



Facultad de Ciencias Económicas y Empresariales - ICADE  
Grado en Business Analytics

**Trabajo Fin de Grado**

**ANÁLISIS COMPARATIVO  
DE LA INFLUENCIA DE LAS  
DECISIONES ARBITRALES:  
CASO REAL MADRID VS.  
ATLÉTICO DE MADRID**

Autor: **Victoria García-Argüelles Fernández-España**

Director: **Jorge Martín Magdalena**

MADRID | Junio 2025

## RESUMEN

Las decisiones arbitrales han sido continuo objeto de investigación y controversia en el fútbol profesional debido a su capacidad de influir significativamente en los resultados de los partidos. Diversos estudios sugieren que el arbitraje no es plenamente imparcial; factores como la ventaja de jugar como local, la influencia del público o la reputación histórica de un equipo pueden condicionar el juicio de los árbitros. En este contexto, los enfrentamientos entre el Real Madrid y el Atlético de Madrid, caracterizados por su naturaleza profundamente competitiva y simbólica, resultan particularmente relevantes para considerar la posible presencia de patrones sistemáticos que reflejen cierta inclinación arbitral. Por lo tanto, el objetivo de este trabajo es investigar si se producen diferencias estadísticamente significativas en determinadas decisiones arbitrales que puedan evidenciar un posible sesgo en estos encuentros. Para ello, se recopilan los datos de los principales derbis oficiales disputados durante la última década, a los que se les aplican contrastes de diferencia de medias y medianas, así como las respectivas correcciones por comparaciones múltiples. Posteriormente, se extiende el análisis al resto de equipos de la liga española a través de algoritmos de *Machine Learning* para contextualizar así las diferencias detectadas. Los resultados manifiestan discrepancias notables en algunas variables arbitrales, lo que invita a reflexionar en torno a la ecuanimidad del arbitraje en escenarios de alta exigencia competitiva.

---

## PALABRAS CLAVE

Arbitraje, Fútbol, Decisiones arbitrales, Árbitros, Real Madrid, Atlético de Madrid

## **ABSTRACT**

Refereeing decisions have long been a subject of continuous inquiry and controversy in professional football due to their capacity to significantly influence match outcomes. Various studies suggest that refereeing is not entirely impartial; factors such as home advantage, crowd influence, or a team's historical reputation can affect referees' judgment. In this context, the matches between Real Madrid and Atlético de Madrid, marked by their highly competitive and symbolic nature, are particularly relevant for considering the possible presence of systematic patterns that may reflect a certain refereeing bias. Therefore, the aim of this study is to examine whether statistically significant differences can be identified in specific refereeing decisions, which could reveal such a bias in these encounters. To this end, data from the main official derbies played over the past decade are collected and subjected to tests for differences in means and medians, along with appropriate corrections given multiple comparisons. Subsequently, the analysis is extended to the rest of the teams in the Spanish league through Machine Learning algorithms to contextualize the detected differences. The results reveal notable discrepancies in certain refereeing variables, prompting a reflection on the neutrality of officiating in highly demanding competitive scenarios.

---

## **KEY WORDS**

Refereeing, Football, Refereeing decisions, Referees, Real Madrid, Atlético de Madrid

## RESUMEN GRÁFICO



**Objetivo:** Investigar la existencia de un sesgo sistemático en las decisiones arbitrales que favorezca o perjudique a uno de los dos equipos en los partidos disputados entre el Real Madrid y el Atlético de Madrid durante la última década.

**Variabes:** Faltas, Tarjetas Amarillas, Tarjetas Rojas, Penaltis, Fuera de Juego

### Fase 1 → Análisis estadístico

**Datos:** 39 partidos de las temporadas 2013/14 - 2023/24

**Herramientas:** Excel, R



- Recopilación manual de datos
- Contraste de diferencia de medias/medianas (prueba t de Student / prueba de Wilcoxon)
- Ajuste de significatividad (correcciones Bonferroni, Holm, Benjamini-Hochberg)
- Tamaño del efecto

### Fase 2 → Algoritmos de *Machine Learning*

**Datos:** 2.280 partidos de las temporadas 2021/22 - 2023/24

**Herramientas:** Python, R



- *Web scraping*
- *Clustering* jerárquico aglomerativo
- Regresión logística
- Árbol de clasificación
- Análisis comparativo top 3 equipos (Real Madrid, FC Barcelona, Atlético de Madrid)



**Resultado Fase 1:** Se detectan diferencias estadísticamente significativas en faltas, tarjetas amarillas y fuera de juego → con más sanciones al Atlético de Madrid. No se observan diferencias significativas en tarjetas rojas ni penaltis.

**Resultado Fase 2:** Los clubes que reciben menos sanciones arbitrales tienden a obtener mejores posiciones en LaLiga.

# ÍNDICE DE CONTENIDOS

Índice de figuras .....	1
<b>CAPÍTULO 1: INTRODUCCIÓN .....</b>	<b>2</b>
1.1    Finalidad y motivos.....	2
1.2    Objetivos de la investigación .....	4
1.3    Estructura y metodología.....	5
<b>CAPÍTULO 2: MARCO TEÓRICO .....</b>	<b>8</b>
2.1    Factores psicológicos en la toma de decisiones arbitrales.....	8
2.2    Evidencia empírica de sesgos arbitrales .....	11
2.3    Fundamentos para el análisis .....	14
<b>CAPÍTULO 3: MATERIAL Y MÉTODOS .....</b>	<b>16</b>
3.1    Planteamiento de hipótesis.....	16
3.2    Muestra de datos.....	16
<b>CAPÍTULO 4: ANÁLISIS Y RESULTADOS .....</b>	<b>18</b>
4.1    Análisis estadístico entre el Real Madrid y el Atlético de Madrid.....	18
4.2    Resultados del análisis entre el Real Madrid y el Atlético de Madrid .....	20
4.3    Algoritmos de <i>Machine Learning</i> para todos los equipos de LaLiga.....	27
4.4    Resultados del análisis para todos los equipos de LaLiga .....	30
<b>CAPÍTULO 5: CONCLUSIONES .....</b>	<b>38</b>
Limitaciones del estudio y futuras líneas de investigación .....	40
Declaración respecto al uso de herramientas IAG.....	41
Bibliografía.....	42
Anexos.....	48

## ÍNDICE DE FIGURAS

<b>Figura 1.</b> Boxplot de la comparación de tarjetas amarillas. ....	24
<b>Figura 2.</b> Boxplot de la comparación de faltas. ....	25
<b>Figura 3.</b> Boxplot de la comparación de fueros de juego.....	26
<b>Figura 4.</b> Matriz de correlación entre las variables. ....	31
<b>Figura 5.</b> Mapa de calor de similitudes entre equipos. ....	32
<b>Figura 6.</b> Gráfico de dispersión según las tarjetas. ....	33
<b>Figura 7.</b> Curva ROC del modelo logit.....	34
<b>Figura 8.</b> Árbol de clasificación para el ranking de equipos. ....	35
<b>Figura 9.</b> Comparación conjunta de variables arbitrales entre equipos. ....	36
<b>Figura 10.</b> Comparación dividida de variables arbitrales entre equipos. ....	37

# CAPÍTULO 1: INTRODUCCIÓN

---

## 1.1 FINALIDAD Y MOTIVOS

El fútbol se ha consolidado indiscutiblemente como el fenómeno deportivo de mayor alcance a nivel global, con una audiencia estimada de más de 3.500 millones de espectadores en todo el mundo (Singh, 2025). Esta dimensión le confiere una trascendencia cultural, social y económica muy superior a la de cualquier otro deporte. Su impacto se extiende más allá de las fronteras del terreno de juego, ejerciendo una influencia directa en la vida de sus seguidores, quienes viven con pasión las múltiples competiciones y se sienten profundamente unidos a los equipos y jugadores. Consiguientemente, el fútbol se ha convertido en una fuente de ingentes ingresos a nivel internacional, sustentada por transmisiones televisivas, ventas de entradas, patrocinios y productos derivados, erigiéndose así como una de las industrias más lucrativas y mediáticas del siglo XXI. Según el informe anual *Football Money League*, los clubes más grandes del mundo en términos económicos generaron un récord de 11.200 millones de euros la temporada pasada (Deloitte, 2025).

En este entorno donde las emociones y los intereses económicos convergen en niveles tan sumamente elevados, garantizar la máxima equidad y justicia en el desarrollo del juego es fundamental para poder mantener la confianza de los aficionados y jugadores, así como de las entidades involucradas. Ello requiere no sólo la correcta aplicación del marco normativo, sino también la toma de decisiones de manera puramente imparcial, especialmente de aquellas que puedan alterar el curso de un partido. La legitimidad del fútbol descansa en el fundamento de que los resultados deben reflejar exclusivamente el rendimiento deportivo, libres de cualquier interferencia ajena al mismo. En este contexto, la figura del árbitro adquiere una importancia vital, pues es el agente externo con mayor capacidad para incidir en el desarrollo y desenlace de los partidos. A diferencia de otros deportes como el tenis o el atletismo —en los cuales la aplicación del reglamento ofrece un margen interpretativo sustancialmente menor—, en disciplinas colectivas como el fútbol, el criterio arbitral debe aplicarse de manera constante, lo que incrementa significativamente su capacidad de influir en el juego (Guillén, 2023).

En consecuencia, las decisiones arbitrales han sido continuo objeto de análisis, debate y controversia en el fútbol profesional. Si bien es cierto que el arbitraje, al ser una actividad mediada por decisiones humanas, está intrínsecamente ligado a la interpretación personal, la ecuanimidad en la aplicación del reglamento es crucial, especialmente en un ámbito donde los márgenes de error son mínimos y el impacto de cada decisión puede ser decisivo. Diversos factores externos como la presión del público, la reputación de los equipos o la ventaja local pueden influir, bien consciente, bien inconscientemente, en las decisiones arbitrales (Dohmen y Sauermann, 2015; Goumas, 2014; Nevill et al., 2002). Pese a que la incorporación de avances tecnológicos —destacando el árbitro asistente de vídeo, comúnmente conocido como VAR— ha permitido reducir errores humanos y rectificar sentencias críticas de forma inmediata (Spitz et al., 2020), no ha logrado eliminar por completo la subjetividad inherente al arbitraje, pues la mayor parte de las decisiones sigue supeditada al criterio individual del árbitro principal o del equipo encargado del videoarbitraje. Esta limitación se hace especialmente patente en aquellas situaciones en las que las reglas del juego conceden cierto grado de discrecionalidad, como es el caso de la imposición de sanciones disciplinarias, la valoración de contactos dentro del área o la revisión de jugadas que podrían derivar en goles anulados o decisiones polémicas (Dadı y Yıldız, 2022).

Estas controversias arbitrales tienden a intensificarse en encuentros de alta rivalidad, en los que las emociones de los jugadores, entrenadores, aficionados y medios de comunicación alcanzan su punto álgido, aumentando aún más la presión sobre los colegiados. En términos generales, los derbis —encuentros disputados entre dos equipos que mantienen una constante rivalidad, mayoritariamente por motivos regionales (Real Academia Española, s.f., definición 2)— se caracterizan por la vorágine y tensión que comprenden, pues están sometidos a un simbolismo tanto histórico como emocional que trasciende lo cotidiano. Esta dimensión simbólica, notoriamente potenciada por una elevada repercusión mediática, exacerba la sensibilidad ante cualquier decisión arbitral (Buraimo et al., 2010). Así, el arbitraje se somete a un análisis minucioso por parte de aficionados, periodistas y analistas, lo que inevitablemente amplifica las críticas y expone posibles errores o percepciones de sesgo. De esta manera, los derbis constituyen un campo de estudio único para investigar la imparcialidad arbitral, pues la presión consustancial a los mismos magnifica las posibles influencias externas que afectan a las decisiones arbitrales.

Entre los derbis más emblemáticos de la historia del fútbol profesional se encuentra el “Derbi Madrileño”, disputado entre el Real Madrid Club de Fútbol y el Club Atlético de Madrid, los dos equipos más laureados de la capital de España. Mientras que el Real Madrid es considerado uno de los equipos más exitosos y prestigiosos a nivel mundial, el Atlético de Madrid ha demostrado ser un contendiente extraordinario, especialmente en la última década. Cabe señalar que, además de esta asimetría en términos de títulos y reconocimiento, diversos argumentos económicos respaldan la tesis de que la desigualdad entre los clubes desvirtúa las competiciones futbolísticas en España (Alonso, 2017). En este contexto, la percepción de que el arbitraje podría favorecer de manera sistemática a un club de mayor renombre, como el Real Madrid, o, por el contrario, simpatizar con un perfil competitivo y aguerrido, como el del Atlético de Madrid, añade una capa adicional de controversia que merece ser analizada con rigor. Así pues, investigar las decisiones arbitrales en los partidos entre estos dos equipos no sólo permite detectar posibles patrones de sesgo, sino que también ofrece una oportunidad para comprender cómo la intensidad y la repercusión de estos enfrentamientos pueden influir en la objetividad de los árbitros.

## **1.2 OBJETIVOS DE LA INVESTIGACIÓN**

El presente trabajo se centra en el análisis de determinadas decisiones arbitrales en los partidos disputados entre el Real Madrid y el Atlético de Madrid en la última década, con el principal objetivo de identificar posibles patrones que puedan evidenciar la presencia de un sesgo arbitral. El estudio se limita a las competiciones españolas más relevantes —esto es, el Campeonato Nacional de Liga de Primera División (LaLiga), la Copa de Su Majestad el Rey y la Supercopa de España—, así como a los torneos europeos de la UEFA en los que ambos clubes se han enfrentado entre sí —concretamente la Liga de Campeones (*Champions League*) y la Supercopa de Europa—.

Para lograr este objetivo general, se procederá a analizar aquellas decisiones arbitrales que representen situaciones críticas en el desarrollo y resultado de los encuentros, tales como el número de faltas señaladas, tarjetas amarillas y rojas, penaltis concedidos y fuera de juego sancionados. Todas estas variables han sido meticulosamente seleccionadas para el análisis por

ser cómodamente cuantificables y menos susceptibles a interpretaciones subjetivas —tal como se ha demostrado en estudios previos (Jiménez-Sánchez et al., 2021; Dawson et al., 2007; Sutter y Kocher, 2004; Gilis et al., 2008)—; aunque puedan depender de la percepción individual del árbitro, son parámetros registrados oficialmente y directamente atribuibles a su criterio, por lo que son menos propensos a verse influenciados por factores externos al mismo. Con ello, se pretende minimizar el ruido y garantizar que las diferencias halladas —de hallarse— reflejen verdaderas tendencias sistemáticas, y no sean fruto del azar o de circunstancias aleatorias del juego. Adicionalmente, se realiza un cruce de estas variables con los datos de todos los clubes de LaLiga, con el fin de contrastar los resultados y minimizar la posibilidad de que cualquier sesgo identificado sea consecuencia de factores ajenos al arbitraje, evitando así atribuciones simplistas que puedan reducir el fenómeno a meras características tácticas de los equipos.

### 1.3 ESTRUCTURA Y METODOLOGÍA

El enfoque metodológico llevado a cabo es esencialmente deductivo y cuantitativo, ya que parte de teorías previas y evidencia empírica existente sobre posibles sesgos arbitrales en el fútbol profesional para contrastarlas con el análisis estadístico de datos reales. La inmensa mayoría de referencias en las que se ha fundamentado este trabajo proceden de artículos e investigaciones publicados en bases académicas especializadas, empleando *Web of Science* y *Scopus* —y, en última instancia, *Google Scholar* (Google Académico)— como principales herramientas de búsqueda bibliográfica. El estudio busca explicar relaciones causales, tratando de identificar posibles correlaciones entre las decisiones arbitrales y los resultados para detectar potenciales patrones de sesgo. Ello se desarrolla en dos fases principales, empleando diversas herramientas para abordar el problema de investigación de manera rigurosa y exhaustiva.

En la primera fase, se realiza un análisis estadístico de muestras pareadas, compuestas por todos los partidos oficiales entre el Real Madrid y el Atlético de Madrid disputados entre las temporadas 2013/14 y 2023/24, correspondientes a competiciones nacionales —LaLiga, Copa del Rey y Supercopa de España— e internacionales —*Champions League* y Supercopa de Europa—. Las principales variables arbitrales a analizar, recopiladas de manera manual, comprenden el número de faltas, tarjetas amarillas, tarjetas rojas, penaltis y fueros de juego. Se

ha decidido no incluir elementos adicionales como la posición en el campo de cada falta o las posibles penalizaciones no señaladas debido a la dificultad de acceso a este tipo de datos y a la subjetividad de los mismos, respectivamente.

Para la efectiva recopilación de datos, se ha llevado a cabo un proceso de verificación cruzada. La principal fuente de información ha sido WhoScored debido a su amplia cobertura estadística. Para garantizar la precisión en los datos, éstos han sido copiosamente verificados con Transfermarkt, mayoritariamente. Sin embargo, ante la detección de discrepancias entre ambas páginas en algunos encuentros, se ha optado por dar mayor peso a la información de Transfermarkt, pues su amplia reputación en el ámbito futbolístico la convierte en una fuente más fiable en términos de consistencia y rigurosidad (Herberger y Wedlich, 2017; He, 2012). Cabe señalar que, dado que no todas las competiciones y variables analizadas están cubiertas en su totalidad por estas plataformas, se ha recurrido a otras fuentes de referencia. Para los partidos de la UEFA, se ha accedido a la base de datos oficial de la misma. A su vez, todo dato ha sido contrastado con diversas páginas complementarias, tales como FBref, BDFutbol, Sofascore, Footballdatabase.eu y ESPN.

Para cada variable, se aplica un contraste de diferencia de medias o de medianas para comparar los promedios de las penalizaciones arbitrales en ambos equipos y determinar si las diferencias entre éstos son significativas desde el punto de vista estadístico. En otras palabras, se decreta si las discrepancias entre las medias o medianas son lo suficientemente considerables como para no deberse al azar. Esto se realiza mediante pruebas estadísticas, principalmente la prueba *t* de Student, estableciendo como hipótesis nula que no existen diferencias entre los dos clubes en cada variable analizada. El nivel de significancia usado es de un *p*-valor de 0,05, lo que significa que se aceptaría una probabilidad del 5% de cometer un error tipo I —es decir, rechazar la hipótesis nula siendo ésta verdadera—. Ahora bien, dado que se llevan a cabo múltiples contrastes, el riesgo de incurrir en errores tipo I aumenta, pues el nivel de confianza general se reduce si no se ajusta el nivel de significancia para cada contraste particular. Por ello, se aplican las correcciones de Bonferroni, Holm y Benjamini-Hochberg, garantizando que las conclusiones sobre las diferencias arbitrales entre los equipos no estén sesgadas por falsos positivos. Finalmente, se cuantifica la magnitud de las diferencias observadas mediante un estudio del tamaño del efecto.

