

Modelado Estadístico del Rendimiento Futbolístico: Un Enfoque Bayesiano Aplicado a Millonarios F.C. en la Temporada 2023-I de la Liga BetPlay

Statistical Modeling of Football Performance: A Bayesian Approach Applied to Millonarios F.C. in the 2023-I BetPlay League Season

Miguel Angel Barrera Moreno^a
mibarreram@unal.edu.co

Nikolle Ramirez Herrera^b
niramirez@unal.edu.co

Julian Hernando Lombo Beltran^c
jlombo@unal.edu.co

Danna Lesley Cruz Reyes^d
dlcruzr@unal.edu.co

Resumen

El presente estudio aborda el análisis del rendimiento futbolístico a través de un enfoque estadístico, aplicando modelos frecuentistas y bayesianos al caso particular de Millonarios F.C. y Atlético Nacional en la temporada 2023–I de la Liga BetPlay. Se desarrolla un modelo de regresión de Poisson como punto de partida y, posteriormente, se plantea un modelo bayesiano jerárquico con efectos aleatorios para estimar parámetros de ataque, defensa y ventaja de localía. Los resultados permiten comparar el comportamiento ofensivo y defensivo de ambos equipos, generar rankings probabilísticos y ofrecer herramientas predictivas útiles en contextos de análisis deportivo.

Palabras clave: Modelos jerárquicos, regresión de Poisson, fútbol colombiano, análisis bayesiano, ventaja de localía.

Abstract

This study focuses on the statistical modeling of football performance, comparing frequentist and Bayesian approaches applied to Millonarios F.C. and Atlético Nacional during the 2023-I BetPlay League season. A Poisson regression model is first developed, followed by a hierarchical Bayesian model with random effects to estimate attack, defense, and home advantage parameters. The results allow for a comparative assessment of both teams' offensive and defensive behavior, probabilistic rankings, and predictive tools relevant for sports analytics.

Keywords: Hierarchical models, Poisson regression, Colombian football, Bayesian analysis, home advantage.

1. Introducción

En las últimas décadas, el modelado estadístico del rendimiento deportivo ha cobrado gran relevancia como herramienta de análisis cuantitativo en disciplinas altamente competitivas, siendo el fútbol profesional una de las más estudiadas. La disponibilidad de registros detallados, el aumento en las capacidades computacionales

^aUniversidad Nacional de Colombia

^bUniversidad Nacional de Colombia

^cUniversidad Nacional de Colombia

^dDepartamento de Estadística, Universidad Nacional de Colombia, Bogotá, Colombia

y el avance de la inferencia estadística han impulsado el desarrollo de modelos cada vez más rigurosos, que permiten no solo describir el comportamiento histórico de los equipos, sino también generar predicciones sobre el resultado de futuros encuentros (Bishop, 2006), (Groll et al., 2018), (Zuccolotto and Manisera, 2020).

A nivel internacional, se han propuesto diversos enfoques para abordar el análisis de datos futbolísticos. Tradicionalmente, la distribución de Poisson ha sido ampliamente utilizada para modelar el número de goles anotados por cada equipo, bajo el supuesto de independencia entre anotaciones del local y el visitante. Sin embargo, esta simplificación ha sido superada por modelos jerárquicos más sofisticados, que permiten capturar estructuras latentes de rendimiento ofensivo y defensivo, así como la ventaja de jugar en casa (Baio and Blangiardo, 2010), (Joseph et al., 2022).

En este contexto, Baio y Blangiardo (Baio and Blangiardo, 2010) propusieron un modelo jerárquico bayesiano que incorpora efectos aleatorios para ataque, defensa y localía, logrando una representación flexible del comportamiento de los equipos en la Serie A italiana. Su propuesta marcó un punto de referencia en el análisis bayesiano de datos deportivos, destacándose tanto por su estructura inferencial como por su capacidad predictiva. Otros autores han extendido este enfoque con redes bayesianas (Joseph et al., 2022), regresiones estructuradas (Goddard, 2005), o modelos bivariados que capturan la correlación entre goles de local y visitante (Karlis and Ntzoufras, 2003), (Karlis and Ntzoufras, 2009), (Koopman and Lit, 2015).

A pesar del crecimiento sostenido de esta línea metodológica en el análisis de ligas europeas, su aplicación en contextos latinoamericanos ha sido limitada. En particular, el fútbol colombiano presenta características competitivas relevantes que aún no han sido exploradas con enfoques bayesianos estructurados, lo que representa una oportunidad para contribuir al desarrollo de la analítica deportiva en la región.

El presente estudio tiene como propósito adaptar e implementar un modelo jerárquico bayesiano al análisis del rendimiento de dos equipos representativos de la liga colombiana: Millonarios F.C. y Atlético Nacional, durante la temporada 2023-I. Estos equipos han sido históricamente protagonistas del torneo, y sus actuaciones constituyen una fuente valiosa de información para evaluar metodologías estadísticas aplicadas al deporte.

Además de evaluar el desempeño mediante un enfoque frecuentista basado en regresión de Poisson, se propone un modelo bayesiano que permite estimar simultáneamente los efectos ofensivos, defensivos y de localía para cada equipo. Estos efectos se consideran latentes, es decir, no observables directamente en los datos. Desde una perspectiva de modelado probabilístico clásico (Bishop, 2006), la incorporación de efectos aleatorios resulta adecuada para capturar esta estructura no directamente observable.

