



Modelación de la relación entre la altura y el desempeño del equipo al final del partido con el resultado y el número de goles anotados

Yeferson Arbey Romo Bastidas

Universidad del Valle
Facultad de Ingeniería, Escuela de Estadística
Cali, Valle del Cauca
2025

Modelación de la relación entre la altura y el desempeño del equipo al final del partido con el resultado y el número de goles anotados

Realizado por

Yeferson Arbey Romo Bastidas

Trabajo de grado para optar por el título de:

Estadístico

Director:

José Rafael Tovar Cuevas

Co- Director:

Álvaro José Flórez Poveda

Universidad del Valle
Facultad de Ingeniería, Escuela de Estadística
Cali, Valle del Cauca
2025

Dedicatoria

Este trabajo de grado lo dedico, en primer lugar, a Dios, por darme la sabiduría y la salud necesarias para alcanzar esta meta.

A mi madre, María, por nunca dejar de apoyarme en todo momento y por creer siempre en mí. Ella es mi fortaleza, gracias a ella aprendí la resiliencia y me hizo confiar en mi potencial, tanto en la vida personal como académica. ¡Es el verdadero significado del amor en mi vida!

A mi padre, por sus valiosos consejos y por inculcarme que lo mejor que se puede hacer es estudiar y esforzarse para alcanzar mejores oportunidades. Este logro es también fruto de su guía y dedicación.

A mi hermana y a mi cuñado, quienes fueron pilares fundamentales con sus consejos oportunos y su apoyo constante, ayudándome a superar desafíos en la vida personal y a alcanzar esta meta.

¡Gracias, familia, por estar siempre conmigo!

Yeferson Arbey Romo Bastidas

Agradecimientos

Quiero expresar mi más sincero agradecimiento al profesor José Rafael Tovar por brindarme la oportunidad de desarrollar este trabajo de grado relacionado con una de mis grandes pasiones: el fútbol. Ha sido un proyecto que disfruté profundamente, y estoy muy agradecido por su forma de enseñar en las clases, que me permitió aprender a ver la Estadística desde otra perspectiva, así como por sus valiosos consejos en la vida personal.

Asimismo, agradezco al profesor Álvaro Flórez por sus oportunas correcciones y sugerencias, que contribuyeron enormemente a mejorar este trabajo.

Mi gratitud también se extiende a los profesores de la Escuela de Estadística, quienes fueron un pilar fundamental para alcanzar esta meta profesional, así como a mis compañeros de carrera, que me brindaron apoyo y compañía durante todo el proceso.

Yeferson Arbey Romo Bastidas

Resumen

El presente trabajo de grado tiene como propósito evaluar la relación entre el desempeño de los equipos de Fútbol Profesional Colombiano al final del partido y la altitud donde se disputa el encuentro, considerando su impacto tanto en el resultado final como en la cantidad de goles anotados por los equipos visitantes. Para medir el desempeño, se toma un índice que sintetiza múltiples indicadores técnicos y tácticos de juego, y también se analizaron los factores que lo componen: Control Ofensivo, Solidez Defensiva y Eficacia en Finalización.

Se ajustaron Modelos Lineales Generalizados: Regresión Poisson y una regresión Poisson Mixta, empleadas para estimar el número esperado de anotaciones del equipo visitante, y modelo ordinal, orientado a estimar la probabilidad de que el equipo visitante pierda, empate o gane, todos en función de la altitud de la ciudad sede, el desempeño que obtuvo al final del partido según el índice y los factores que componen este índice.

En cuanto al número de goles anotados por el equipo visitante, se observó que cuando el equipo debe ascender en promedio 2420 msnm, su capacidad anotadora disminuye. Esta reducción también se presenta cuando el equipo finaliza el partido con un nivel de rendimiento medio o bajo. Respecto a los factores que conforman el índice de rendimiento, se encontró que tanto el Control Ofensivo como la Eficacia en Finalización son determinantes y contribuyen de manera significativa a que un equipo logre anotar.

Por otra parte, la probabilidad de que el equipo visitante gane el partido disminuye cuando debe ascender en promedio 2420 msnm. Además, el nivel de rendimiento al final del partido influye directamente en sus probabilidades de victoria. Entre los factores del índice, la Eficacia en Finalización se identificó como el único factor consistentemente relevante para explicar el resultado final del partido en el Fútbol Profesional Colombiano .

Palabras clave: Índice de rendimiento, Indicadores técnicos-tácticos en el fútbol, Altitud, Resultado del partido, Modelos Lineales Generalizados, Regresión Multinomial Ordinal, Regresión Poisson, Fútbol Profesional Colombiano.

Tabla de contenido

1	Planteamiento del problema	1
1.1	Definición del problema	1
1.2	Pregunta de Investigación	2
1.3	Justificación	3
1.4	Objetivos	4
1.4.1	Objetivo General	4
1.4.2	Objetivos Específicos	4
2	Antecedentes	5
3	Marco teórico	11
3.1	Marco Contextual	11
3.1.1	Altitud	11
3.1.2	Categorización de la altitud	11
3.1.3	Índice	12
3.1.4	Indicador de desempeño	12
3.2	Marco Estadístico	13
3.2.1	Modelo de Regresión de Poisson	13
3.2.2	Estimación del parámetro de sobredispersión	16
3.2.3	Indicadores de calidad de ajuste del modelo	17
3.2.4	Modelo de Regresión Multinomial	19
3.2.5	Modelo de Regresión Ordinal	21
3.2.6	Modelo Poisson GLMM para datos de conteo correlacionados	23
3.2.7	Razón de Verosimilitudes	24
3.2.8	Test de hipótesis e intervalos de confianza de Wald	26
4	Metodología	27
4.1	Archivo de datos	27
4.2	Variables incluidas en el análisis	27
4.3	Formación de los grupos	29
4.3.1	Interpretación de los grupos	30
4.3.2	Altitud del equipo visitante	31

4.4	Modelos propuestos	33
4.4.1	Modelos para la variable respuesta: Número de Goles Anotados	33
4.4.2	Bonda de ajuste y calidad del modelo Poisson	34
4.4.3	Modelo mixto	34
4.4.4	Modelos para la variable respuesta: Resultado Final del Partido	36
4.4.5	Bonda de ajuste y calidad del modelo Ordinal	39
4.4.6	Comparación de modelos	39
5	Resultados	40
5.1	Análisis exploratorio	40
5.1.1	Número de goles anotados del equipo visitante	40
5.1.2	Resultado final del equipo visitante	41
5.2	Análisis exploratorio de covariables y su relación con las variables respuesta	42
5.2.1	Rendimiento del equipo visitante al final del partido según el índice	42
5.2.2	Relación entre el rendimiento del equipo visitante con el re- sultado final y el número de goles anotados: análisis mediante prueba de chi-cuadrado	43
5.2.3	Factores del índice de rendimiento	45
5.2.4	Relación entre el resultado final y los factores del índice de rendimiento: análisis mediante prueba de chi-cuadrado	46
5.2.5	Relación entre los goles anotados del visitante y los factores del índice de rendimiento: análisis mediante prueba de chi- cuadrado	47
5.2.6	Goles Anotados, Resultado Final según la condición de de- splazamiento del equipo visitante (altitud): análisis mediante prueba de chi-cuadrado	49
5.3	Resultado 1: Modelación de la relación entre la altura, el desempeño del equipo al final del partido y su influencia en el número de goles anotados	51
5.3.1	Modelo 1: Regresión Poisson para los goles del equipo vi- sitante en función del desplazamiento y el rendimiento del equipo al final del partido	52
5.3.2	Bondad de ajuste y calidad del modelo 1	52
5.3.3	Modelo 2: Regresión Poisson para los goles del equipo visi- tante en función de los factores del índice de rendimiento . .	56
5.3.4	Bonda de ajuste y calidad del modelo 2	56
5.3.5	Comparación de indicadores entre el Modelo Poisson Mixto y el Modelo Poisson	59

5.4	Resultado 2: Modelación de la relación entre la altura, el desempeño del equipo al final del partido y su influencia en el resultado	60
5.4.1	Modelo 3: Regresión ordinal del resultado del equipo visitante en función de la altura y el rendimiento del equipo al final del partido	60
5.4.2	Bondad de ajuste y calidad del Modelo Ordinal 1	61
5.4.3	Modelo 4: Regresión ordinal del resultado del equipo visitante en función de los factores del índice de rendimiento	65
5.4.4	Bondad de ajuste y calidad del Modelo Ordinal 2	65
6	Conclusiones	69
7	Recomendaciones	71
8	Anexos	72
	Bibliografía	76

Lista de Tablas

2.1	Construcción de variables a partir de indicadores de juego	6
2.2	Cargas factoriales del factor 1 (Visitante)	7
2.3	Cargas factoriales del factor 2 (Visitante)	7
2.4	Cargas factoriales del factor 3 (Visitante)	7
3.1	Categoría de altitud para el estadio según sus metros sobre el nivel del mar.	11
3.2	Descripción de los diferentes indicadores del rendimiento que han sido utilizados en el análisis de los deportes en equipos, según Lago-Peñas (2008)	12
4.1	Descripción de las variables utilizadas en el estudio	28
4.2	Combinaciones de altitud entre equipos local y visitante	29
4.3	Descripción de los grupos según el desplazamiento altitudinal del equipo visitante	30
4.4	Altura de estadios del FPC y categoría según clasificación de Bartsch.	31

4.5	Descripción de los grupos y número de partidos jugados según el desplazamiento del equipo visitante	32
5.1	Distribución de goles anotados por el equipo visitante	41
5.2	Distribución categorizada de goles anotados por el equipo visitante .	41
5.3	Distribución de resultados del equipo visitante	42
5.4	Categorización del índice de rendimiento	42
5.5	Distribución del índice de rendimiento	43
5.6	Categorización de los factores del índice de rendimiento	45
5.7	Distribución de los niveles de rendimiento en cada factor	46
5.8	Distribución de factores por resultado del equipo visitante con p-valores	47
5.9	Distribución de los factores de rendimiento en función de los goles anotados por el equipo visitante con p-valores	48
5.10	Parámetro de dispersión	52
5.11	Evaluación del ajuste mediante desviaciones y prueba de razón de verosimilitud del modelo 1	53
5.12	Estimación del modelo Poisson para goles del visitante según categoría de desplazamiento y rendimiento	53
5.13	Parámetro de dispersión	56
5.14	Evaluación del ajuste mediante desviaciones y prueba de razón de verosimilitud del modelo 2	56
5.15	Estimación del modelo Poisson para los goles del visitante según los factores del índice de rendimiento	57
5.16	Comparación de AIC entre modelo lineal generalizado Poisson simple y modelo lineal generalizado Poisson mixto	59
5.17	Prueba del supuesto de Odds Proporcionales para el modelo de regresión ordinal, utilizando la prueba de Brant	60
5.18	Prueba de razón de verosimilitud (LRT) para el modelo logit ordinal (Modelo 3)	61
5.19	Bondad de ajuste del modelo logit ordinal (Modelo 3)	62
5.20	Estimación del modelo ordinal para el resultado del partido según la categoría de desplazamiento y el rendimiento	62
5.21	Prueba del supuesto de Odds Proporcionales para el modelo de regresión ordinal, utilizando la prueba de Brant	65
5.22	Prueba de razón de verosimilitud (LRT) para el modelo logit ordinal (Modelo 4)	66
5.23	Bondad de ajuste del modelo logit ordinal (Modelo 4)	66
5.24	Estimación del modelo ordinal para el resultado del partido según los factores del índice de rendimiento	67
8.1	Estadios, ciudades y departamentos del fútbol profesional colombiano.	72

8.2	Distribución de goles por equipo visitante.	73
8.3	Resultado final según el equipo visitante	74
8.4	Nivel de rendimiento según el equipo visitante	75
8.5	Resumen descriptivo de goles por nivel del índice de rendimiento. .	75

Lista de Figuras

2.1	Modelo formativo visualizado mediante un gráfico.	5
5.1	Resultado del equipo visitante según el nivel del rendimiento al final del partido	44
5.2	Goles anotados por el equipo visitante según el nivel del rendimiento al final del partido	44
5.3	Goles anotados por el equipo visitante según su desplazamiento . .	50
5.4	Resultado del equipo visitante según su desplazamiento	50
5.5	Razones de tasas con intervalos de confianza del 95% del Modelo 1.	54
5.6	Razones de tasas con intervalos de confianza del 95% del Modelo 2.	57
5.7	Razones de tasas con intervalos de confianza del 95% del Modelo 3.	63
5.8	Razones de tasas con intervalos de confianza del 95% del Modelo 4.	67

Capítulo 1

Planteamiento del problema

1.1 Definición del problema

Aceptar que la suerte desempeña un papel relevante en los resultados de los equipos al final de una temporada implica minimizar la importancia del rendimiento y, por tanto, del entrenamiento. Por el contrario, si se considera que el azar tiene una influencia limitada en la posición final del campeonato, se reconoce que rendimiento y resultado están estrechamente vinculados, tal como señala Lago-Peñas (2008)

En el contexto del Fútbol Profesional Colombiano (FPC), surge entonces la necesidad de evaluar si el desempeño al final del partido, medido en este trabajo mediante un índice sintético construido a partir de indicadores técnico-tácticos de jugadas ofensivas, defensivas y a balón parado creado por Chávez-Avendaño (2025), se relaciona con el resultado y el número de goles anotados que obtienen los equipos visitantes. No obstante, se reconoce que dicha relación no es estable, porque aunque podría asumirse que registrar el éxito o fracaso de un equipo en un partido equivale a observar su verdadero nivel de prestación, los deportes de equipo siempre implican un margen de aleatoriedad y factores impredecibles: un empate o una derrota por 1–0 puede ocultar una clara superioridad en el juego de un conjunto sobre otro, como señala Lago-Peñas (2008). Por tanto, desempeño y resultado no siempre son equivalentes, lo que abre la posibilidad de modelar el resultado a partir del rendimiento observado.

Por otra parte, si se aceptara que el azar es el componente determinante del resultado, entonces el desenlace de un partido no podría modelarse en absoluto. Lago-Peñas (2008) señala que no tiene sentido ni es posible separar de manera estricta los efectos del azar y los del rendimiento. Ambos coexisten, y precisamente por ello resulta pertinente estudiar hasta qué punto el rendimiento de los equipos visitantes permite explicar los resultados y el número de goles que anotan.

Dejar atrás la resignación expresada en frases como “cada partido es un mundo”, “el baloncesto es así” o “fútbol es fútbol” es fundamental. Estas expresiones suponen considerar el fenómeno deportivo como inabordable y renunciar a comprenderlo con rigor.

