11.533, 11.707) \end{aligned}$$

Por lo tanto, el intervalo de confianza del 95 % para el tiempo promedio de carrera del atleta es entre 11.533 y 11.707 segundos.

### **Ejemplo 2: intervalo de confianza para el número de goles**

En otro caso, un entrenador de fútbol quiere estimar el número promedio de goles que un delantero podría marcar en un próximo partido. Analiza los datos de los últimos 12 partidos, en los cuales el delantero ha marcado los siguientes números de goles: 1, 2, 0, 1, 2, 3, 2, 2, 1, 0, 1, 2. La media muestral  $\bar{x}$  es 1.4167 y la desviación estándar  $\sigma$  es 0.831.

Para calcular un intervalo de confianza del 95 %, se utiliza un valor crítico  $z$  de 1.96 (para un nivel de confianza del 95 %):

$$\begin{aligned} \text{Intervalo de confianza} &= 1.4167 \pm 1.96 * \frac{0.831}{\sqrt{12}} = 1.4167 \pm 0.470 \\ &= (0.9467, 1.8867) \end{aligned}$$

Por lo tanto, el intervalo de confianza del 95 % para el número promedio de goles que podría marcar el delantero es entre 0.9467 y 1.8867.

### **Cambio mínimo detectable**

El cambio mínimo detectable (CMD) o *smallest worthwhile change* (SWC) es una herramienta estadística importante en el deporte, que permite a los entrenadores, analistas y atletas evaluar cambios significativos en el rendimiento. El CMD indica el cambio más pequeño en una medida que puede ser considerado relevante o significativo, es decir, un cambio que no se debe únicamente a la variabilidad natural o al error de medición.



## Conceptos de cambio mínimo detectable (CMD)

El CMD se utiliza para determinar si un cambio observado en el rendimiento de un atleta es lo suficientemente grande como para ser considerado significativo. La fórmula general del CMD es similar a la de un intervalo de confianza, pero con una estructura que se centra en el promedio y el rango de variación:

$$CMD = \underline{x} \pm (k * \sigma)$$

Donde:

$\underline{x}$  = es la media de las mediciones,

$k$  = es una constante que varía dependiendo de cuán robusto sea el dato, generalmente se utiliza  $k = 0.2$ ,

$\sigma$  = es la desviación estándar de las mediciones.

La constante  $k$  se ajusta en función del nivel de confianza deseado y la centralidad del dato. Si el dato es robusto, la constante será pequeña, pero si el dato es muy descentralizado, la constante será mayor.

## Uso práctico del CMD

### Ejemplo 1: velocidad en carrera

Supongamos que un atleta está entrenando para mejorar su velocidad en los 100 metros. Si el promedio de sus tiempos de carrera es de 11.5 segundos y la desviación estándar de sus tiempos es de 0.15 segundos, podemos calcular el CMD con  $k = 0.2$ :

$$CMD = 11.5 \pm (0.2 * 0.15) = 11.5 \pm 0.03$$

Esto significa que, para este atleta, cualquier cambio en el tiempo de carrera que sea menor a 0.03 segundos probablemente se deba a la variabilidad natural, mientras que un cambio mayor a 0.03 segundos sería significativo y podría indicar una mejora real en el rendimiento.

### Ejemplo 2: puntuación en baloncesto

Consideremos un jugador de baloncesto con un promedio de 20 puntos por partido y una desviación estándar de 2 puntos. Para evaluar si un cambio en su puntuación es significativo, podemos calcular el CMD con  $k = 0.2$ :

$$CMD = 20 \pm (0.2 * 2) = 20 \pm 0.4$$

En este caso, un cambio de más de 0.4 puntos en el rendimiento del jugador sería considerado significativo. Esto puede ayudar a los entrenadores a evaluar si el



jugador está en una tendencia de mejora o si un cambio en el rendimiento es simplemente parte de la variabilidad natural del juego.

#### Utilidad del CMD en el deporte

El CMD es especialmente útil para el monitoreo del rendimiento a lo largo del tiempo, ya que ayuda a identificar cambios significativos en el rendimiento de los atletas. Por ejemplo, en el fútbol, si un delantero tiene un promedio de 0.5 goles por partido con una desviación estándar de 0.1, el CMD puede ser utilizado para evaluar si un aumento en su rendimiento es significativo o si simplemente se debe a la variabilidad natural.

#### Contraste de hipótesis

El contraste de hipótesis es otra herramienta importante en la estadística inferencial, aunque en este curso no se va a desarrollar en profundidad por la complejidad que implica, vamos a dar una breve explicación con un ejemplo para que se vea su potencialidad. Se utiliza para probar afirmaciones o hipótesis sobre una población, basándose en el análisis de una muestra de datos. El proceso general para realizar un test de hipótesis es el siguiente.