En la segunda fase, se extiende el alcance del estudio, trabajando con una base de datos notablemente más amplia que contiene información sobre las mismas variables anteriormente seleccionadas en la primera parte, pero, en este caso, aplicadas a todos los equipos que han competido en la LaLiga durante las últimas tres temporadas (de 2021/22 a 2023/24). De este modo, se pretende contextualizar el trato arbitral recibido por el Real Madrid y el Atlético de Madrid dentro del conjunto de equipos de la competición doméstica. Para la recopilación de estos datos, se ha realizado un proceso de *web scraping* sobre la plataforma FBref; la elección de esta fuente futbolística se debe a que su formato de datos es altamente accesible para la extracción automatizada, pues la información se presenta de manera tabular en HTML.

Una vez extraído y estructurado este segundo grupo de datos, se utiliza una técnica de *clustering* jerárquico para agrupar los equipos en función de las respectivas decisiones arbitrales. En concreto, el *clustering* aglomerativo genera una estructura de agrupación en forma de dendrograma que permite visualizar si las observaciones correspondientes a uno u otro club tienden a posicionarse continuamente en los clústeres que reflejen un estilo de juego más agresivo o un arbitraje más estricto o, por el contrario, un estilo de juego menos físico o un arbitraje más permisivo. Seguidamente, se desarrolla un modelo de aprendizaje supervisado, particularmente una regresión logística, con el propósito de ver si las variables disciplinarias pueden llegar a predecir la identidad o trato arbitral de un equipo. Ello se complementa a su vez con un árbol de clasificación, el cual permite representar de forma jerárquica el peso y la relevancia de cada variable en base a la distinción entre equipos. Por último, se lleva a cabo un análisis comparativo en torno a los tres equipos más representativos de LaLiga —Fútbol Club Barcelona, Real Madrid y Atlético de Madrid— para investigar y profundizar en las posibles similitudes y diferencias en el trato recibido por parte del arbitraje.

Para realizar el presente trabajo, se emplean diversas herramientas especializadas en el procesamiento, análisis y visualización de datos. Si bien es cierto que la base de datos de la primera fase del estudio se crea inicialmente en Excel, R se utiliza como programa principal para realizar todos los análisis pertinentes, así como para generar visualizaciones que faciliten la interpretación de los resultados obtenidos. Además, el *web scraping* se lleva a cabo con Python debido a la concisión del lenguaje y a la eficiencia de su librería Pandas, especializada en la extracción y manipulación de datos tabulares.

## CAPÍTULO 2: MARCO TEÓRICO

---

### 2.1 FACTORES PSICOLÓGICOS EN LA TOMA DE DECISIONES ARBITRALES

“El fútbol no sería lo que es sin la inevitable falibilidad arbitral. Lo que distingue al fútbol de los demás deportes es la casi infinita posibilidad que nos da de debatir si un resultado fue justo o no” (Carlin, 2012). Esta afirmación revela uno de los elementos más distintivos del fútbol: la presencia de decisiones arbitrales sujetas a interpretación, las cuales, pese a la introducción de tecnologías de asistencia como el VAR (*Video Assistant Referee*), continúan suscitando un espacio de incertidumbre y polémica, lo que plantea cuestiones sustanciales en torno a la imparcialidad en la diligencia de las reglas del juego.

Resulta importante destacar que un juez de fútbol toma, aproximadamente, diez mil decisiones por partido (Porte, 2020). Esta cifra, impactante en sí misma, refleja no sólo la complejidad inherente al arbitraje, sino también la magnitud de la responsabilidad que recae sobre los colegiados. Cada una de sus decisiones incide, en mayor o menor medida, tanto en el desarrollo del juego como en la percepción de justicia por parte de jugadores, entrenadores, aficionados y medios de comunicación.

En el mundo del deporte, el arbitraje no sólo demanda una sólida preparación técnica, sino también una fortaleza psicológica significativa para combatir los altos niveles de presión supeditados a la toma de decisiones en tiempo real. Tal como señalan Weinberg y Richardson, “el arbitraje de élite supone una actividad donde las variables psicológicas desempeñan un papel fundamental en el desempeño del árbitro y en la precisión de sus decisiones” (1990). Diversos estudios señalan la naturaleza vulnerable del juicio arbitral ante la presión social del público, jugadores y hasta entrenadores; esta influencia tiende a condicionar las decisiones de los colegiados, pues tienen un impacto directo sobre su estado emocional y, en consecuencia, sobre su rendimiento y precisión (Debien et al., 2014; Goumas, 2014; Pettersson-Lidbom y Priks, 2010).

Uno de los elementos psicológicos más determinantes —y, por ello, más estudiados— en el ámbito del arbitraje es el estrés (Nogueira et al., 2022). Éste se manifiesta como una respuesta congénita al desafío continuo de tomar decisiones necesariamente apresuradas y acertadas bajo el escrutinio de miles de espectadores, en contextos en los que cualquier error puede influir directamente en el resultado del partido. Un estudio de Taylor y Daniel (1988) recoge las fuentes de estrés más citadas experimentadas por árbitros y oficiales deportivos, de entre las cuales destacan el miedo a fracasar o cometer errores, exponerse a abusos verbales por parte de jugadores, entrenadores o espectadores y recibir amenazas a su integridad física. A su vez, son susceptibles a la presión derivada de la gestión del tiempo y del ritmo de juego en cada encuentro. Se entiende que, ante decisiones de semejante magnitud, es tan lógico como natural que afloren miedos, inseguridades, bloqueos, sentimientos de responsabilidad y emociones que puedan sobrepasar a los árbitros. El temor a tomar una decisión equivocada refleja una serie de mecanismos psicológicos complejos que dificultan la toma de decisiones trascendentales, las cuales han de resolverse con rapidez y con escaso margen de reflexión previa para minimizar la angustia que genera afrontarlas.

En esta línea, Weinberg y Richardson (1990) identifican tres fuentes potenciales de miedo para los árbitros. En primer lugar, el miedo al fracaso: las equivocaciones arbitrales pueden derivar en errores irreparables, cuyas consecuencias podrían ser graves e imposibles de enmendar —de ahí que éste sea el factor que predice con mayor fiabilidad el agotamiento de los jueces deportivos (Taylor et al., 1990)—. En segundo lugar, el miedo a la incapacidad: el estrés puede erosionar la autoconfianza, lo que da lugar a una autocrítica excesiva y a la merma en la percepción de competencia personal. En tercer lugar, el miedo a la impresión de pérdida de control: la sensación de descontrol puede desencadenar en un fenómeno conocido como “compensación arbitral” (*make-up call*), en el cual el colegiado, tras haber cometido un error, se siente impulsado a favorecer al equipo perjudicado en decisiones posteriores. Este intento de corregir retroactivamente el fallo inicial produce un efecto en cadena que corrompe la imparcialidad del arbitraje (Schrödter et al., 2022), pues, en lugar de aplicar las reglas de manera objetiva, el árbitro actúa forzosamente condicionado por la necesidad de rectificar, lo que puede derivar en decisiones que, lejos de subsanar el error anterior, introduzcan un sesgo adicional y alteren el desarrollo natural del partido.

Cabe destacar que las situaciones de estrés que enfrentan los árbitros no se limitan al propio partido, sino que abarcan tanto los momentos previos como los posteriores al mismo. Durante las fases precedentes a un encuentro —desde el momento de designación al partido en cuestión hasta el vestuario, calentamiento y salida al campo—, los colegiados ya sufren una intensa carga emocional (Vela y Arbinaga, 2018). Ésta no sólo se extiende al transcurso del partido, sino que persiste después, una vez atienden a la crítica mediática y social en torno a las repercusiones del encuentro. Por añadidura, los árbitros deben someterse a evaluaciones técnicas de su desempeño y a las políticas establecidas por las respectivas federaciones u órganos encargados de ello (Fuentes, 2021).

Por otra parte, la implementación de tecnologías de asistencia, como es el caso del VAR, ha introducido una dualidad en la presión recaída sobre los árbitros. Por un lado, el uso del asistente de vídeo ha logrado minimizar el margen de error, lo que reduce notablemente la responsabilidad de los colegiados (Karafil, 2023). No obstante, por otro lado, también puede incrementar la ansiedad en su toma de decisiones, pues han de equilibrar su criterio personal con las imágenes proporcionadas. Es decir, se ven obligados a evaluar objetivamente si las imágenes confirman o refutan su decisión inicial, ello bajo un escrutinio público magnificado por las repeticiones en pantalla y los comentarios de jugadores, entrenadores, aficionados y periodistas deportivos. Esta exposición pública convierte cada decisión en un momento de alta tensión, ya que cualquier fallo será visto, analizado y debatido al instante por millones de personas (Vicandi, 2022). Asimismo, a nivel emocional, la revisión mediante el VAR puede percibirse como un cuestionamiento implícito de la competencia de los árbitros para juzgar de manera rápida y correcta, lo que puede mermar su autoestima y confianza, aumentando así el riesgo de que tomen decisiones más conservadoras para evitar que sean objeto de revisión.

Todos estos niveles de presión pueden derivar no sólo en fallos involuntarios en las resoluciones, sino también en predisposiciones que afecten la objetividad de los árbitros. En encuentros de alta tensión, como pueden ser los enfrentamientos entre equipos rivales de gran magnitud, esta predisposición puede materializarse en una inclinación instintiva hacia uno u otro equipo —aun sin una intención consciente de hacerlo— como mecanismo psicológico de defensa para los jueces, quienes, bajo las múltiples presiones que experimentan mencionadas con anterioridad, podrían tomar decisiones que reduzcan posibles represalias o críticas. Del mismo modo, el miedo a incurrir en un fallo que tenga una gran repercusión en el marcador puede llevar a los árbitros a adoptar decisiones indebidamente cautelosas o permisivas, sobre

todo en sentencias decisivas como pueden ser los penaltis o las tarjetas rojas. En este tipo de circunstancias, el juez podría no penalizar determinadas infracciones o, por el contrario, ser más riguroso de lo habitual, con el objetivo de disminuir el riesgo de errores en su actuación (Wallace et al., 2005). Si bien es cierto que estas predisposiciones no nacen de una intención reflexiva, sino como respuesta al entorno, sus efectos, además de influir en la dinámica del juego, pueden tener un impacto en la opinión pública respecto a la justicia y transparencia del fútbol profesional.

## 2.2 EVIDENCIA EMPÍRICA DE SESGOS ARBITRALES

Múltiples estudios han evidenciado que el arbitraje, pese a su vocación de neutralidad, no alcanza un nivel de plena imparcialidad; factores como la ventaja local, la presión ejercida por el público o la reputación histórica de un equipo pueden condicionar —consciente o inconscientemente— la toma de decisiones por parte de los árbitros.

Un ejemplo destacado es el estudio de Goumas (2014), quien documenta múltiples dimensiones del sesgo arbitral en beneficio de los equipos locales en competiciones de la UEFA. En términos de ventaja local, medida por puntos obtenidos durante la fase de grupos, los equipos locales lograron el 57,8 % de los puntos en la Liga de Campeones y el 59,2 % en la Europa League. Cabe señalar que estos registros se sitúan por debajo de los ya hallados en varias ligas domésticas europeas, como la Premier League inglesa —61% (Seckin y Pollard, 2008)—, LaLiga española —62,1 % (Sánchez et al., 2009)— o la Süper Lig turca —61,5 % (Seckin y Pollard, 2008)—. Además, el análisis reveló que los equipos visitantes reciben un 25% más de tarjetas amarillas que los anfitriones en la *Champions League*, aun tras controlar variables como la mera superioridad ofensiva durante el partido. Estas diferencias sugieren la existencia de una tendencia sistemática que favorece a los equipos que juegan en casa, lo que corrobora investigaciones previas que ya detectaron este fenómeno, conocido como “ventaja local”. En particular, se ha evidenciado cierto favoritismo hacia el conjunto anfitrión en la concesión de penaltis (Dohmen, 2008; Nevill et al., 1996), la asignación de tiempo añadido (Dohmen, 2008; Garicano et al., 2005; Sutter y Kocher, 2004) y la imposición de sanciones disciplinarias (Dawson et al., 2007; Reilly y Witt, 2013).

Esta predisposición parece estar particularmente influenciada por el ímpetu de la afición local (Boyko et al., 2007; Garicano et al., 2005); diversos factores como la intensidad del ruido, el tamaño de la multitud y su proximidad al campo aumentan la probabilidad de obtener un arbitraje favorable al conjunto local (Poolton et al., 2011). En estadios con niveles de presión sumamente altos, los colegiados pueden verse afectados —si bien de manera no deliberada— por las reacciones del entorno, sobre todo ante jugadas ambiguas que exijan una valoración subjetiva. Según la perspectiva de Haselton y Buss (2000) conocida como la *error management theory*, en circunstancias de incertidumbre, los árbitros tienden a minimizar el coste percibido del error en el que podrían estar incurriendo, lo que da lugar a una inclinación sistemática hacia decisiones que resulten menos disruptivas para el entorno inmediato —esto es, beneficiar al equipo local cuando el juicio no es inequívoco (Haselton y Nettle, 2006)—.

En este contexto, Pollard (2008) llevó a cabo un trabajo de investigación exhaustivo con el propósito de indagar en los factores que contribuyen a la cuestión de la ventaja local. Según el autor, este fenómeno, ampliamente reconocido en el ámbito deportivo, repercute no sólo en el rendimiento de los equipos, sino también en las decisiones arbitrales. Pese a que existen evidencias empíricas que apuntan a un sesgo arbitral en favor del equipo local, sigue siendo objeto de debate si dicho sesgo actúa favoreciendo directamente al local, perjudicando al visitante o mediante una combinación de ambas dinámicas. Asimismo, cabe mencionar el papel de la familiaridad del equipo anfitrión con su entorno —incluyendo aspectos como el estadio, el clima o las condiciones del terreno de juego— como un componente adicional de la ventaja local. Pollard insiste en la interacción entre estos factores, indicando que no operan de forma individual, sino que tienden a armonizarse e interactuar de maneras que resultan complicadas de investigar y cuantificar.

Adicionalmente, diversos estudios sugieren la existencia de un sesgo arbitral influido por la reputación de los equipos, ya sea de forma positiva o negativa. Por un lado, los equipos con mayor prestigio institucional tienden a beneficiarse de un tratamiento más favorable por parte de los árbitros; esto se debería, en parte, a la presión social y mediática que recae sobre los encuentros en los que participan clubes de alto perfil, pues éstos generan expectativas más elevadas de éxito. Por otro lado, los equipos conocidos por su juego físico o agresivo suelen ser más susceptibles a decisiones arbitrales en su contra, debido a la percepción previa que los árbitros tienen de los mismos.

En un estudio realizado a las cinco grandes ligas europeas (la Premier League inglesa, LaLiga española, la Serie A italiana, la Bundesliga alemana y la Ligue 1 francesa) desde la temporada 2011/12 hasta la 2017/18 con el fin de investigar el presunto trato diferencial hacia los clubes con mayor número de títulos o de ingresos, se hallaron evidencias de la existencia de dicho favoritismo en cuatro de las cinco ligas elegidas (Audrino, 2020). En estas competiciones, los jugadores de estos equipos fueron sancionados con menor frecuencia; concretamente, en LaLiga, la Serie A y la Bundesliga, recibieron entre 0,2 y 0,4 tarjetas amarillas menos por partido. Este análisis puede suscitar inseguridades acerca de la equidad en el fútbol profesional, pues clubes que ya de por sí parten con ventajas estructurales en cuanto a inversión económica y calidad de plantilla, podrían beneficiarse adicionalmente de resoluciones arbitrales favorables.

En esta línea, una investigación realizada por Erikstad y Johansen (2020) se centra en el análisis relativo a la concesión de penaltis para evidenciar la idea de que los equipos más exitosos ejercen una influencia social significativa sobre el arbitraje. Los autores constataron que los clubes con mayor reconocimiento recibieron más penaltis de los que un panel experto —a modo de VAR— hubiera otorgado. En concreto, tras tomar una temporada de la Primera División de Noruega en calidad de muestra, los equipos históricamente dominantes resultaron beneficiados con 11 penaltis, pese a que el panel de expertos únicamente consideró 10 de ellos como acertados. Así pues, se les concedió un 110% de los penaltis frente al 73,3% de media general; por el contrario, sus respectivos oponentes apenas recibieron el 12,5% de los penaltis que el mismo panel consideró que debían haber sido señalados. En el artículo, estos resultados se explican mediante la teoría del impacto social de Latané (1981), la cual sostiene que el impacto de un grupo sobre un individuo depende de tres factores: fuerza, inmediatez y cantidad; la fuerza indicaría la relevancia trascendental y éxito de los jugadores, la inmediatez aludiría a la proximidad física entre los jugadores y el árbitro y la cantidad haría referencia al número de personas que influyen en el entorno. En este contexto, los dos primeros elementos tienen una magnitud mucho mayor en los jugadores que en el público, por lo que la influencia de los miembros de un equipo exitoso sobre el arbitraje puede ser incluso superior a la de la afición local.

Por otra parte, el efecto halo también presenta una dinámica inversa; un efecto cuerno puede recaer sobre clubes con reputaciones relativamente negativas, particularmente aquellos

cuyo estilo de juego resulta más físico o agresivo. Este prejuicio cognitivo puede influir en la percepción arbitral, de manera que, por ejemplo, una misma falta podría ser interpretada con distinta valoración si la comete un jugador que ha transgredido el reglamento anteriormente a que si la comete otro jugador que no ha incurrido en infracciones previas (Caracuel et al., 1995). De la misma forma, aquellos equipos más combativos pueden ser objeto de decisiones más severas por parte de los árbitros debido al estatus negativo asociado a su estilo de juego. Un estudio de Jones et al. (2002) analizó cómo el conocimiento sobre la reputación agresiva de un equipo influía en los colegiados. Para ello, 38 árbitros fueron asignados aleatoriamente a dos grupos, uno experimental y otro de control, y se les presentaron 50 incidentes ocurridos en el campo. Los resultados mostraron que, aunque el número total de decisiones fue similar en ambos grupos, los árbitros del grupo experimental —que habían sido informados sobre el estatus agresivo del club— sancionaron con un mayor número de tarjetas rojas y amarillas a los indisciplinados de dicho club.