También se reconoce que diversos factores contextuales influyen en el resultado del partido y la cantidad de goles, entre ellos la altitud. Numerosos expertos coinciden en que, si bien no existe un riesgo grave para la salud de los deportistas de élite no aclimatados, sí existe una desventaja considerable para los equipos visitantes que compiten en altura, como señala Matijasevic (2017).

Por la geografía colombiana, la diferencia en altitud entre sedes puede ser extrema. Los equipos de la Costa Caribe, como Junior, Jaguares o Magdalena, juegan prácticamente a nivel del mar, mientras que los equipos de Bogotá (Millonarios, Santa Fe) y Boyacá (Patriotas, Boyacá Chicó) compiten por encima de los 2600 msnm. En consecuencia, un equipo visitante que viaja desde el nivel del mar para disputar un partido en Bogotá o Tunja enfrenta condiciones ambientales y fisiológicas radicalmente distintas, ya que el ascenso a grandes altitudes reduce la disponibilidad de oxígeno, lo que puede provocar fatiga prematura y una disminución del rendimiento físico en equipos no aclimatados, afectando su desempeño competitivo. Esta heterogeneidad hace imprescindible considerar la altitud en cualquier intento por modelar o explicar el resultado y los goles anotados por el equipo visitante, un ejemplo ilustrativo es el caso de la selección de Brasil, que logró un triunfo histórico el 29 de marzo de 2022 al vencer 4–0 a Bolivia en La Paz (Estadio Hernando Siles), su primera victoria en ese estadio en 37 años, como reporta Infobae (2022).

En conjunto, estas diferencias estructurales justifican incorporar la altitud como un factor clave, junto con el rendimiento, en los modelos destinados a explicar el resultado del partido y la capacidad anotadora de los equipos visitantes del Fútbol Profesional Colombiano.

1.2 Pregunta de Investigación

¿Hay relación entre el desempeño del equipo al final del partido, medido a través de un índice, y la altura, con el resultado y los goles anotados en el Fútbol Profesional Colombiano?

1.3 Justificación

Castellano (2018) afirma que existe un interés creciente por identificar qué indicadores de rendimiento representan mejor, o se asocian de manera más consistente, con el resultado final en partidos o campeonatos de fútbol. En el presente trabajo se parte de un índice que mide el desempeño construido a partir de diversos indicadores técnicos y tácticos, dado que el análisis del rendimiento se ha consolidado como un componente fundamental para alcanzar el éxito deportivo en el fútbol de élite, como señalan Lago-Peñas, Casáis, Dellal, Rey e Domínguez (2010).

No obstante, conviene recordar que el resultado final y el número de goles anotados de un equipo no depende únicamente de su desempeño técnico y táctico. También la altitud desempeña un papel relevante, pues diversos autores como Bartsch et al. (2008), Faude et al. (2011), Gore et al. (2008) y McSharry (2007) han señalado que los cambios de altitud pueden afectar la condición física y la salud de los jugadores.

Lago-Peñas (2008) indica que, en los deportes de equipo la identificación de las causas es esencial para comprender los fenómenos y construir explicaciones válidas. En términos generales, el concepto de causa se fundamenta en comparar lo que efectivamente ocurrió bajo ciertas condiciones con lo que habría podido ocurrir bajo condiciones alternativas.

Encontrar los factores que muestran una mayor relación con el resultado final y con el número de goles anotados resulta relevante para que los entrenadores identifiquen fortalezas que pueden potenciarse mediante el entrenamiento, así como debilidades que deben corregirse antes de la competición. De este modo, es posible orientar el trabajo táctico y estratégico hacia el escenario ideal: ganar el partido y, en lo posible, anotar la mayor cantidad de goles.

Abordar la complejidad inherente al fútbol requiere apoyarse en herramientas analíticas rigurosas, entre las cuales la estadística desempeña un papel central. En este marco, la literatura distingue dos enfoques empíricos principales para modelar el comportamiento competitivo en el fútbol: uno orientado a modelar directamente el resultado del partido (victoria, empate o derrota) y otro centrado en modelar el número de goles anotados y recibidos según Lee (2005). En el presente trabajo se adopta una aproximación que incorpora ambos enfoques, al modelar tanto el resultado final como los goles anotados exclusivamente, por los equipos en condición de visitante

Aunque este trabajo inicia con un modelo Poisson simple para evaluar la relación entre rendimiento, altitud y goles anotados por los visitantes, el supuesto de independencia puede verse comprometido, ya que los goles de un mismo equipo podrían estar correlacionados a lo largo de la temporada. Coxe, West e Aiken (2009) señalan que la regresión Poisson, como miembro de la familia de los modelos

lineales generalizados, es ampliamente utilizada para variables de conteo y puede proporcionar estimaciones precisas en este tipo de datos. Sin embargo, cuando existe dependencia entre las observaciones, este enfoque resulta insuficiente. En este contexto, los modelos lineales generalizados mixtos (GLMM), el modelo Poisson mixto para ser preciso, se convierten en una alternativa adecuada, pues permiten capturar esta estructura no independiente mediante la incorporación de efectos aleatorios que recogen la variabilidad propia entre equipos.

De forma complementaria, cuando se busca modelar el resultado del partido (ganar, empatar o perder), los GLM con enlace logit multinomial permiten capturar la naturaleza categórica del desenlace. Dado que el número de resultados posibles del partido (victoria, derrota y empate) es mayor a dos, se utiliza el enfoque de regresión logística multinomial para modelar dichos resultados (Bland e Altman 2000). Estos modelos pueden compararse con el enfoque ordinal para determinar cuál describe mejor la relación entre el resultado y los factores de interés.

Aunque existen diversos estudios que modelan indicadores de resultado, como los goles anotados o el desenlace del partido, la literatura no ha considerado un índice sintético que mide el desempeño de los equipos al final de partido y que integre múltiples aspectos del juego como variable explicativa. Incorporar este tipo de índice permite una representación más completa del comportamiento competitivo y favorece la construcción de modelos más parsimoniosos.

1.4 Objetivos

1.4.1 Objetivo General

Evaluar la relación entre el desempeño de los equipos de fútbol al final del partido, medido a través de un índice, y la altura, con el resultado y la cantidad de goles anotados.

1.4.2 Objetivos Específicos

- Modelar la relación entre altura y el desempeño del equipo medido a través de un índice al final del partido y su influencia en el número de goles anotados.
- Modelar la relación entre altura y el desempeño del equipo medido a través de un índice al final del partido y su influencia en el resultado.

Capítulo 2

Antecedentes

En este capítulo se presenta una revisión bibliográfica de investigaciones previas que abordan temas similares, considerando tanto el enfoque contextual como el metodológico.

1. Construcción de un indicador para evaluar el rendimiento de equipos de fútbol usando variables latentes.

Chávez-Avendaño (2025) desarrolló un índice para medir el rendimiento de los equipos del Fútbol Profesional Colombiano tanto locales como visitantes al final de partido, empleando Análisis Factorial con un enfoque formativo. El rendimiento se conceptualizó como un constructo de segundo orden formado por tres dimensiones latentes: Control Ofensivo, Solidez Defensiva y Eficacia en la Finalización, cada una medida mediante indicadores técnico-tácticos (Figura 2.1). En particular, en este trabajo se utiliza el índice de rendimiento en condición de visitante, derivado de la estructura formativa propuesta por lo autores.

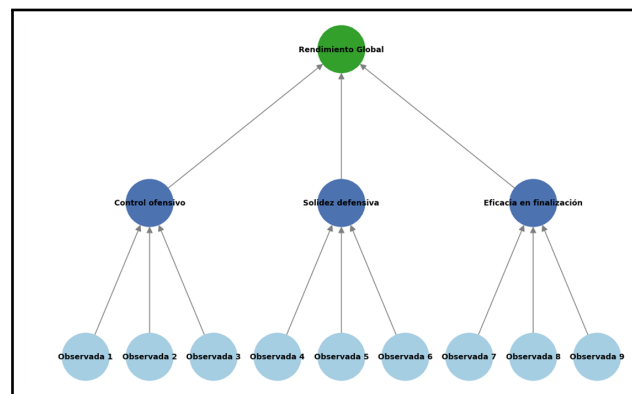


Figura 2.1: Modelo formativo visualizado mediante un gráfico.

Para medir las variables latentes de primer orden, se seleccionaron indicadores que aportan positivamente al rendimiento y se transformaron en proporciones de éxito entre 0 y 1, facilitando su interpretación y reduciendo la multicolinealidad. Las variables resultantes de esta construcción se presentan en la Tabla 2.1, las cuales corresponden a las variables observadas que conforman el modelo formativo ilustrado en la Figura 2.1.

ID	Variable	Construcción
X1	Tiros a portería	$\frac{\text{Tiros a portería}}{\text{Tiros totales}}$
X2	Pases efectivos	$\frac{\text{Pases efectivos}}{\text{Pases totales}}$
X3	Pases de finalización efectivos	$\frac{\text{Pases de finalización efectivos}}{\text{Pases de finalización}}$
X4	Centros efectivos	$\frac{\text{Centros efectivos}}{\text{Centros totales}}$
X5	Disputas en ataque ganadas	$\frac{\text{Disputas ofensivas ganadas}}{\text{Disputas ofensivas}}$
X6	Posesión del balón	Variable directa del conjunto de datos
X7	Progresión al último tercio	$\frac{\text{Entradas al último tercio}}{\text{Entradas a campo rival}}$
X8	Penetración al área	$\frac{\text{Entradas al área rival}}{\text{Entradas al último tercio}}$
X9	Ataques en área rival finalizados en tiro	$\frac{\text{Ataques por izquierda, centro y derecha finalizados en tiro}}{\text{Total de ataques por izquierda, centro y derecha}}$
X10	Disputas por arriba ganadas	$\frac{\text{Duelos aéreos ganados}}{\text{Duelos aéreos}}$
X11	Disputas defensivas ganadas	$\frac{\text{Duelos defensivos ganados}}{\text{Duelos defensivos}}$
X12	Entradas efectivas	$\frac{\text{Entradas exitosas para robar balón}}{\text{Entradas intentadas para robar balón}}$
X13	Intercepciones	$\frac{\text{Intercepciones}}{\text{Pases del rival}}$
X14	Tiros libres rematados	$\frac{\text{Tiros libres rematados}}{\text{Tiros libres a favor}}$
X15	Corners rematados	$\frac{\text{Corners rematados}}{\text{Corners a favor}}$
X16	Faltas graves	$\frac{\text{Penaltis (rival)} + \text{Tiros libres directos (rival)}}{\text{Total de faltas cometidas por rival}}$

Tabla 2.1: Construcción de variables a partir de indicadores de juego

El uso de la técnica de Análisis Factorial Exploratorio (AFE), permitió considerar en el análisis de rendimiento un gran conjunto de indicadores, y no limitarse a los más tradicionales en el análisis del rendimiento de los equipos fútbol según la literatura, como la posesión de balón, los puntos obtenidos, el hecho de marcar el primer gol, los remates directos al arco o las tarjetas amarillas, entre otros. El AFE permitió reducir la información a tres factores, los cuales explican el 33.17 % de la variabilidad del rendimiento del equipo en condición de visitante. Los resultados del Análisis Factorial Exploratorio (AFE),

incluyendo los factores rotados RC1, RC2 y RC3 obtenidos mediante una rotación Varimax para mejorar su interpretabilidad, se presentan en las Tablas 2.2, 2.3 y 2.4, de acuerdo con las definiciones propuestas por Chávez-Avendaño (2025).

Tabla 2.2: Cargas factoriales del factor 1 (Visitante)

Variable	RC1
Pases efectivos (X2)	0.1174
Disputas en ataque ganadas (X5)	0.2052
Posesión del balón (X6)	0.0979
Progresión al último tercio (X7)	0.1105
Ataques en área rival finalizados en tiro (X9)	0.1447
Disputas por arriba ganadas (X10)	0.1880
Tiros libres a favor rematados (X14)	0.0785
Corners a favor rematados (X15)	0.0578

El factor 1 de la Tabla 2.2 llamado **Control Ofensivo** incluye variables que reflejan cómo un equipo mantiene y controla el juego en la ofensiva, desde la posesión hasta las oportunidades de crear jugadas ofensivas.

Tabla 2.3: Cargas factoriales del factor 2 (Visitante)

Variable	RC2
Disputas defensivas ganadas (X11)	0.5534
Entradas efectivas (X12)	0.5239
Faltas graves (X16)	-0.0773

El factor 2 de la Tabla 2.3 llamado **Solidez Defensiva** agrupa variables que capturan la capacidad del equipo para defender de manera sólida y evitar que el rival genere situaciones de peligro.

Tabla 2.4: Cargas factoriales del factor 3 (Visitante)

Variable	RC3
Tiros a portería (X1)	0.1393
Pases de finalización efectivos (X3)	0.2919
Centros efectivos (X4)	0.2810
Penetración al área (X8)	0.2657
Interceptaciones (X13)	0.0222

El factor 3 de la Tabla 2.4 llamado **Eficacia en Finalización** incluye variables que reflejan la habilidad del equipo para crear y convertir oportunidades de gol.

Finalmente, el índice de rendimiento visitante se define como:

$$IRVM_{\text{visitante},i} = 0.040X_1 + 0.045X_2 + 0.085X_3 + 0.082X_4 + 0.078X_5 + 0.037X_6 + 0.042X_7 + 0.077X_8 + 0.055X_9 + 0.072X_{10} + 0.179X_{11} + 0.169X_{12} + 0.006X_{13} + 0.030X_{14} + 0.022X_{15} - 0.025X_{16}$$

Luego se normaliza a una escala de 0 a 100 mediante:

$$IRVM_{\text{norm},i} = \frac{IRVM_i - IRVM_{\min}}{IRVM_{\max} - IRVM_{\min}}$$

donde $IRVM_{\min}$ y $IRVM_{\max}$ son, respectivamente, el valor mínimo y máximo del índice calculado para todos los equipos.

Con esta normalización, 1 representa el máximo rendimiento y 0 el mínimo. Este trabajo utiliza únicamente el índice visitante, calculado para los 1898 partidos una vez finalizado cada encuentro, como variable explicativa junto con la altura sobre el nivel del mar con el fin de analizar su relación en el resultado del partido y con el número de goles anotados por el equipo visitante.

2. Factores asociados con el resultado del partido y el número de goles anotados y recibidos en la Premier League de Inglaterra.

Bilek e Aygün (2022) analizaron los partidos de la Premier League 2017–18 para identificar qué factores explican el resultado del partido (ganar, empatar o perder) y los goles anotados/recibidos. Para ello consideraron variables relacionadas con rendimiento ofensivo y defensivo, localía, calidad del oponente y antecedentes del equipo (como tiros, posesión, goles previos, despejes, entre otras).