1. Formular la hipótesis nula ( $H_0$ ) y la hipótesis alternativa ( $H_1$ ): la hipótesis nula es la afirmación que se quiere probar o refutar, mientras que la hipótesis alternativa es lo que se quiere demostrar. Por ejemplo, si se quiere saber si un jugador de baloncesto tiene un rendimiento superior a la media, la hipótesis nula podría ser que el rendimiento del jugador es igual a la media, mientras que la hipótesis alternativa sería que el rendimiento del jugador es superior a la media.
2. Seleccionar un nivel de significancia ( $\alpha$ ): el nivel de significancia es el umbral para decidir si se rechaza o no la hipótesis nula. Es común utilizar  $\alpha = 0.05$ , lo que significa que hay un 5 % de probabilidad de rechazar la hipótesis nula cuando es verdadera.
3. Calcular el valor del estadístico de prueba: el estadístico de prueba se calcula a partir de los datos de la muestra y se compara con un valor crítico para decidir si se rechaza o no la hipótesis nula. El tipo de estadístico de prueba depende de la hipótesis y de los datos disponibles.
4. Tomar una decisión: si el estadístico de prueba es mayor que el valor crítico, se rechaza la hipótesis nula y se acepta la hipótesis alternativa. De lo contrario, no se puede rechazar la hipótesis nula.

Ejemplo: comparación de rendimiento entre jugadores



Como ejemplo práctico, considera dos jugadores de baloncesto, A y B, que juegan en el mismo equipo. Quieres saber si el rendimiento de A es significativamente mayor que el de B, basado en sus puntuaciones promedio por partido. La hipótesis nula sería que el rendimiento de A y B es el mismo, mientras que la hipótesis alternativa sería que el rendimiento de A es mayor que el de B.

Los datos disponibles son los siguientes.

Jugador A:

Puntuaciones en 10 partidos: 15, 18, 14, 20, 25, 22, 18, 19, 21, 24

Media:  $\bar{x}_A = 19.6$

Desviación estándar:  $\sigma_A = 3.5$

Tamaño de muestra:  $n = 10$

Jugador B:

Puntuaciones en 10 partidos: 12, 15, 14, 16, 18, 17, 14, 15, 14, 13

Media:  $\bar{x}_B = 14.8$

Desviación estándar:  $\sigma_B = 1.7$

Tamaño de muestra:  $n = 10$

La hipótesis nula sería:

$$H_0: \bar{x}_A = \bar{x}_B$$

y la hipótesis alternativa sería:

$$H_1: \bar{x}_A > \bar{x}_B$$

Para comparar las medias de dos grupos independientes, puedes usar un test t para muestras independientes. El estadístico de prueba se calcula como:

$$t = \frac{\bar{x}_A - \bar{x}_B}{\sqrt{\frac{\sigma_A^2}{n_A} + \frac{\sigma_B^2}{n_B}}} = \frac{19.6 - 14.8}{\sqrt{\frac{3.5^2}{10} + \frac{1.7^2}{10}}} = 3.94$$

Con  $n_A + n_B - 2 = 18$  grados de libertad, el valor crítico para  $t$  a un nivel de significancia de  $\alpha=0.05$  es aproximadamente 1.73. Dado que  $t=3.94$  es mayor que 1.73, se rechaza la hipótesis nula y se concluye que el rendimiento del jugador A es significativamente mayor que el del jugador B.



## Series temporales

El análisis de series temporales es una técnica estadística que permite analizar datos recolectados en intervalos de tiempo uniformes. Es una herramienta especialmente útil en el ámbito deportivo, donde el rendimiento y la carga de entrenamiento de los atletas y equipos pueden variar a lo largo del tiempo. El análisis de series temporales es relevante para identificar patrones, tendencias y fluctuaciones en los datos, y proporciona información valiosa para entrenadores, analistas y otros profesionales del deporte, ayudando a mejorar el rendimiento y a tomar decisiones estratégicas informadas.

### Introducción a las series temporales y su relevancia en deportes

Las series temporales son secuencias de datos ordenadas en función del tiempo, lo que significa que cada dato se asocia con un punto específico en el tiempo. En el contexto deportivo, las series temporales se pueden utilizar para analizar la carga de entrenamiento de los atletas, evaluar el progreso de los equipos a lo largo de una temporada, y predecir resultados futuros basados en datos pasados.

El análisis de series temporales es relevante en deportes por varias razones.

