



La XG y su asociación con los goles marcados en el fútbol de élite

The XG and their association with goals scored in elite football

Murillo, CM^{1ABCDEF}

¹ Corporación Universitaria del Caribe - CECAR, Colombia, cristian.murillo@cecar.edu.co

Responsabilidades. (A Diseño de la investigación; B Recolector de datos; C Redactor del trabajo; D Tratamiento estadístico; E Apoyo económico; F Idea original y coordinador de toda la investigación)

Recibido el 27 de junio de 2025

Aceptado el 24 de noviembre de 2025

DOI: 10.24310/riccafd.14.2.2025.21310

Correspondencia: Cristian Murillo. cristian.murillo@cecar.edu.co

RESUMEN

El propósito de este estudio fue analizar la asociación entre xG y goles marcados en partidos de fútbol. Este estudio fue cuantitativo, correlacional y longitudinal. Se analizaron 333 partidos de liga. Los datos se obtuvieron de Sofascore un sitio web que ofrece estadísticas y resultados de diversos eventos deportivos. Se planteó la hipótesis de que existe una correlación positiva entre el xG y los goles marcados. La muestra consistió en 333 partidos oficiales de competiciones internacionales, incluyendo la Liga de Campeones 2013/24, la Copa Libertadores 2024, la Eurocopa 2024 y la Copa América 2024. Estos torneos fueron seleccionados por su gran relevancia internacional. Los resultados de esta investigación confirman la validez del uso del xG como indicador para analizar el rendimiento ofensivo de un equipo, aunque su capacidad predictiva suele variar según el contexto o la competición ($r = 0,488$, $p < 0,001$). En los clubes, los jugadores tienen mayor continuidad al entrenar juntos, lo que facilita establecer un patrón de juego, lo que podría explicar un xG más alto en comparación con los torneos de selecciones nacionales. Los resultados obtenidos en este estudio confirman la relación entre xG y goles marcados en partidos de fútbol, estableciéndolo así como un indicador válido para medir el rendimiento ofensivo en el fútbol. Se encontró que, en los torneos de clubes, la cantidad de asociación, en la Copa Libertadores ($r = 0,537$) teniendo la mejor asociación entre las variables, en la Champions League la asociación fue menor ($r = 0,403$). Mientras tanto, en los torneos de selecciones nacionales, los valores para la Copa América y la Eurocopa fueron ($r = 0,475$) y ($r = 0,479$), respectivamente, donde se puede observar cierta similitud en la

asociación de variables, que se atribuye a la pobre cohesión grupal y al pobre ajuste táctico de las selecciones nacionales en comparación con los clubes.

PALABRAS CLAVE: fútbol, goles, indicador, métrica, resultado

ABSTRACT

The purpose of this study was to analyze the association between xG and goals scored in football matches. This was a quantitative, correlational, and longitudinal study. A total of 333 league matches were analyzed. Data were obtained from Sofascore, a website that provides statistics and results for various sporting events. The hypothesis was that there is a positive correlation between xG and goals scored. The sample consisted of 333 official matches from international competitions, including the 2013/24 Champions League, the 2014 Copa Libertadores, the 2014 UEFA European Championship, and the 2014 Copa América. These tournaments were selected due to their high international relevance. The results of this research confirm the validity of using xG as an indicator to analyze a team's offensive performance, although its predictive capacity tends to vary depending on the context or competition ($r = 0.488$, $p < 0.001$). At club level, players have greater continuity in training together, which facilitates the establishment of a playing pattern. This could explain a higher xG (xG) compared to national team tournaments. The results obtained in this study confirm the relationship between xG and goals scored in football matches, thus establishing it as a valid indicator for measuring offensive performance in football. It was found that in club tournaments, the amount of association was highest in the Copa Libertadores ($r = 0.537$), with the strongest association between variables, while the association was lowest in the Champions League ($r = 0.403$). Meanwhile, in national team tournaments, the values for the Copa América and the European Championship were ($r = 0.475$) and ($r = 0.479$), respectively, where a certain similarity in the association of variables can be observed. This is attributed to the poor group cohesion and tactical adjustment of national teams compared to clubs.