Utilizaron Regresión de Poisson para modelar goles anotados y recibidos, y Regresión Logística Multinomial junto con Máquinas de Vectores de Soporte (SVM) para predecir el resultado del partido.

Los principales hallazgos fueron:

Goles anotados: dependen principalmente de anotar primero, los tiros al arco y los goles recibidos.

Goles recibidos: disminuyen si el equipo anota primero, juega de local, realiza más tiros y más despejes; aumentan si el rival es de mayor calidad.

Resultado del partido: influyen anotar primero, la localía, la cantidad de tiros, los tiros al arco, los despejes, la calidad del oponente y la posesión del balón.

El resultado más relevante del estudio es que anotar el primer gol fue la única variable significativa en todos los modelos, convirtiéndose en el factor más determinante del desempeño y del éxito en la Premier League.

3. Copa Libertadores de la CONMEBOL: Impacto de la altitud en goles y resultados en 16 años de partidos de fútbol.

Cabrera-Hernández, García-Corzo, Povea-Combariza et al. (2023) realizaron un estudio sobre los partidos de fútbol de clubes jugados entre los años 2000 y 2015 en la Copa Libertadores de la CONMEBOL (se analizaron 2039 partidos), con el objetivo de describir el efecto de la altitud en los resultados y goles de los partidos.

Para ello, los autores clasificaron a los equipos según las categorías de altitud propuestas por Bärtsch et al. (2008)

Nivel del mar (A): 0 a 500 m.s.n.m.

Nivel bajo (B): 501 a 2000 m.s.n.m.

Altitud moderada (C): 2001 a 3000 m.s.n.m.

Altitud alta (D): 3001 a 5500 m.s.n.m.

Altitud extrema (E): Más de 5500 m.s.n.m.

Aunque la clasificación original llega hasta los 5500 m s. n. m., en la Copa Libertadores no existen estadios a esas altitudes, por lo que solo se consideraron las categorías entre 0 y 3500 metros.

Se definieron diversas combinaciones, clasificadas según el cambio en la categoría del equipo visitante: mantenerse en la misma categoría (MA), ascender 1, 2 o 3 categorías (Asc1, Asc2, Asc3) o descender 1, 2 o 3 categorías (Des1, Des2, Des3).

Para evaluar la influencia de la altitud, se utilizó un Modelo Lineal Generalizado de Poisson que permitió estudiar la relación entre la altitud de los equipos (local y visitante) y el número de goles marcados por cada uno, considerando su categoría. Además, para estimar la probabilidad de que un equipo local ganara, empatara o perdiera un partido, se ajustó un modelo de regresión para variables ordinales bajo la suposición de una distribución multinomial y utilizando el enlace logístico.

Finalmente, uno de los principales hallazgos de este estudio fue que el descenso de altitud tuvo un mayor impacto negativo en el rendimiento del equipo visitante. En particular, las probabilidades de perder aumentaron considerablemente cuando el equipo visitante debía descender dos o tres categorías

de altitud, o cuando ascendía tres categorías. Esto se tradujo en una menor probabilidad de anotar goles y una mayor probabilidad de recibirlos, reflejando tanto el impacto fisiológico de los cambios de altitud como el de las condiciones ambientales asociadas.

Capítulo 3

Marco teórico

3.1 Marco Contextual

3.1.1 Altitud

La altitud se define como la distancia vertical medida desde el nivel medio del mar hasta un punto determinado. Esta se expresa en metros sobre el nivel del mar (m.s.n.m.) o en metros bajo el nivel del mar (m.b.n.m.), dependiendo de si el lugar se encuentra por encima o por debajo de dicho nivel (Barrera Bassols y Palma Ruiz, 2008).

3.1.2 Categorización de la altitud

La entidad que rige el fútbol mundial FIFA (Federation Internationale de Football Association) trabaja con la siguiente categorización de la altitud (Bartsch et al., 2008):

Tabla 3.1: Categoría de altitud para el estadio según sus metros sobre el nivel del mar.

Altitud (m.s.n.m.)	Categoría de Altitud
0 a 500 m.s.n.m	Nivel del mar (A)
501 a 2000 m.s.n.m	Nivel bajo (B)
2001 a 3000 m.s.n.m	Altitud moderada (C)
3001 a 5500 m.s.n.m	Altitud alta (D)
Mayor que 5500 m.s.n.m	Altitud extrema (E)

3.1.3 Índice

Según Silva (1997), los índices son construcciones teóricas utilizadas para medir conceptos complejos mediante la agregación de múltiples indicadores o variables. En estadística, un índice sintetiza la información de diversas variables para proporcionar una medida única y representativa de un fenómeno específico.

3.1.4 Indicador de desempeño

En el contexto del fútbol, un indicador de desempeño es una medida utilizada comúnmente para evaluar ya sea el rendimiento de jugadores, del equipo, de los entrenadores, o el conjunto de todos estos. Existe una amplia variedad de métricas para el desarrollo de estos indicadores, como lo son la posesión del balón, pases efectivos, tiros a portería, número de goles, entre otros. Gracias a estos indicadores, se puede realizar una evaluación objetiva y cuantificable del rendimiento, permitiendo así identificar deficiencias en tácticas, y tomar las respectivas decisiones para finalizar un partido con resultado positivo Gómez y Hernández (2016).

Tabla 3.2: Descripción de los diferentes indicadores del rendimiento que han sido utilizados en el análisis de los deportes en equipos, según Lago-Peñas (2008)

Categoría	Indicadores
Clasificación del partido (Match classification)	Goles anotados, número de tiros a portería, tiros fuera, saques de esquina, centros, entre otros (Hughes (1993)).
Biomecánicos (Biomechanical)	<i>Golpeo (Kicking)</i> : velocidad y efecto del balón, cinemática y cinética de la pierna de golpeo, transferencias de energía, secuenciación de acciones articulares, fuerzas y momentos articulares netos. <i>Saques de banda (Throw-in)</i> : velocidad de liberación del balón, cinemática de los brazos, secuencia de velocidades segmentarias máximas (Lee (1998), Polmam (1993)).
Técnicos (Technical)	Pases al oponente, entradas ganadas y pérdidas, tiros fuera, regates, pérdidas de control, centros a portería y fuera, entre otros (Hughes (1993), Pettit (2001)).
Tácticos (Tactical)	Pases/posesión, ritmo de ataque, tiros, entradas ganadas y perdidas, distribución de pases, longitud de los pases, regates, entre otros (Hughes (1993), Pettit (2001)).

3.2 Marco Estadístico

3.2.1 Modelo de Regresión de Poisson

Se supondrá que Y_i , la variable de respuesta, es un conteo, es decir, una variable aleatoria que toma valores enteros no negativos, tales como $Y_i = 0, 1, 2, \dots$. Un modelo probabilístico razonable para este tipo de datos es, con frecuencia, la distribución de Poisson, Montgomery, Peck e Vining (2012).

La función de masa de probabilidad de una variable aleatoria $Y \sim \text{Poisson}(\mu)$ está dada por:

$$P(Y = y) = \frac{e^{-\mu} \mu^y}{y!}, \quad y = 0, 1, 2, \dots,$$

donde el parámetro $\mu > 0$ representa la media o número esperado de ocurrencias del evento.

La distribución de Poisson presenta la propiedad de que su media y su varianza coinciden, es decir:

$$\mathbb{E}(Y) = \mu, \quad \mathbb{V}(Y) = \mu.$$

Esta característica implica que tanto el valor esperado como la variabilidad del conteo están determinados por el mismo parámetro μ .

Especificación del modelo de regresión

El modelo de regresión de Poisson se formula para un conjunto de n observaciones independientes, indexadas por $i = 1, 2, \dots, n$. Se asume que la variable respuesta Y_i sigue una distribución de Poisson con media μ_i , de modo que su valor esperado puede escribirse como

$$\mathbb{E}(Y_i) = \mu_i.$$

En el marco de los modelos lineales generalizados, se supone que existe una función $g(\cdot)$, denominada función de enlace, que relaciona la media condicional de la variable respuesta con un predictor lineal de la forma

$$g(\mu_i) = \eta_i = \beta_0 + \beta_1 x_{i1} + \dots + \beta_k x_{ik} = \mathbf{x}_i^\top \boldsymbol{\beta}.$$

De manera equivalente, la relación entre la media de la respuesta y el predictor lineal puede expresarse como

$$\mu_i = g^{-1}(\eta_i) = g^{-1}(\mathbf{x}_i^\top \boldsymbol{\beta}).$$

Existen diversas funciones de enlace que pueden emplearse con la distribución de Poisson. No obstante, la más utilizada es la función de enlace logarítmica, definida como

$$g(\mu_i) = \log(\mu_i) = \mathbf{x}_i^\top \boldsymbol{\beta}.$$

Bajo esta especificación, la media condicional de la variable respuesta se expresa como

$$\mu_i = \exp(\mathbf{x}_i^\top \boldsymbol{\beta}).$$

La función de enlace logarítmica resulta especialmente atractiva en el modelo de regresión de Poisson, ya que garantiza que los valores predichos de la media condicional sean estrictamente positivos, lo cual es coherente con la naturaleza de los datos de conteo. Este enfoque es ampliamente discutido en la literatura clásica sobre modelos de regresión para datos discretos (Montgomery, Peck e Vining 2012).

Estimación de los parámetros

Los parámetros del modelo de regresión de Poisson se estiman mediante el método de máxima verosimilitud. Supóngase que se dispone de una muestra aleatoria de n observaciones independientes de la variable respuesta Y_i , $i = 1, 2, \dots, n$, donde cada $Y_i \sim \text{Poisson}(\mu_i)$.

Bajo este supuesto, la función de verosimilitud del modelo está dada por

$$L(\boldsymbol{\beta}) = \prod_{i=1}^n \frac{e^{-\mu_i} \mu_i^{y_i}}{y_i!},$$

donde $\mu_i = g^{-1}(\mathbf{x}_i^\top \boldsymbol{\beta})$ y $g(\cdot)$ es la función de enlace.

Al tomar logaritmos, se obtiene la función de log-verosimilitud

$$\ell(\boldsymbol{\beta}) = \sum_{i=1}^n [y_i \log(\mu_i) - \mu_i - \log(y_i!)] .$$

Dado que las ecuaciones de verosimilitud no admiten una solución analítica cerrada, los estimadores de máxima verosimilitud de los parámetros se obtienen mediante procedimientos numéricos iterativos, tales como el algoritmo de Newton–Raphson o el método de mínimos cuadrados iterativamente ponderados (*Iteratively Reweighted Least Squares*, IRLS), de forma análoga a lo utilizado en la regresión logística (Montgomery, Peck e Vining 2012).

Una vez obtenidos los estimadores $\hat{\beta}$, el modelo ajustado de regresión de Poisson se expresa como

$$\hat{\mu}_i = \exp(\mathbf{x}_i^\top \hat{\beta}).$$

Inferencia estadística

La inferencia sobre el modelo de regresión de Poisson y sus parámetros se basa en la teoría de máxima verosimilitud. La desviación del modelo constituye una medida global de bondad de ajuste y permite la comparación entre modelos anidados mediante pruebas de razón de verosimilitud, evaluando la diferencia de desviaciones entre un modelo completo y uno reducido.

Asimismo, para realizar pruebas de hipótesis e intervalos de confianza sobre parámetros individuales del modelo, se emplea la inferencia de Wald, la cual se sustenta en las propiedades asintóticas de los estimadores de máxima verosimilitud bajo el supuesto de muestras grandes (Montgomery, Peck e Vining 2012).

Interpretación de los parámetros

En el modelo de regresión de Poisson con función de enlace logarítmica, la interpretación de los parámetros es de naturaleza multiplicativa. En particular, el término $e^{\hat{\beta}_0}$ representa el valor esperado de la variable respuesta correspondiente a la categoría de referencia o cuando todas las variables explicativas toman el valor cero.

Según, Gómez Gallego (2019) para los coeficientes se tiene que:

- Si $\beta_j > 0$, un incremento unitario en la variable explicativa x_j , manteniendo constantes las demás variables, produce un aumento porcentual de $(e^{\hat{\beta}_j} - 1) \times 100$ en la media condicional de la respuesta.
- Si $\beta_j < 0$, dicho incremento genera una disminución porcentual de $(e^{\hat{\beta}_j} - 1) \times 100$ en la media condicional de la respuesta.

En el caso de variables cualitativas, esta interpretación se realiza en términos de la comparación entre la categoría considerada y la categoría de referencia, manteniendo constantes las demás covariables del modelo.

3.2.2 Estimación del parámetro de sobredispersión

En una distribución de Poisson, la media y la varianza son iguales, pero cuando se trabaja con recuentos reales no suele cumplirse este supuesto. Con frecuencia, la varianza es mayor que la media; a esto se le llama sobredispersión. Este nivel de dispersión a menudo se caracteriza por un parámetro de dispersión ϕ , que es igual a 1 si no hay sobredispersión, es mayor que 1 cuando hay sobredispersión, y es menor que 1 en caso de subdispersión (Cohen, 2014).

En el contexto del Fútbol Profesional Colombiano, la presencia de una cantidad considerable de partidos en los que los equipos visitantes no anotan goles genera una distribución asimétrica y con exceso de ceros, lo cual incrementa la probabilidad de sobredispersión. Ignorar este fenómeno puede conducir a una subestimación de los errores estándar de los coeficientes, aumentando artificialmente la significancia estadística de las covariables y afectando la correcta selección e interpretación del modelo.

La sobredispersión ha sido ampliamente discutida en la literatura como un problema frecuente en modelos de conteo. King (1989) señala que una de sus principales causas es la omisión de variables explicativas relevantes que, aun cuando no son observadas o medidas, influyen directamente sobre la variable respuesta. Adicionalmente, Osgood (2000) destaca que la presencia de dependencia entre las observaciones constituye otra fuente importante de sobredispersión, ya que dicha situación viola el supuesto de independencia inherente a la distribución de Poisson. En el análisis de datos deportivos, ambas condiciones resultan plausibles, dado que el rendimiento de los equipos puede verse afectado por factores no observados y por dinámicas propias de la competencia.

Por estas razones, la estimación del parámetro de sobredispersión se incorpora como un paso clave en el análisis, permitiendo evaluar la adecuación del modelo Poisson y, de ser necesario, considerar ajustes metodológicos que garanticen una inferencia estadística válida.