Esta estrategia no solo incorpora la incertidumbre asociada a las estimaciones, sino que también posibilita comparaciones probabilísticas entre equipos y la formulación de predicciones basadas en distribuciones posteriores (Gelman et al., 2013), (McElreath, 2020).

El trabajo se inscribe dentro de una línea de investigación aplicada que busca fortalecer la analítica deportiva en América Latina, aportando herramientas metodológicas robustas y reproducibles que puedan ser empleadas en escenarios reales de toma de decisiones en el ámbito deportivo.

2. Descripción del conjunto de datos

La información utilizada en este estudio fue recopilada a partir de la plataforma especializada **colombia.as**, la cual ofrece datos detallados y actualizados sobre los encuentros disputados en la Liga BetPlay, primera división del fútbol profesional colombiano. Con base en esta fuente, se construyó una base de datos estructurada que permitió realizar tanto el análisis descriptivo como la implementación de los modelos estadísticos propuestos.

El conjunto de datos comprende exclusivamente los partidos en los que participaron Millonarios F.C. y Atlético Nacional durante la temporada 2023-I, independientemente de si actuaron como locales o visitantes. Para cada encuentro se registraron variables clave como el identificador del partido, el nombre de los equipos involucrados, su rol en el juego (local o visitante), y el número de goles anotados por cada uno. Asimismo,

mo, se incluyó un identificador numérico para cada equipo, el cual facilita la codificación en el marco del modelado jerárquico. Finalmente, se incorporó la jornada correspondiente del torneo, permitiendo ordenar cronológicamente los partidos y explorar posibles efectos asociados a la progresión de la competencia.

3. Modelo Bayesiano

Con el objetivo de incorporar las características ofensivas y defensivas específicas de cada equipo, así como el efecto de jugar como local, se propone un modelo jerárquico bayesiano con efectos aleatorios. Este modelo sigue la formulación log-lineal por (Gelman et al., 2013), (Bishop, 2006), adecuada para conteos de goles en contextos deportivos. Este tipo de modelo ha sido previamente utilizado con éxito en el análisis bayesiano de resultados futbolísticos, como en Baio y Blangiardo (Baio and Blangiardo, 2010), quienes aplicaron una estructura jerárquica similar para predecir marcadores de partidos en ligas europeas.

$$\begin{aligned}\log \theta_{g1} &= home + att_{h(g)} + def_{a(g)}, \\ \log \theta_{g2} &= att_{a(g)} + def_{h(g)},\end{aligned}$$

donde θ_{g1} y θ_{g2} representan las tasas esperadas de goles del equipo local y visitante, respectivamente, en el partido g . Las observaciones y_{g1} y y_{g2} corresponden al número real de goles anotados por los equipos en dicho encuentro, modelados como variables Poisson condicionales:

$$y_{gj} \mid \theta_{gj} \sim \text{Poisson}(\theta_{gj}), \quad j = 1, 2.$$

Aquí, *home* capta el efecto sistemático de la localía; los términos att_t y def_t representan los efectos aleatorios de ataque y defensa del equipo t ; y las funciones $h(g)$ y $a(g)$ indican qué equipos jugaron como local y visitante, respectivamente, en el partido g .

Para completar la especificación bayesiana, se asignan distribuciones previas a cada uno de los parámetros del modelo:

$$\begin{aligned}home &\sim \mathcal{N}(0, 0.0001), \\ att_t &\sim \mathcal{N}(\mu_{att}, \tau_{att}^2), & def_t &\sim \mathcal{N}(\mu_{def}, \tau_{def}^2),\end{aligned}$$

donde $t = 1, \dots, T$ y T es el número total de equipos. Para garantizar la identificabilidad del modelo, se imponen restricciones de centrado:

$$\sum_{t=1}^T att_t = 0, \quad \sum_{t=1}^T def_t = 0.$$

Finalmente, las hiperpriors para los parámetros de ataque y defensa son:

$$\begin{aligned}\mu_{att} &\sim \mathcal{N}(0, 0.0001), & \mu_{def} &\sim \mathcal{N}(0, 0.0001), \\ \tau_{att}^2 &\sim \text{Inv-Gamma}(0.1, 0.1), & \tau_{def}^2 &\sim \text{Inv-Gamma}(0.1, 0.1).\end{aligned}$$

4. Resultados

En la Figura 1 se presentan las frecuencias absolutas de goles anotados por los equipos Millonarios y Nacional, diferenciando entre los partidos jugados como locales y como visitantes. Este gráfico proporciona una primera aproximación visual al rendimiento ofensivo condicionado por la localía.