Este hallazgo puede explicarse por el hecho de que el conocimiento previo acerca de la reputación de un equipo actúa como una heurística de representatividad, lo que afecta al juicio y a las decisiones de los colegiados. Gerd Nufer revela que, dada la gran cantidad de información disponible y el tiempo limitado para la toma de decisiones, resulta imposible que la mente humana procese y analice por completo todos los estímulos que recibe (2019). En este sentido, los árbitros, sometidos a condiciones de alta presión y máxima celeridad, suelen emplear atajos mentales para simplificar sus valoraciones; no obstante, a pesar de que este mecanismo heurístico permite actuar rápidamente, también aumenta la susceptibilidad a sesgos cognitivos (Nufer y Alesi, 2018).

### **2.3 FUNDAMENTOS PARA EL ANÁLISIS**

Una conclusión particularmente relevante al comparar los estudios mencionados con anterioridad es que el sesgo arbitral favorable hacia los equipos más laureados y el sesgo punitivo hacia aquellos con una identidad más agresiva, lejos de ser mutuamente excluyentes, tienden a reforzarse entre sí, generando un efecto compuesto que intensifica su impacto sobre las decisiones arbitrales. Así pues, en encuentros en los que un club con elevado prestigio se

enfrenta a otro caracterizado por un estilo de juego más físico, el arbitraje puede verse doblemente condicionado, lo que incrementa de manera significativa la probabilidad de fallos favorables al equipo históricamente exitoso. Ello puede afectar no sólo a los resultados deportivos, sino a la percepción de justicia e imparcialidad en las competiciones.

De esta manera, el Real Madrid y el Atlético de Madrid constituyen un escenario de estudio idóneo para investigar la interacción entre estos sesgos. Por un lado, el Real Madrid, como club deportivo con un palmarés ampliamente reconocido a nivel internacional y como líder del ranking europeo por valor de mercado (Football Benchmark, 2025), puede encarnar el perfil característico de equipo beneficiado por su reputación y peso institucional. Por otro lado, el Atlético de Madrid ha sido históricamente asociado a un estilo de juego más físico con un enfoque defensivo, frecuentemente descrito en la literatura deportiva como agresivo o de contacto (Uribarri, 2016). Así, esta dualidad proporciona un marco empírico conveniente para evaluar posibles patrones de trato diferencial por parte de los árbitros, así como decretar si la coexistencia de ambos presuntos sesgos puede llegar a suscitar efectos acumulativos en el arbitraje, especialmente en un contexto de alta rivalidad y visibilidad mediática como el de los derbis madrileños (Reoyo, 2025). Cabe señalar, además, que a pesar de la gran cantidad de investigaciones dedicadas a la ventaja local o al favoritismo hacia los equipos de élite, la correlación entre el estatus de éxito y la percepción negativa de cierto estilo de juego ha sido escasamente abordada en la literatura académica. Asimismo, no se han identificado hasta la fecha estudios que analicen específicamente la influencia del arbitraje en los partidos entre el Real Madrid y el Atlético de Madrid, pese a la relevancia deportiva, mediática y simbólica que rodea estos enfrentamientos.

## **CAPÍTULO 3: MATERIAL Y MÉTODOS**

---

### **3.1 PLANTEAMIENTO DE HIPÓTESIS**

La pregunta central que orienta esta investigación es si las decisiones arbitrales en los derbis entre el Real Madrid y el Atlético de Madrid han favorecido de manera sistemática a uno de los dos clubes. Esta cuestión se complementa con preguntas secundarias, tales como si determinados tipos de sanciones se aplican con distinta frecuencia a cada equipo o, si en caso de detectarse un posible patrón de trato desigual, éste persiste o se diluye al ampliar la muestra al resto de equipos participantes en LaLiga.

Este trabajo se fundamenta en la siguiente hipótesis: existen diferencias significativas en las decisiones arbitrales en los partidos oficiales entre el Real Madrid y el Atlético de Madrid. En este sentido, se especula que uno de estos dos equipos ha recibido un mayor número de penalizaciones —bien sea en forma de faltas, tarjetas, penaltis o fueros de juego— en comparación con el otro, en una proporción que no puede ser atribuida únicamente al desarrollo del juego. Asimismo, se especula que la reputación histórica de un club no es un factor neutro, sino que, como se ha observado en investigaciones previas, incide en el grado de severidad de las sanciones impuestas. En esta línea, se conjetura que la condición de equipo históricamente exitoso, como es el caso del Real Madrid, o la asociación a un estilo de juego más físico, como es el caso del Atlético de Madrid, pueden condicionar de manera diferencial el arbitraje.

### **3.2 MUESTRA DE DATOS**

La muestra de datos del estudio se divide en dos partes, en correspondencia con las dos fases diferenciadas del análisis: la primera se centra únicamente en los derbis madrileños, mientras que la segunda incluye a todos los equipos de LaLiga en las últimas tres temporadas.

En la primera fase, la muestra comprende la totalidad de los enfrentamientos oficiales celebrados entre el Real Madrid y el Atlético de Madrid durante la última década, abarcando desde la temporada 2013/14 hasta la 2023/24, con un total de 39 encuentros. Se incluyen las competiciones nacionales de LaLiga, la Copa del Rey y la Supercopa de España, así como la Liga de Campeones de la UEFA y la Supercopa de Europa a nivel internacional. La elección de estos partidos se justifica por su alto nivel competitivo —además de la histórica rivalidad entre ambos equipos—, haciendo de éstos un campo de estudio idóneo para explorar posibles sesgos arbitrales. Se determinan las siguientes variables clave: número de faltas sancionadas, tarjetas amarillas y rojas mostradas, penaltis concedidos y fuera de juego señalados. Estos datos se recopilan desde fuentes fiables como Whoscored y FBref, así como reportes oficiales de las competiciones en cuestión; se asegura su precisión con la validación cruzada entre las diversas plataformas.

En la segunda fase, la muestra se amplía a todos los equipos participantes en LaLiga durante las temporadas 2021/22, 2022/23 y 2023/24. Esta base de datos recoge las mismas variables que la primera fase, pero agregadas como promedios anuales por equipo, de modo que la estructura de datos resultante presenta tres observaciones por equipo, una por cada año analizado. Esta ampliación permite aplicar técnicas de análisis exploratorio para identificar patrones en las decisiones arbitrales, así como agrupar los equipos según las similitudes en el comportamiento de las variables analizadas.

## CAPÍTULO 4: ANÁLISIS Y RESULTADOS

---

### 4.1 ANÁLISIS ESTADÍSTICO ENTRE EL REAL MADRID Y EL ATLÉTICO DE MADRID

Primeramente, se lleva a cabo un estudio estadístico de los partidos disputados entre el Real Madrid y el Atlético de Madrid a lo largo de la última década. La muestra comprende un total de 39 partidos; ésta fue elaborada inicialmente en Excel y posteriormente importada a RStudio para su tratamiento y análisis estadístico.

Para examinar adecuadamente la muestra, se realiza un análisis de muestras pareadas, entendiendo como tales aquellas en las que se comparan observaciones dependientes —esto es, dos registros correspondientes a un mismo partido, uno para cada equipo—. Se consideran cinco variables arbitrales clave: número de penaltis, tarjetas amarillas, tarjetas rojas, faltas cometidas y fueros de juego. La estructura de la base de datos recoge todas estas variables de manera emparejada para cada enfrentamiento, asignando al Real Madrid la denominación de *RMA* y al Atlético de Madrid *ATM*, manteniendo ese orden de forma sistemática en todo registro. La construcción de la base se lleva a cabo manualmente, introduciendo cada dato de manera rigurosa tras su correcta verificación; a continuación, ésta es procesada en el entorno de programación estadística R para poder realizar los contrastes y respectivas visualizaciones.

Antes de comenzar el análisis, se procede a eliminar todas las variables y funciones cargadas en la memoria y a establecer el directorio de trabajo en el cual se encuentra el *script* actual. Para poder ejecutar la base de datos, es necesario cargar la librería necesaria para leer archivos Excel (*readxl*). Una vez cargados los datos, se explora la base mediante una serie de comprobaciones iniciales con el objetivo de verificar que los datos han sido importados de manera correcta y que éstos son compatibles con los procedimientos estadísticos a aplicar. A continuación, se realizan los contrastes de diferencia de medias y medianas de cada variable.

Primeramente, debe revisarse que los datos sigan una distribución normal mediante el test de Shapiro-Wilk (*shapiro.test*). El estadístico  $W$  cuantifica la correspondencia entre los valores observados y los esperados bajo una distribución normal. Puede oscilar entre 0 y 1; cuanto más cercano a 1, mayor ajuste a la normalidad, mientras que los valores próximos a 0 sugieren desviaciones considerables. Además, en esta prueba, se evalúa el valor  $p$  asociado. Si éste es superior al umbral convencional de significación de 0,05, no se rechaza la hipótesis nula, por lo que los datos se consideran normales. En cambio, si el valor es inferior a 0,05, se concluye que no hay normalidad. Este paso es clave, pues sirve para determinar si se puede aplicar una prueba paramétrica, como la prueba  $t$  de Student —en el caso de una distribución normal—, o si es mejor emplear una alternativa no paramétrica, como la prueba de Wilcoxon —si los datos no son normales—. Cabe recordar que el número de observaciones es reducido, por lo que, en aquellas situaciones en las que no se alcanza normalidad, se recurre a pruebas no paramétricas, puesto que no asumen una distribución específica, además de resultar más adecuadas para muestras pequeñas.

A pesar de no ser estrictamente necesario, resulta conveniente realizar la prueba  $F$  (*var.test*) para evaluar la homogeneidad de varianzas. Si bien es cierto que tanto la prueba  $t$  para muestras pareadas (*t.test*) como el test de Wilcoxon (*wilcox.test*) no requieren el supuesto de igualdad de varianzas, la prueba  $F$  contribuye a interpretar la dispersión de los datos, así como garantizar la validez de los resultados estadísticos, particularmente en muestras pequeñas en las que una desviación puede tener un gran impacto. En cuanto a la interpretación, si el valor  $p$  es mayor que 0,05, no se rechaza la hipótesis nula de que las varianzas son iguales; contrariamente, si éste es menor que 0,05, se entiende que las varianzas son diferentes.

Al ejecutar múltiples comparaciones simultáneamente, aumenta el riesgo de obtener resultados que, aun siendo estadísticamente significativos, se deben al azar; éstos se conocen como falsos positivos o errores de tipo I. Ello supone rechazar incorrectamente una o varias hipótesis nulas, lo que podría resultar en conclusiones erróneas de los datos. Para abordar este problema de la multiplicidad en las pruebas de hipótesis y lograr mantener la fiabilidad de la investigación, se lleva a cabo una serie de correcciones para ajustar los valores  $p$  obtenidos en cada prueba individual. En primer lugar, el método Bonferroni ajusta el nivel de significancia estadística según la cantidad de pruebas realizadas, multiplicando cada  $p$ -valor original por el número total de comparaciones. Si bien es eficiente, esta corrección puede resultar demasiado conservadora, dando lugar a errores de tipo II, mediante los cuales se acepta incorrectamente

una o varias hipótesis nulas. El método Holm, por su parte, ordena los valores  $p$  de manera ascendente y los ajusta secuencialmente. Es considerado menos conservador que Bonferroni, pues ajusta los  $p$ -valores de forma progresiva y adaptativa, en lugar de implementar la misma corrección rígida a todos. Finalmente, el método de Benjamini-Hochberg (BH) se emplea para controlar la tasa de falsos descubrimientos (FDR), entendida como la proporción de falsos positivos. En todos estos métodos, se compara el  $p$ -valor ajustado en cada caso con el umbral clásico de 0,05, de tal manera que, si el valor  $p$  ajustado es menor que el umbral, se concluye que el resultado continúa siendo estadísticamente significativo.

Una vez aplicadas las correcciones y comprobadas las significancias estadísticas, se evalúa la magnitud de las diferencias halladas. De esta manera, se puede estipular el tamaño del impacto, así como identificar la variable que mayor efecto presenta. En función del tipo de prueba, se emplean dos indicadores diferentes. Para aquellas variables que cumplen con los supuestos de normalidad y llevan a cabo pruebas paramétricas, se opta por la  $d$  de Cohen (*cohen.d*). Funciona de la siguiente manera: los valores de la  $d$  cercanos a 0,2 representan un efecto pequeño, los similares a 0,5 muestran un efecto medio y los mayores de 0,8 señalan un efecto grande. Por otro lado, para aquellas variables que no siguen una distribución normal y requieren pruebas no paramétricas, se utiliza el estadístico  $A$  de Vargha y Delaney (*VD.A*), donde los valores de la  $A$  inferiores a 0,44 o superiores a 0,56 sugieren un efecto pequeño, los inferiores a 0,36 o superiores a 0,64 implican un efecto medio y los inferiores a 0,29 o superiores a 0,71 suponen un efecto grande.

## **4.2 RESULTADOS DEL ANÁLISIS ENTRE EL REAL MADRID Y EL ATLÉTICO DE MADRID**

Así pues, se procede a analizar cada variable según la estructura previamente definida.

En cuanto al número de penaltis, se evalúa, en primer lugar, si esta variable sigue una distribución normal para ambos equipos. La prueba de Shapiro-Wilk devuelve para el Real Madrid un estadístico  $W = 0,449$  con un valor  $p < 0,001$ , así como un  $W = 0,395$  y  $p < 0,001$

para el Atlético de Madrid. Estos resultados indican una desviación respecto a la normalidad, lo que justifica el rechazo de la hipótesis nula y, por ende, se desaconseja el uso de pruebas paramétricas. Adicionalmente, se contrasta la homogeneidad de varianzas en ambos grupos, obteniéndose un estadístico  $F = 1,777$  y un valor  $p = 0,081$ . De esta manera, las varianzas se consideran homogéneas. Finalmente, se aplica la prueba no paramétrica de Wilcoxon, cuyos resultados ( $V = 38,5$  y  $p = 0,627$ ) manifiestan que no existen diferencias estadísticamente significativas en el número de penaltis señalados a ambos equipos en los partidos analizados.

Respecto a la variable de tarjetas amarillas, el test de Shapiro-Wilk arroja para el Real Madrid unos valores de  $W = 0,910$  y  $p = 0,004$ , rechazando la hipótesis nula de normalidad; sin embargo, para el Atlético de Madrid, son de  $W = 0,955$  y  $p = 0,123$ . En cuanto a la homogeneidad de varianzas, la prueba F devuelve un estadístico  $F = 0,738$  y un  $p = 0,352$ , lo que significa que las varianzas para ambos clubes se consideran homogéneas. Dada la falta de normalidad en al menos uno de los dos grupos, se recomienda usar una prueba no paramétrica como Wilcoxon. El resultado es de  $V = 130$  y  $p = 0,004$ , lo que revela que existe una diferencia significativa en el número de tarjetas amarillas. Ello sugiere que, en los partidos elegidos, el Atlético de Madrid recibe considerablemente más tarjetas amarillas que el Real Madrid.

En términos de tarjetas rojas, los resultados de la prueba de normalidad resultan ser estadísticamente significativos para ambas muestras, con valores  $p < 0,001$  tanto para el Real Madrid ( $W = 0,233$ ) como para el Atlético de Madrid ( $W = 0,395$ ). Respecto a la igualdad de varianzas, se obtiene un estadístico  $F = 0,435$  y un  $p$ -valor =  $0,012$ , luego se concluye que las varianzas son considerablemente diferentes. Por último, se realiza el test de Wilcoxon, cuyos resultados ( $V = 8$  y  $p = 0,299$ ) indican que no se observan grandes diferencias significativas en esta variable.

En cuanto al número de faltas, al aplicar la prueba de Shapiro-Wilk, tanto para el Real Madrid ( $W = 0,955$  y  $p = 0,122$ ) como para el Atlético de Madrid ( $W = 0,984$  y  $p = 0,830$ ), se obtiene un  $p$ -valor superior al umbral de  $0,05$ , lo cual sugiere que ambas variables siguen una distribución normal. Posteriormente, se verifica la homogeneidad de varianzas, obteniendo un valor de  $F = 1,295$  y  $p = 0,429$ , luego ambas varianzas se pueden considerar iguales. Debido a la normalidad en los datos de los dos grupos, se efectúa la prueba t para muestras pareadas, cuyos valores  $t = -3,957$  y  $p < 0,001$  permiten rechazar la hipótesis nula. En consecuencia, se asume la existencia de una diferencia significativa en el número de faltas cometidas por el Real

Madrid y el Atlético de Madrid, con una diferencia media de -3,77 —esto es, el Atlético de Madrid fue pitado, en promedio, 3,77 faltas más que el Real Madrid—.

Respecto a la variable de fuera de juego, ambos equipos muestran un  $p$ -valor inferior a 0,05 para el test de Shapiro-Wilk (Real Madrid:  $W = 0,690$  y  $p < 0,001$ ; Atlético de Madrid:  $W = 0,872$  y  $p < 0,001$ ), por lo que no se cumple la normalidad en ninguno de los dos grupos. En cuanto a la homogeneidad de varianzas, se obtiene un estadístico  $F = 0,522$  y un  $p = 0,048$ , luego no se asume igualdad de varianzas para ambos equipos. Finalmente, llevando a cabo la prueba de Wilcoxon, los resultados ( $V = 64$  y  $p < 0,001$ ) reflejan que existen diferencias significativas en los fuera de juego cometidos por ambos equipos, con el Atlético de Madrid incurriendo en un número considerablemente mayor de posiciones adelantadas.

Posteriormente, se recogen los  $p$ -valores correspondientes a las pruebas de Wilcoxon y  $t$ -test para todas las variables, siendo éstos 0,6270171300 para los penaltis, 0,0038469306 para las tarjetas amarillas, 0,2986176744 para las tarjetas rojas, 0,0003210775 para las faltas y 0,0007337915 para los fuera de juego. Al realizar múltiples comparaciones, se incrementa el riesgo de cometer un error tipo I; para controlar la tasa de error, se aplican tres métodos de corrección: Bonferroni, Holm y Benjamini-Hochberg.