1. **Rendimiento fluctuante:** los atletas y equipos no tienen un rendimiento constante a lo largo del tiempo. El análisis de series temporales permite identificar patrones de mejora o disminución en el rendimiento y ajustar la carga de entrenamiento en consecuencia.
2. **Planificación estratégica:** los equipos pueden usar series temporales para planificar estrategias, identificando tendencias y adaptándose a cambios en el rendimiento y la carga de entrenamiento.
3. **Prevención de lesiones:** el análisis de series temporales puede ayudar a identificar patrones de fatiga o sobrecarga, permitiendo a los entrenadores ajustar los regímenes de entrenamiento para prevenir lesiones.

### Análisis de carga de entrenamiento a lo largo de una temporada

El rendimiento de los atletas y equipos puede variar significativamente a lo largo de una temporada, debido a factores como el estado físico, el nivel de competencia y la carga de trabajo. El análisis de series temporales permite evaluar cómo cambia el rendimiento y la carga de entrenamiento a lo largo del tiempo, identificando tendencias y fluctuaciones que pueden influir en la planificación y la estrategia.

El análisis de series temporales se puede realizar de diversas formas, desde métodos simples como gráficos de rendimiento hasta técnicas más complejas como el análisis de tendencias y estacionalidad. Una de las técnicas más comunes es el uso de promedios móviles, o *rolling averages*, que suavizan las fluctuaciones a corto plazo y resaltan las tendencias a más largo plazo.



### *Rolling average*

El *rolling average*, o promedio móvil, es una técnica que se utiliza para suavizar las fluctuaciones en una serie temporal y ayudar a identificar tendencias a corto y largo plazo. Se calcula tomando el promedio de un número fijo de puntos de datos consecutivos y luego desplazando el periodo de promediado a través de la serie temporal.

El *rolling average* es especialmente útil en deportes para identificar tendencias en el rendimiento o evaluar la consistencia de los jugadores. Existen dos tipos principales de *rolling average*.

#### **1. Rolling average a corto plazo**

El *rolling average* a corto plazo utiliza un periodo de promediado más corto, lo que permite identificar tendencias recientes o cambios rápidos en el rendimiento. Por ejemplo, si quieres evaluar la carga de entrenamiento de un jugador de baloncesto en sus últimos 3 partidos, puedes calcular el *rolling average* de 3 partidos, lo que proporcionará una visión actualizada del rendimiento reciente del jugador. Puedes hacer lo mismo para analizar cargas de entrenamiento si coges los datos de una semana.

#### **2. Rolling average a largo plazo**

El *rolling average* a largo plazo utiliza un periodo de promediado más largo, lo que permite identificar tendencias más estables o a largo plazo. Por ejemplo, si quieres evaluar la carga de entrenamiento de un jugador a lo largo de una temporada, puedes calcular el *rolling average* de 10 partidos, lo que proporcionará una visión más estable del rendimiento del jugador. Para hacer la comparativa con las cargas de entrenamiento, puedes coger los datos de 4 o 6 semanas.

Aplicaciones prácticas del cambio mínimo detectable y los intervalos de confianza  
El cambio mínimo detectable (CMD) y los intervalos de confianza son herramientas útiles para detectar fluctuaciones en series temporales y evaluar si los cambios en el rendimiento o la carga de entrenamiento son significativos.

#### **1. CMD para fluctuaciones significativas**

El CMD es útil para identificar cambios significativos en la carga de entrenamiento a lo largo del tiempo. Por ejemplo, si un jugador de baloncesto tiene una carga de entrenamiento promedio de 200 unidades con una desviación estándar de 20, el CMD se puede calcular como:

$$CMD = 200 \pm (0.2 * 20) = 200 \pm 4$$



Esto significa que un cambio de más de 4 unidades en la carga de entrenamiento sería significativo. Si el jugador tiene una carga de entrenamiento de 210 unidades en un día, esto indicaría un aumento significativo en la carga de entrenamiento.

En el análisis de series temporales, el CMD se puede usar para evaluar si los cambios en la carga de entrenamiento de un jugador o equipo en diferentes puntos en el tiempo son significativos. Si un jugador tiene una serie de cargas de entrenamiento de 200, 205, 190, 215 y 220 unidades en 5 días consecutivos, puedes usar el CMD para identificar si alguno de estos cambios es significativo.

## 2. Intervalos de confianza para fluctuaciones estadísticas

Los intervalos de confianza proporcionan un rango de valores dentro del cual se espera que se encuentre el verdadero promedio de la carga de entrenamiento de un jugador o equipo. Si el promedio del jugador es 200 unidades con una desviación estándar de 20, y se han recolectado datos durante 10 días, el intervalo de confianza del 95 % sería:

$$\text{Intervalo de confianza} = 200 \pm 1.96 * \frac{20}{\sqrt{10}} = 200 \pm 12.4 = (187.6, 212.4)$$

Esto significa que hay un 95 % de probabilidad de que el promedio verdadero de la carga de entrenamiento se encuentre entre 187.6 y 212.4 unidades.