De acuerdo con Bingham (2010), Faraway (2016) y Yan (2009), los residuales de Pearson se definen como la diferencia entre el valor observado y el valor esperado bajo el modelo, escalada por la desviación estándar teórica de la variable respuesta.

$$e_{P_i} = \frac{y_i - \hat{\mu}_i}{\sqrt{\text{Var}(\hat{\mu}_i)}} \quad (3.1)$$

donde y_i representa el valor observado y $\hat{\mu}_i$ corresponde a la estimación del modelo para la i -ésima observación. A partir de estos residuales, el estadístico chi-cuadrado de Pearson se define, siguiendo a Bingham (2010), Faraway (2016) y McCullagh (1989), como:

$$\chi^2 = \sum_{i=1}^n \frac{(y_i - \hat{\mu}_i)^2}{\text{Var}(\hat{\mu}_i)} \quad (3.2)$$

donde $\text{Var}(Y_i) = \phi \cdot \text{Var}(\hat{\mu}_i)$.

El estadístico chi-cuadrado escalado de Pearson se define como:

$$\chi_s^2 = \frac{\chi^2}{\phi} \quad (3.3)$$

Si el modelo está correctamente especificado, entonces asintóticamente:

$$\chi_s^2 \approx \frac{\chi^2}{n - p}$$

Como la media de una variable aleatoria χ_{n-p}^2 es $n - p$, se usa la aproximación:

$$\chi_s^2 \approx n - p$$

En consecuencia, la estimación del parámetro de dispersión se define de acuerdo con McCullagh (1989), Wedderburn (1974) y Tutz (2011), como:

$$\hat{\phi} = \frac{\chi^2}{n - p} = \frac{1}{n - p} \sum_{i=1}^n \frac{(y_i - \hat{\mu}_i)^2}{\text{Var}(\hat{\mu}_i)} \quad (3.4)$$

3.2.3 Indicadores de calidad de ajuste del modelo

Criterio de información de Akaike (AIC)

Según Dobson y Barnett (2018), el AIC constituye un estadístico alternativo de bondad de ajuste para modelos estadísticos, el cual combina la información proveniente de la función de log-verosimilitud (ℓ) y el número de parámetros

estimados (p). Formalmente, este criterio se define siguiendo a Dobson y Barnett (2018) y Yan (2009) como:

$$\text{AIC} = -2\ell + 2p \quad (3.5)$$

Por lo cual, un valor pequeño de este estadístico indica que el modelo está bien ajustado.

Pseudo- R^2

Según Dobson y Barnett (2018), se define otro estadístico utilizado para evaluar la bondad del ajuste del modelo, el cual es el Pseudo- R^2 o Pseudo- R^2 de McFadden, de acuerdo con Fernández y Fernández (2004). Este estadístico viene dado por:

$$R_{\text{MF}}^2 = \text{Pseudo-}R^2 = 1 - \frac{D_1}{D_0}, \quad (3.6)$$

donde D_1 y D_0 son las desviaciones del modelo ajustado y del modelo nulo (mínimo), respectivamente. Estas desviaciones se definen en términos de la log-verosimilitud como

$$D_1 = -2\ell_1 = -2 \log L_1, \quad D_0 = -2\ell_0 = -2 \log L_0,$$

siendo ℓ_1 la log-verosimilitud del modelo de interés (ajustado con las covariables) y ℓ_0 la log-verosimilitud del modelo nulo es decir, el modelo que contiene únicamente el intercepto (sin covariables). Así, la razón D_1/D_0 cuantifica la proporción de la desviación del modelo nulo que permanece en el modelo ajustado, y por tanto $1 - D_1/D_0$ mide la mejora proporcional en la log-verosimilitud debida a las covariables incluidas.

El rango teórico es $0 \leq R_{\text{MF}}^2 \leq 1$, aunque en la práctica raramente se observan valores cercanos a 1; típicamente se considera que $0.2 \leq R_{\text{MF}}^2 \leq 0.4$ indica un ajuste razonable en modelos de elección discreta.

Fernández y Fernández (2004) señalan que otros autores prefieren coeficientes basados directamente en la verosimilitud y no únicamente en la razón de desviaciones. Una alternativa común es el Pseudo- R^2 de Nagelkerke (1991), definido como:

$$R_N^2 = \frac{1 - \exp\left(-\frac{D_0 - D_1}{n}\right)}{1 - \exp\left(-\frac{D_0}{n}\right)}, \quad (3.7)$$

donde n es el tamaño de la muestra. Obsérvese que en la literatura a veces aparece la forma equivalente usando log-verosimilitudes ℓ_0 y ℓ_1 :

$$R_N^2 = \frac{1 - \exp\left(\frac{2(\ell_0 - \ell_1)}{n}\right)}{1 - \exp\left(\frac{2\ell_0}{n}\right)}.$$

El rango de R_N^2 también es $0 \leq R_N^2 \leq 1$ y su interpretación es análoga a la del R^2 en regresión lineal, aunque igualmente es poco frecuente que se acerque a 1.

3.2.4 Modelo de Regresión Multinomial

De acuerdo con Dobson (2018), la distribución multinomial se utiliza para modelar variables aleatorias con múltiples categorías. En este contexto, se considera una variable respuesta Y con J categorías, cuyas probabilidades asociadas son $\pi_1, \pi_2, \dots, \pi_J$, sujetas a la restricción

$$\pi_1 + \pi_2 + \dots + \pi_J = 1.$$

Si se dispone de n observaciones independientes de Y , con y_1 observaciones en la categoría 1, y_2 en la categoría 2, y así sucesivamente, entonces el vector de conteos puede expresarse como:

$$\mathbf{y} = [y_1, y_2, \dots, y_J]^T \quad \text{con} \quad \sum_{j=1}^J y_j = n$$

La distribución multinomial está dada por:

$$f(\mathbf{y}; n) = \frac{n!}{y_1! y_2! \dots y_J!} \pi_1^{y_1} \pi_2^{y_2} \dots \pi_J^{y_J} \quad (3.8)$$

La cual se denota como $M(n; \pi_1, \pi_2, \dots, \pi_J)$. Aunque la distribución multinomial no pertenece estrictamente a la familia exponencial, puede establecerse

una relación con la distribución Poisson. Sean $Y_1, \dots, Y_J$ variables aleatorias independientes con $Y_j \sim \text{Po}(\lambda_j)$, su función de probabilidad conjunta es:

$$f(\mathbf{y}) = \prod_{j=1}^J \frac{\lambda_j^{y_j} e^{-\lambda_j}}{y_j!} \quad (3.9)$$

Sea $n = \sum_{j=1}^J Y_j$, entonces n sigue una distribución $\text{Poisson}(\lambda_1 + \dots + \lambda_J)$. Por lo tanto, la distribución condicional de $\mathbf{y}$ dado n es:

$$f(\mathbf{y} | n) = \frac{n!}{y_1! y_2! \dots y_J!} \left(\frac{\lambda_1}{\sum \lambda_j} \right)^{y_1} \dots \left(\frac{\lambda_J}{\sum \lambda_j} \right)^{y_J} \quad (3.10)$$

Si definimos $\pi_j = \lambda_j / \sum \lambda_j$, entonces la ecuación anterior coincide con la expresión de la distribución multinomial en (3.8).

De acuerdo con Chatterjee (2013), cuando la variable respuesta es de tipo nominal, es decir, no presenta un orden natural entre sus categorías, la regresión logística multinomial constituye una extensión natural del modelo de regresión logística binaria. En este enfoque, se construyen $J - 1$ relaciones logísticas tomando una de las categorías como referencia (usualmente la categoría J). En particular, para $j = 1, \dots, J - 1$, la relación se expresa como:

$$\log \left(\frac{\pi_j(x)}{\pi_J(x)} \right) = \beta_{0j} + \beta_{1j}x_1 + \dots + \beta_{pj}x_n = \mathbf{x}^T \boldsymbol{\beta}_j \quad (3.11)$$

Una vez estimados los coeficientes $\boldsymbol{\beta}_j$, se pueden calcular los predictores lineales $\mathbf{x}^T \boldsymbol{\beta}_j$, y las probabilidades estimadas son:

$$\pi_j(x) = \frac{\exp(\mathbf{x}^T \boldsymbol{\beta}_j)}{1 + \sum_{k=1}^{J-1} \exp(\mathbf{x}^T \boldsymbol{\beta}_k)}, \quad j = 1, \dots, J - 1 \quad (3.12)$$

$$\pi_J(x) = \frac{1}{1 + \sum_{k=1}^{J-1} \exp(\mathbf{x}^T \boldsymbol{\beta}_k)} \quad (3.13)$$

Es común interpretar los efectos de las variables explicativas en términos de odds ratios (OR). Supongamos una variable binaria x que indica exposición ($x = 1$) o no ($x = 0$). El odds ratio de exposición para la categoría j frente a la de referencia (J) es:

$$\text{OR}_j = \frac{\pi_j^{(1)} / \pi_J^{(1)}}{\pi_j^{(0)} / \pi_J^{(0)}} = \exp(\beta_{1j}) \quad (3.14)$$

Por lo tanto: - Si $\beta_{1j} = 0 \Rightarrow \text{OR}_j = 1$ (sin efecto de la exposición). - Si $\beta_{1j} > 0 \Rightarrow \text{OR}_j > 1$ (efecto positivo). - Si $\beta_{1j} < 0 \Rightarrow \text{OR}_j < 1$ (efecto negativo).

Como señala Chatterjee (2013), el valor $\exp(\beta_{jk})$ representa el cambio multiplicativo en los odds de estar en la categoría j frente a la de referencia J asociado a un incremento de una unidad en x_k , manteniendo las demás variables constantes.

La elección de la categoría de referencia es arbitraria y no afecta la validez del modelo. Si en lugar de J , se elige otra categoría M como referencia, la ecuación correspondiente es:

$$\log \left(\frac{\pi_j(x)}{\pi_M(x)} \right) = (\beta_{0j} - \beta_{0M}) + (\beta_{1j} - \beta_{1M})x_1 + \cdots + (\beta_{pj} - \beta_{pM})x_p \quad (3.15)$$

Esto implica que los coeficientes dependen de la categoría de referencia utilizada, pero el modelo subyacente es el mismo. Finalmente, el modelo multinomial asume el supuesto de independencia de alternativas irrelevantes (IIA), lo que implica que la elección entre dos categorías no debe verse afectada por la presencia de otras categorías.

Este supuesto es razonable cuando las categorías de respuesta son claramente distintas entre sí. En casos donde este supuesto no se cumple, se recomienda utilizar modelos alternativos como el modelo logit anidado o modelos de clases latentes.

3.2.5 Modelo de Regresión Ordinal

Las variables de respuesta ordinal se caracterizan por poseer categorías con un orden natural, aunque sin una distancia cuantificable entre ellas. Para este tipo de datos, los modelos ordinales permiten capturar tendencias monótonas, de manera que la probabilidad de ubicarse en categorías superiores (o inferiores) varía sistemáticamente conforme lo hacen las covariables. Estos modelos resultan más parsimoniosos que los utilizados para respuestas nominales, al requerir un menor número de parámetros al incorporar el orden inherente de la variable respuesta Agresti (2015).

Modelo Logit Acumulativo con Odds Proporcionales

Sea y_i la respuesta ordinal observada para el individuo i , con categorías ordenadas

$$1 < 2 < \cdots < c.$$

El modelo se formula a partir de las probabilidades acumuladas

$$P(y_i \leq j) = \pi_{i1} + \cdots + \pi_{ij}, \quad j = 1, \dots, c-1,$$

las cuales representan la probabilidad de que la respuesta se encuentre en la categoría j o en alguna inferior. A partir de estas probabilidades se define el logit acumulado:

$$\text{logit}[P(y_i \leq j)] = \log \left(\frac{P(y_i \leq j)}{P(y_i > j)} \right), \quad j = 1, \dots, c-1.$$

Cada uno de estos logits utiliza todas las categorías de la respuesta, convirtiendo el problema ordinal en una serie de comparaciones binarias acumuladas.

El modelo logit acumulativo, también conocido como modelo de odds proporcionales, se expresa como

$$\text{logit}[P(y_i \leq j)] = \alpha_j + \mathbf{x}_i^\top \boldsymbol{\beta}, \quad j = 1, \dots, c-1.$$

En esta formulación, los parámetros α_j representan los puntos de corte asociados a cada categoría acumulada y satisfacen $\alpha_1 < \alpha_2 < \cdots < \alpha_{c-1}$, dado que las probabilidades acumuladas aumentan con j . El vector $\boldsymbol{\beta}$ contiene los efectos de las covariables, los cuales son comunes a todos los logits acumulativos.

Interpretación. El modelo implica que los efectos de las covariables son constantes para todos los puntos de corte, propiedad conocida como *odds proporcionales*. Para dos niveles u y v de una covariable continua, se cumple

$$\text{logit}[P(y_i \leq j \mid x_i = u)] - \text{logit}[P(y_i \leq j \mid x_i = v)] = (u - v)\boldsymbol{\beta},$$

lo que significa que las *odds* acumuladas a nivel u son $\exp[(u - v)\boldsymbol{\beta}]$ veces las correspondientes a nivel v . De manera equivalente, un incremento de una unidad en la covariable x_{ik} multiplica las odds de ubicarse en la categoría j o alguna inferior por $\exp(\beta_k)$, ajustando por las demás covariables. Esta relación permanece constante para todos los valores de j .

Propiedades. El modelo respeta el orden de las categorías: si se invierte su orden, el modelo sigue siendo válido ajustando el signo de $\boldsymbol{\beta}$. Sin embargo, la estructura ordinal se pierde si las categorías se permutan arbitrariamente. Adicionalmente, como todos los logits comparten los mismos coeficientes, las

curvas resultantes no se cruzan, y el modelo impone una relación monótonamente creciente o decreciente entre la variable explicativa y la probabilidad de responder en categorías superiores.

En conjunto, el modelo logit acumulativo constituye la herramienta estándar para el análisis de respuestas ordinales, proporcionando interpretaciones directas mediante odds-ratios y un ajuste eficiente gracias al supuesto de proporcionalidad de efectos.

3.2.6 Modelo Poisson GLMM para datos de conteo correlacionados

Según Agresti (2015), en datos de conteo, los modelos de mezcla proporcionan una forma flexible de manejar la sobredispersión. Al mezclar la distribución Poisson con una distribución gamma para su media, se obtiene marginalmente la distribución binomial negativa. Una alternativa consiste en mezclar la media log-Poisson con un efecto aleatorio normal. La estructura del GLMM con enlace log es

$$\log [\mathbb{E}(y_{ij} \mid u_i)] = x_{ij}\beta + z_{ij}u_i,$$

donde $u_i \sim N(0, \Sigma_u)$. Condicionado en u_i , y_{ij} sigue una distribución Poisson. También es posible usar el enlace identidad, pero presenta un problema estructural: con un intercepto aleatorio y $\sigma_u > 0$, existe probabilidad positiva de que el predictor lineal sea negativo. El modelo binomial negativo se obtiene cuando $\exp(u_i)$ sigue una distribución gamma.