KEY WORDS: football, goals, indicator, metric, result

INTRODUCCIÓN

El fútbol es el deporte más reconocido del mundo y uno de los que genera mayores ingresos para la FIFA. En la temporada 2022/23, la Premier League inglesa alcanzó aproximadamente tres mil millones de libras en derechos de transmisión y comerciales (1). Esta importante inversión financiera requiere un enfoque para optimizar el rendimiento mediante métodos innovadores (2). A medida que aumenta el crecimiento global del deporte, también aumenta la necesidad de estrategias y herramientas tecnológicas para apoyar el rendimiento de equipos y jugadores.

A medida que el fútbol global crece, también aumenta la necesidad de implementar herramientas tecnológicas para el rendimiento grupal y colectivo. Por ello, el análisis del rendimiento ofensivo ha cobrado gran importancia y ha evolucionado gracias al uso de métricas, como la diferencia de goles esperada (xG), que se ha convertido en una herramienta para evaluar la calidad de las ocasiones generadas por un jugador o equipo (3).

El xG mide la probabilidad de que un tiro resulte en gol. Se basa en variables como la posición del tiro, la distancia a la portería y la presión defensiva rival (4). Además, esta herramienta permite analizar la eficiencia ofensiva de un equipo y comparar el rendimiento esperado y el real, lo que la convierte en una herramienta válida para entrenadores, analistas y científicos deportivos (4).

A pesar del auge del xG como herramienta analítica, aún persisten muchas dudas sobre su capacidad predictiva en diferentes contextos competitivos. Un estudio reveló que la asociación entre el xG y los goles marcados depende del torneo, el nivel y la dinámica del juego (5).

Por otro lado, los avances en el análisis deportivo han propiciado el desarrollo de métricas que ayudan a los equipos a gestionar el rendimiento de sus jugadores. Una métrica como los goles esperados (xG) se ha convertido en parte integral de la toma de decisiones de los clubes (6). Sin embargo, se carece de un enfoque integral para evaluar a jugadores y equipos, ya que las métricas conocidas no ofrecen una visión completa de las contribuciones de los jugadores ni de la efectividad del equipo (7).

En consecuencia, esta brecha se deriva de la ausencia de características cruciales y las limitaciones inherentes de la xG. Los analistas de rendimiento deben reconocer la importancia del contexto; basarse únicamente en una métrica para el análisis estadístico puede ser perjudicial (8). Si bien quienes utilizan la xG reconocen sus ventajas, también es necesario reconocer sus limitaciones. Si bien los datos utilizados en cualquier modelo son importantes, su interpretación es vital. Una comprensión integral que abarque tanto las ventajas como las desventajas de la xG es necesaria para un desarrollo óptimo (9).

Aunque el fútbol es un deporte muy popular, históricamente no ha habido mucha conexión entre el deporte y el análisis de datos estadísticos. Sin embargo, la métrica xG mitiga esto al utilizar métodos estadísticos para estimar la probabilidad de que un tiro resulte en gol (10).

Además, estudios han explorado la aplicación de modelos xG para incorporar diferentes características, basadas en dos variables clave: la distancia a la portería y el ángulo de tiro (11). Un estudio reveló que, utilizando estadísticas de las temporadas 2012/13 de las ligas inglesa y alemana, el estudio implicó dividir el campo en ocho zonas y analizar la probabilidad de gol según los tiros realizados desde cada zona. Los resultados mostraron que la distancia y el ángulo influyen significativamente en la probabilidad de gol (1).

De manera similar, otro estudio tuvo como objetivo investigar el impacto de la distancia y el ángulo de tiro. Este estudio se basó en eventos presentados en una liga de fútbol durante la temporada 2017/18. Este estudio utilizó un enfoque probabilístico para analizar las oportunidades de gol. Este estudio destacó la complejidad y el potencial del modelado xG en el análisis (9). Se debe tener en cuenta el tipo de tiro, que proporciona información contextual sobre los tiros. Los tiros se pueden dividir en: segmento corporal utilizado (pie derecho/pie izquierdo o cabeza) y la situación de juego en la que se produjo el tiro. Los autores incorporaron estas características en su modelo y descubrieron que ambos aspectos influyen en la efectividad del tiro (12).