En primer lugar, el método Bonferroni ajusta los valores  $p$  al dividir el nivel de significancia (en este caso, 0,05) entre el número total de comparaciones realizadas. De esta manera, los  $p$ -valores corregidos son los siguientes: 1 para los penaltis, 0,019234653 para las tarjetas amarillas, 1 para las tarjetas rojas, 0,001605388 para las faltas y 0,003668957 para los fuera de juego. En este análisis, los valores ajustados iguales a 1 indican que no existe evidencia suficiente para poder rechazar la hipótesis nula. Por tanto, las variables de penaltis y tarjetas rojas no alcanzan un nivel de significancia tras el ajuste, mientras que el número de tarjetas amarillas, faltas y fuera de juego continúan considerándose significativas.

En segundo lugar, el método Holm modifica los  $p$ -valores de manera secuencial. Los valores ajustados son de 0,627017130 para los penaltis, 0,011540792 para las tarjetas amarillas, 0,597235349 para las tarjetas rojas, 0,001605388 para las faltas y 0,002935166 para los fuera de juego. Este procedimiento considera significativos aquellos resultados con un  $p$ -valor inferior a 0,05, por lo que coincide con el método de Bonferroni en señalar las tarjetas amarillas, faltas y fuera de juego como variables destacadas.

En tercer y último lugar, el método Benjamini-Hochberg controla la tasa de falsos descubrimientos, proporcionando  $p$ -valores de 0,627017130 para los penaltis, 0,006411551 para las tarjetas amarillas, 0,373272093 para las tarjetas rojas, 0,001605388 para las faltas y 0,001834479 para los fuera de juego. De igual modo, toma en consideración los valores  $p$  menores al umbral de significación, luego mantiene las tarjetas amarillas, faltas y fuera de juego como variables con diferencias estadísticamente significativas.

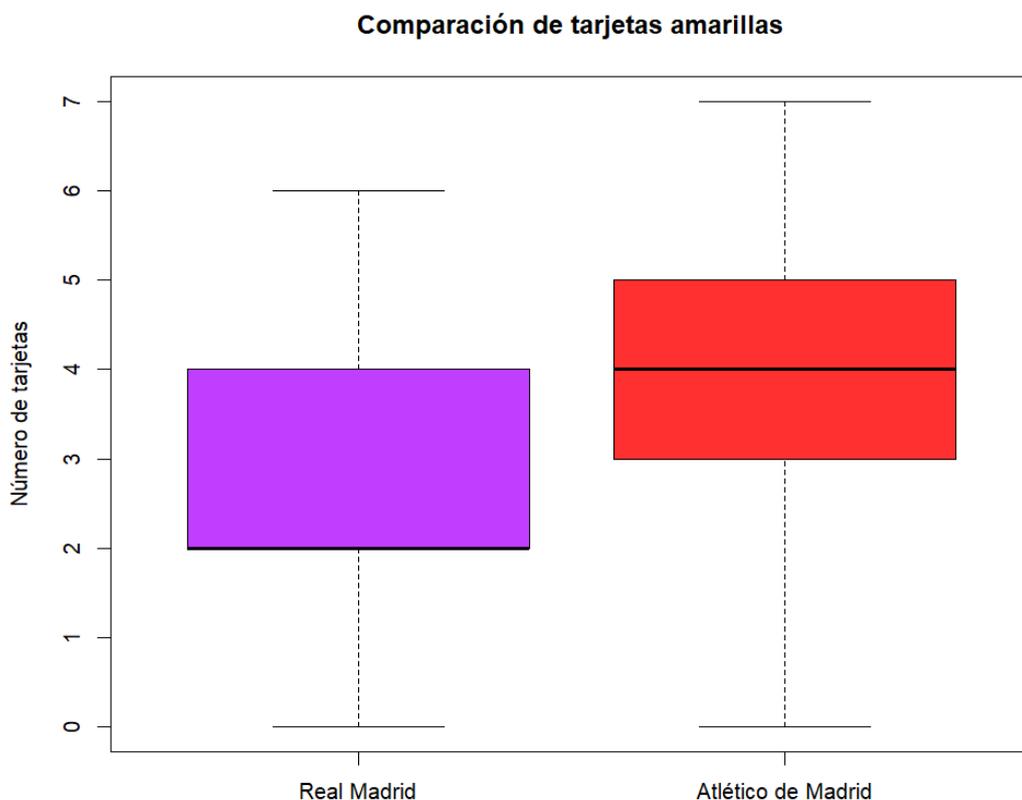
Con el fin de complementar los análisis de significancia estadística, se evalúa a su vez el tamaño del efecto de las diferencias halladas entre el Real Madrid y el Atlético de Madrid en aquellas variables que resultaron significativas tras los ajustes de los valores  $p$ . De este modo, se valora cuán importante o relevante son estas diferencias, determinando si tienen un impacto tangible en el contexto estudiado. Para ello, se utilizan dos métricas diferentes en base al tipo de prueba realizada previamente: la  $d$  de Cohen para el t-test y el estadístico  $A$  de Vargha y Delaney para las pruebas de Wilcoxon.

En el caso de las tarjetas amarillas, se aplica la medida de Vargha y Delaney, que estima la probabilidad de que un valor tomado aleatoriamente del Real Madrid sea mayor que uno del Atlético de Madrid. El resultado del estadístico es de  $A = 0,311$ , lo que representa una diferencia de tamaño medio según los umbrales convencionales ( $A < 0,36$  o  $A > 0,64$ ). En esta línea, un valor  $A < 0,5$  indica que el segundo grupo analizado tiene valores más altos; por lo tanto, existe una tendencia moderadamente marcada que indica que es más probable que el Atlético de Madrid reciba un mayor número de tarjetas amarillas que el Real Madrid en los encuentros de la muestra.

Para la variable correspondiente al número de faltas, se emplea la  $d$  de Cohen. El valor calculado es  $d = -0,868$ , lo que muestra un efecto muy grande, ya que supera el umbral de 0,8. El símbolo negativo del estadístico señala que el primer grupo comete menos faltas que el segundo; dicho de otro modo, el resultado apunta a que el Real Madrid comete, en promedio, un número significativamente menor de faltas que el Atlético de Madrid. Cabe mencionar que el intervalo de confianza del 95% ( $[-1,381; -0,355]$ ) no contiene el valor 0, lo que implica que existe un alto grado de certeza de que esta diferencia no se debe al azar.

Por último, en cuanto a los fueras de juego, el estadístico de Vargha y Delaney arroja un valor de  $A = 0,284$ , el cual, al ser inferior a 0,36 pero no a 0,29 —valor convencional para indicar efecto grande—, muestra un efecto de tamaño mediano. De esta forma, se figura que el Atlético de Madrid incurre con mayor frecuencia en fueras de juego que el Real Madrid.

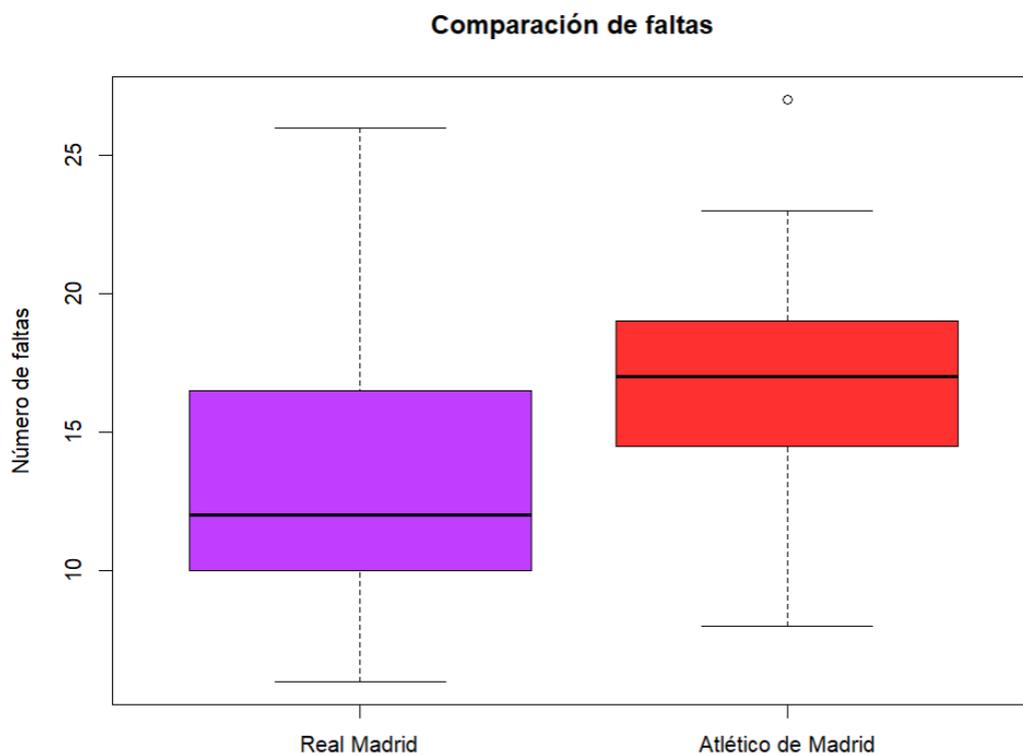
Finalmente, con el objetivo de proporcionar una representación visual de lo hallado, se utilizan diagramas de cajas (*boxplots*) para cada variable con diferencias estadísticamente significativas. Esta herramienta resulta conveniente para comparar la distribución de los datos de ambos equipos, mostrando no sólo las diferencias en las medianas, sino también el grado de dispersión, simetría de los datos y la presencia de valores atípicos. Si bien estas gráficas no constituyen una prueba estadística por sí mismas, sirven como recurso comunicativo para sintetizar e ilustrar los principales hallazgos de una manera accesible a una audiencia general. Se establece una codificación cromática diferenciada —morado para el Real Madrid y rojo para el Atlético de Madrid— para facilitar la claridad visual de las comparaciones.



**Figura 1.** *Boxplot de la comparación de tarjetas amarillas.*

Fuente: Elaboración propia (en R).

Este primer diagrama muestra la distribución del número de tarjetas amarillas recibidas por el Real Madrid y el Atlético de Madrid en los partidos analizados. Primeramente, se aprecia que la mediana del Atlético de Madrid es superior a la del Real Madrid, en una proporción de 4 frente a 2, lo que indica que, en la mayoría de partidos, el Atlético recibe el doble de tarjetas amarillas que su rival. El rango intercuartílico —esto es, la distancia entre el primer y el tercer cuartil— del Real Madrid abarca de aproximadamente 2 a 4 tarjetas, con un mínimo de 0 y un máximo de 6; por otro lado, el del Atlético de Madrid va de 3 a 5 tarjetas, con el mismo mínimo de 0 y recibiendo hasta 7 tarjetas de máximo en los derbis analizados. Además, el valor máximo del Atlético es superior al del Real Madrid, si bien ello no consta una diferencia significativa.

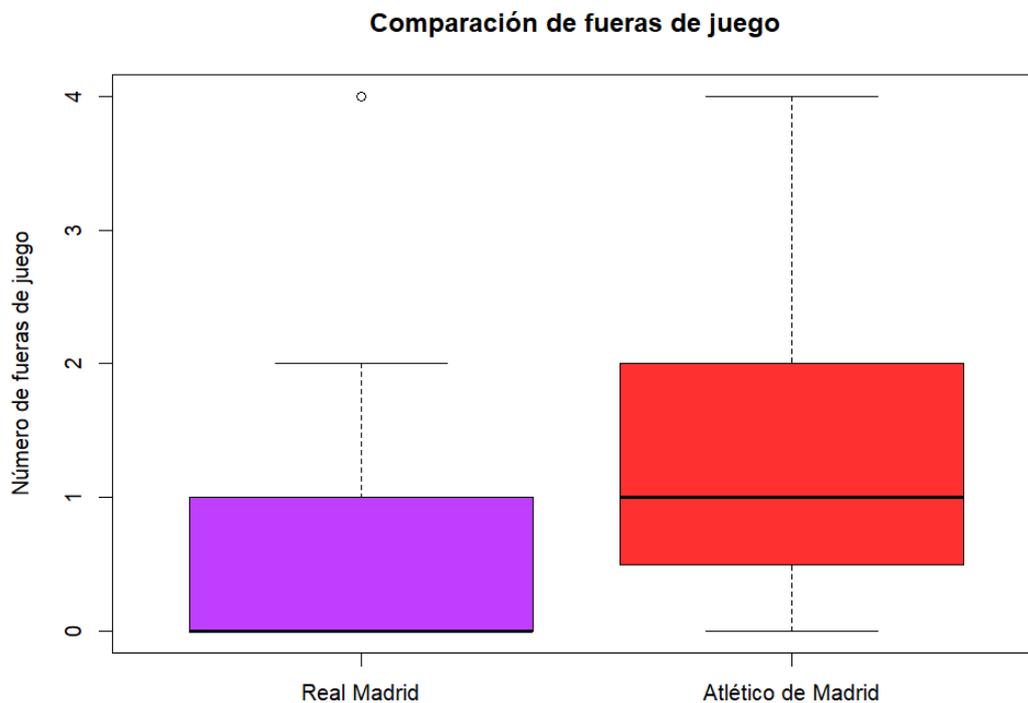


**Figura 2.** *Boxplot de la comparación de faltas.*

Fuente: Elaboración propia (en R).

El segundo *boxplot* expone las diferencias en la cantidad de faltas cometidas por el Real Madrid y el Atlético de Madrid, evidenciando que el Atlético de Madrid presenta una mediana más alta en comparación con el Real Madrid. Del mismo modo, el rango intercuartílico también es notablemente superior en el Atlético, así como el valor mínimo. En cambio, el valor máximo del Real Madrid es más elevado, alcanzando las 25 faltas; no obstante, la gran separación entre el tercer cuartil y el límite superior sugiere que las observaciones más altas de faltas son menos

frecuentes. Además, se aprecia un valor atípico (*outlier*) —esto es, una observación que excede el umbral que define el *boxplot* como valor extremo— en el caso del Atlético, señalando que, en al menos un partido, el número de faltas pitadas en contra del Atlético fue particularmente elevado en comparación con el resto.



**Figura 3.** *Boxplot de la comparación de fueros de juego.*

Fuente: Elaboración propia (en R).

En cuanto al tercer gráfico referente a la comparación en el número de fueros de juego, se observa que el Real Madrid presenta una mediana de 0, en oposición a la mediana de 1 del Atlético de Madrid. Esto significa que, en el caso del Real Madrid, la mitad de sus partidos como mínimo finalizaron sin incurrir en ninguna posición adelantada, mientras que, en el caso del Atlético, en al menos la mitad de sus encuentros cometió una o más infracciones por fuera de juego. Asimismo, el rango intercuartílico del Real Madrid oscila entre 0 y 1 y el del Atlético entre 0,5 y 2. Los bigotes son más largos para el Atlético de Madrid; su valor máximo refleja 4 fueros de juego por partido, el doble que el del Real Madrid. En este caso, también se observa un *outlier* en el Real Madrid, aunque, observando su distribución de los datos, parece tratarse de una excepción puntual.

### 4.3 ALGORITMOS DE *MACHINE LEARNING* PARA TODOS LOS EQUIPOS DE LALIGA

A continuación, se realizan múltiples algoritmos de *Machine Learning* que incluyen a todo equipo que ha participado en la liga española en las últimas tres temporadas. El objetivo de este estudio es contextualizar las diferencias arbitrales observadas en los encuentros entre el Real Madrid y el Atlético de Madrid, con el propósito de determinar si los patrones arbitrales encontrados son específicos de estos clubes o, por el contrario, responden a una tendencia más general que también existe en otros equipos.

Para la recopilación de datos, se ha realizado un proceso de *web scraping* en Google Colab, una herramienta gratuita de Google para escribir y ejecutar código online en Python. En esta plataforma, se cargan las tablas clasificatorias en base a los resultados en LaLiga, así como las de las infracciones —integrando las cinco variables elegidas anteriormente— cometidas por y a los equipos en cuestión, todo ello para cada temporada a analizar. Para que el lenguaje de programación identifique correctamente las distintas URLs de cada temporada, se emplea la función *read\_html* de la librería Pandas, pues permite extraer tablas HTML desde páginas web para convertirlas en *DataFrames* —estructuras de datos bidimensionales a modo de tablas—. Para esta fase, se accede al portal FBref, ya que su organización digital facilita la extracción automatizada. Además, todas las tablas estadísticas para una misma temporada de LaLiga se hallan bajo una única URL. De esta forma, el acceso a diferentes datos —en este caso, penaltis, tarjetas amarillas, tarjetas rojas, faltas, fueras de juego y *ranking*— se consigue modificando únicamente el atributo “id” relativo a cada tabla, sin necesidad de cambiar la dirección del enlace principal, lo que optimiza el proceso de recolección.

En primer lugar, se extrae la tabla referente a la clasificación de los clubes de LaLiga para cada año. Una vez obtenida dicha tabla, se normalizan los nombres de todos los equipos, añadiendo el sufijo *\_2324* —o el correspondiente a la temporada en cuestión— al nombre de cada club para poder distinguir los datos por año en los análisis posteriores. Posteriormente, se depura el conjunto de datos para conservar exclusivamente las columnas necesarias —*RL* para la posición final en la tabla, *Equipo* para el nombre del club, *PC* para penaltis, *TA* para tarjetas amarillas, *TR* para tarjetas rojas, *Fls* para faltas y *PA* para fueras de juego— y eliminar así

información redundante o no relevante para los fines específicos del trabajo. Cabe señalar que, con el objetivo de evitar ambigüedades en la integración de los datos, se renombran las columnas relativas a las infracciones incorporando los sufijos *\_a* y *\_b* para indicar que éstas han sido cometidas por y a los equipos, respectivamente. Así, se dispone de tres tablas para cada temporada: una con la clasificación final de los clubes, otra con las principales sanciones cometidas por cada club y una tercera con las sanciones que estos mismos clubes han recibido por parte de sus rivales en su contra.

Finalmente, una vez extraídos y procesados todos los datos necesarios para el estudio, se consolidan en una única tabla conjunta. Para ello, los tres *DataFrames* anuales referentes a la clasificación, las sanciones cometidas por los equipos y las sanciones sufridas se agrupan en tres listas. Similarmente, cada lista se concatena verticalmente, preparando tres *DataFrames* globales, uno por cada tipo de información; éstos se fusionan en un *DataFrame* unificado que recoge todos los datos relevantes para el análisis. Por último, esta base de datos final se exporta en formato CSV y se descarga para su consiguiente uso en R. Si bien Python también dispone de librerías sólidas para diversos modelos de *Machine Learning*, se recurre a R debido a la experiencia previa con este lenguaje, lo que además permite hacer uso de los dos lenguajes de programación más utilizados en el ámbito académico.