El Poisson GLMM con efectos aleatorios normales tiene la ventaja, respecto al GLM binomial negativo, de permitir de forma sencilla efectos aleatorios multivariados y modelos multinivel.

La versión con intercepto aleatorio del Poisson GLMM implica que el modelo marginal cumple

$$\mathbb{E}(y_{ij}) = \mathbb{E} [\mathbb{E}(y_{ij} \mid u_i)] = \mathbb{E} [\exp(x_{ij}\beta + u_i)] = \exp(x_{ij}\beta + \sigma_u^2/2).$$

Esto se debe a que $\mathbb{E}[\exp(u_i)] = \exp(\sigma_u^2/2)$, dado que una variable normal $N(0, \sigma_u^2)$ tiene función generadora de momentos $\mathbb{E}[\exp(tu_i)] = \exp(t^2\sigma_u^2/2)$.

Así, el logaritmo de la media condicional es $x_{ij}\beta + u_i$, y el logaritmo de la media marginal es $x_{ij}\beta + \sigma_u^2/2$. Se mantiene una estructura loglineal, y los efectos marginales de las covariables son iguales a los efectos específicos del

sujeto. Por tanto, la razón de medias entre dos valores de x_{ij} es la misma condicional y marginalmente.

La varianza marginal del modelo es:

$$\begin{aligned}\text{Var}(y_{ij}) &= \mathbb{E}[\text{Var}(y_{ij} \mid u_i)] + \text{Var}[\mathbb{E}(y_{ij} \mid u_i)] \\ &= \mathbb{E}[\exp(x_{ij}\boldsymbol{\beta} + u_i)] + \exp(2x_{ij}\boldsymbol{\beta}) \text{Var}[e^{u_i}] \\ &= \exp(x_{ij}\boldsymbol{\beta} + \sigma_u^2/2) + \exp(2x_{ij}\boldsymbol{\beta}) (e^{2\sigma_u^2} - e^{\sigma_u^2}) \\ &= \mathbb{E}(y_{ij}) + [\mathbb{E}(y_{ij})]^2 (e^{\sigma_u^2} - 1).\end{aligned}$$

Aquí $\text{Var}(e^{u_i}) = e^{2\sigma_u^2} - e^{\sigma_u^2}$, usando la función generadora de momentos en $t = 2$ y $t = 1$. Al igual que en el modelo binomial negativo, la varianza marginal es una función cuadrática de la media marginal.

El modelo Poisson ordinario se obtiene cuando $\sigma_u = 0$. Cuando $\sigma_u > 0$, la distribución marginal ya no es Poisson, y la diferencia entre varianza y media aumenta con σ_u . Marginalmente, las observaciones dentro de un mismo clúster presentan correlación no negativa, como en el GLMM binario.

3.2.7 Razón de Verosimilitudes

Hay varias maneras de probar las hipótesis sobre un vector de parámetros de tamaño p . Una de estas es comparar la bondad de ajuste de dos modelos. Dobson and Barnett (2018) definen que los modelos deben ser anidados o jerárquicos, es decir, tienen la misma distribución de probabilidad y la misma función de enlace, pero el componente lineal del modelo más simple M_0 es un caso especial del componente lineal del modelo más general M_1 .

Considere la hipótesis nula:

$$H_0 : \boldsymbol{\beta} = \boldsymbol{\beta}_0 = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_q \end{bmatrix}$$

correspondiente al modelo M_0 y una hipótesis más general:

$$H_1 : \boldsymbol{\beta} = \boldsymbol{\beta}_1 = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_p \end{bmatrix}$$

correspondiente a M_1 , donde $q < p < n$.

Se puede probar H_0 contra H_1 usando como estadístico la diferencia de las devianzas:

$$D = D_0 - D_1 = 2 [\ell(\hat{\beta}_1; y) - \ell(\hat{\beta}_0; y)]$$

Si el valor de D es consistente con la distribución $\chi^2_{(p-q)}$, generalmente se elige el modelo M_0 correspondiente a H_0 porque es más simple. Este resultado se utiliza para probar la hipótesis H_1 de la siguiente manera: si el valor de D está en la región crítica

$$\text{RR: } D > \chi^2_{(p-q), \alpha}$$

entonces rechazamos H_0 a favor de H_1 , debido a que el modelo M_1 proporciona una descripción significativamente mejor de los datos (aunque también podría no ajustarse particularmente bien).

Por otra parte, Dobson y Barnett (2018) indican que la desviación D es considerada como el estadístico de razón de verosimilitudes, el cual tiene una distribución muestral χ^2 con $m - p$ grados de libertad, donde m es el número de parámetros del modelo saturado y p es el número de parámetros del modelo de interés. Para las situaciones donde se tienen n observaciones Y_i , $i = 1, \dots, n$, todas con valores potencialmente diferentes para x_i^T , entonces el modelo saturado puede ser especificado con n parámetros.

Cuando las observaciones Y_i tienen distribución diferente a la Normal, la distribución de D puede ser solo aproximadamente χ^2 . Sin embargo, para las distribuciones binomial y Poisson, por ejemplo, D puede ser calculada y usarse directamente como estadística de bondad de ajuste. Debido a esto, puede establecerse la siguiente prueba de hipótesis:

H_0 : El modelo se ajusta bien a los datos

H_1 : El modelo no se ajusta bien a los datos

$$\text{RC: } D > \chi^2_{(N-p), \alpha}$$

3.2.8 Test de hipótesis e intervalos de confianza de Wald

Yan y Su (2009) indican que la máxima verosimilitud es un sistema bien establecido para la inferencia estadística. Por esto, los estimadores de máxima verosimilitud tienen muchas propiedades atractivas. Por ejemplo, son asintóticamente consistentes y eficientes. La normalidad cuando hay muestras grandes también está disponible en condiciones regulares:

$$\hat{\beta} \xrightarrow{d} N(\beta, I^{-1}) \quad \text{o} \quad N\{\beta, (X^\top W X)^{-1}\} \quad \text{cuando } n \rightarrow \infty$$

donde la notación $\xrightarrow{d}$ significa “converge en distribución a” y la matriz de varianzas y covarianzas es:

$$\text{Cov}(\hat{\beta}) = I^{-1} = (X^\top W X)^{-1}$$

El estadístico de Wald para probar la hipótesis nula $H_0 : \Lambda\beta = b$, donde Λ es $p \times (q + 1)$ de rango p , toma una forma similar al estadístico de la regresión lineal ordinaria:

$$G = (\Lambda\hat{\beta} - b)^\top \left\{ \Lambda(X^\top W X)^{-1} \Lambda^\top \right\} (\Lambda\hat{\beta} - b)$$

La distribución nula de G es una $\chi^2(p)$. Como un caso especial, para $H_0 : \beta_j = b$ contra $H_1 : \beta_j \neq b$, la prueba z se puede usar:

$$z = \sqrt{G} = \frac{\hat{\beta}_j - b}{\text{s.e.}(\hat{\beta}_j)} \sim N(0,1) \quad \text{bajo } H_0$$

Por otra parte, los intervalos de confianza de Wald para β_j vienen dados por:

$$\hat{\beta}_j \pm z_{\alpha/2} \times \text{s.e.}(\hat{\beta}_j)$$

Capítulo 4

Metodología

4.1 Archivo de datos

La base de datos proporcionada por Sergio Andrés García Corzo, profesor de la Escuela Nacional del Deporte, contiene mediciones detalladas de equipos de fútbol profesional colombiano en los años 2016 - 2020, abarcando indicadores tácticos, técnicos, resultado del equipo en cada partido (victoria, derrota o empate), número de goles anotado e información sobre la altura donde se juega el partido. La recolección de los datos se realizó manualmente mediante la observación de videos de los partidos. Este método implica una labor detallada de ingresar la información relevante a medida que se analizan los videos, permitiendo una recolección precisa de los indicadores necesarios para el estudio.

4.2 Variables incluidas en el análisis

En la Tabla 4.1 se presenta las variables utilizadas en este trabajo. Estas incluyen identificadores del equipo visitante, como el nombre del equipo; indicadores de resultado, tales como la cantidad de goles anotados y el resultado final del partido; métricas de desempeño representadas por un índice normalizado y sus tres factores (Control Ofensivo, Solidez Defensiva y Eficacia en Finalización); así como variables contextuales que capturan características del entorno, como la magnitud del desplazamiento altitudinal (delta de altitud) y la categoría de desplazamiento correspondiente al equipo visitante.

Tabla 4.1: Descripción de las variables utilizadas en el estudio

Variable	Definición	Tipo - Escala
Nombre de equipo visitante	Nombre del equipo que juega como visitante	Cualitativa - Nominal
Goles de equipo visitante	Número de goles anotados por el equipo visitante en el partido	Cuantitativa discreta - Razón
Resultado visitante	Resultado del equipo visitante (Victoria, Empate o Derrota)	Cualitativa - Ordinal
Índice de Rendimiento	Índice continuo que resume el rendimiento del equipo visitante	Cuantitativa continua - Razón
Índice de Rendimiento categórico	Versión categorizada del índice de rendimiento (bajo, medio, alto)	Cualitativa - Ordinal
Control Ofensivo	Componente del índice que representa acciones ofensivas del equipo	Cuantitativa continua - Razón
Control Ofensivo categórico	Versión categorizada del índice de Control Ofensivo (bajo, medio, alto)	Cualitativa - Ordinal
Solidez Defensiva	Componente del índice que refleja acciones defensivas del equipo	Cuantitativa continua - Razón
Solidez Defensiva categórica	Versión categorizada del índice de Solidez Defensiva (bajo, medio, alto)	Cualitativa - Ordinal
Eficacia en Finalización	Componente del índice asociado con la conversión de jugadas en goles	Cuantitativa continua - Razón
Eficacia en Finalización categórica	Versión categorizada del índice de Eficacia en Finalización (bajo, medio, alto)	Cualitativa - Ordinal
Categoría de desplazamiento del visitante	Categoría que define la distancia del equipo visitante para enfrentar al local	Cualitativa - Ordinal
Delta de altitud	Diferencia de altitud entre el estadio del local y el estadio del visitante	Cuantitativa continua - Razón

4.3 Formación de los grupos

Con base en la clasificación de altitud propuesta por Bartsch et al. (ver tabla 3.1), se estableció una categorización adaptada al contexto del Fútbol Profesional Colombiano. En este, solo se presentan partidos en ciudades cuya altitud no supera los 2822 m.s.n.m. Por lo tanto, únicamente se consideraron las tres primeras categorías:

A: Cerca del nivel del mar (0–500 m.s.n.m)

B: Baja altitud (500–2000 m.s.n.m)

C: Altitud moderada (2000–3000 m.s.n.m)

A partir de esta categorización, se construyeron combinaciones entre las altitudes donde se encuentra el estadio del equipo local y del equipo visitante según la ciudad.

En total, se identificaron las $3^2 = 9$ combinaciones posibles entre las tres categorías de altitud (A, B y C), correspondientes al producto cartesiano entre las categorías de altitud del equipo local y las del equipo visitante. Las combinaciones fueron clasificadas según la dirección del desplazamiento del equipo visitante y la altitud donde juega el equipo local.

Tabla 4.2: Combinaciones de altitud entre equipos local y visitante

Grupo	Combinaciones (Local, Visitante)	Tipo de desplazamiento
MA	(A, A), (B, B), (C, C)	Misma altitud (No hay desplazamiento)
Asc 1	(B, A), (C, B)	Ascenso de 1 categoría
Asc 2	(C, A)	Ascenso de 2 categorías
Desc 1	(A, B), (B, C)	Descenso de 1 categoría
Desc 2	(A, C)	Descenso de 2 categoría

Cada partido de los equipos visitantes en la base de datos fue asignado a uno de estos grupos

4.3.1 Interpretación de los grupos

Los grupos fueron definidos en función del tipo de desplazamiento altitudinal del equipo visitante, el cual se establece a partir de la comparación entre la categoría de altitud de la ciudad del equipo local y la del visitante (tabla 4.2). De esta forma, se distinguen desplazamientos de ascenso, de descenso o la ausencia de desplazamiento cuando ambos equipos juegan en ciudades con altitudes similares.

Una vez contruidos los grupos, se utilizó la variable delta de altitud como una medida cuantitativa para medir la magnitud promedio del cambio altitudinal dentro de cada uno. El delta corresponde a la diferencia (en metros sobre el nivel del mar) entre la altitud de la ciudad del local y la del visitante: valores positivos indican que el visitante asciende, valores negativos que desciende, y valores cercanos a cero que no existe desplazamiento relevante.

El cálculo del promedio de delta en cada grupo permite expresar, en términos de metros, la magnitud típica del cambio altitudinal que enfrentan los equipos visitantes al disputar sus partidos.

Tabla 4.3: Descripción de los grupos según el desplazamiento altitudinal del equipo visitante

Grupo	Interpretación
Asc 2	El equipo visitante sube en promedio 2420 metros.
Asc 1	El equipo visitante sube en promedio 1221 metros.
MA	El equipo visitante no se traslada (delta de altitud ≈ 0).
Desc 1	El equipo visitante baja en promedio 1225 metros.
Desc 2	El equipo visitante baja en promedio 2419 metros.

4.3.2 Altitud del equipo visitante

La Tabla 4.4 muestra la altitud (msnm) de los estadios donde juegan los equipos del Fútbol Profesional Colombiano (FPC), categorizados según la clasificación de altitud adaptada de Bartsch (Tabla 3.1).

Tabla 4.4: Altura de estadios del FPC y categoría según clasificación de Bartsch.

Equipo	Estadio	Altitud (m.s.n.m)	Categoría
Águilas	Alberto Grisales	2125	C
Alianza	Daniel Villa Zapata	75	A
América	Pascual Guerrero	1018	B
Boyacá Chicó	La Independencia	2811	C
Bucaramanga	Américo Montanini	959	B
Caldas	Palogrande	2200	C
Cali	Deportivo Cali	1001	B
Cortuluá	Doce de Octubre	966	B
Cúcuta	General Santander	1176	B
Envigado	Polideportivo Sur	1675	B
Equidad	Metropolitano de Techo	2640	C
Fortaleza	Varios estadios	2640	C
Huila	Plazas Alcid	442	A
Jaguares	No principal (Montería)	18	A
Junior	Metropolitano Roberto Meléndez	18	A
Leones	Metropolitano Ciudad de Itagüí	1550	B
Magdalena	Sierra Nevada	15	A
Medellín	Atanasio Girardot	1495	B
Millonarios	El Campín	2640	C
Nacional	Atanasio Girardot	1495	B
Pasto	Departamental Libertad	2527	C
Patriotas	La Independencia	2822	C
Pereira	Hernán Ramírez Villegas	1411	B
Santa Fe	El Campín	2640	C
Tigres	El Campín	2640	C
Tolima	Manuel Murillo Toro	1285	B

Los equipos del FPC compiten en ciudades que varían ampliamente en altitud, desde 15–18 msnm (Magdalena, Junior, Jaguares) hasta más de 2800 msnm (Patriotas, Boyacá Chicó). La distribución por categorías de altitud de los equipos es la siguiente, en la categoría A (0 - 500 msnm) están en total 5 (19%) equipos, en la categoría B (501 - 2000 msnm) están en total 11 (42%) equipos y en la categoría C hay 10 (39%) equipos. 3 estadios son compartidos por más de un equipo, 7 equipos comparten estadio con al menos otro club y el estadio El Campín es el más compartido por 3 equipos (Millonarios, Santa Fe y Tigres).