Además, descubrieron que los tiros libres y los penales tienen más probabilidades de resultar en goles en comparación con los tiros en campo abierto y los tiros con los pies; por el contrario, los cabezazos tienen menos probabilidades de resultar en goles (13).

Si bien diversos autores han demostrado la capacidad del xG para predecir el rendimiento ofensivo, carecen de validez de datos o no consideran variables contextuales como el torneo, el partido en casa o fuera, el tipo de partido, entre otras (9); (7). En base a esto, se plantea la hipótesis de que existe una correlación positiva y estadísticamente significativa entre los goles esperados (xG) y los goles marcados en el fútbol de élite.

El propósito de este estudio fue analizar la asociación entre xG y goles marcados en partidos de fútbol.

MATERIAL Y METODOS

Este estudio fue cuantitativo, correlacional y longitudinal. Se analizaron 333 partidos de liga. El conjunto de datos se obtuvo de Sofascore, un sitio web que ofrece estadísticas y resultados de diversos eventos deportivos. Se planteó la hipótesis de que existe una correlación positiva entre el xG y los goles marcados.

La muestra consistió en 333 partidos oficiales de competiciones internacionales, incluyendo la Liga de Campeones 2013/24, la Copa Libertadores 2024, la Eurocopa 2024 y la Copa América 2024. Estos torneos fueron seleccionados por su gran relevancia internacional.

Se analizaron dos variables principales:

- **Variable independiente:** xG, que corresponde a una métrica utilizada para estimar la probabilidad de que un tiro termine en gol, la cual se basa en factores como: distancia, ángulo de tiro, presión defensiva, asistencia, entre otros.
- **Variable dependiente:** Goles marcados, goles conseguidos en cada partido.

Los datos utilizados en esta investigación fueron tomados de la plataforma SofaScore, que proporciona datos para analizar el rendimiento de determinados deportes y también estadísticas en tiempo real.

Análisis estadístico

Para el análisis de los datos se utilizó el software estadístico SPSS, se procesaron los partidos, goles marcados y xG de cada partido, seguido de la creación de una tabla de muestras pareadas para determinar si existían diferencias significativas entre las medias y una correlación para indicar el grado de relación de las variables utilizadas.

Los datos se obtuvieron de los registros de partidos disponibles en la página web de SofaScore. Posteriormente, se exportaron a una base de datos para su depuración. Previo al análisis inferencial, se evaluó la normalidad de los datos mediante la prueba de Shapiro-Wilk, confirmando una distribución normal ($p > 0,05$). Asimismo, se verificó la fiabilidad de los datos de SofaScore comparando aleatoriamente el 10 % de los registros con los informes oficiales de la CONMEBOL y la UEFA, que mostraron una alta consistencia ($ICC = 0,91$). Estos procedimientos garantizaron la fiabilidad y la validez de la base de datos analizada.

RESULTADOS

La tabla muestra los resultados del xG promedio y el número de goles marcados en cada competición, lo que demuestra que no hay diferencias

significativas en las medias (Tabla 1). Por un lado, el xG en la Copa Libertadores promedió $2,30 \pm 0,95$, por el lado de los goles el promedio fue de $2,40 \pm 1,43$, en la Champions League el xG promedio fue de $3,00 \pm 1,04$ y el promedio de goles fue de $2,99 \pm 1,70$. En cuanto a los torneos de selecciones, la Copa América tuvo un promedio de $2,41 \pm 0,84$, pero con un promedio de goleadores menor ($2,18 \pm 1,55$), en la Eurocopa se obtuvo un promedio de $2,61 \pm 0,91$ en cuanto a xG y el promedio de goles fue de $2,29 \pm 1,36$.