Previamente al análisis en cuestión, se eliminan todos los datos del espacio de trabajo y se carga la base de datos unificada, así como la librería requerida para leerla (*readr*). Una vez leído el conjunto de datos inicial, se eliminan aquellas columnas que resulten irrelevantes o no comparables y se estandarizan todas las variables, otorgando a los datos una media de 0 y una desviación típica de 1 y permitiendo que todas las variables contribuyan equitativamente al estudio. Adicionalmente, se evalúa la idoneidad de desarrollar un Análisis de Componentes Principales (PCA) —técnica empleada para reducir la cantidad de variables— calculando la matriz de correlación, la cual devuelve el grado de relación lineal entre cada par de variables. Para confirmar la conveniencia o falta de la misma de reducir la dimensionalidad, se aplica la prueba de adecuación de Kaiser-Meyer-Olkin (KMO). Se asume que los valores entre 0,8 y 1 reflejan una alta adecuación de la muestra, mientras que aquellos inferiores a 0,6 sugieren que los resultados del análisis factorial podrían no ser fiables o útiles. Finalmente, se convierte el *DataFrame* a una matriz numérica, requisito necesario para las funciones posteriores en R.

En primer lugar, se lleva a cabo un *clustering* para clasificar a los equipos en grupos homogéneos en función de las acciones señaladas por los árbitros. Para ello, debe estimarse el número óptimo de clústeres mediante el paquete *NbClust*, el cual explora múltiples índices de validación interna, considera distintas opciones para la cantidad de clústeres y recomienda aquella que más veces se repita, siguiendo la regla de la mayoría. De esta forma, el algoritmo *hclust* crea una estructura jerárquica de grupos representada mediante un dendrograma, un gráfico en forma de árbol que muestra visualmente el proceso de agrupamiento. Éste se corta en el número de clústeres sugeridos anteriormente, asignando así cada observación a uno de ellos. Esta clasificación es incorporada al conjunto de datos original como una nueva variable para facilitar análisis comparativos en etapas posteriores del trabajo. Finalmente, se compara el rendimiento medio entre ambos clústeres para poder identificar, en cierto modo, si el grupo de equipos caracterizado por poseer una mayor o menor carga disciplinaria también presenta diferencias en términos de resultados competitivos. En esta línea, se lleva a cabo un análisis de la varianza (ANOVA) unifactorial, tomando como variable dependiente el *ranking* final y como variable explicativa el clúster asignado a cada club.

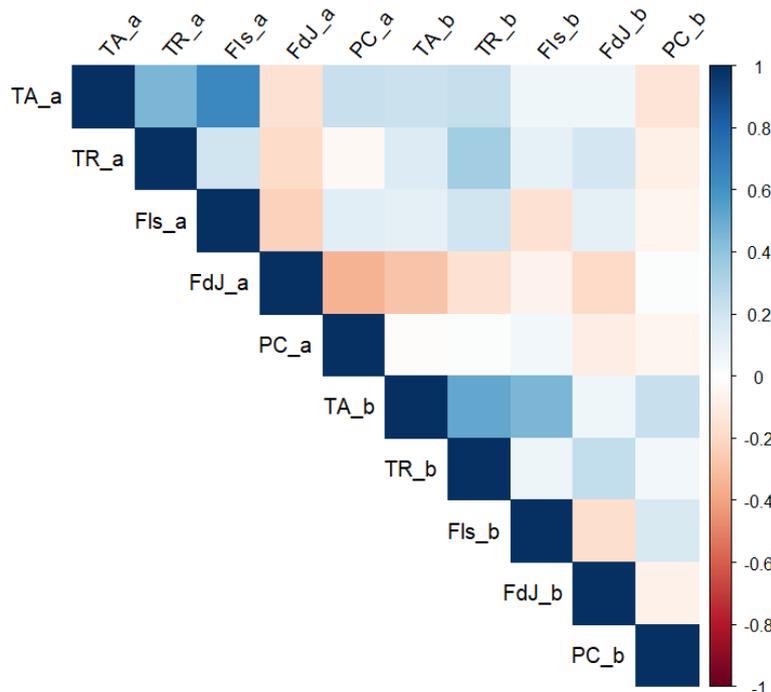
Una vez evaluada la posible capacidad explicativa de los patrones de comportamiento arbitrales sobre el rendimiento final de los clubes en la clasificación, se desarrolla un modelo de regresión logística (*logit*). Con esta técnica de aprendizaje supervisado, se logra estimar la probabilidad de que un equipo pertenezca a una determinada categoría —en este caso, acabar la temporada o bien en la mitad superior o bien en la mitad inferior de la tabla— en función de variables explicativas, tanto individuales —las acciones de cada equipo— como compuestas —la pertenencia a un clúster determinado—. De esta manera, el *logit* permite determinar hasta qué punto estos elementos pueden ser considerados predictores representativos del rendimiento deportivo. Así pues, se diseñan tres modelos: el primero considera exclusivamente la variable correspondiente a la pertenencia a uno u otro clúster, el segundo incluye de manera individual todas las variables originales referentes al estilo de juego y el tercero integra ambos enfoques. Posteriormente, se comparan los tres modelos con el propósito de fijar el predictor más eficaz, siguiendo el criterio de evaluación de la curva ROC y el área bajo dicha curva (AUC); esta área adopta valores comprendidos entre 0,5 y 1, donde un valor de 1 representa una capacidad perfecta de discriminación entre categorías, mientras que uno de 0,5 refleja un poder predictivo nulo, equivalente al del azar.

A continuación, se lleva a cabo un modelo de árbol de clasificación con el objetivo de detectar, de forma jerárquica, las variables más influyentes en la predicción del rendimiento final de los equipos. A tal efecto, se ejecuta la función *rpart* del paquete homónimo, la cual construye árboles de clasificación binaria a partir de un conjunto de variables explicativas. En este modelo, se incluyen tanto la pertenencia a un clúster determinado como los distintos indicadores arbitrales. La representación gráfica del árbol resultante facilita en gran medida la interpretación del modelo, visualizando las divisiones clave y localizando posibles umbrales críticos en las variables.

Finalmente, para enriquecer el trabajo y profundizar en la comparación de los estilos de juego y el trato arbitral recibido por los clubes con mayor presencia en la parte superior de la clasificación, se realiza un último análisis dedicado a los tres equipos más representativos de LaLiga en los últimos años: Real Madrid, Atlético de Madrid y Fútbol Club Barcelona. Para ello, se depura la base de datos original, conservando exclusivamente las observaciones correspondientes a dichos clubes. Se reorganizan los datos mediante la función *melt*, la cual permite convertir un formato ancho a uno largo, donde todas las observaciones relativas a una misma variable se encuentran en una única columna. Con esta estructura, se genera un conjunto de representaciones gráficas individuales para cada variable, organizadas en subgráficos dentro de una misma visualización. Ello mejora la claridad visual, ya que evita que las discrepancias en las magnitudes entre variables distorsionen la interpretación de las mismas.

#### **4.4 RESULTADOS DEL ANÁLISIS PARA TODOS LOS EQUIPOS DE LALIGA**

Una vez preparados los datos y confirmada la ausencia de valores perdidos, se calcula la matriz de correlaciones. Esta matriz muestra la intensidad y dirección de la relación lineal entre las variables, valorando la pertinencia de aplicar técnicas de reducción dimensional como el PCA, pues estos métodos resultan más efectivos cuando existen correlaciones significativas entre las variables, pudiendo condensar la información en un menor número de componentes.



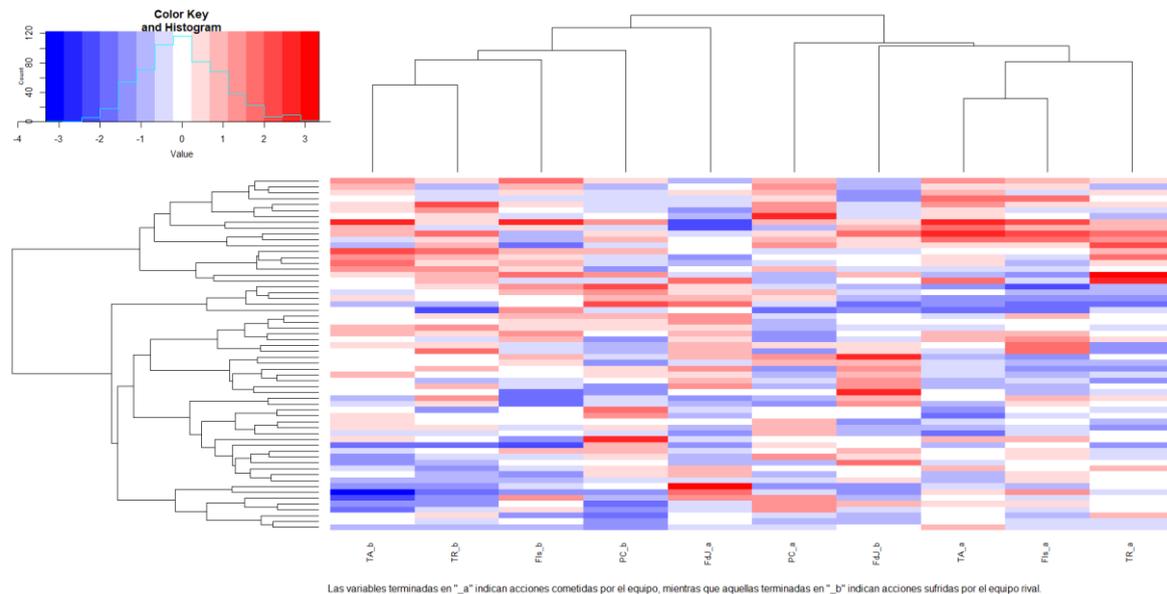
**Figura 4.** Matriz de correlación entre las variables.

Fuente: Elaboración propia (en R).

Esta matriz ilustra asociaciones moderadas entre algunas variables, principalmente la correlación entre tarjetas amarillas y faltas cometidas ( $r \approx 0,65$ ), lo que sugiere una posible redundancia en la información. Esto significa que ambas variables podrían estar capturando distintos aspectos de un mismo fenómeno dentro del análisis, lo que implica que una de ellas podría ser reemplazada o presentar un poder explicativo similar al de la otra. Sin embargo, para evaluar la viabilidad del PCA, se realiza la prueba de Kaiser-Meyer-Olkin (KMO), cuyo valor global es de 0,47; este resultado se halla por debajo del umbral mínimo recomendado de 0,6, por lo que se descarta que las correlaciones parciales sean lo suficientemente fuertes como para poder justificar una reducción de dimensiones. Por tanto, se decanta por mantener el conjunto completo de variables estandarizadas en los análisis posteriores.

Para agrupar los equipos según sus características mediante un *clustering* jerárquico, se debe especificar con anterioridad el número óptimo de clústeres. Atendiendo a la regla de la mayoría con la función *NbClust*, la partición más adecuada resulta en dos grupos. Se usa la distancia euclidiana como medida de similitud entre equipos, cuantificando cuán similares o distintos son los clubes según las variables numéricas estandarizadas. Asimismo, se emplea el enlace de *ward.d2* —variante del enlace de Ward clásico— como método de aglomeración,

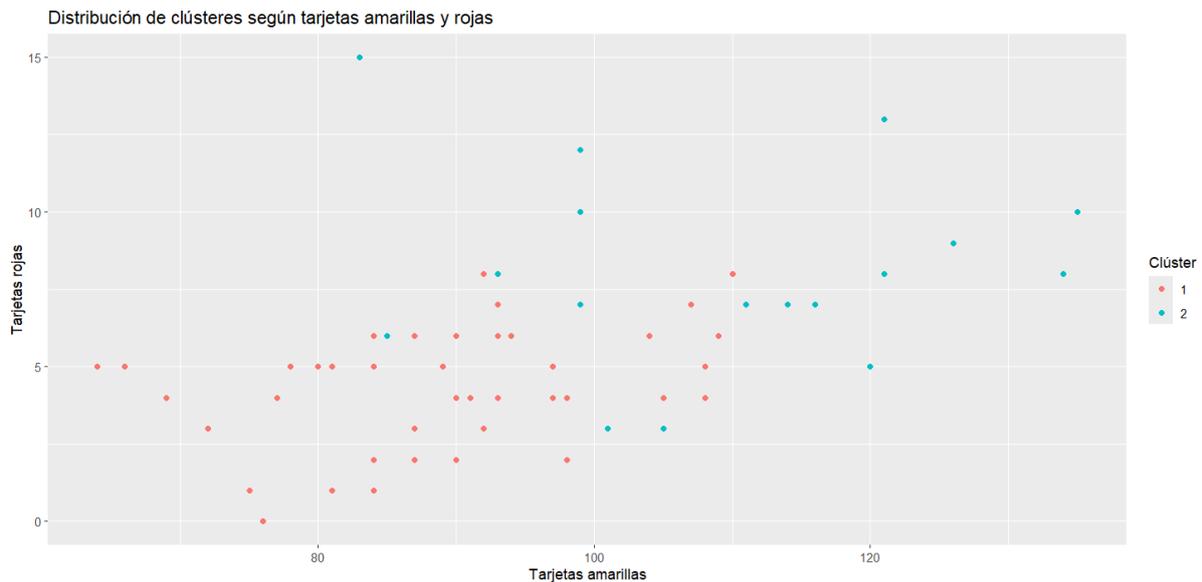
buscando que las observaciones dentro de un mismo clúster sean lo más homogéneas posible entre sí. Por añadidura, se elabora un mapa de calor con el propósito de visualizar los niveles de similitud entre los equipos. En él, las filas corresponden a cada observación —un equipo concreto en una temporada determinada— y las columnas a las distintas variables analizadas. El dendrograma resultante de la agrupación jerárquica manifiesta dos grupos principales, en concordancia con la segmentación de clústeres anterior.



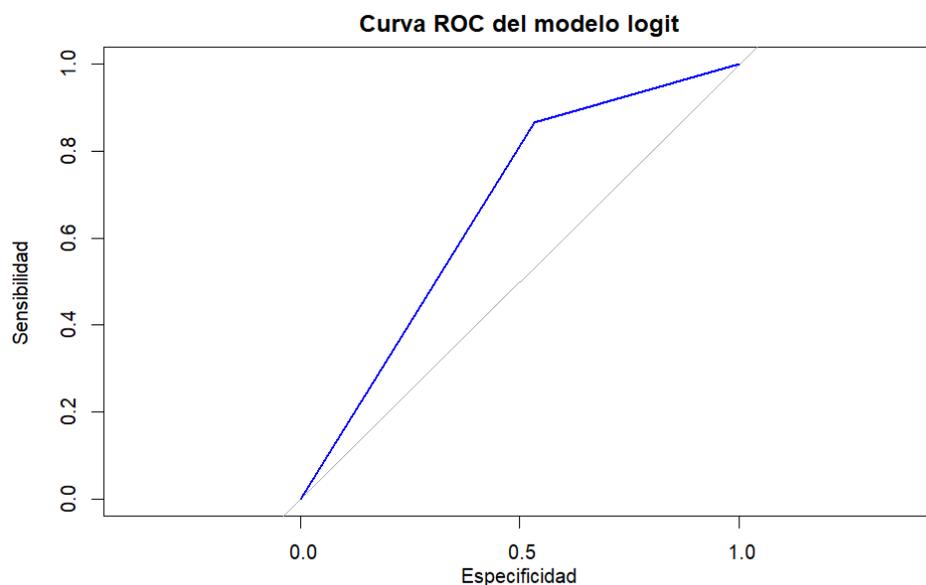
**Figura 5.** Mapa de calor de similitudes entre equipos.

Fuente: Elaboración propia (en R).

De esta manera, se realiza un análisis de *clustering* jerárquico, empleado los mismos métodos de distancia euclidiana y enlace *ward.d2*. Este procedimiento logra identificar dos grupos considerablemente diferenciados; 42 equipos pertenecen al clúster 1 y 18 al clúster 2. Esta asignación de clústeres se integra como una nueva variable al conjunto de datos original, convirtiéndola en un factor para su efectivo análisis posterior. Al calcular la diferencia media por variable entre ambos clústeres, se verifica que las variables —a excepción de los fuera de juego— presentan una distribución variada en función del clúster, siendo mayor el número de tarjetas, faltas y penaltis pitados a los equipos del segundo clúster en comparación con los del primero. Adicionalmente, se elabora un gráfico de dispersión centrado exclusivamente en las tarjetas amarillas y rojas, mostrando cómo se distribuyen los clubes en función de las tarjetas.



Posteriormente, se estiman tres modelos *logit* con el fin de explicar la probabilidad de que un equipo finalice la temporada en la mitad superior o inferior de la clasificación. A tal efecto, se transforma la variable referente al *ranking* en una variable binaria, donde el valor de 1 comprende a los equipos en la mitad superior y 0 a los equipos en la mitad inferior. El primer *logit* considera únicamente la pertenencia a un clúster; tomando en consideración el símbolo negativo del coeficiente y el p-valor de 0,007, se entiende que pertenecer al clúster 2 reduce notoriamente la probabilidad de situarse en la mitad superior. El segundo modelo se centra exclusivamente en las variables que definen el estilo de juego o el arbitraje, destacando tan sólo el número de penaltis como predictor significativo (valor  $p = 0,015$ ). En el tercer *logit*, el cual combina ambas fuentes de información, la variable relativa al rendimiento pierde drásticamente relevancia estadística, manteniéndose los penaltis señalados como el único predictor relevante (valor  $p = 0,025$ ). Para la comparación y correcta elección de los modelos, se establece el criterio de información de Akaike (AIC), optándose por el *logit* que presente menor AIC. De esta manera, se favorece el primer modelo, pues éste recibe un AIC de 78,89, lo que sugiere que la pertenencia a un clúster u otro resume mejora la probabilidad de bajo rendimiento que el resto de modelos. Para evaluar la capacidad predictiva del *logit*, se genera la curva ROC, así como el AUC correspondiente. El resultado,  $AUC = 0,667$ , refleja un poder discriminativo moderado, reforzando la relevancia de la variable del *ranking* como factor explicativo del rendimiento relativo de los clubes.