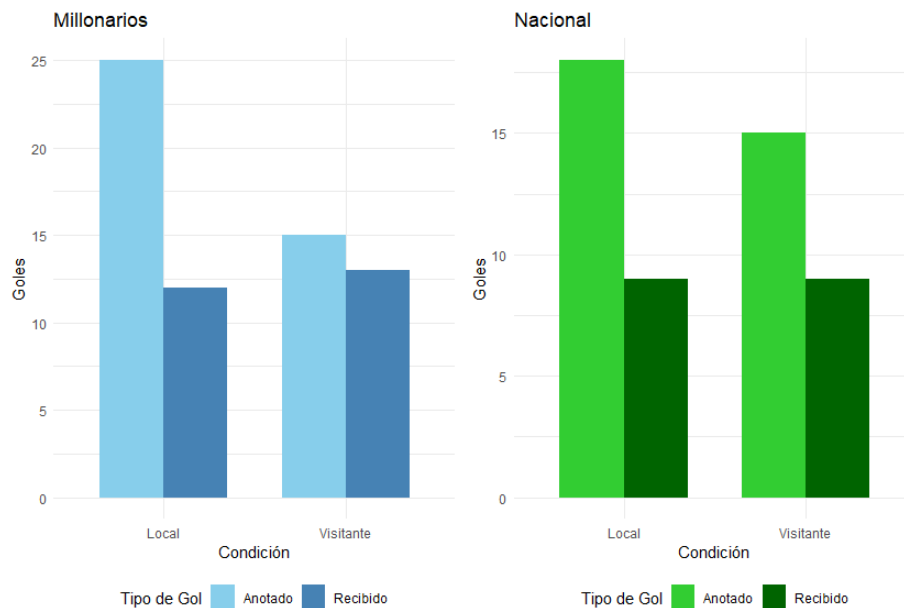


Figura 1: Goles anotados por Millonarios y Nacional según condición de local o visitante.

En el caso de Millonarios, se observa un rendimiento ofensivo claramente superior en condición de local, evidenciando una mayor capacidad anotadora frente a su desempeño como visitante. Esta diferencia sugiere un aprovechamiento significativo del factor localía. En cuanto a la defensa, el equipo muestra cierta estabilidad en la cantidad de goles recibidos, con una ligera tendencia a permitir más goles cuando juega fuera de casa.

Atlético Nacional, por su parte, presenta un comportamiento más uniforme entre localía y visita. Aunque también anota más goles como local, la diferencia con su rendimiento como visitante es menos marcada que en el caso de Millonarios. En términos defensivos, Nacional muestra una frecuencia similar de goles recibidos en ambas condiciones.

Estos patrones sugieren que ambos equipos tienden a maximizar su rendimiento ofensivo en condición de local, aunque Millonarios presenta una variabilidad más acentuada en su producción de goles. Aun así, el equipo mantiene cierta regularidad en defensa. Nacional, en contraste, exhibe una mayor estabilidad en ambos frentes.

La Figura 2 muestra histogramas que describen la frecuencia de goles anotados y recibidos por partido para ambos equipos. Millonarios destaca por una mayor concentración de partidos en los que anota entre 1 y 2 goles, independientemente de la localía. En defensa, su rendimiento mejora notoriamente cuando juega en casa. Nacional, por otro lado, también anota con frecuencia 1 gol por partido, pero su comportamiento muestra mayor variabilidad, especialmente como visitante. Defensivamente, exhibe una mayor proporción de partidos con cero o un gol recibido en casa, lo que respalda su solidez como local.

Las estimaciones posteriores obtenidas mediante inferencia bayesiana permiten cuantificar el rendimiento ofensivo y defensivo de cada equipo. En la Tabla 1 se presentan las medias posteriores para los efectos de ataque y defensa estimados para cada equipo que participó en la Liga BetPlay 2023-I. Estas estimaciones representan los efectos aleatorios obtenidos a partir del modelo jerárquico descrito previamente, centrados alrededor del promedio de la liga.

Además, el parámetro asociado al efecto de localía fue estimado con una media posterior de 0.245, con un intervalo de credibilidad del 95 % entre 0.127 y 0.357. Este resultado indica una ventaja significativa de jugar como local, ya que dicho efecto incrementa sistemáticamente la tasa de anotación del equipo local respecto al visitante.