De acuerdo con la Tabla 4.5, se observan los grupos formados del archivo de datos con los partidos jugados.

Hay más partidos en el grupo MA ya que como se evidenció hay equipos que juegan en el mismo estadio y la alturas de los estadios en la mayoría de los casos son cercanas, por ende caen en las combinaciones (A, A), (B, B) Y (C,C) según la (tabla 4.2).

Tabla 4.5: Descripción de los grupos y número de partidos jugados según el desplazamiento del equipo visitante

Grupo	Partidos jugados	Definición
MA	653	El equipo visitante no se traslada.
Asc 1	474	El equipo visitante sube en promedio 1221 metros.
Asc 2	149	El equipo visitante sube en promedio 2420 metros.
Desc 1	468	El equipo visitante baja en promedio 1225 metros.
Desc 2	154	El equipo visitante baja en promedio 2419 metros.

4.4 Modelos propuestos

4.4.1 Modelos para la variable respuesta: Número de Goles Anotados

Dado que la variable Número de goles anotados es una medida cuantitativa discreta que toma valores enteros no negativos, resulta adecuado modelarla mediante un enfoque de regresión de Poisson.

Sea Y_i el número de goles anotados por el equipo visitante en el i -ésimo partido. Entonces, para los Modelos 1 y 2 se asume que

$$Y_i \sim \text{Poisson}(\lambda_i), \quad \log(\lambda_i) = x_i' \beta,$$

donde x_i' representa el vector de covariables asociado al i -ésimo partido y β es el vector de parámetros del modelo.

Modelo 1: Número de goles anotados en función del desplazamiento y el índice de rendimiento

El modelo que se muestra en la Ecuación (4.1), tiene como objetivo evaluar si el número de goles que anota un equipo visitante se relaciona con el tipo de desplazamiento que ha tenido que hacer para llegar al estadio y su desempeño a final del partido.

$$\begin{aligned} \log(\lambda_i) = & \beta_0 + \beta_1 \text{Asc1}_i + \beta_2 \text{Asc2}_i + \beta_3 \text{Des1}_i + \beta_4 \text{Des2}_i \\ & + \beta_5 \text{RendBajo}_i + \beta_6 \text{RendMedio}_i \end{aligned} \quad (4.1)$$

donde:

- $\text{Asc1}_i, \text{Asc2}_i, \text{Des1}_i, \text{Des2}_i$ variables indicadoras que toman el valor 1 si el equipo visitante pertenece a la categoría de desplazamiento correspondiente (Asc1 , Asc2 , Des1 o Des2) y 0 en caso contrario. La categoría de referencia es misma altitud (MA).
- RendBajo_i y RendMedio_i variables indicadoras que toman el valor 1 si el equipo visitante presenta un rendimiento bajo o medio al final del partido respectivamente, y 0 en caso contrario. La categoría de referencia es rendimiento alto.

Modelo 2: Número de goles anotados en función de los factores del índice de rendimiento

El modelo que se muestra en la Ecuación (4.2), evalúa si el número de goles que anota un equipo visitante puede explicarse a partir de los tres factores que conforman el índice de rendimiento: Control Ofensivo, Solidez Defensiva y Eficacia en Finalización. Estos factores fueron previamente clasificadas en tres niveles de rendimiento: bajo, medio y alto.

$$\begin{aligned} \log(\lambda_i) = & \beta_0 + \beta_1 CO_{\text{bajo},i} + \beta_2 CO_{\text{medio},i} \\ & + \beta_3 SD_{\text{bajo},i} + \beta_4 SD_{\text{medio},i} \\ & + \beta_5 EF_{\text{bajo},i} + \beta_6 EF_{\text{medio},i} \end{aligned} \quad (4.2)$$

donde:

- $CO_{\text{bajo},i}$, $CO_{\text{medio},i}$ variables indicadoras del nivel de rendimiento en el factor Control Ofensivo.
- $SD_{\text{bajo},i}$, $SD_{\text{medio},i}$ variables indicadoras del nivel de rendimiento en el factor Solidez Defensiva.
- $EF_{\text{bajo},i}$, $EF_{\text{medio},i}$ variables indicadoras del nivel de rendimiento en el factor Eficacia en Finalización.

El nivel de referencia para cada factor es el rendimiento alto.

4.4.2 Bonda de ajuste y calidad del modelo Poisson

Para probar el ajuste de los modelos se realizaron las pruebas de hipótesis de Razón de Verosimilitudes, donde se detecta si las covariables utilizadas mejoran el ajuste del modelo a los datos, y si el modelo se ajusta bien a los datos. También, se aplicó la prueba de hipótesis de Wald para la significancia de los parámetros. Se estimó el parámetro de sobredispersión, para la validación de supuestos.

4.4.3 Modelo mixto

Se explora la posibilidad de dependencia entre los goles anotados por el equipo visitante, la cual no es capturada por el modelo Poisson simple. Aunque el modelo Poisson asume que cada gol es un evento independiente, es plausible que existan diferencias sistemáticas entre equipos, de modo que algunos visitantes tengan un nivel base de anotación distinto al de otros. Para evaluar esta

hipótesis y relajar el supuesto de independencia, se ajustan modelos lineales generalizados mixtos (GLMM) con interceptos aleatorios por equipo visitante.

Sea Y_{ij} el número de goles anotados por el equipo visitante j en el i -ésimo partido. El modelo mixto asume que, condicionados al efecto aleatorio del equipo visitante,

$$Y_{ij} \mid u_{0j} \sim \text{Poisson}(\lambda_{ij}), \quad \log(\lambda_{ij} \mid u_{0j}) = \eta_{ij},$$

donde $u_{0j} \sim \mathcal{N}(0, \sigma_u^2)$ representa el efecto aleatorio del equipo j .

Modelo mixto 1:

$$\begin{aligned} \log(\lambda_{ij} \mid u_{0j}) = & \beta_0 + \beta_1 \text{Asc1}_{ij} + \beta_2 \text{Asc2}_{ij} + \beta_3 \text{Des1}_{ij} + \beta_4 \text{Des2}_{ij} \\ & + \beta_5 \text{RendBajo}_{ij} + \beta_6 \text{RendMedio}_{ij} + u_{0j} \end{aligned} \quad (4.3)$$

Modelo mixto 2:

$$\begin{aligned} \log(\lambda_{ij} \mid u_{0j}) = & \beta_0 + \beta_1 \text{CO}_{\text{bajo},ij} + \beta_2 \text{CO}_{\text{medio},ij} \\ & + \beta_3 \text{SD}_{\text{bajo},ij} + \beta_4 \text{SD}_{\text{medio},ij} \\ & + \beta_5 \text{EF}_{\text{bajo},ij} + \beta_6 \text{EF}_{\text{medio},ij} + u_{0j} \end{aligned} \quad (4.4)$$

Interpretación del efecto aleatorio

En los modelos mixtos se incluye el término

$$u_{0j} \sim \mathcal{N}(0, \sigma_u^2),$$

que representa el efecto aleatorio del equipo visitante j . Este término permite capturar diferencias propias de cada equipo que no son explicadas por las covariables del modelo, como estilos de juego o capacidad ofensiva general.

El efecto aleatorio ajusta el nivel base de anotación de cada equipo:

- $u_{0j} > 0$: el equipo tiende a anotar más que el promedio.
- $u_{0j} < 0$: el equipo tiende a anotar menos que el promedio.

Su inclusión permite modelar la dependencia entre partidos del mismo equipo y relajar la suposición de independencia del modelo Poisson simple.

4.4.4 Modelos para la variable respuesta: Resultado Final del Partido

La variable Resultado Final del Partido se considera una variable categórica de tipo ordinal, dado que los posibles resultados pueden ordenarse de acuerdo con el desempeño del equipo visitante en el encuentro. Sea Y_i la variable aleatoria que representa el resultado del partido i , definida como:

$$Y_i = \begin{cases} 0, & \text{si el equipo visitante pierde el partido,} \\ 1, & \text{si el equipo visitante empata el partido,} \\ 2, & \text{si el equipo visitante gana el partido.} \end{cases}$$

Dada la naturaleza ordinal de la variable respuesta, se empleó un modelo de regresión logística ordinal bajo el supuesto de odds proporcionales. Este modelo permite analizar la relación entre el resultado del partido y un conjunto de covariables explicativas, a partir de las probabilidades acumuladas $\Pr(Y_i \leq k)$, para $k = 0, 1$.

En particular, los valores de k corresponden a los distintos puntos de corte de la variable respuesta, de modo que:

- para $k = 0$, la probabilidad acumulada $\Pr(Y_i \leq 0)$ representa la probabilidad de que el equipo visitante pierda el partido
- para $k = 1$, la probabilidad acumulada $\Pr(Y_i \leq 1)$ representa la probabilidad de que el equipo visitante pierda o empate el partido.

La forma general del modelo se expresa como:

$$\text{logit}(\Pr(Y_i \leq k \mid \mathbf{X}_i)) = \alpha_k - \mathbf{X}_i^\top \boldsymbol{\beta}, \quad k = 0, 1, \quad (4.5)$$

donde α_k representan los parámetros de corte asociados a cada categoría acumulada, $\mathbf{X}_i$ es el vector de covariables del partido i y $\boldsymbol{\beta}$ es el vector de coeficientes comunes a todas las categorías, lo cual refleja el supuesto de odds proporcionales.

Modelo 3: Resultado final del partido en función del desplazamiento y el índice de rendimiento

Con el fin de evaluar la relación entre el resultado del equipo visitante, el desplazamiento asociado a la altitud y el nivel de rendimiento del equipo al final del partido, se ajustó un modelo de regresión logística ordinal bajo el supuesto de odds proporcionales.

Como variables explicativas se consideraron la categoría de desplazamiento del equipo visitante y el índice de rendimiento, ambos tratados como variables categóricas. Para la variable de desplazamiento se tomó como categoría de referencia la condición de misma altitud (MA), mientras que para el índice de rendimiento la categoría de referencia fue el nivel alto al final del partido.

$$\begin{aligned} \text{logit}(\Pr(Y_i \leq 0 \mid \mathbf{X}_i)) = & \alpha_0 - \beta_1 \text{Asc1}_i - \beta_2 \text{Asc2}_i - \beta_3 \text{Des1}_i \\ & - \beta_4 \text{Des2}_i - \beta_5 \text{RendMedio}_i - \beta_6 \text{RendBajo}_i \end{aligned} \quad (4.6)$$

$$\begin{aligned} \text{logit}(\Pr(Y_i \leq 1 \mid \mathbf{X}_i)) = & \alpha_1 - \beta_1 \text{Asc1}_i - \beta_2 \text{Asc2}_i - \beta_3 \text{Des1}_i \\ & - \beta_4 \text{Des2}_i - \beta_5 \text{RendMedio}_i - \beta_6 \text{RendBajo}_i \end{aligned} \quad (4.7)$$

donde:

- Asc1_i , Asc2_i , Des1_i y Des2_i son variables indicadoras asociadas a las categorías de desplazamiento del equipo visitante, tomando como referencia la condición de misma altitud (MA)
- RendMedio_i y RendBajo_i corresponden a variables indicadoras del índice de rendimiento normalizado, siendo la categoría de referencia el nivel alto
- α_0 y α_1 son los parámetros de corte del modelo
- $\boldsymbol{\beta} = (\beta_1, \dots, \beta_6)^\top$ es el vector de coeficientes asociados a las covariables explicativas.

En este modelo, un coeficiente positivo indica un aumento en las odds acumuladas de que el resultado del partido sea menor o igual a una categoría k , es decir, una mayor probabilidad de obtener resultados menos favorables (perder o empatar). Por el contrario, un coeficiente negativo se asocia con una disminución de dichas odds, lo que implica una mayor probabilidad de alcanzar resultados más favorables (empatar o ganar), en comparación con las categorías de referencia.

Modelo 4: Resultado final del partido en función de los factores del índice de rendimiento

Con el propósito de analizar la relación entre el resultado final del partido del equipo visitante y los factores del índice de rendimiento, se ajustó un modelo de regresión logística ordinal bajo el supuesto de odds proporcionales.

Como variables explicativas se consideraron los componentes del índice de rendimiento: Control Ofensivo, Solidez Defensiva y Eficacia en Finalización, los cuales fueron tratados como variables categóricas ordinales con tres niveles (bajo, medio y alto). Para cada uno de los componentes, se tomó como categoría de referencia el nivel alto.

$$\begin{aligned} \text{logit}(\Pr(Y_i \leq 0 \mid \mathbf{X}_i)) = & \alpha_0 - \beta_1 \text{CO}_{\text{Medio},i} - \beta_2 \text{CO}_{\text{Bajo},i} - \beta_3 \text{SD}_{\text{Medio},i} \\ & - \beta_4 \text{SD}_{\text{Bajo},i} - \beta_5 \text{EF}_{\text{Medio},i} - \beta_6 \text{EF}_{\text{Bajo},i} \quad (4.8) \end{aligned}$$

$$\begin{aligned} \text{logit}(\Pr(Y_i \leq 1 \mid \mathbf{X}_i)) = & \alpha_1 - \beta_1 \text{CO}_{\text{Medio},i} - \beta_2 \text{CO}_{\text{Bajo},i} - \beta_3 \text{SD}_{\text{Medio},i} \\ & - \beta_4 \text{SD}_{\text{Bajo},i} - \beta_5 \text{EF}_{\text{Medio},i} - \beta_6 \text{EF}_{\text{Bajo},i} \quad (4.9) \end{aligned}$$

- $\text{CO}_{\text{Medio},i}$ y $\text{CO}_{\text{Bajo},i}$ son variables indicadoras asociadas a los niveles medio y bajo del componente Control Ofensivo, tomando como referencia el nivel alto
- $\text{SD}_{\text{Medio},i}$ y $\text{SD}_{\text{Bajo},i}$ corresponden a variables indicadoras del componente Solidez Defensiva, siendo el nivel alto la categoría de referencia
- $\text{EF}_{\text{Medio},i}$ y $\text{EF}_{\text{Bajo},i}$ son variables indicadoras del componente Eficacia en Finalización, con nivel alto como categoría de referencia
- α_0 y α_1 representan los parámetros de corte del modelo
- $\boldsymbol{\beta} = (\beta_1, \dots, \beta_6)^\top$ es el vector de coeficientes asociado a los componentes del índice.