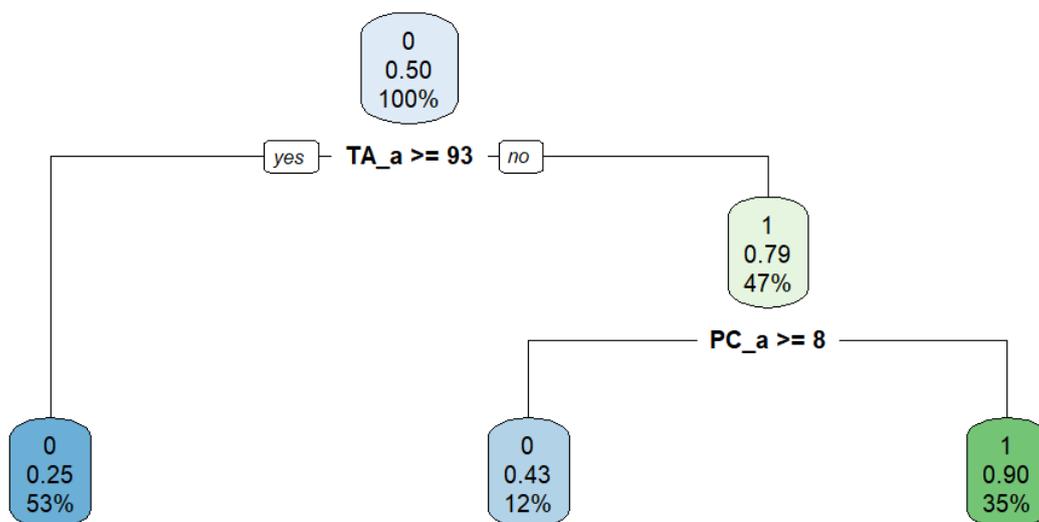


**Figura 7.** Curva ROC del modelo logit.

Fuente: Elaboración propia (en R).

Seguidamente, se procede a implementar un árbol de decisión, más específicamente de clasificación, para predecir qué variables explican mejor si un equipo finaliza en la mitad superior o inferior de la tabla clasificatoria. Este árbol señala las tarjetas amarillas mostradas a los equipos en cuestión como el primer nodo de partición, de tal forma que los equipos que reciben 93 o más tarjetas tienden a estar en la mitad inferior del *ranking*. En oposición, para los equipos con menos de 93 tarjetas amarillas, los penaltis sancionados resultan clave, pues aquellos que cometen menos de 8 penaltis tienen una probabilidad elevada (90%) de terminar en la mitad superior; contrariamente, los clubes con 8 o más penaltis —cabe señalar que un 43% de los equipos en este nodo pertenecen a esta clase— suelen acabar la temporada en los puestos más bajos de la clasificación. En suma, se puede concluir que, en términos generales, a mayor número de tarjetas amarillas, menor rendimiento en la tabla. El árbol primero divide por tarjetas amarillas y se observa que, dentro del grupo menos amonestado, los penaltis son los que más discriminan el rendimiento deportivo; ergo, parece ser que las tarjetas amarillas y los penaltis concedidos son las variables más influyentes en la predicción del rendimiento. Se ejecuta un vector para confirmar el grado de influencia de cada variable en la construcción del árbol, obteniendo que las tarjetas amarillas (8,571) constituyen, con diferencia, la variable más relevante de predicción, seguida de los penaltis (4,218).

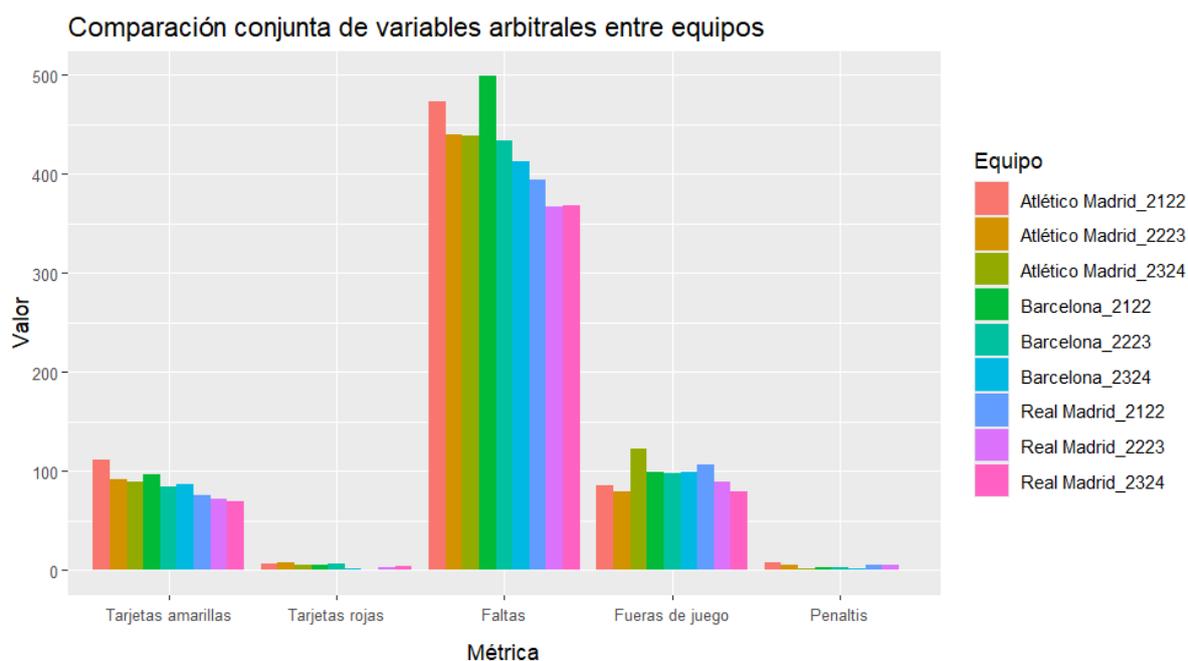
**Árbol de clasificación para el ranking de equipos**



**Figura 8.** *Árbol de clasificación para el ranking de equipos.*

Fuente: Elaboración propia (en R).

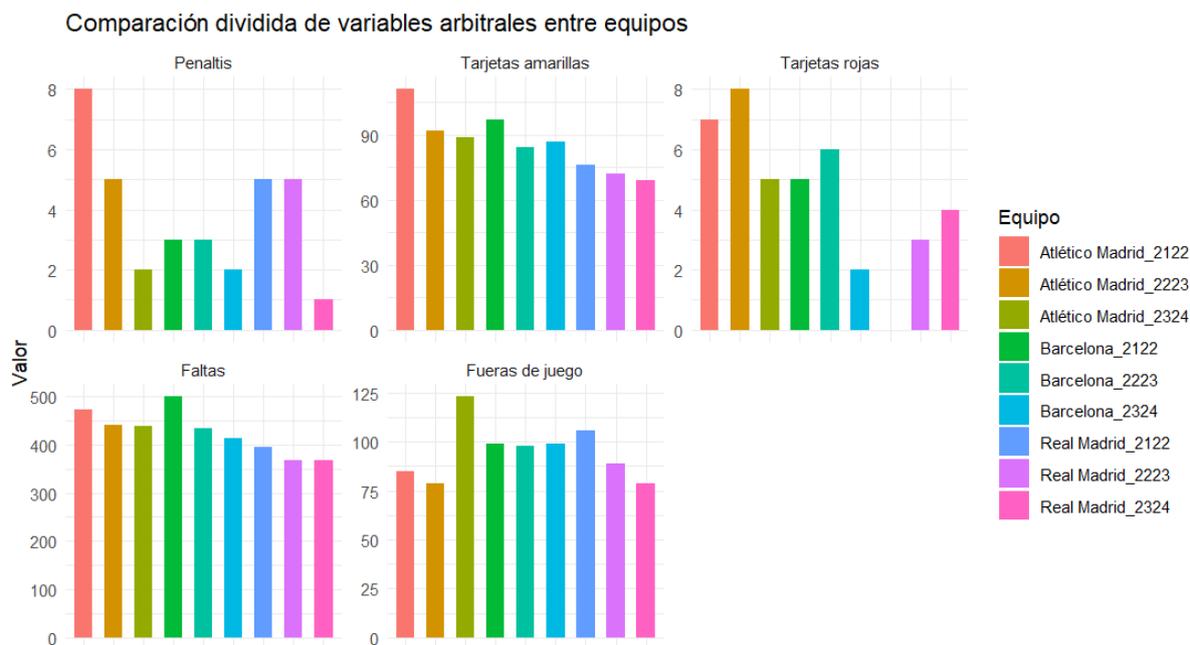
Finalmente, se lleva a cabo un análisis comparativo entre los tres equipos con mayor protagonismo histórico de LaLiga: Real Madrid, Fútbol Club Barcelona y Atlético de Madrid. Para ello, se filtra el conjunto de datos de tal modo que, para las cinco variables arbitrales clave, éste devuelva exclusivamente las observaciones correspondientes a estos tres clubes. Las nueve observaciones resultantes —tres por cada equipo para cada una de las tres temporadas— son reorganizadas mediante la función *melt*, creando un gráfico de barras comparativo para cada métrica.



**Figura 9.** Comparación conjunta de variables arbitrales entre equipos.

Fuente: Elaboración propia (en R).

Para facilitar la claridad visual y evitar que las diferencias en las magnitudes entre las distintas variables alteren su correcta interpretación, se genera un nuevo gráfico que presente de manera individualizada una visualización para cada variable. A través de esta segmentación, se puede comparar a los equipos más fácilmente en cada dimensión, sin que el impacto visual de una variable en particular domine sobre las demás. De esta manera, se logra entender mejor las discrepancias en el estilo de juego de los clubes en relación con las decisiones arbitrales.



**Figura 10.** Comparación dividida de variables arbitrales entre equipos.

Fuente: Elaboración propia (en R).

En primer lugar, en cuanto al número de penaltis, se resalta que el Atlético de Madrid en la temporada 2021/22 fue pitado más del doble de penaltis que la media de ninguno de los otros dos equipos, cuyos datos presentan una cantidad de penaltis más homogénea. Por otro lado, en el panel de tarjetas amarillas, se observa una ligera tendencia decreciente para todos los equipos a nivel temporal. Nuevamente, el Atlético de Madrid presenta el máximo valor para cada temporada, mientras que el Real Madrid contiene el mínimo. En cuanto a la cantidad de tarjetas rojas, el Atlético muestra el máximo de 8 en la temporada 2022/23, así como el Real Madrid obtiene el mínimo de 0. En cambio, respecto a las faltas cometidas, es el Fútbol Club Barcelona el que muestra un máximo de 499 faltas; de nuevo, el Atlético incurre en un mayor número de faltas que el Real Madrid. Adicionalmente, se vuelve a identificar cierta tendencia declinante con el tiempo. Por último, en posiciones adelantadas, el Atlético de Madrid lidera en esta variable en la última temporada, si bien mostrando una variabilidad superior a la del resto de los equipos.

## CAPÍTULO 5: CONCLUSIONES

---

El presente trabajo ha verificado la existencia de posibles patrones diferenciales en las decisiones arbitrales tomadas en los partidos disputados entre el Real Madrid y el Atlético de Madrid durante la última década, si bien ello no implica causalidad. A través de un análisis estadístico minucioso, complementado con técnicas de aprendizaje automático, se ha buscado no sólo distinguir posibles discrepancias cuantificables, sino también interpretarlas dentro del contexto general de LaLiga.

En primer lugar, los análisis pareados realizados exclusivamente sobre los datos de los derbis madrileños han revelado diferencias estadísticamente notables en tres variables clave: tarjetas amarillas, faltas y fuera de juego. En concreto, se ha manifestado que el Atlético de Madrid acumula más tarjetas amarillas, comete un mayor número de faltas y cae en posición adelantada con más frecuencia que el Real Madrid. Estas diferencias se mantienen incluso tras haber aplicado correcciones por comparaciones múltiples, lo cual incrementa su validez estadística. Además, el tamaño del efecto observado, medido mediante la  $d$  de Cohen y el estadístico  $A$  de Vargha y Delaney, parece apuntar a que dichas discrepancias no sólo son estadísticamente significativas, sino también relevantes en cuanto a magnitud —en especial, en el caso de las faltas, se identificó un efecto de gran tamaño—. Por el contrario, no se han detectado diferencias significativas en la cantidad de penaltis o de tarjetas rojas. Cabe señalar, no obstante, que esta ausencia de evidencia no implica necesariamente una equidad absoluta, ya que podría deberse a la baja frecuencia de este tipo de sanciones.

En segundo lugar, con el propósito de valorar si estas diferencias son excepcionales o forman parte de una tendencia más generalizada en el fútbol español, se extendió el alcance del análisis. A través de técnicas de *clustering* jerárquico, se identificaron dos tipos de clubes según su perfil disciplinario. Se comprobó que aquellos equipos encuadrados en el clúster con mayor carga de sanciones —entre los que se encuentra un año del Atlético de Madrid— tienden a ocupar posiciones más bajas en la clasificación. Este hallazgo resulta especialmente llamativo si se parte de la hipótesis, comúnmente referida en la literatura futbolística, de que los clubes de mayor prestigio institucional suelen verse favorecidos en las decisiones arbitrales. En este

sentido, sorprende que un equipo de la dimensión del Atlético de Madrid aparezca, al menos en una de las temporadas analizadas, en el clúster que recibe un trato más severo, rompiendo así con las expectativas habituales sobre el sesgo a favor de los más grandes. Adicionalmente, la relación entre el número de amonestaciones y el *ranking* en la tabla a final de temporada fue respaldada mediante un ANOVA y reforzada por modelos *logit* y árboles de clasificación, los cuales señalaron las tarjetas amarillas y los penaltis como las variables más determinantes en la predicción del rendimiento deportivo. Por último, la comparación entre el Real Madrid, el Fútbol Club Barcelona y el Atlético de Madrid corrobora estas conclusiones; el Atlético de Madrid tiene el mayor número de sanciones en prácticamente todas las variables —a excepción de los penaltis—, mientras que el Real Madrid figura las cifras más bajas en todas las métricas —nuevamente, a excepción de los penaltis—. Esta disparidad podría apuntar, más allá del contexto específico de los derbis, a una posible asimetría estructural en el trato arbitral que merece ser estudiada con mayor profundidad.

En conjunto, los resultados obtenidos pueden conducir a pensar que sí existen indicios estadísticos de un trato diferenciado entre el Real Madrid y el Atlético de Madrid en aspectos clave del arbitraje, especialmente en decisiones que inciden directamente en el desarrollo del juego. Al extender el foco al resto de clubes de LaLiga, se observa que estos patrones no son únicos, sino que parecen formar parte de una dinámica más amplia en la que factores como el estilo de juego, la reputación institucional o incluso posibles sesgos inconscientes por parte de los árbitros podrían estar influyendo en su comportamiento. Cabe mencionar que este trabajo no busca emitir un juicio concluyente sobre la intencionalidad o legitimidad de las decisiones arbitrales, sino aportar evidencia empírica que contribuya al debate académico y social sobre la equidad en el fútbol profesional. En último término, el objetivo final es crear conciencia en los colegiados en torno a la posible existencia de sesgos que pueden manifestarse, máxime en encuentros de grandes magnitudes como son los derbis entre estos dos gigantes del fútbol español.

## LIMITACIONES DEL ESTUDIO

Entre las grandes limitaciones del estudio, destaca el tamaño relativamente reducido de la primera muestra, al centrarse en enfrentamientos específicos a lo largo de una década. Pese a haber utilizado técnicas estadísticas consistentes y ampliarse el análisis, la naturaleza limitada de los derbis madrileños puede no ser plena y fielmente representativa del comportamiento arbitral. Adicionalmente, las decisiones arbitrales están contextualizadas por múltiples factores situacionales —como la influencia de la afición, la reputación de los jugadores en cuestión, el minuto del partido o el marcador— que no se han incluido en el modelo, lo que podría afectar a la interpretación de los resultados.

Además, otra limitación considerable es la subjetividad inherente a muchas decisiones arbitrales, donde el criterio propio del colegiado juega un papel fundamental. Aunque el uso de herramientas de aprendizaje automático permite captar patrones básicos, estas tecnologías no son capaces de recoger el marco cualitativo en el que se dan estas decisiones. Asimismo, factores como el estilo de juego o el porcentaje de posesión de cada equipo pueden generar un mayor número de amonestaciones de manera natural, por lo que podrían estar influyendo en las cifras observadas.

## FUTURAS LÍNEAS DE INVESTIGACIÓN

Como líneas futuras de investigación, sería interesante ampliar el análisis a un mayor número de temporadas y competiciones, con el fin de comprobar si los patrones observados se mantienen en contextos diferentes. Además, variables contextuales como el marcador en el momento de la decisión, la posición en el campo de la sanción, el minuto del partido, el campo de juego —si se es local o visitante— o la presión por la posición en el *ranking* de los equipos en cada encuentro podría enriquecer sustancialmente la investigación. Similarmente, se podría llevar a cabo un análisis longitudinal a raíz de la introducción del VAR, evaluando si esta herramienta ha logrado reducir las discrepancias arbitrales.

## DECLARACIÓN RESPECTO AL USO DE HERRAMIENTAS IAG

Por la presente, yo, Victoria García-Argüelles Fernández-España, estudiante de Business Analytics y Relaciones Internacionales de la Universidad Pontificia Comillas, al presentar mi Trabajo Fin de Grado titulado “Análisis comparativo de la influencia de las decisiones arbitrales: Caso Real Madrid vs. Atlético de Madrid”, declaro que he utilizado la herramienta de Inteligencia Artificial Generativa ChatGPT u otras similares de IAG de código sólo en el contexto de las actividades descritas a continuación:

1. Brainstorming de ideas de investigación: Idear y esbozar posibles áreas de investigación.
2. Metodólogo: Descubrir métodos aplicables a problemas específicos de investigación.
3. Interpretador de código: Realizar análisis de datos preliminares.
4. Corrector de estilo literario y de lenguaje: Mejorar la calidad lingüística y estilística del texto.
5. Revisor: Recibir sugerencias sobre cómo mejorar y perfeccionar el trabajo con diferentes niveles de exigencia.

Afirmo que toda la información y contenido presentados en este trabajo son producto de mi investigación y esfuerzo individual, excepto donde se ha indicado lo contrario y se han dado los créditos correspondientes (he incluido las referencias adecuadas en el TFG y he explicitado para qué se ha usado ChatGPT u otras herramientas similares). Soy consciente de las implicaciones académicas y éticas de presentar un trabajo no original y acepto las consecuencias de cualquier violación a esta declaración.