Tabla 1. Estadísticas de muestras emparejadas

		Media	N	Desv. Desviación	Desv. Error promedio
Par 1	xG Copa libertadores	2,30	125	,95	,08
	Goles Libertadores	2,40	125	1,43	,12
Par 2	xG Champions league	3,00	125	1,04	,09
	Goles Champions	2,99	125	1,70	,15
Par 3	xG Copa América	2,41	32	,84	,14
	Goles Copa América	2,18	32	1,55	,27
Par 4	xG Euro	2,61	51	,91	,12
	Goles Euro	2,29	51	1,36	,19

En la tabla se pueden observar los resultados de la media de xG y la cantidad de goles marcados

Tabla 2. Correlaciones de muestras emparejadas

		N	Correlación	Sig.
Par 1	xG Copa libertadores & Goles Libertadores	125	,537	,000
Par 2	xG Champions league & Goles Champions	125	,403	,000
Par 3	xG Copa América & Goles Copa América	32	,475	,006
Par 4	xG Euro & Goles Euro	51	,479	,000

La tabla indica una correlación positiva. La significancia asintótica tuvo un valor $<0,00$, por lo que se acepta la hipótesis del investigador, en otras palabras, se acepta que a mayor xG, mayor número de goles marcados. Además, se obtuvo un coeficiente de correlación de $r = 0,488$, lo que indica que existe una relación moderada y positiva entre ambas variables analizadas. La significancia ($p = 0,000$) confirma que es significativa, indicando que xG es un predictor válido para predecir el número de goles en la muestra analizada. A pesar de ser una correlación moderada, al no ser alta, se afirma que existen otros factores que influyen en la marcación de goles, como la calidad del disparo, la calidad del portero, el rendimiento de la defensa, entre otros.

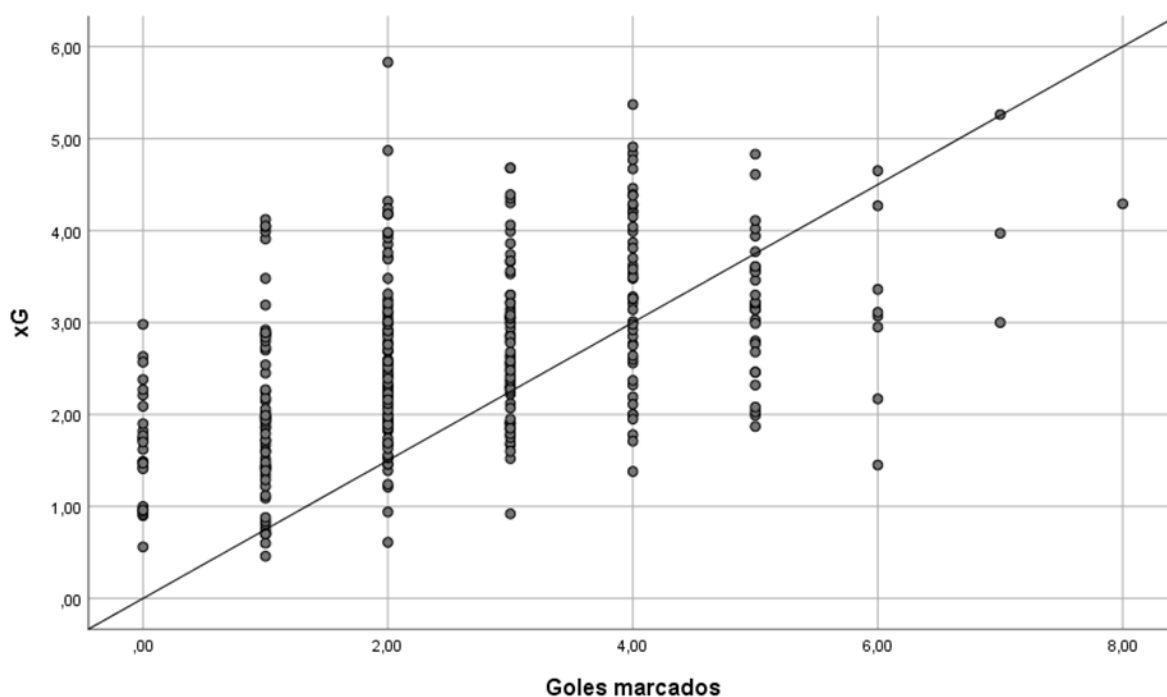
Tabla 3. Correlaciones

		xG Goles marcados	
xG	Correlación de Pearson	1	,488**
	Sig. (bilateral)		,000
	N	33	333
Goles marcados	Correlación de Pearson	,4**	1
	Sig. (bilateral)	,00	
	N	33	333