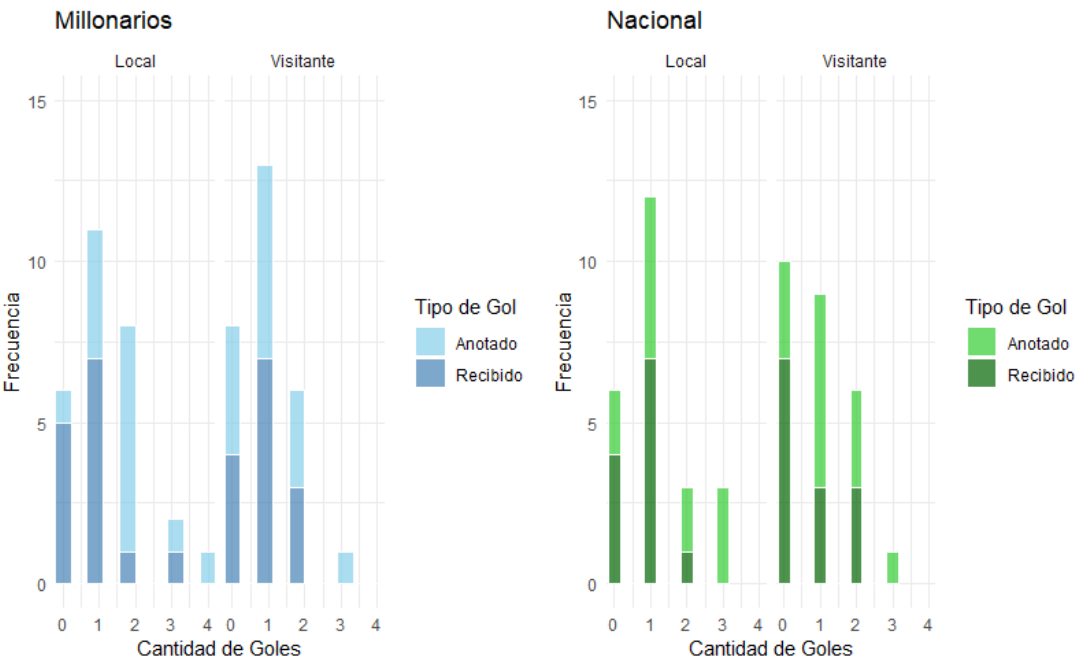


Figura 2: Distribución de goles anotados y recibidos por Millonarios y Nacional, según condición.

Código	Equipo	Ataque	Defensa
1	Atlético Bucaramanga	-0.308	-0.054
2	Unión Magdalena	-0.249	0.295
3	Rionegro Águilas	0.268	0.102
4	Chicó	0.055	-0.215
5	Tolima	0.051	0.187
6	Nacional	0.094	-0.495
7	Jaguars FC	-0.264	0.145
8	La Equidad	-0.069	-0.218
9	Deportivo Cali	-0.060	0.138
10	Millonarios	0.268	-0.247
11	Envigado	-0.399	-0.091
12	Huila	-0.043	0.445
13	Deportivo Pereira	0.095	0.157
14	Junior	-0.222	-0.221
15	Once Caldas	-0.205	0.085
16	América de Cali	0.335	-0.011
17	Alianza FC	0.215	-0.066
18	Santa Fe	0.284	0.142
19	Medellín	0.153	0.057
20	Pasto	0.001	-0.135

Tabla 1: Medias posteriores para habilidades de equipo

4.1. Análisis del Rendimiento Ofensivo

La Tabla 2 presenta las estimaciones posteriores del parámetro de ataque junto con sus intervalos de credibilidad del 95 %. Estos resultados permiten identificar los equipos con mejor desempeño ofensivo a lo largo

de la temporada.

Tabla 2: Ranking de Ataque - Estimaciones Bayesianas

Equipo	Parámetros de Ataque		
	Estimación	LI (2.5 %)	LS (97.5 %)
América de Cali	0.335	0.029	0.618
Santa Fe	0.284	-0.078	0.616
Rionegro Águilas	0.268	-0.053	0.568
Millonarios	0.268	-0.041	0.559
Alianza FC	0.215	-0.109	0.520
Medellín	0.153	-0.186	0.478
Deportivo Pereira	0.095	-0.278	0.455
Nacional	0.094	-0.247	0.407
Chicó	0.055	-0.296	0.388
Tolima	0.051	-0.352	0.411
Pasto	0.001	-0.361	0.334
Huila	-0.043	-0.447	0.335
Deportivo Cali	-0.060	-0.476	0.328
La Equidad	-0.069	-0.486	0.322
Once Caldas	-0.205	-0.652	0.197
Junior	-0.222	-0.673	0.187
Unión Magdalena	-0.249	-0.707	0.174
Jaguars FC	-0.264	-0.712	0.168
Atlético Bucaramanga	-0.308	-0.784	0.135
Envigado	-0.399	-0.890	0.043

Entre los resultados más destacados se encuentra América de Cali, cuya estimación de ataque (0.335) presenta un intervalo de credibilidad completamente por encima de cero, lo que indica evidencia fuerte de un rendimiento ofensivo superior al promedio. Millonarios, Santa Fe y Rionegro Águilas también presentan estimaciones positivas, aunque sus intervalos incluyen el valor cero, lo cual limita la certeza estadística de su superioridad ofensiva.

La Figura 3 complementa este análisis al mostrar visualmente las estimaciones y los intervalos de credibilidad del ataque por equipo. Los equipos con puntos verdes son aquellos cuya media es positiva; los puntos rojos indican estimaciones negativas.

4.2. Análisis del Rendimiento Defensivo

La Tabla 3 reporta las estimaciones del parámetro de defensa. En este caso, valores más negativos corresponden a una mayor solidez defensiva, es decir, menor propensión a recibir goles.

Atlético Nacional se destaca como el equipo más sólido defensivamente, con una estimación de -0.495 y un intervalo de credibilidad completamente negativo. Le siguen equipos como Millonarios, La Equidad y Junior, aunque sus intervalos cruzan el cero, lo cual introduce mayor incertidumbre.

La Figura 4 presenta los intervalos de credibilidad para las estimaciones de defensa. De forma análoga al gráfico anterior, los colores representan la dirección del efecto estimado.