Fecha: 15 de junio de 2025

## BIBLIOGRAFÍA

- Alonso, P. (2017). El fútbol español: Luces y sombras. *Razón y Fe*, 276(1425-1426), 17-24.  
<https://revistas.comillas.edu/index.php/razonyfe/article/view/9325/8747>
- Audrino, F. (2020). Strongest team favoritism in European national football: Myth or reality? *International Journal of Sport Finance*, 15(3), 137-152.  
<https://doi.org/10.32731/jsf.2020.a927063>
- Boyko, R. H., Boyko, A. R., & Boyko, M. G. (2007) Referee bias contributes to home advantage in English Premiership football. *Journal of Sports Sciences*, 25(11), 1185-1194.  
<https://doi.org/10.1080/02640410601038576>
- Buraimo, B., Forrest, D., & Simmons, R. (2010). The 12th man?: Refereeing bias in English and German soccer. *Journal of the Royal Statistical Society Series A (Statistics in Society)*, 173(2), 431–449. <https://doi.org/10.1111/j.1467-985X.2009.00604.x>
- Caracuel, J. C., Andreu, R., & Pérez, E. (1995). Análisis psicológico del arbitraje y juicio deportivos: Una aproximación desde el modelo interconductual. *Motricidad: European Journal of Human Movement*, 1.
- Carlin, J. C. (4 de noviembre de 2012). Conspiración de silencio. *El País*.  
[https://elpais.com/deportes/2012/11/03/actualidad/1351966232\\_651386.html](https://elpais.com/deportes/2012/11/03/actualidad/1351966232_651386.html)
- Dadı, M. M., & Yıldız, O. (2022). Opinions of football referees on the VAR system and VAR training. *Turkish Journal of Sport and Exercise*, 24(1), 52-65. <https://doi.org/10.15314/tsed.1072304>
- Dawson, P., Dobson, S., Goddard, J., & Wilson, J. (2007). Are football referees really biased and inconsistent?: Evidence on the incidence of disciplinary sanction in the English Premier League. *Journal of the Royal Statistical Society Series A (Statistics in Society)*, 170(1), 231-250. <https://doi.org/10.1111/j.1467-985X.2006.00451.x>
- Debien, P. B., Noce, F., Debien, J. B. P., & da Costa, V. T. (2014). Stress in rhythmic gymnastics refereeing: A systematic review. *Revista da Educação Física/UEM*, 25(3), 489-500.  
<https://doi.org/10.4025/reveducfis.v25i3.22031>

- Deloitte. (23 de enero de 2025). *Deloitte Football Money League 2025*.  
<https://www.deloitte.com/uk/en/services/financial-advisory/analysis/deloitte-football-money-league.html>
- Dohmen, T. J. (2008). The influence of social forces: Evidence from the behavior of football referees. *Economic Inquiry*, 46(3), 411-424. <https://doi.org/10.1111/j.1465-7295.2007.00112.x>
- Dohmen, T., & Sauermaun, J. (2015). Referee bias. *Journal of Economic Surveys*, 30(4), 679-695. <https://doi.org/10.1111/joes.12106>
- Erikstad, M. K., & Johansen, B. T. (2020). Referee bias in professional football: Favoritism toward successful teams in potential penalty situations. *Frontiers in Sports and Active Living*, 2(19). <https://doi.org/10.3389/fspor.2020.00019>
- Football Benchmark. (28 de mayo de 2025). *Football club's valuation: The European elite 2025*.  
<https://footballbenchmark.com/documents/d/guest/football-benchmark-the-european-elite-2025>
- Fuentes, R. (2 de febrero de 2021). *Así se evalúa actualmente a los árbitros de Primera y Segunda*. Iusport.  
<https://iusport.com/archive/54470/asi-se-evalua-actualmente-a-los-arbitros-de-primera-y-segunda>
- Garicano, L., Palacios-Huerta, I., & Prendergast, C. (2005). Favoritism under social pressure. *The Review of Economics and Statistics*, 87(2), 208-216. <https://doi.org/10.1162/0034653053970267>
- Gilis, B., Helsen, W., Catteeuw, P., & Wagemans, J. (2008). Offside decisions by expert assistant referees in association football: Perception and recall of spatial positions in complex dynamic events. *Journal of Experimental Psychology: Applied*, 14(1), 21-35. <https://doi.org/10.1037/1076-898X.14.1.21>
- Goumas, C. (2014). Home advantage and referee bias in European football. *European Journal of Sport Science*, 14(Suppl. 1), S243–S249. <https://doi.org/10.1080/17461391.2012.686062>

- Guillén, A. (2023). Algunos apuntes y debates sobre la reputación del colectivo arbitral en el fútbol español. *Revista Española de Derecho Deportivo*, 52, 91-113.  
<https://doi.org/10.30462/redp-2023-02-04-933>
- Haselton, M. G., & Buss, D. M. (2000). Error management theory: A new perspective on biases in cross-sex mind reading. *Journal of Personality and Social Psychology*, 78(1), 81-91.  
<https://doi.org/10.1037/0022-3514.78.1.81>
- Haselton, M. G., & Nettle, D. (2006). The paranoid optimist: An integrative evolutionary model of cognitive biases. *Personality and Social Psychology Review*, 10(1), 47-66.  
[https://doi.org/10.1207/s15327957pspr1001\\_3](https://doi.org/10.1207/s15327957pspr1001_3)
- He, Y. (2012). *Predicting market value of soccer players using linear modeling techniques*. University of California Berkeley Department of Statistics.  
[https://www.stat.berkeley.edu/~aldous/Research/Ugrad/Yuan\\_He.pdf](https://www.stat.berkeley.edu/~aldous/Research/Ugrad/Yuan_He.pdf)
- Herberger, T. A., & Wedlich, F. (2017). Does selection bias matter in football players' valuation? A crowdsourced valuation approach on players' athletic characteristics. *Journal of Global Sport Management*, 2(3), 196-214. <https://doi.org/10.1080/24704067.2017.1350593>
- Jiménez-Sánchez, Á., Lavín, J. M., & Endara, D. F. (2021). Repercusiones de jugar sin público en la ventaja local, las decisiones arbitrales y en los componentes del juego. *Cuadernos de Psicología del Deporte*, 21(2), 198-212. <https://doi.org/10.6018/cpd.444811>
- Jones, M. V., Paull, G. C., & Erskine, J. (2002). The impact of a team's aggressive reputation on the decisions of association football referees. *Journal of Sports Sciences*, 20(12), 991-1000.  
<https://doi.org/10.1080/026404102321011751>
- Karafil, A. Y. (2023). Examination of football referees' attitudes towards Video Assistant Referee system (VAR) by Q method. *Spor Bilimleri Araştırmaları Dergisi*, 8(1), 99-112.  
<https://doi.org/10.25307/jssr.1178303>
- Latané, B. (1981). The psychology of social impact. *American Psychologist*, 36(4), 343-356.  
<https://doi.org/10.1037/0003-066X.36.4.343>
- Nevill, A. M., Newell, S. M., & Gale, S. (1996). Factors associated with home advantage in English and Scottish Soccer matches. *Journal of Sports Sciences*, 14(2), 181-186.

<https://doi.org/10.1080/02640419608727700>

Nevill, A., Balmer, N., & Williams, A. M. (2002). The influence of crowd noise and experience upon refereeing decisions in football. *Psychology of Sport and Exercise*, 3(4), 261-272. [https://doi.org/10.1016/s1469-0292\(01\)00033-4](https://doi.org/10.1016/s1469-0292(01)00033-4)

Nogueira, D., Fontes, L., Gomes, A. R., & Resende, R. (2022). Emociones y percepción de rendimiento en árbitros: Importancia del estrés y la evaluación cognitiva. *Cuadernos de Psicología del Deporte*, 22(2), 156-170.

Nufer, G., & Alesi, N. (2018). The Halo Effect in sports. *International Journal of Business and Social Science*, 9(2). <https://doi.org/10.30845/ijbss>

Nufer, G. (2019). "Say hello to Halo": The halo effect in sports. *Innovative Marketing*, 15(3), 116-129. [https://doi.org/10.21511/im.15\(3\).2019.09](https://doi.org/10.21511/im.15(3).2019.09)

Pettersson-Lidbom, P., & Priks, M. (2010). Behavior under social pressure: Empty Italian stadiums and referee bias. *Economics Letters*, 108(2), 212-214. <https://doi.org/10.1016/j.econlet.2010.04.023>

Pollard, R. (2008). Home advantage in football: A current review of an unsolved puzzle. *The Open Sports Sciences Journal*, 1(1), 12-14. <https://doi.org/10.2174/1875399X00801010012>

Poolton, J. M., Chan, M. S., & Masters, R. (2011). The home team advantage gives football referees something to ruminate about. *International Journal of Sports Science & Coaching*, 6(4), 545-552. <https://doi.org/10.1260/1747-9541.6.4.545>

Porte, E. G. (2020). Nivel de estrés percibido en árbitros de fútbol del Alto Valle. *Universidad de Flores*. <https://doi.org/20.500.14340/530>

Real Academia Española. (s.f.). Derbi. En *Diccionario de la lengua española*. Recuperado el 23 de febrero de 2025, de <https://dle.rae.es/derbi>

Reilly, B., & Witt, R. (2013). Red cards, referee home bias and social pressure: Evidence from English Premiership Soccer. *Applied Economics Letters*, 20(7), 710-714. <https://doi.org/10.1080/13504851.2012.734591>

- Reoyo, A. (2025). El derbi madrileño en un mar de polémica. *La Escuadra*.  
<https://diariolaescuadra.com/derbi-madrileno-cargado-de-polemica/>
- Sánchez, P. A., García-Calvo, T., Leo, F. M., Pollard, R., & Gómez, M. A. (2009). An analysis of home advantage in the top two Spanish professional football leagues. *Perceptual and Motor Skills*, 108(3), 789-797. <https://doi.org/10.2466/pms.108.3.789-797>
- Schrödter, R., Noël, B., & Klatt, S. (2022). Game management by referees to compensate for errors in judgement: A decision flow model. *International Journal of Sport and Exercise Psychology*, 22(3), 612-631. <https://doi.org/10.1080/1612197X.2022.2161105>
- Seckin, A., & Pollard, R. (2008). Home advantage in Turkish professional soccer. *Perceptual and Motor Skills*, 107(1), 51-54. <https://doi.org/10.2466/pms.107.1.51-54>
- Singh, A. (20 de febrero de 2025). *The most popular sports in the world*. WorldAtlas. Recuperado el 3 de marzo de 2025 de  
<https://www.worldatlas.com/articles/what-are-the-most-popular-sports-in-the-world.html>
- Spitz, J., Wagemans, J., Memmert, D., Williams, A. M., & Helsen, W. F. (2020). Video assistant referees (VAR): The impact of technology on decision making in association football referees. *Journal of Sports Sciences*, 39(2), 147-153. <https://doi.org/10.1080/02640414.2020.1809163>
- Sutter, M., & Kocher, M. G. (2004). Favoritism of agents - The case of referees' home bias. *Journal of Economic Psychology*, 25(4), 461-469. [https://doi.org/10.1016/S0167-4870\(03\)00013-8](https://doi.org/10.1016/S0167-4870(03)00013-8)
- Taylor, A. H., & Daniel, J. V. (1988). Sources of stress in soccer officiating: An empirical study. En T. Reilly, A. Lees, K. Davids & W. J. Murphy (Eds.), *Science and football: Proceedings of the First World Congress of Science and Football* (pp. 538-544). Routledge.
- Taylor, A. H., Daniel, J. V., Leith, L., & Burke, R. J. (1990). Perceived stress, psychological burnout and paths to turnover intentions among sport officials. *Journal of Applied Sport Psychology*, 2(1), 84-97. <https://doi.org/10.1080/10413209008406422>
- Uribarri, R. (2016). La patraña del equipo violento. *Contexto y Acción*.  
<https://ctxt.es/es/20161012/Deportes/8919/Atletico-de-Madrid-faltas-tarjetas-equipo-violento.htm>

- Vela, D., & Arbinaga, F. (2018). Aplicación del listado de situaciones estresantes para el arbitraje (LISEA) en árbitros de fútbol. *Revista Iberoamericana de Psicología del Ejercicio y el Deporte*, 13(2), 281-287.
- Vicandi, I. (2022). El impacto del VAR en los órganos disciplinarios sancionadores del fútbol profesional en España. *Revista Aranzadi de Derecho de Deporte y Entretenimiento*, 77.
- Wallace, H. M., Baumeister, R. F., & Vohs, K. D. (2005). Audience support and choking under pressure: A home disadvantage? *Journal of Sports Sciences*, 23(4), 429-438.  
<https://doi.org/10.1080/02640410400021666>
- Weinberg, R. S., & Richardson, P. A. (1990). *Psychology of officiating*. Leisure Press.  
<https://doi.org/10.1123/tsp.5.2.196>

## ANEXOS

### 1. Código del análisis estadístico (R)

```
# Pasos previos
remove(list = ls()) # eliminar las variables y funciones en memoria
setwd(dirname(rstudioapi::getActiveDocumentContext())$path) # establecer el directorio de trabajo
getwd() # verificar el directorio de trabajo

# Cargar los datos
library(readxl)
df1 = read_excel("BD1.xlsx")

# Revisar los datos
View(df1)
summary(df1)
str(df1)

### ANÁLISIS DE PENALTIS

# Comprobar la normalidad
shapiro.test(df1$Penaltis_RMA)
shapiro.test(df1$Penaltis_ATM)

# Comprobar la igualdad de varianzas
var.test(df1$Penaltis_RMA, df1$Penaltis_ATM)

# Realizar el contraste
wilcox.test(df1$Penaltis_RMA, df1$Penaltis_ATM, paired = TRUE, alternative =
"two.sided") # paired = TRUE por ser datos pareados; alternative = "two.sided" por asumir una hipótesis bilateral, investigando tanto si hay diferencias como si no

### ANÁLISIS DE TARJETAS AMARILLAS

shapiro.test(df1$Tarjetas_Amarillas_RMA)
shapiro.test(df1$Tarjetas_Amarillas_ATM)
```

```

var.test(df1$Tarjetas_Amarillas_RMA, df1$Tarjetas_Amarillas_ATM)

wilcox.test(df1$Tarjetas_Amarillas_RMA, df1$Tarjetas_Amarillas_ATM, paired = TRUE,
alternative = "two.sided")

### ANÁLISIS DE TARJETAS ROJAS

shapiro.test(df1$Tarjetas_Rojas_RMA)
shapiro.test(df1$Tarjetas_Rojas_ATM)

var.test(df1$Tarjetas_Rojas_RMA, df1$Tarjetas_Rojas_ATM)

wilcox.test(df1$Tarjetas_Rojas_RMA, df1$Tarjetas_Rojas_ATM, paired = TRUE, alternative =
"two.sided")

### ANÁLISIS DE FALTAS

shapiro.test(df1$Faltas_RMA)
shapiro.test(df1$Faltas_ATM)

var.test(df1$Faltas_RMA, df1$Faltas_ATM)

t.test(df1$Faltas_RMA, df1$Faltas_ATM, var.equal = TRUE, paired = TRUE) # var.equal =
TRUE por haber igualdad de varianzas

### ANÁLISIS DE FUERAS DE JUEGO

shapiro.test(df1$Fueras_Juego_RMA)
shapiro.test(df1$Fueras_Juego_ATM)

var.test(df1$Fueras_Juego_RMA, df1$Fueras_Juego_ATM)

wilcox.test(df1$Fueras_Juego_RMA, df1$Fueras_Juego_ATM, paired = TRUE, alternative =
"two.sided")

### CORRECCIONES

# Ver los p-valores
p_valores = c(

```

```

wilcox.test(df1$Penaltis_RMA, df1$Penaltis_ATM, paired = TRUE)$p.value,
wilcox.test(df1$Tarjetas_Amarillas_RMA, df1$Tarjetas_Amarillas_ATM, paired =
TRUE)$p.value,
wilcox.test(df1$Tarjetas_Rojas_RMA, df1$Tarjetas_Rojas_ATM, paired = TRUE)$p.value,
t.test(df1$Faltas_RMA, df1$Faltas_ATM, paired = TRUE)$p.value,
wilcox.test(df1$Fuera_Juego_RMA, df1$Fuera_Juego_ATM, paired = TRUE)$p.value
)
p_valores

# Método Bonferroni
p_valores_ajustados_bonferroni = p.adjust(p_valores, method = "bonferroni")
p_valores_ajustados_bonferroni

# Método Holm
p_valores_ajustados_holm = p.adjust(p_valores, method = "holm")
p_valores_ajustados_holm

# Método Benjamini-Hochberg
p_valores_ajustados_BH = p.adjust(p_valores, method = "BH")
p_valores_ajustados_BH

### TAMAÑO DEL IMPACTO DE LAS DIFERENCIAS SIGNIFICATIVAS

# Tarjetas Amarillas
install.packages("effsize")
library(effsize)
VD.A(df1$Tarjetas_Amarillas_RMA, df1$Tarjetas_Amarillas_ATM)

# Faltas
cohen.d(df1$Faltas_RMA, df1$Faltas_ATM, paired = TRUE)

# Fuera de juego
VD.A(df1$Fuera_Juego_RMA, df1$Fuera_Juego_ATM)

### VISUALIZACIONES

# Tarjetas Amarillas
boxplot(df1$Tarjetas_Amarillas_RMA, df1$Tarjetas_Amarillas_ATM,
names = c("Real Madrid", "Atlético de Madrid"),
col = c("#BF3EFF", "#FF3030"),
main = "Comparación de tarjetas amarillas",

```

```

        ylab = "Número de tarjetas")

# Faltas
boxplot(df1$Faltas_RMA, df1$Faltas_ATM,
        names = c("Real Madrid", "Atlético de Madrid"),
        col = c("#BF3EFF", "#FF3030"),
        main = "Comparación de faltas",
        ylab = "Número de faltas")

# Fueras de juego
boxplot(df1$Fuera_Juego_RMA, df1$Fuera_Juego_ATM,
        names = c("Real Madrid", "Atlético de Madrid"),
        col = c("#BF3EFF", "#FF3030"),
        main = "Comparación de fueros de juego",
        ylab = "Número de fueros de juego")