En el gráfico, se puede observar la clara relación lineal positiva entre los datos: a medida que aumentaba el xG, también lo hacía el número de goles marcados. Por lo tanto, se puede establecer que el número de goles marcados estaba directamente relacionado con el xG o que eran variables directamente proporcionales (Figura 1). La dispersión indica que, si bien existe una relación entre las dos variables analizadas, esta no es del todo perfecta. Hay puntos por encima y por debajo de la línea de regresión, lo que indica que algunos clubes o selecciones nacionales superaron o no alcanzaron el xG en algunos partidos. Esto puede atribuirse a factores como la falta de definición, el rendimiento del portero rival o una efectividad deficiente o excesiva en la creación de ocasiones de gol.

La correlación positiva moderada ($r=0,488$) indicó que, aunque el xG explica en gran medida la variabilidad de los goles marcados, hay otros factores importantes, como la calidad de la finalización, el portero y la organización defensiva del equipo contrario.

Figura 1. Dispersión simple



Los resultados de esta investigación confirman la validez del uso del xG como indicador para analizar el rendimiento ofensivo de un equipo, aunque su capacidad predictiva suele variar según el contexto o la competición. En los clubes, los jugadores tienen mayor continuidad al entrenar juntos, lo que facilita establecer un patrón de juego, lo que podría explicar un xG más alto en comparación con los torneos de selecciones nacionales.

DISCUSIÓN

El estudio sobre la métrica estadística moderna de goles esperados (xG) y su impacto en los goles marcados arrojó varios hallazgos significativos (14). Uno de los principales resultados de esta investigación fue la fuerte asociación

entre los goles marcados y el xG. (14). Un estudio coincidió con los hallazgos de este, ya que los equipos con promedios de xG más altos tuvieron promedios de goles más altos a lo largo de una temporada (15).

Por otro lado, un estudio determinó que el xG predice el rendimiento de un equipo con un alto nivel de fiabilidad. Los equipos con un xG alto en partidos anteriores tendían a marcar un gran número de goles en partidos posteriores. Por lo tanto, esta herramienta ayuda a entrenadores y analistas a implementar estrategias para la toma de decisiones tácticas (16). Otro estudio indicó que, debido a la asociación del xG con los goles marcados, los entrenadores podrían utilizar la información proporcionada por el xG para ajustar su trabajo táctico ofensivo, planificando sesiones donde se ayuda a los jugadores a encontrar posiciones más ventajosas para marcar (17).

El xG es un indicador útil para indicar el número de goles marcados por los equipos, lo que ayuda a mejorar el rendimiento individual y colectivo de forma objetiva (18). En los entrenamientos, los entrenadores pueden usar los datos de xG para diseñar ejercicios que simulen escenarios (19). Por ejemplo, crear ejercicios centrados en rematar desde posiciones centrales dentro del área puede ayudar a los jugadores a desarrollar su capacidad para convertir ocasiones de alta calidad. Además, los datos de xG pueden utilizarse para evaluar el rendimiento, proporcionando objetivos claros y basados en datos para trabajar en los entrenamientos.

LIMITACIONES Y CAMINOS FUTUROS

Aunque el xG es un predictor útil del rendimiento ofensivo, existen otros factores que pueden influir en la puntuación en un partido. Algunos factores que pueden influir en los goles son: calidad individual, equipos con delanteros destacados tienen mayor posibilidad de superar su xG, por otro lado, a menor calidad de los jugadores mayor dificultad para marcar.

Rendimiento del portero rival, cuando un portero tiene un buen rendimiento se caracteriza por su capacidad de evitar goles en su contra.

Tácticas defensivas, Los equipos que dan alta prioridad a la solidez defensiva son capaces de reducir las opciones de su oponente y por tanto su xG

Los hallazgos de esta investigación tienen aplicaciones prácticas para analizar el rendimiento y tomar decisiones tácticas. Entrenadores y analistas pueden usar la métrica xG no solo para evaluar el rendimiento ofensivo de sus equipos o del rival, sino también para identificar áreas de mejora en la definición de jugadas y la creación de oportunidades de gol.