4.4.5 Bonda de ajuste y calidad del modelo Ordinal

Inicialmente se probó el supuesto de Odds Proportional, posterior a ello se verifica el ajuste del modelo utilizando el test de hipótesis de Razón de Verosimilitudes, donde se detecta si las covariables utilizadas mejoran el ajuste del modelo a los datos, y si el modelo se ajusta bien a los datos. Además, se realiza la prueba de significancia individual de los parámetros. Finalmente para establecer la calidad del modelo se hizo uso del Pseudo-R² de Nagelkerke.

4.4.6 Comparación de modelos

Se comparan los modelos mixtos de Poisson con el modelo Poisson simple. De acuerdo con Chatterjee y Simonoff (2013), modelos distintos pueden ser comparados utilizando criterios de información como el AIC o el AICc.

Capítulo 5

Resultados

5.1 Análisis exploratorio

El análisis exploratorio se realizó con el objetivo de caracterizar las variables involucradas y explorar posibles patrones o relaciones entre el rendimiento del equipo visitante, el resultado del partido, los goles anotados y la altitud.

Descripción de las variables respuesta

En primer lugar, se describieron las variables respuesta: Número de Goles Anotados del equipo visitante y Resultado Final del equipo visitante.

5.1.1 Número de goles anotados del equipo visitante

La distribución de goles anotados por el equipo visitante presentada en la Tabla 5.1 muestra que, en la mayoría de los casos, los equipos visitantes anotan 0 o 1 gol. Esto representa un 76.77% del total de observaciones, lo cual indica que hacer más de un gol para el equipo visitante no fue muy usual en el fútbol colombiano en ese período, pues los partidos en los que el visitante anota por ejemplo 4 o más goles son extremadamente poco frecuentes (apenas 24 casos en total).

Tabla 5.1: Distribución de goles anotados por el equipo visitante

Goles	Frecuencia	Porcentaje
0	754	39.73%
1	703	37.04%
2	322	16.97%
3	95	5.01%
4	21	1.11%
5	2	0.11%
6	1	0.05%

La Tabla 5.2 muestra una versión categorizada de los goles del equipo visitante, creada de su distribución original Tabla 5.1. Dado que la mayoría de los partidos terminan con 0 o 1 gol visitante, y las frecuencias disminuyen notablemente a partir de 2 goles, se agruparon todas las observaciones con dos o más goles en una sola categoría. Esto permite evitar celdas con frecuencias bajas en análisis multivariados posteriores.

Tabla 5.2: Distribución categorizada de goles anotados por el equipo visitante

Goles	Frecuencia	Porcentaje
Sin goles	754	39.73%
Un gol	703	37.04%
Dos o más goles	441	23.23%

5.1.2 Resultado final del equipo visitante

La información presentada en la Tabla 5.3 muestra que, del total de partidos en los que el equipo es visitante, el 47.15% culminaron en derrota, siendo esta la categoría más frecuente. Los resultados indican una clara desventaja para los equipos que juegan fuera de casa.

Tabla 5.3: Distribución de resultados del equipo visitante

Resultado	Frecuencia	Porcentaje
Ganar	438	23.08%
Empatar	565	29.77%
Perder	895	47.15%

5.2 Análisis exploratorio de covariables y su relación con las variables respuesta

Se describen las covariables (Índice de Rendimiento, Factores del Índice de Rendimiento y Desplazamiento del Equipo Visitante en función de la altitud), analizadas de manera individual y en relación con las variables respuesta.

5.2.1 Rendimiento del equipo visitante al final del partido según el índice

Recordemos que, el índice de rendimiento del equipo visitante se calculó para cada uno de los 1 898 partidos una vez finalizado el encuentro, a partir de una combinación lineal de las variables $X_1, \dots, X_{16}$, previamente definidas en la Tabla 2.1. Dicho índice se evaluó exclusivamente en los partidos en los que el equipo actuó en condición de visitante y se define como:

$$IRVM_{\text{Visitante},i} = 0.040X_1 + 0.045X_2 + 0.085X_3 + 0.082X_4 + 0.078X_5 + 0.037X_6 + 0.042X_7 + 0.077X_8 + 0.055X_9 + 0.072X_{10} + 0.179X_{11} + 0.169X_{12} + 0.006X_{13} + 0.030X_{14} + 0.022X_{15} - 0.025X_{16}$$

Para facilitar la interpretación del rendimiento, se categorizó en tres niveles de desempeño según los puntos de corte propuestos por Chávez-Avendaño (2025), los cuales fueron calculados mediante un algoritmo k-means. Los rangos definidos fueron los siguientes:

Tabla 5.4: Categorización del índice de rendimiento

Nivel de rendimiento	Rango del índice (IR)
Rendimiento Bajo	$0 \leq IR < 0.4024$
Rendimiento Medio	$0.4024 \leq IR < 0.5688$
Rendimiento Alto	$0.5688 \leq IR \leq 1$

Según la Tabla 5.5 la distribución del índice de rendimiento por categorías (alto, medio y bajo) sugiere que, en general, el rendimiento de los equipos visitantes al finalizar el encuentro tiende a ubicarse en niveles medios.

Tabla 5.5: Distribución del índice de rendimiento

Categoría	Frecuencia	Porcentaje
Rendimiento Alto	488	25.71%
Rendimiento Medio	867	45.68%
Rendimiento Bajo	543	28.61%

5.2.2 Relación entre el rendimiento del equipo visitante con el resultado final y el número de goles anotados: análisis mediante prueba de chi-cuadrado

Las Figuras 5.1 y 5.2 muestran, respectivamente, la distribución porcentual del resultado final y del número de goles anotados por el equipo visitante según el nivel de rendimiento al finalizar el partido. Se observa que los equipos visitantes con rendimiento alto presentan una mayor proporción de victorias y una menor frecuencia de derrotas en comparación con aquellos con rendimiento medio y bajo. Asimismo, la probabilidad de no anotar goles es mayor cuando el rendimiento es medio o bajo, mientras que la proporción de partidos con dos o más goles aumenta notablemente en escenarios de rendimiento alto.

Estas diferencias fueron estadísticamente significativas, según la prueba de chi-cuadrado de independencia, tanto para el resultado del partido ($p < 0.001$) como para el número de goles anotados por el visitante ($p < 0.001$), lo que evidencia una asociación entre el resultado final y el número de goles anotados con el rendimiento del equipo visitante al finalizar el encuentro.

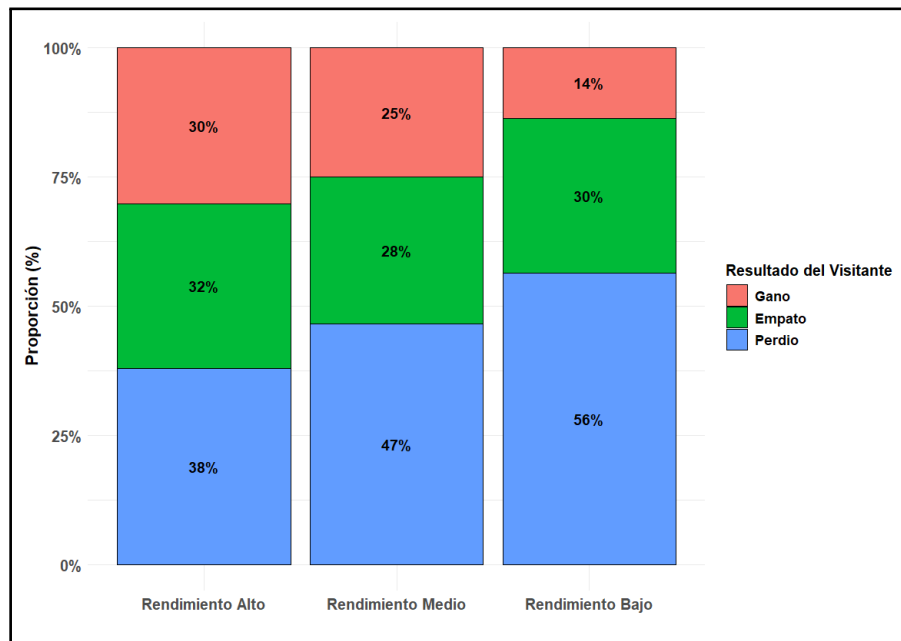


Figura 5.1: Resultado del equipo visitante según el nivel del rendimiento al final del partido

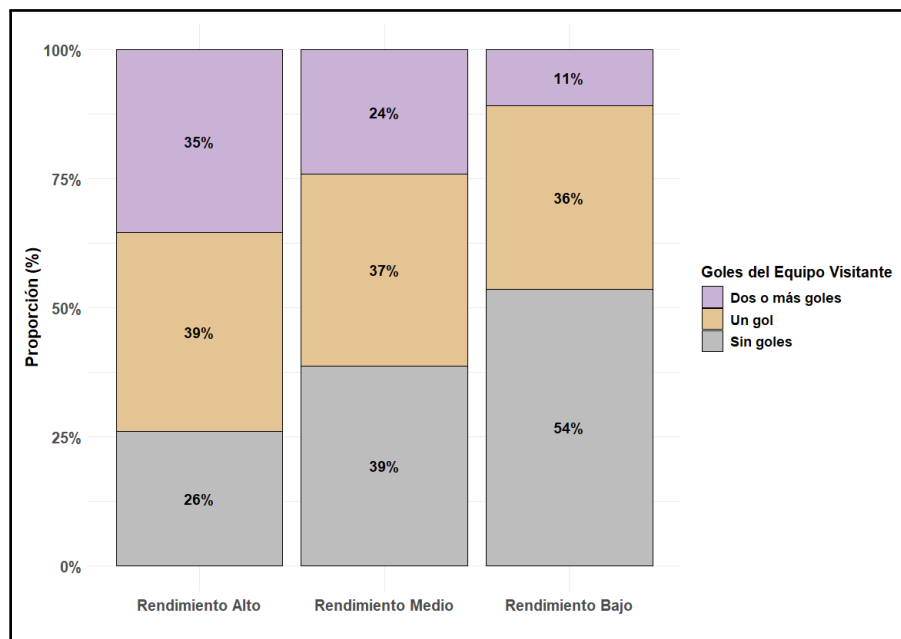


Figura 5.2: Goles anotados por el equipo visitante según el nivel del rendimiento al final del partido

5.2.3 Factores del índice de rendimiento

Con el fin de facilitar la interpretación de los factores que componen el índice de rendimiento, cada uno de ellos fue categorizado en tres niveles de desempeño (bajo, medio y alto), de acuerdo con los puntos de corte propuestos por el autor Chávez-Avendaño (2025). Dado que los factores presentan distribuciones y escalas diferenciadas, los rangos de categorización se definieron de manera específica para cada uno de ellos: Control Ofensivo, Solidez Defensiva y Eficacia en la Finalización.

Tabla 5.6: Categorización de los factores del índice de rendimiento

Factor	Nivel de rendimiento	Rango del índice
Control Ofensivo	Bajo	$0 \leq CO < 0.3905$
	Medio	$0.3905 \leq CO < 0.5863$
	Alto	$0.5863 \leq CO \leq 1$
Solidez Defensiva	Bajo	$0 \leq SD < 0.3825$
	Medio	$0.3825 \leq SD < 0.5532$
	Alto	$0.5532 \leq SD \leq 1$
Eficacia en la Finalización	Bajo	$0 \leq EF < 0.3605$
	Medio	$0.3605 \leq EF < 0.5631$
	Alto	$0.5631 \leq EF \leq 1$

Esta categorización permitió analizar de manera comparativa el comportamiento de cada factor del índice de rendimiento en relación con el resultado del partido, número de goles y las condiciones de desplazamiento del equipo visitante.

Los resultados de la Tabla 5.7 muestran que la mayoría de los equipos se clasifican en el nivel medio en las tres dimensiones del índice de rendimiento: Control Ofensivo, Solidez Defensiva y Eficacia en Finalización. En cada componente, cerca del 44% al 46% de los casos presentan un rendimiento medio, mientras que los niveles alto y bajo están más equilibrados entre sí. Esto indica un comportamiento relativamente homogéneo entre los equipos en cuanto a su desempeño táctico y capacidad para ejecutar acciones clave en defensa, ataque y finalización de jugadas.

Tabla 5.7: Distribución de los niveles de rendimiento en cada factor

Variable	n = 1898
Control Ofensivo	
Rendimiento Alto	463 (24%)
Rendimiento Medio	869 (46%)
Rendimiento Bajo	566 (30%)
Solidez Defensiva	
Rendimiento Alto	501 (26%)
Rendimiento Medio	874 (46%)
Rendimiento Bajo	523 (28%)
Eficacia en Finalización	
Rendimiento Alto	477 (25%)
Rendimiento Medio	841 (44%)
Rendimiento Bajo	580 (31%)

5.2.4 Relación entre el resultado final y los factores del índice de rendimiento: análisis mediante prueba de chi-cuadrado

Según los resultados obtenidos, los equipos que ganaron muestran una mayor proporción de casos con rendimiento alto en los tres factores: Control Ofensivo (30%), Solidez Defensiva (26%) y, especialmente, Eficacia en Finalización (36%). Este último destaca como el factor con mayor diferenciación respecto a los equipos que no ganaron, lo que sugiere que la capacidad para concretar oportunidades ofensivas fue un factor determinante en sus victorias.

En contraste, los equipos que perdieron presentan una mayor proporción de rendimiento bajo en Eficacia en Finalización (38%), así como valores relativamente altos en la categoría baja de Control Ofensivo (31%) y Solidez Defensiva (28%). Esto indica limitaciones tanto en su capacidad de generar acciones ofensivas efectivas como en la conversión de oportunidades en goles.

Se realizaron pruebas de *chi-cuadrado* para evaluar la asociación entre el nivel de rendimiento en cada factor y el resultado del equipo visitante. Los resultados muestran que existe una asociación estadísticamente significativa para Control Ofensivo ($\chi^2 = 7.64$, $p = 0.022$) y, de manera más marcada, para Eficacia en Finalización ($\chi^2 = 50.12$, $p < 0.001$), mientras que Solidez Defensiva no presentó una asociación significativa ($\chi^2 = 0.21$, $p = 0.90$).