4.3. Evaluación Conjunta: Ataque y Defensa

Más allá de los rankings univariados, resulta útil analizar conjuntamente la intensidad ofensiva y defensiva de los equipos. La Figura 5 muestra la posición relativa de cada equipo en el plano definido por sus efectos

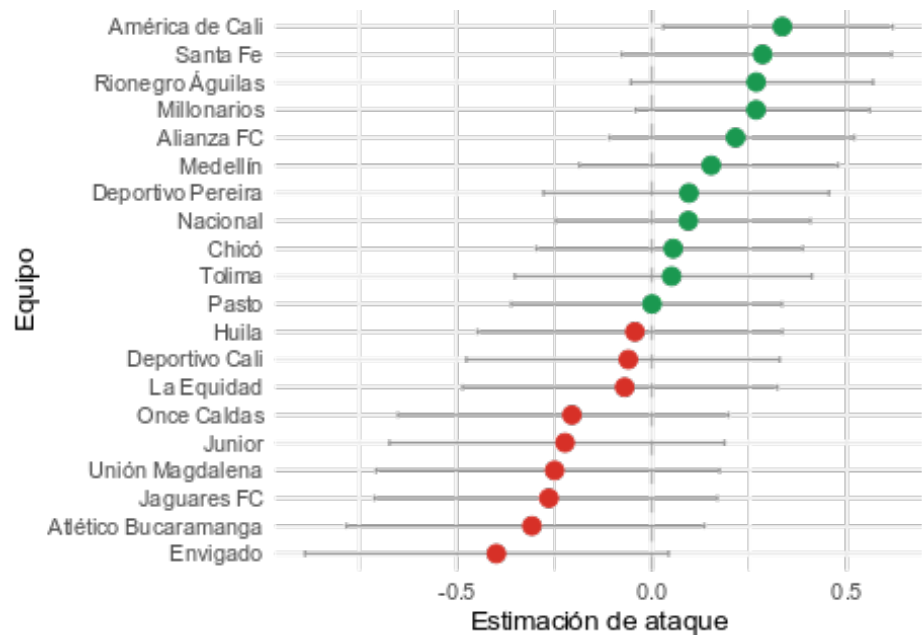


Figura 3: Intervalos de Credibilidad para los Parámetros de Ataque

Equipo	Parámetros de Defensa		
	Estimación	LI (2.5 %)	LS (97.5 %)
Nacional	-0.495	-0.906	-0.102
Millonarios	-0.247	-0.638	0.106
Junior	-0.221	-0.667	0.190
La Equidad	-0.218	-0.670	0.204
Chicó	-0.215	-0.610	0.137
Pasto	-0.135	-0.529	0.213
Envigado	-0.091	-0.519	0.301
Alianza FC	-0.066	-0.440	0.291
Atlético Bucaramanga	-0.054	-0.466	0.320
América de Cali	-0.011	-0.372	0.330
Medellín	0.057	-0.272	0.365
Once Caldas	0.085	-0.299	0.438
Rionegro Águilas	0.102	-0.233	0.419
Deportivo Cali	0.138	-0.245	0.480
Santa Fe	0.142	-0.248	0.483
Jaguares FC	0.145	-0.219	0.487
Deportivo Pereira	0.157	-0.218	0.512
Tolima	0.187	-0.182	0.525
Unión Magdalena	0.295	-0.064	0.619
Huila	0.445	0.111	0.755

Tabla 3: Ranking de Defensa - Estimaciones Bayesianas

centrados de ataque (eje horizontal) y defensa (eje vertical). En este gráfico, Millonarios y Nacional aparecen resaltados en azul y verde, respectivamente, reflejando su perfil balanceado y competitivo.

Este tipo de representación gráfica permite identificar tanto a los equipos más completos (como Nacional y Millonarios), como a aquellos con desequilibrios importantes entre ataque y defensa. Por ejemplo, América de Cali se sitúa entre los mejores ataques, mientras que Envigado y Jaguares FC aparecen rezagados en