Una de las limitaciones del estudio es que los datos se extrajeron de una única fuente (SofaScore), lo que podría dar lugar a una recopilación de datos sesgada. Además, solo se analizaron indicadores ofensivos, sin tener en cuenta otras variables como el equipo local o visitante, el estilo de juego del rival, etc. Por lo tanto, se recomienda que futuros estudios integren modelos multivariantes para identificar patrones de predicción del rendimiento más complejos.

REFERENCIAS

1. Bradlow E, Gangwar M, Kopalle P, Voleti S. The role of big data and predictive analytics in retailing. *J Retail*. 2017;5(93):79-95.
2. Davis J, Robberechts P. How data availability affects the ability to learn good xG models. *Int Workshop Mach Learn Data Min Sports Anal Ghent*. 2020;13(24):17-27.
3. Fernández-Navarro J, Fradua L, Zubillaga A, Ford P, McRobert A. Influence of contextual variables on styles of play in soccer. *Int J Perform Anal Sport*. 2019;19(6):747-59.
4. Hoernig M, Link D. Individual ball possession in socce. *PLoS ONE*. 2017;12(7):1-15.
5. Gómez A, Mitrotasios M, Armatas V, Lago-Peñas C. Analysis of playing styles according to team quality and match location in Greek professional soccer. *Int J Perform Anal Sport*. 2018;18(3):986-97.
6. Zeng Z, Pan B. A Machine Learning Model to Predict Player's Positions based on Performance. *icSPORTS [Internet]*. 2021;1(17):36-42. Disponible en: <https://www.scitepress.org/PublishedPapers/2021/106533/106533.pdf>
7. Mead J, O'Hare A, McMenemy P. Expected goals in football: Improving model performance and demonstrating value. *PLoS ONE*. 2023;18(5):20-8.
8. Brechot M, Flepp L. Dealing with randomness in match outcomes: How to rethink performance evaluation in European club football using expected goals. *J Sport Econ*. 2020;4(21):335-62.
9. Anzer G, Bauer P. A Goal Scoring Probability Model for Shots Based on Synchronized Positional and Event Data in Football (Soccer). *Front Sports Act Living*. 2021;3(62):75-82.
10. Brechot M, Flepp R. Dealing With Randomness in Match Outcomes: How to Rethink Performance Evaluation in European Club Football Using Expected Goals. *J Sports Econ*. 2020;21(4):335-62.
11. Antipov E, Pokryshevskaya E. Interpretable machine learning for demand modeling with high-dimensional data using gradient boosting machines and shapley values. *J Rev Pricing Manage*. 2020;19(7):355-64.
12. Fairchild A, Pelechrinis K, Kokkodis M. Spatial analysis of shots in MLS: A model for expected goals and fractal dimensionality. *J Sports Anal*. 2018;4(3):165-74.

13. Schulze E, Mendes B, Maurício N, Furtado B, Cesário N, Carriço S, et al. Effects of positional variables on shooting outcome in elite football. *Sci Med Footb.* 2017;2(2):93-100.
14. Khrapach V, Siryi O. Statistical Metric Expected Goals in Football and Its Impact on Scoring Performance: a review article. *Health Technol.* 2024;2(3):47-54.
15. Sabin R. Estimating player value in American football using plus-minus models. *J Quant Anal Sports.* 2021;4(17):313-64.
16. Mead J, O'Hare A, McMenemy P. Expected goals in football: Improving model performance and demonstrating value. *PLoS ONE.* 2023;18(4):45-55.
17. Zi M, Gao D. A study of football goal trajectories of trained players through a kinetic model. *J Eng-Joe.* 2023;12(2):4-6.
18. Anzer G, Bauer P. A goal scoring probability model for shots based on synchronized positional and event data in football (soccer). *Front Sports Act Living.* 2021;3(21):624-475.
19. Cavus M, Biecek P. Explainable expected goal models for performance analysis in football analytics. *Data Sci Adv Anal.* 2022;4(9):1-9.