En el análisis de series temporales, los intervalos de confianza se pueden utilizar para evaluar si la carga de entrenamiento de un jugador o equipo en un día específico es significativamente diferente del promedio. Si el jugador tiene una carga de entrenamiento de 220 unidades en un día, esto estaría fuera del intervalo de confianza, lo que sugiere que la carga de entrenamiento en ese día fue significativamente superior al promedio.

Ejemplo práctico: evolución de la carga de entrenamiento de un jugador

Para ilustrar el uso del análisis de series temporales en el deporte, consideremos el ejemplo de un jugador de baloncesto cuya carga de entrenamiento se mide en términos de unidades por día a lo largo de una temporada.

### 1. Datos iniciales

El jugador tiene las siguientes cargas de entrenamiento en los primeros 10 días: 200, 210, 190, 215, 220, 205, 195, 200, 210, 220.



## 2. Análisis del *rolling average* a corto plazo

Calculamos el *rolling average* de 3 días para evaluar la carga de entrenamiento reciente del jugador. Esto nos dará una visión más actualizada del rendimiento a corto plazo del jugador. Los valores se calculan de la siguiente manera:

1.  $(200 + 210 + 190)/3 = 200$
2.  $(210 + 190 + 215)/3 = 205$
3.  $(190 + 215 + 220)/3 = 208.33$
4.  $(215 + 220 + 205)/3 = 213.33$
5.  $(220 + 205 + 195)/3 = 206.67$
6.  $(205 + 195 + 200)/3 = 200$
7.  $(195 + 200 + 210)/3 = 201.67$
8.  $(200 + 210 + 220)/3 = 210$

El *rolling average* a corto plazo muestra que el jugador ha mantenido una carga de entrenamiento relativamente consistente, con ligeras variaciones.

## 3. Análisis del *rolling average* a largo plazo

Para evaluar la carga de entrenamiento a lo largo de la temporada, puedes calcular el *rolling average* de 10 días. Esto te dará una visión más estable del rendimiento del jugador. Si el jugador continúa con una carga de entrenamiento de 210 unidades en cada uno de los próximos 5 días, el *rolling average* se calculará de la siguiente manera:

1.  $(200 + 210 + 190 + 215 + 220 + 205 + 195 + 200 + 210 + 220)/10 = 206.5$
2.  $(210 + 190 + 215 + 220 + 205 + 195 + 200 + 210 + 220 + 210)/10 = 207.5$
3.  $(190 + 215 + 220 + 205 + 195 + 200 + 210 + 220 + 210 + 210)/10 = 207.5$
4.  $(215 + 220 + 205 + 195 + 200 + 210 + 220 + 210 + 210 + 210)/10 = 209.5$
5.  $(220 + 205 + 195 + 200 + 210 + 220 + 210 + 210 + 210 + 210)/10 = 209$

El *rolling average* a largo plazo muestra que el jugador ha tenido una carga de entrenamiento consistente, pero con una ligera disminución hacia el final, lo que podría ser una señal de fatiga o disminución del rendimiento.

## 4. Relación del *rolling average* a corto y largo plazo. Ejemplo con las cargas de entrenamiento

El *rolling average* (RA) es una técnica utilizada para suavizar las fluctuaciones en una serie temporal, permitiendo a los entrenadores y analistas deportivos observar las tendencias en el rendimiento o las cargas de entrenamiento.



El *rolling average* a corto plazo, generalmente de 3 a 7 días, ayuda a identificar tendencias recientes o cambios rápidos en las cargas de entrenamiento. Por ejemplo, si un atleta aumenta repentinamente su carga de entrenamiento, el RA a corto plazo reflejará este cambio más rápido que el RA a largo plazo. Esto puede ser útil para detectar picos o caídas repentinas en la carga de entrenamiento, lo que puede alertar a los entrenadores sobre el riesgo de fatiga o lesión.

El *rolling average* a largo plazo, que puede abarcar 10 días o más, ofrece una visión más estable y a largo plazo de las cargas de entrenamiento. Es útil para identificar tendencias sostenidas y evaluar la consistencia del entrenamiento. El RA a largo plazo es especialmente útil para monitorear el volumen de entrenamiento durante una temporada y asegurarse de que los atletas no estén sobreentrenando o subentrenando.

Ambos enfoques son importantes para la planificación estratégica y la prevención de lesiones, ya que proporcionan información crítica sobre cómo varía la carga de entrenamiento a lo largo del tiempo y cómo se pueden ajustar las estrategias de entrenamiento para optimizar el rendimiento y la salud de los atletas.