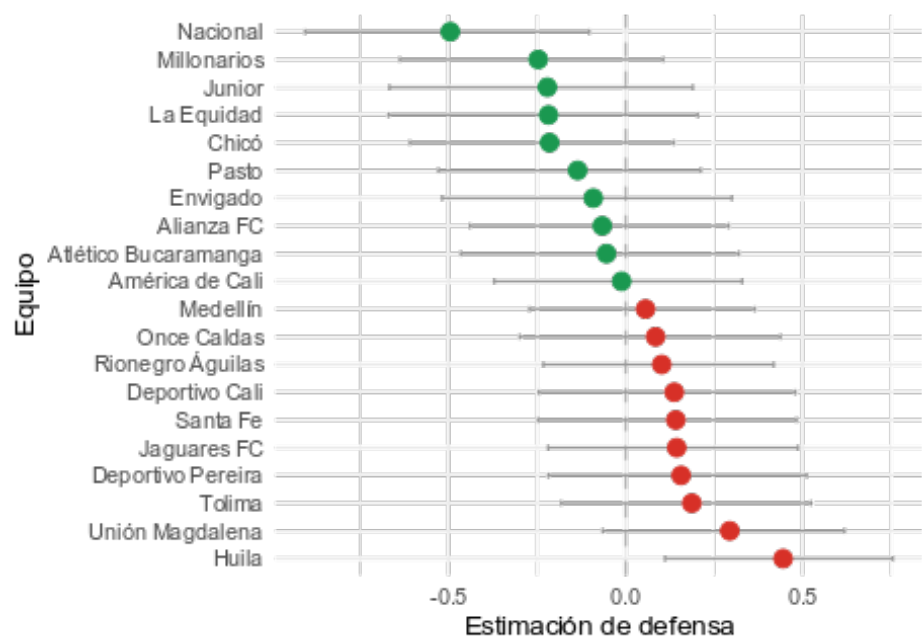


Figura 4: Intervalos de Credibilidad para los Parámetros de Defensa

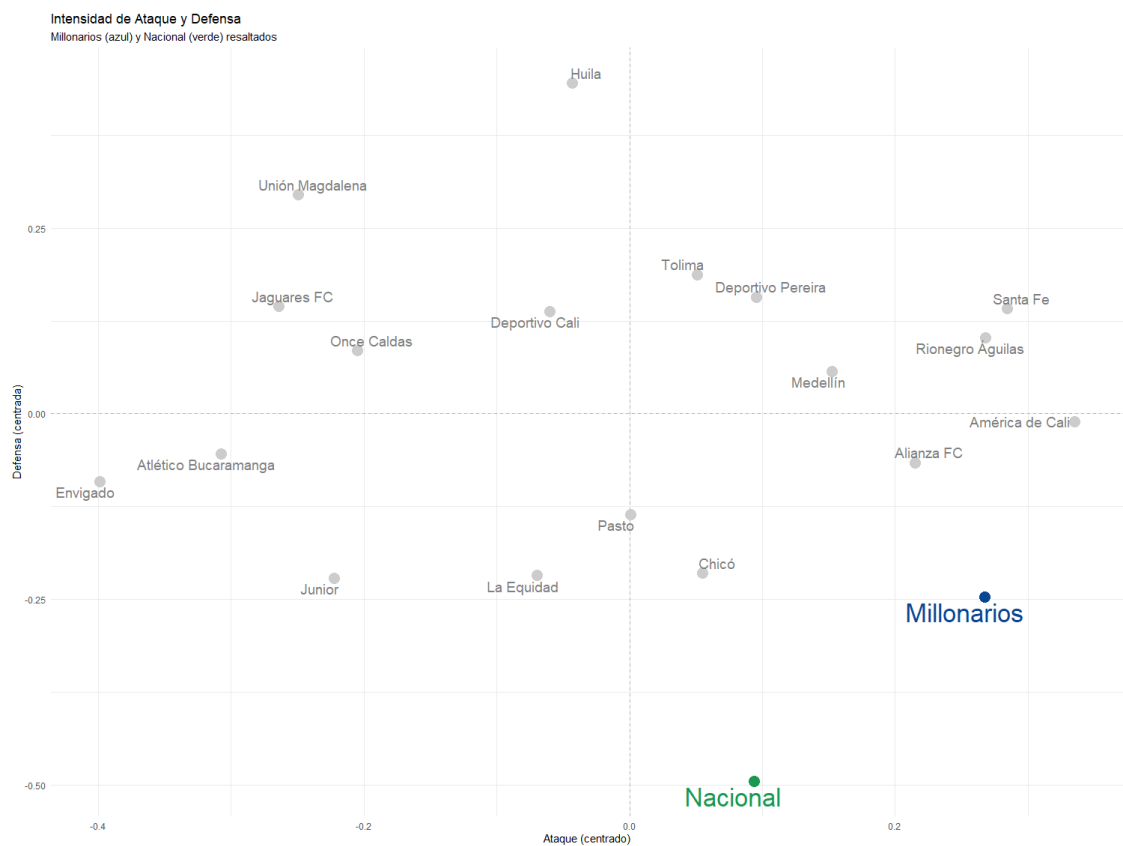


Figura 5: Intensidad de Ataque y Defensa de los equipos, con Millonarios (azul) y Nacional (verde) resaltados. Los ejes representan las métricas centradas con respecto al promedio de la liga.

ambos aspectos.

Finalmente, las Tablas 4 y 5 presentan un resumen de los equipos que destacaron por su rendimiento ofensivo y defensivo, incluyendo sus intervalos de credibilidad al 95 %.

Equipo	Estimación Ataque	LI 0.025	LS 0.975
América de Cali	0.3396	0.0307	0.6292
Santa Fe	0.2790	-0.0779	0.6069
Millonarios	0.2712	-0.0476	0.5609

Tabla 4: Estimaciones de ataque y sus intervalos de credibilidad

Equipo	Estimación Defensa	LI 0.025	LS 0.975
Nacional	-0.4887	-0.9177	-0.0977
Millonarios	-0.2447	-0.6284	0.1106
La Equidad	-0.2204	-0.6822	0.1899

Tabla 5: Estimaciones de defensa y sus intervalos de credibilidad

Estas estimaciones resumen el desempeño relativo de los equipos. Millonarios F.C. destaca por su consistencia tanto ofensiva como defensiva, mientras que América de Cali y Nacional sobresalen en sus respectivas dimensiones. El enfoque bayesiano permite así una evaluación más matizada del rendimiento futbolístico, integrando inferencias probabilísticas con capacidad explicativa.