```

## 2. Código del *web scraping* (Python)

```

# Cargar la librería
import pandas as pd

## **Tabla LaLiga 23/24**

# Leer la tabla de la página HTML de FBref
df0_2324 = pd.read_html("https://fbref.com/es/comps/12/2023-2024/Estadisticas-2023-2024-
La-Liga",
                        attrs = {"id":"results2023-2024121_overall"}) [0] # extraer la primera
tabla y transformarla en un diccionario, con el id especificado como key

# Incluir el sufijo "_2324" a los equipos para identificar cada temporada
df0_2324["Equipo"] = df0_2324["Equipo"] + "_" + "2324"

# Seleccionar únicamente las columnas "RL" (posición al final de temporada) y "Equipo"
df0_2324 = df0_2324[["RL", "Equipo"]]

df0_2324.head()

## **Faltas cometidas por los equipos 23/24**

```

```

dfa_2324 = pd.read_html("https://fbref.com/es/comps/12/2023-2024/Estadisticas-2023-2024-
La-Liga",
                        attrs = {"id":"stats_squads_misc_for"}) [0]

# Unificar las dos primeras columnas de encabezado en una sola línea para deshacerse de
la estructura multinivel
dfa_2324.columns = ["_".join(col).strip() for col in dfa_2324.columns.values]

dfa_2324["Unnamed: 0_level_0_Equipo"] = dfa_2324["Unnamed: 0_level_0_Equipo"] + "_" +
"2324"

# Seleccionar únicamente las columnas "TA" (Tarjetas Amarillas), "TR" (Tarjetas Rojas),
"Fls" (Número de faltas), "PA" (Fuera de juego) y "Penal concedido"
dfa_2324 = dfa_2324[["Unnamed: 0_level_0_Equipo", "Rendimiento_TA", "Rendimiento_TR",
"Rendimiento_Fls", "Rendimiento_PA", "Rendimiento_Penal concedido"]]

# Cambiar el nombre de las columnas
dfa_2324 = dfa_2324.rename(columns = {
    "Unnamed: 0_level_0_Equipo": "Equipo",
    "Rendimiento_TA": "TA_a",
    "Rendimiento_TR": "TR_a",
    "Rendimiento_Fls": "Fls_a",
    "Rendimiento_PA": "FdJ_a",
    "Rendimiento_Penal concedido": "PC_a"
})

dfa_2324.head()

## **Faltas cometidas a los equipos 23/24**

dfb_2324 = pd.read_html("https://fbref.com/es/comps/12/2023-2024/Estadisticas-2023-2024-
La-Liga",
                        attrs = {"id":"stats_squads_misc_against"}) [0]

dfb_2324.columns = ["_".join(col).strip() for col in dfb_2324.columns.values]

dfb_2324["Unnamed: 0_level_0_Equipo"] = dfb_2324["Unnamed: 0_level_0_Equipo"] + "_" +
"2324"

dfb_2324 = dfb_2324[["Unnamed: 0_level_0_Equipo", "Rendimiento_TA", "Rendimiento_TR",
"Rendimiento_Fls", "Rendimiento_PA", "Rendimiento_Penal concedido"]]

```

```

dfb_2324 = dfb_2324.rename(columns = {
    "Unnamed: 0_level_0_Equipo": "Equipo",
    "Rendimiento_TA": "TA_b",
    "Rendimiento_TR": "TR_b",
    "Rendimiento_Fls": "Fls_b",
    "Rendimiento_PA": "FdJ_b",
    "Rendimiento_Penal concedido": "PC_b"
})

# Quitar el "vs. " de todas las observaciones en la columna "Equipo" para poder unificar
# las filas posteriormente
dfb_2324["Equipo"] = dfb_2324["Equipo"].str.replace("vs. ", "", regex = False)
dfb_2324.head()

## **Tabla LaLiga 22/23**

df0_2223 = pd.read_html("https://fbref.com/es/comps/12/2022-2023/Estadisticas-2022-2023-
La-Liga",
                        attrs = {"id": "results2022-2023121_overall"}) [0]
df0_2223["Equipo"] = df0_2223["Equipo"] + "_" + "2223"
df0_2223 = df0_2223[["RL", "Equipo"]]
df0_2223.head()

## **Faltas cometidas por los equipos 22/23**

dfa_2223 = pd.read_html("https://fbref.com/es/comps/12/2022-2023/Estadisticas-2022-2023-
La-Liga",
                        attrs = {"id": "stats_squads_misc_for"}) [0]
dfa_2223.columns = ["_".join(col).strip() for col in dfa_2223.columns.values]
dfa_2223["Unnamed: 0_level_0_Equipo"] = dfa_2223["Unnamed: 0_level_0_Equipo"] + "_" +
"2223"
dfa_2223 = dfa_2223[["Unnamed: 0_level_0_Equipo", "Rendimiento_TA", "Rendimiento_TR",
"Rendimiento_Fls", "Rendimiento_PA", "Rendimiento_Penal concedido"]]
dfa_2223 = dfa_2223.rename(columns = {
    "Unnamed: 0_level_0_Equipo": "Equipo",
    "Rendimiento_TA": "TA_a",
    "Rendimiento_TR": "TR_a",
    "Rendimiento_Fls": "Fls_a",
    "Rendimiento_PA": "FdJ_a",
    "Rendimiento_Penal concedido": "PC_a"
})

```

```

})
dfa_2223.head()

## **Faltas cometidas a los equipos 22/23**

dfb_2223 = pd.read_html("https://fbref.com/es/comps/12/2022-2023/Estadisticas-2022-2023-
La-Liga",
                        attrs = {"id": "stats_squads_misc_against"}) [0]
dfb_2223.columns = ["_".join(col).strip() for col in dfb_2223.columns.values]
dfb_2223["Unnamed: 0_level_0_Equipo"] = dfb_2223["Unnamed: 0_level_0_Equipo"] + "_" +
"2223"
dfb_2223 = dfb_2223[["Unnamed: 0_level_0_Equipo", "Rendimiento_TA", "Rendimiento_TR",
"Rendimiento_Fls", "Rendimiento_PA", "Rendimiento_Penal concedido"]]
dfb_2223 = dfb_2223.rename(columns = {
    "Unnamed: 0_level_0_Equipo": "Equipo",
    "Rendimiento_TA": "TA_b",
    "Rendimiento_TR": "TR_b",
    "Rendimiento_Fls": "Fls_b",
    "Rendimiento_PA": "FdJ_b",
    "Rendimiento_Penal concedido": "PC_b"
})
dfb_2223["Equipo"] = dfb_2223["Equipo"].str.replace("vs. ", "", regex = False)
dfb_2223.head()

## **Tabla LaLiga 21/22**

df0_2122 = pd.read_html("https://fbref.com/es/comps/12/2021-2022/Estadisticas-2021-2022-
La-Liga",
                        attrs = {"id": "results2021-2022121_overall"}) [0]
df0_2122["Equipo"] = df0_2122["Equipo"] + "_" + "2122"
df0_2122 = df0_2122[["RL", "Equipo"]]
df0_2122.head()

## **Faltas cometidas por los equipos 21/22**

dfa_2122 = pd.read_html("https://fbref.com/es/comps/12/2021-2022/Estadisticas-2021-2022-
La-Liga",
                        attrs = {"id": "stats_squads_misc_for"}) [0]
dfa_2122.columns = ["_".join(col).strip() for col in dfa_2122.columns.values]

```

```

dfa_2122["Unnamed: 0_level_0_Equipo"] = dfa_2122["Unnamed: 0_level_0_Equipo"] + "_" +
"2122"
dfa_2122 = dfa_2122[["Unnamed: 0_level_0_Equipo", "Rendimiento_TA", "Rendimiento_TR",
"Rendimiento_Fls", "Rendimiento_PA", "Rendimiento_Penal concedido"]]
dfa_2122 = dfa_2122.rename(columns = {
    "Unnamed: 0_level_0_Equipo": "Equipo",
    "Rendimiento_TA": "TA_a",
    "Rendimiento_TR": "TR_a",
    "Rendimiento_Fls": "Fls_a",
    "Rendimiento_PA": "FdJ_a",
    "Rendimiento_Penal concedido": "PC_a"
})
dfa_2122.head()

## **Faltas cometidas a los equipos 21/22**

dfb_2122 = pd.read_html("https://fbref.com/es/comps/12/2021-2022/Estadisticas-2021-2022-
La-Liga",
                        attrs = {"id": "stats_squads_misc_against"}) [0]
dfb_2122.columns = ["_".join(col).strip() for col in dfb_2122.columns.values]
dfb_2122["Unnamed: 0_level_0_Equipo"] = dfb_2122["Unnamed: 0_level_0_Equipo"] + "_" +
"2122"
dfb_2122 = dfb_2122[["Unnamed: 0_level_0_Equipo", "Rendimiento_TA", "Rendimiento_TR",
"Rendimiento_Fls", "Rendimiento_PA", "Rendimiento_Penal concedido"]]
dfb_2122 = dfb_2122.rename(columns = {
    "Unnamed: 0_level_0_Equipo": "Equipo",
    "Rendimiento_TA": "TA_b",
    "Rendimiento_TR": "TR_b",
    "Rendimiento_Fls": "Fls_b",
    "Rendimiento_PA": "FdJ_b",
    "Rendimiento_Penal concedido": "PC_b"
})
dfb_2122["Equipo"] = dfb_2122["Equipo"].str.replace("vs. ", "", regex = False)
dfb_2122.head()

## **Descarga**

# Crear listas para cada df
df0_lista = [df0_2324, df0_2223, df0_2122]
dfa_lista = [dfa_2324, dfa_2223, dfa_2122]
dfb_lista = [dfb_2324, dfb_2223, dfb_2122]

```

```

# Concatenar todos los dfs de cada tipo en un sólo df
df0 = pd.concat(df0_lista, ignore_index=True)
dfa = pd.concat(dfa_lista, ignore_index=True)
dfb = pd.concat(dfb_lista, ignore_index=True)

# Unir dfa y dfb en un sólo df con sufijos diferenciadores
df_faltas = dfa.merge(dfb, on="Equipo", how="outer", suffixes=("_a", "_b"))

# Unir con df0 para recuperar RL
df_final = df0.merge(df_faltas, on="Equipo", how="outer")
print(df_final.head())

# Transformar en formato CSV
df_final.to_csv("df_final.csv")

# Exportar el archivo
from google.colab import files
files.download("df_final.csv")

```

### 3. Código de los algoritmos de *Machine Learning* (R)

```

# Pasos previos
remove(list = ls())
setwd(dirname(rstudioapi::getActiveDocumentContext()$path))
getwd()

# Cargar los datos
library(readr)
df2 = read_csv("BD2.csv")

# Revisar los datos
View(df2)
summary(df2)
str(df2)

### PREPARACIÓN DE LOS DATOS

# Crear un nuevo df eliminando las 3 primeras columnas

```

```

df2_tipificado = df2[, -(1:3)]

# Estandarizar las variables
df2_tipificado = as.data.frame(scale(df2_tipificado))

# Confirmar que han sido correctamente estandarizadas
summary(df2_tipificado)
str(df2_tipificado)

# Verificar que no hay NAs
sum(is.na(df2_tipificado))

# Calcular la matriz de correlación
cor_matrix = cor(df2_tipificado, use = "pairwise.complete.obs")
print(cor_matrix)

# Visualizar la matriz de correlación
library(corrplot)
corrplot(cor_matrix, method = "color", type = "upper", tl.col = "black", tl.srt = 50)

# Examinar la correlación parcial entre variables
library(psych)
KMO(cor_matrix)

# Convertir el df a una matriz
df2_tipificado = as.matrix(df2_tipificado)
summary(df2_tipificado)

### CLUSTERING

# Obtener número óptimo de clústeres
library(NbClust)
NbClust(df2_tipificado, diss = NULL, distance = "euclidean",
        min.nc = 2, max.nc = 15, method = "ward.D2")

# Generar un mapa de calor
library("gplots")
heatmap.2(x = df2_tipificado, scale = "none", # no se escalan los datos porque ya están
          estandarizados
          distfun = function(x){dist(x, method = "euclidean")},
          hclustfun = function(x){hclust(x, method = "ward.D2")},
          trace = "none",

```

```

col = bluered(15),
margins = c(8,3),
cexRow = 0.1,
cexCol = 0.8,
labCol = NULL,
Rowv = TRUE, # agrupar los equipos
labRow = rep("", nrow(df2_tipificado))) # eliminar las etiquetas de las filas
(para que no se solapen)

# Añadir leyenda
mtext("Las variables terminadas en '_a' indican acciones cometidas por el equipo,
mientras que aquellas terminadas en '_b' indican acciones sufridas por el equipo
rival.",
side = 1, line = 6.5, adj = 1.3, cex = 0.8)

# Crear el clustering jerárquico
res.dist = dist(df2_tipificado, method = "euclidean")
res.hc = hclust(d = res.dist, method = "ward.D2")
HC = cutree(res.hc, k = 2)
table(HC)

# Añadir el clustering al df original
df2$HC = HC
df2$HC = as.factor(df2$HC)
summary(df2)
View(df2)

# Calcular la media de cada variable según el clúster
aggregate(. ~ HC, data = df2[, c("HC", "TA_a", "TR_a", "Fls_a", "FdJ_a", "PC_a")], mean)

# Visualizar la distribución de los clústeres
library(ggplot2)
ggplot(df2, aes(x = TA_a, y = TR_a, color = as.factor(HC))) +
  geom_point() +
  labs(title = "Distribución de clústeres según tarjetas amarillas y rojas",
       color = "Clúster") +
  xlab("Tarjetas amarillas") +
  ylab("Tarjetas rojas")

# Gráfico de dispersión 3D
install.packages("plotly")
library(plotly)
fig = plot_ly(df2, x = ~TA_a, y = ~TR_a, z = ~Fls_a, color = ~as.factor(HC),

```

```

        type = "scatter3d", mode = "markers")
fig

# Comparar rendimiento entre clústeres
tapply(df2$RL, df2$HC, mean)

# Realizar ANOVA
anova_result = aov(RL ~ HC, data = df2)
summary(anova_result)

### LOGIT

# Transformar la variable RL en binaria
df2$RL_binario = ifelse(df2$RL <= median(df2$RL), 1, 0)

# Primer logit: pertenencia a un clúster
logit_HC = glm(RL_binario ~ HC,
              family = binomial(), data = df2)
summary(logit_HC)

# Segundo logit: variables individuales
logit_variables = glm(RL_binario ~ TA_a + TR_a + Fls_a + FdJ_a + PC_a +
                    TA_b + TR_b + Fls_b + FdJ_b + PC_b,
                    family = binomial(), data = df2)
summary(logit_variables)

# Tercer logit: combinación de ambos
logit_total = glm(RL_binario ~ HC + TA_a + TR_a + Fls_a + FdJ_a + PC_a +
                TA_b + TR_b + Fls_b + FdJ_b + PC_b,
                family = binomial(), data = df2)
summary(logit_total)

# Comparar los tres modelos
AIC(logit_HC, logit_variables, logit_total)

# Evaluar el modelo
library(pROC)
predicted_probs = predict(logit_HC, type = "response")
roc_curve = roc(df2$RL_binario, predicted_probs)
plot(roc_curve, main = "Curva ROC del modelo logit", col = "blue", legacy.axes = TRUE,
     xlab = "", ylab = "") # las últimas funciones son para eliminar los títulos en inglés
title(xlab = "Especificidad", ylab = "Sensibilidad") # poner los títulos en español

```

```

auc(roc_curve)

### ÁRBOL DE CLASIFICACIÓN

# Crear árbol de clasificación
library(rpart)
tree_model = rpart(RL_binario ~ HC + TA_a + TR_a + Fls_a + FdJ_a + PC_a +
                  TA_b + TR_b + Fls_b + FdJ_b + PC_b,
                  data = df2, method = "class")

# Visualizar el árbol de clasificación
library(rpart.plot)
rpart.plot(tree_model, main = "Árbol de clasificación para el ranking de equipos")

# Ver importancia de cada variable
tree_model$variable.importance

### ANÁLISIS COMPARATIVO: RMA, ATM Y FCB

# Filtrar por Real Madrid, Atlético de Madrid y FC Barcelona
df2_top3 = df2[grep("Real Madrid|Atlético Madrid|Barcelona", df2$Equipo), ]
df2_top3

# Visualizar gráfico de barras
library(reshape2)
df2_melt = melt(df2_top3,
               id.vars = c("Equipo", "RL"),
               measure.vars = c("TA_a", "TR_a", "Fls_a", "FdJ_a", "PC_a")) #
reorganizar el df
df2_melt
ggplot(df2_melt, aes(x = variable, y = value, fill = Equipo)) +
  geom_bar(stat = "identity", position = "dodge") +
  labs(title = "Comparación conjunta de variables arbitrales entre equipos",
       x = "Métrica",
       y = "Valor") +
  scale_x_discrete(labels = c(
    "PC_a" = "Penaltis",
    "TA_a" = "Tarjetas amarillas",
    "TR_a" = "Tarjetas rojas",
    "Fls_a" = "Faltas",

```

```

    "FdJ_a" = "Fueras de juego")) +
  theme(axis.text.x = element_text(angle = 0, vjust = 0.5, size = 8),
        axis.title.x = element_text(margin = margin(t = 10)),
        axis.text.y = element_text(angle = 0, vjust = 0.5, size = 8),
        axis.title.y = element_text(margin = margin(t = 10))) # axis.text para ajustar
los subtítulos del eje; axis.title para aumentar la separación entre el título y los
subtítulos

# Visualizar gráfico de barras por variable
df2_melt$variable = factor(df2_melt$variable,
                          levels = c("PC_a", "TA_a", "TR_a", "Fls_a", "FdJ_a"),
                          labels = c("Penaltis", "Tarjetas amarillas", "Tarjetas
rojas", "Faltas", "Fueras de juego")) # cambiar los nombres de las variables
ggplot(df2_melt, aes(x = Equipo, y = value, fill = Equipo)) +
  geom_bar(stat = "identity", position = position_dodge(), width = 0.6) +
  facet_wrap(~ variable, scales = "free_y") +
  labs(title = "Comparación dividida de variables arbitrales entre equipos",
       x = NULL, y = "Valor", fill = "Equipo") +
  theme_minimal() +
  theme(
    axis.text.x = element_blank(), # quitar los nombres del eje X
    axis.ticks.x = element_blank(), # quitar las marcas del eje X
    axis.title.y = element_text(margin = margin(t = 10)),
    strip.text = element_text(size = 9)) # ajustar el tamaño de los títulos de los
subgráficos

```