5. Conclusiones

Recientemente, se han utilizado enfoques bayesianos más complejos como redes probabilísticas para modelar partidos de fútbol (Joseph et al., 2022), lo que demuestra el creciente interés por metodologías probabilísticas en este campo. El modelo jerárquico bayesiano desarrollado en este trabajo permitió estimar de manera conjunta y coherente los efectos ofensivos, defensivos y de localía en el contexto del fútbol profesional colombiano, específicamente en la Liga BetPlay 2023-I. A diferencia del modelo frecuentista simplificado, centrado únicamente en la ventaja de jugar como local, el enfoque bayesiano incorporó explícitamente la heterogeneidad entre equipos mediante efectos aleatorios, capturando así diferencias estructurales en el rendimiento.

Una de las principales fortalezas del modelo es su capacidad para producir intervalos de credibilidad asociados a cada estimación, lo que facilita una evaluación más rigurosa de la incertidumbre en el análisis. Esto permite distinguir no solo a los equipos con mejores desempeños medios, sino también a aquellos cuyos efectos estimados son estadísticamente significativos.

Los resultados obtenidos muestran que equipos como América de Cali, Nacional y Millonarios destacaron por su rendimiento ofensivo, defensivo o balanceado, respectivamente. Además, se confirma el efecto positivo de la localía sobre la capacidad ofensiva de los equipos, con alta certeza estadística.

Este enfoque ofrece un marco analítico sólido y flexible, con alto potencial para ser extendido a tareas de predicción, comparación entre temporadas, y evaluación de cambios tácticos o estructurales dentro de los equipos. Su implementación puede servir como herramienta para la toma de decisiones deportivas basada en evidencia estadística, tanto para cuerpos técnicos como para analistas, medios y fanáticos del deporte.

En futuras investigaciones, este modelo podría ser ampliado para incluir efectos dinámicos entre fechas, interacciones jugador - equipo o información contextual adicional (por ejemplo, tarjetas, posesión o calidad de rivales), lo que abriría nuevas posibilidades en el análisis cuantitativo del rendimiento futbolístico.

Trabajos Futuros

Este trabajo puede extenderse en varias direcciones relevantes para el análisis estadístico del rendimiento futbolístico:

- **Predicción de resultados:** Una línea natural de extensión consiste en utilizar el modelo bayesiano ajustado para predecir resultados futuros de partidos, tanto en el corto como en el mediano plazo. Esto permitiría evaluar el poder predictivo del modelo e identificar posibles mejoras en su estructura.
- **Revisión de los hiperparámetros:** Se ha observado que las varianzas asociadas a los efectos aleatorios (ataque, defensa y localía) tienden a valores extremadamente pequeños, lo cual podría estar limitando la flexibilidad del modelo. Será necesario revisar los hiperparámetros de las distribuciones normales empleadas como priors, y considerar alternativas más robustas como el uso de priors de tipo *spike-and-slab*, que permitan una mejor regularización y detección de efectos significativos.

Códigos

El código en R utilizado en este trabajo puede consultarse en:

<https://github.com/DannaCruz/BETPLAY>.

Recibido: Agosto de 2025

Aceptado: Diciembre de 2025

Referencias

- Gianluca Baio and Marta Blangiardo. Bayesian hierarchical model for the prediction of football results. *Journal of Applied Statistics*, 37(2):253–264, 2010. doi: 10.1080/02664760802684177.
- Christopher M. Bishop. *Pattern Recognition and Machine Learning*. Springer, 2006.
- Andrew Gelman, John B. Carlin, Hal S. Stern, David B. Dunson, Aki Vehtari, and Donald B. Rubin. *Bayesian Data Analysis*. Chapman and Hall/CRC, 3rd edition, 2013.
- John Goddard. Regression models for forecasting goals and match results in association football. *International Journal of Forecasting*, 21(2):331–340, 2005.
- Andreas Groll, Christophe Ley, Gunther Schaubberger, and Hans Van Eetvelde. Prediction of football match outcomes: A comparison of state-of-the-art statistical models. *International Journal of Forecasting*, 35(2):712–722, 2018.
- Anthony Joseph, Norman Fenton, and Martin Neil. Bayesian networks for modelling football match outcomes. *European Journal of Operational Research*, 299(2):618–631, 2022.
- Dimitris Karlis and Ioannis Ntzoufras. Analysis of sports data by using bivariate poisson models. *The Statistician*, 52(3):381–393, 2003.
- Dimitris Karlis and Ioannis Ntzoufras. Bayesian modelling of football outcomes: Using the skellam’s distribution for the goal difference. *IMA Journal of Management Mathematics*, 20(2):133–145, 2009.
- Siem Jan Koopman and Roger Lit. Dynamic bivariate poisson models for analysing and forecasting match results in sports. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 178(1):167–186, 2015.
- Richard McElreath. *Statistical Rethinking: A Bayesian Course with Examples in R and Stan*. CRC Press, 2nd edition, 2020.
- Paolo Zuccolotto and Marica Manisera. *Basketball Data Science*. CRC Press, 2020.

A. Distribuciones Posteriores

La inferencia de los parámetros del modelo bayesiano se realiza a través de sus distribuciones posteriores. A continuación, se detallan las expresiones de estas distribuciones, las cuales son proporcionales al producto de la verosimilitud de los datos (proveniente de la distribución Poisson) y la distribución a priori asignada a cada parámetro.

A.1. Posterior para el Efecto de Jugar en Casa (home)

La distribución posterior para el parámetro **home**, que representa la ventaja de jugar como local, se define como:

$$p(\text{home} \mid \dots) \propto \prod_{g=1}^G \text{Poisson}(y_{g1} \mid \theta_{g1}) \times \mathcal{N}(\text{home} \mid 0, 0.0001)$$

Donde G es el número total de partidos en la temporada, y_{g1} son los goles anotados por el equipo local en el partido g , y θ_{g1} es la tasa de goles esperados para el equipo local, que depende de **home** y de las habilidades de ataque y defensa de los equipos involucrados.

A.2. Posteriores para las Habilidades de Ataque y Defensa

Para cada equipo $t = 1, \dots, T$ (donde T es el número total de equipos), las distribuciones posteriores para sus habilidades de ataque (att_t) y defensa (def_t) son las siguientes:

Ataque (att_t):

$$p(\text{att}_t \mid \dots) \propto \prod_{g: h(g)=t} \text{Poisson}(y_{g1} \mid \theta_{g1}) \times \prod_{g: a(g)=t} \text{Poisson}(y_{g2} \mid \theta_{g2}) \times \mathcal{N}(\text{att}_t \mid \mu_{\text{att}}, \tau_{\text{att}}^2)$$

Aquí, la probabilidad de att_t se ve influenciada por los goles anotados cuando el equipo t es local ($h(g) = t$) y los goles recibidos por el oponente cuando el equipo t es visitante ($a(g) = t$), así como por su prior normal con media μ_{att} y varianza τ_{att}^2 .

Defensa (def_t):

$$p(\text{def}_t \mid \dots) \propto \prod_{g: a(g)=t} \text{Poisson}(y_{g1} \mid \theta_{g1}) \times \prod_{g: h(g)=t} \text{Poisson}(y_{g2} \mid \theta_{g2}) \times \mathcal{N}(\text{def}_t \mid \mu_{\text{def}}, \tau_{\text{def}}^2)$$

La posterior de def_t depende de los goles recibidos cuando el equipo t es local ($a(g) = t$) y los goles anotados por el oponente cuando el equipo t es local ($h(g) = t$), en conjunto con su prior normal.

A.3. Posteriores Completamente Condicionales para Hiperparámetros

Las distribuciones posteriores para los hiperparámetros de ataque y defensa ($\mu_{\text{att}}, \mu_{\text{def}}, \tau_{\text{att}}^2, \tau_{\text{def}}^2$) son cruciales para la inferencia completa del modelo jerárquico.

Medias ($\mu_{\text{att}}, \mu_{\text{def}}$): Las distribuciones posteriores de las medias de los efectos de ataque y defensa son distribuciones normales cerradas:

$$p(\mu_{\text{att}}|\cdots) \sim \mathcal{N}\left(\frac{\sum_{t=1}^T \text{att}_t/\tau_{\text{att}}^2}{T/\tau_{\text{att}}^2 + 1/0.0001}, \left(\frac{T}{\tau_{\text{att}}^2} + \frac{1}{0.0001}\right)^{-1}\right)$$

$$p(\mu_{\text{def}}|\cdots) \sim \mathcal{N}\left(\frac{\sum_{t=1}^T \text{def}_t/\tau_{\text{def}}^2}{T/\tau_{\text{def}}^2 + 1/0.0001}, \left(\frac{T}{\tau_{\text{def}}^2} + \frac{1}{0.0001}\right)^{-1}\right)$$

Varianzas ($\tau_{\text{att}}^2, \tau_{\text{def}}^2$): Las distribuciones posteriores de las varianzas de los efectos de ataque y defensa son distribuciones Gamma Inversa, que son las conjugadas de la normal para la varianza:

$$p(\tau_{\text{att}}^2|\cdots) \sim \text{Inv-Gamma}\left(0.1 + \frac{T}{2}, 0.1 + \frac{1}{2} \sum_{t=1}^T (\text{att}_t - \mu_{\text{att}})^2\right)$$

$$p(\tau_{\text{def}}^2|\cdots) \sim \text{Inv-Gamma}\left(0.1 + \frac{T}{2}, 0.1 + \frac{1}{2} \sum_{t=1}^T (\text{def}_t - \mu_{\text{def}})^2\right)$$

A.4. Implementación Práctica

La obtención de muestras de estas distribuciones posteriores requiere el uso de métodos de Cadenas de Markov Monte Carlo (MCMC). Específicamente, se emplearía un algoritmo de *Gibbs sampling* para los parámetros de media ($\mu_{\text{att}}, \mu_{\text{def}}$) y varianza ($\tau_{\text{att}}^2, \tau_{\text{def}}^2$) debido a sus formas conjugadas. Sin embargo, para los efectos de **home** y las habilidades de ataque (att_t) y defensa (def_t), se necesitaría un algoritmo de *Metropolis-Hastings* debido a:

1. La **no conjugación completa** de estas posteriores con las distribuciones de los datos.
2. La necesidad de imponer **restricciones de suma cero** sobre los parámetros de ataque y defensa, lo cual complica la derivación de formas cerradas.
3. La **forma no estándar** de algunas de estas distribuciones posteriores que impide un muestreo directo.