Tabla 5.8: Distribución de factores por resultado del equipo visitante con p-valores

Variable	Ganar	Empatar	Perder	p-valor
Control Ofensivo				0.022
Rend. Alto	129 (30%)	139 (25%)	195 (22%)	
Rend. Medio	198 (45%)	251 (44%)	420 (47%)	
Rend. Bajo	111 (25%)	175 (31%)	280 (31%)	
Solidez Defensiva				0.9
Rend. Alto	113 (26%)	155 (28%)	233 (26%)	
Rend. Medio	209 (48%)	251 (44%)	414 (46%)	
Rend. Bajo	116 (26%)	159 (28%)	248 (28%)	
Eficacia en Finalización				< 0.001
Rend. Alto	156 (36%)	151 (27%)	170 (19%)	
Rend. Medio	207 (47%)	252 (45%)	382 (43%)	
Rend. Bajo	75 (17%)	162 (29%)	343 (38%)	

5.2.5 Relación entre los goles anotados del visitante y los factores del índice de rendimiento: análisis mediante prueba de chi-cuadrado

La Tabla 5.9 presenta los resultados de los goles anotados por el equipo visitante según cada factor del índice de rendimiento.

Por ejemplo, en el factor **Control Ofensivo** se observa una tendencia creciente: a medida que aumenta el número de goles, también lo hace la proporción de equipos con rendimiento alto en este factor (18% sin goles, 26% con un gol y 34% con dos o más goles). Paralelamente, disminuye la proporción de rendimiento bajo (de 35% a 20%). La prueba de *chi-cuadrado* confirma que esta relación es estadísticamente significativa ($\chi^2 = 45.21$, $p < 0.001$), lo que sugiere que un mejor Control Ofensivo está asociado con la cantidad de goles anotados por el equipo visitante.

En el caso de **Solidez Defensiva**, no se identifican cambios importantes entre los grupos. Las proporciones de cada nivel se mantienen estables, y la prueba de chi-cuadrado indica que no existe una asociación significativa ($\chi^2 = 1.82$, $p = 0.4$), lo que sugiere que este componente no tiene una relación directa con la cantidad de goles anotados, siendo su aporte más relevante en la capacidad de evitar goles.

La **Eficacia en Finalización** muestra una asociación muy clara y positiva con los goles anotados. La proporción de equipos con rendimiento alto se incrementa de manera notable (15% sin goles, 26% con un gol y 41% con

dos o más goles), mientras que el rendimiento bajo disminuye (de 45% a 13%). La prueba de chi-cuadrado confirma que esta relación es altamente significativa ($\chi^2 = 72.18$, $p < 0.001$), evidenciando que la capacidad de concretar oportunidades es determinante en la generación de goles.

En conclusión, estos resultados indican que, si bien el Control Ofensivo contribuye a la producción de goles, es la Eficacia en Finalización el factor que más explica las diferencias en la cantidad de goles anotados por el equipo visitante, mientras que la Solidez Defensiva no muestra un efecto relevante en este aspecto.

Tabla 5.9: Distribución de los factores de rendimiento en función de los goles anotados por el equipo visitante con p-valores

Variable	Sin goles	Un gol	Dos o más goles	p-valor
Control Ofensivo				< 0.001
Rend. Alto	132 (17%)	180 (26%)	151 (35%)	
Rend. Medio	359 (48%)	310 (44%)	200 (45%)	
Rend. Bajo	263 (35%)	213 (30%)	90 (20%)	
Solidez Defensiva				0.4
Rend. Alto	203 (27%)	190 (27%)	108 (24%)	
Rend. Medio	354 (47%)	324 (46%)	196 (45%)	
Rend. Bajo	197 (26%)	189 (27%)	137 (31%)	
Eficacia en Finalización				< 0.001
Rend. Alto	112 (15%)	182 (26%)	183 (41%)	
Rend. Medio	304 (40%)	336 (48%)	201 (46%)	
Rend. Bajo	338 (45%)	185 (26%)	57 (13%)	

5.2.6 Goles Anotados, Resultado Final según la condición de desplazamiento del equipo visitante (altitud): análisis mediante prueba de chi-cuadrado

Goles del anotados por el visitante: Al analizar la cantidad de goles anotados por el equipo visitante según la categoría de desplazamiento en altitud (Figura 5.3), se observa una asociación estadísticamente significativa entre ambas variables (prueba chi-cuadrado, $p\text{-valor} = 0.034$). En particular, los equipos que ascienden dos categorías (Asc 2) presentan la mayor proporción de partidos sin goles (52%) y la menor proporción de encuentros con dos o más goles (14%), lo que sugiere que los ascensos pronunciados en altitud limitan la capacidad ofensiva del equipo visitante. En contraste, los equipos que no se desplazaron (MA), descendieron una categoría (Desc 1) o ascendieron una categoría (Asc 1) muestran distribuciones más equilibradas entre las tres categorías de goles, con proporciones similares de partidos sin goles, con un gol y con dos o más goles. Este patrón indica que los desplazamientos leves en altitud tienen un impacto menos marcado sobre la producción ofensiva.

Resultado del visitante: Respecto al resultado final del partido, la distribución de victorias, empates y derrotas según la condición de desplazamiento en altitud se presenta en la Figura 5.4. Aunque se observan diferencias descriptivas entre algunas categorías (Asc 2 por ejemplo), la prueba chi-cuadrado no evidencia una asociación estadísticamente significativa entre el resultado del visitante y el desplazamiento en altitud ($p\text{-valor} = 0.20$).

No obstante, desde un punto de vista descriptivo, los equipos que ascienden dos categorías (Asc 2) presentan la mayor proporción de derrotas (58%), seguidos por aquellos que descienden dos categorías (Desc 2) con un 50%. Por el contrario, los equipos que no se desplazaron (MA) y aquellos con desplazamientos moderados (Asc 1 y Desc 1) muestran mayores proporciones de victorias y empates, lo que sugiere un mejor desempeño relativo cuando los cambios de altitud no son extremos.

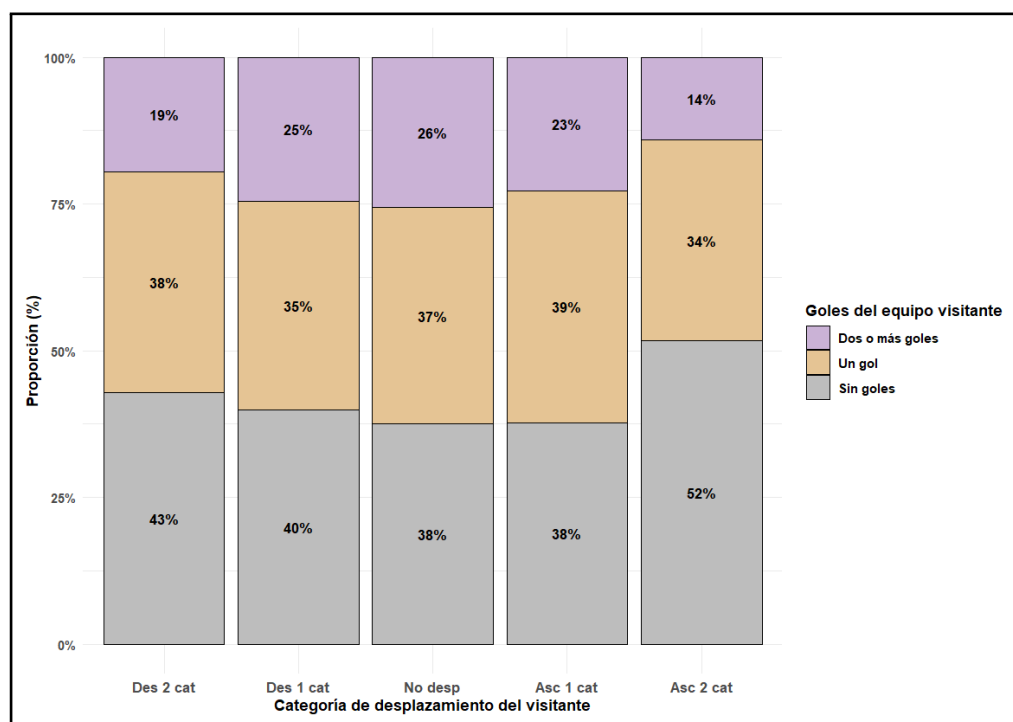


Figura 5.3: Goles anotados por el equipo visitante según su desplazamiento

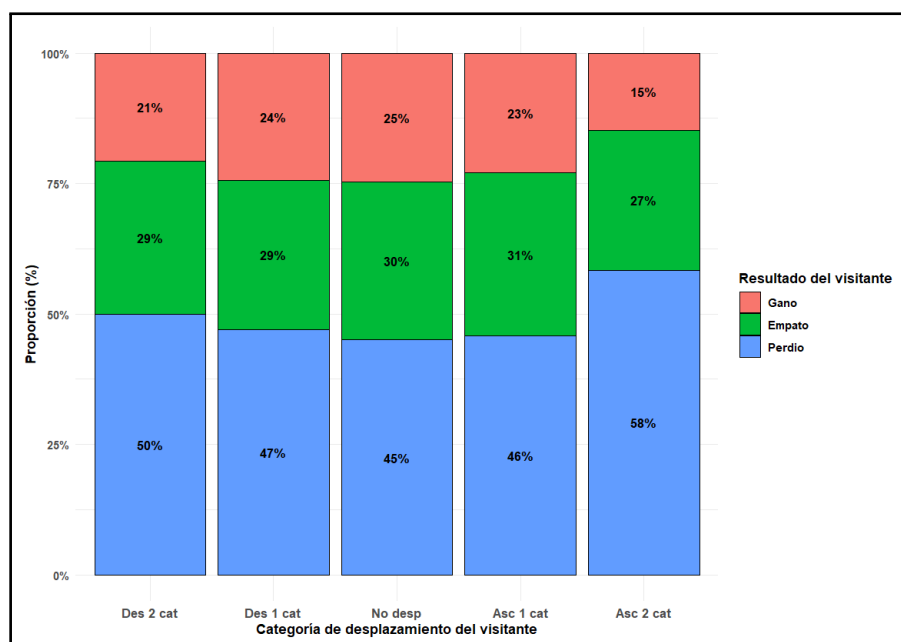


Figura 5.4: Resultado del equipo visitante según su desplazamiento

5.3 Resultado 1: Modelación de la relación entre la altura, el desempeño del equipo al final del partido y su influencia en el número de goles anotados

Verificación de supuestos en la distribución Poisson

Supuesto de independencia:

El modelo Poisson asume independencia entre los conteos de goles observados en distintos partidos. Sin embargo, en la práctica este supuesto puede no cumplirse, ya que los goles anotados por un mismo equipo a lo largo de la temporada pueden presentar correlación debido a factores persistentes como la calidad del equipo, el estilo de juego o el estado de forma.

Para considerar esta posible dependencia, se ajustan modelos mixtos con interceptos aleatorios por equipo visitante. Aun así, inicialmente se estima el modelo Poisson simple asumiendo independencia.

Supuesto de inflación de ceros:

Según García-Artiles, Gómez-Déniz e Dávila Cárdenes (2014), el índice del inflado de ceros se define como el cociente entre el número de elementos de la muestra con valor cero y el número de los mismos con valor distinto de cero. Si este cociente es mayor que 1, la muestra presenta inflado de ceros. En nuestro caso tenemos el siguiente cociente:

$$\text{Índice de inflado de ceros} = \frac{754}{1.144} = 0,66$$

Por lo cual podemos afirmar que la variable respuesta número de goles anotados por el equipo visitante no presenta inflación de ceros.

5.3.1 Modelo 1: Regresión Poisson para los goles del equipo visitante en función del desplazamiento y el rendimiento del equipo al final del partido

$$\log(\lambda_i) = \beta_0 + \beta_1 \text{Asc1}_i + \beta_2 \text{Asc2}_i + \beta_3 \text{Des1}_i + \beta_4 \text{Des2}_i + \beta_5 \text{RendBajo}_i + \beta_6 \text{RendMedio}_i \quad (5.1)$$

donde:

- $\text{Asc1}_i, \text{Asc2}_i, \text{Des1}_i, \text{Des2}_i$ variables indicadoras que toman el valor 1 si el equipo visitante pertenece a la categoría de desplazamiento correspondiente ($\text{Asc1}, \text{Asc2}, \text{Des1}$ o Des2) y 0 en caso contrario. La categoría de referencia es misma altitud (MA).
- RendBajo_i y RendMedio_i variables indicadoras que toman el valor 1 si el equipo visitante presenta un rendimiento bajo o medio, respectivamente, y 0 en caso contrario. La categoría de referencia es rendimiento alto.

5.3.2 Bondad de ajuste y calidad del modelo 1

De acuerdo con Cohen (2014) y Dobson (2018), en los modelos de regresión de Poisson existe sobredispersión cuando el parámetro de dispersión, estimado como la razón entre la desviación residual y los grados de libertad, es mayor que uno. En la Tabla 5.10 se reportan las estimaciones correspondientes al Modelo 1 ajustado. Los resultados indican que el parámetro de dispersión es aproximadamente igual a uno, lo que sugiere que no existe evidencia de sobredispersión. En consecuencia, se considera que el supuesto de equidispersión se cumple y que el Modelo presenta un ajuste adecuado a los datos.

Estadístico	Valor
Parámetro de dispersión ϕ	0.899

Tabla 5.10: Parámetro de dispersión

Para evaluar la calidad del ajuste del modelo Poisson ajustado, se analizaron la desviación nula y la desviación residual, así como la prueba de razón de verosimilitud (Likelihood Ratio Test).

Tabla 5.11: Evaluación del ajuste mediante desviaciones y prueba de razón de verosimilitud del modelo 1

Medida	Valor	Grados de libertad
Desviación nula (D_{null})	2122.6	1897
Desviación residual (D_{res})	1979.8	1891
Diferencia de desviaciones (D)	142.76	6
p -valor	≈ 0	–

Como el p -valor es muy pequeño (menor a 0.001), el modelo con co-variables mejora significativamente frente al modelo nulo. Esto indica que el desplazamiento del visitante y el índice de rendimiento sí aportan información útil para explicar los goles anotados del equipo visitante.

Tabla 5.12: Estimación del modelo Poisson para goles del visitante según categoría de desplazamiento y rendimiento

Variable	Coeficiente e^{β_j}		Std. Error	z value	p-value
Intercept (MA, Rend. Alto)	0.2541	1.29	0.0514	4.94	$p < 0.001$ ***
Asc 1	-0.0244	0.98	0.0622	-0.39	0.695
Asc 2	-0.3435	0.71	0.1097	-3.13	0.002 **
Desc 1	-0.0067	0.99	0.0620	-0.11	0.914
Desc 2	-0.1564	0.86	0.0987	-1.59	0.113
Rendimiento Bajo	-0.7525	0.47	0.0696	-10.81	$p < 0.001$ ***
Rendimiento Medio	-0.2774	0.76	0.0537	-5.16	$p < 0.001$ ***

Nota: *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$.

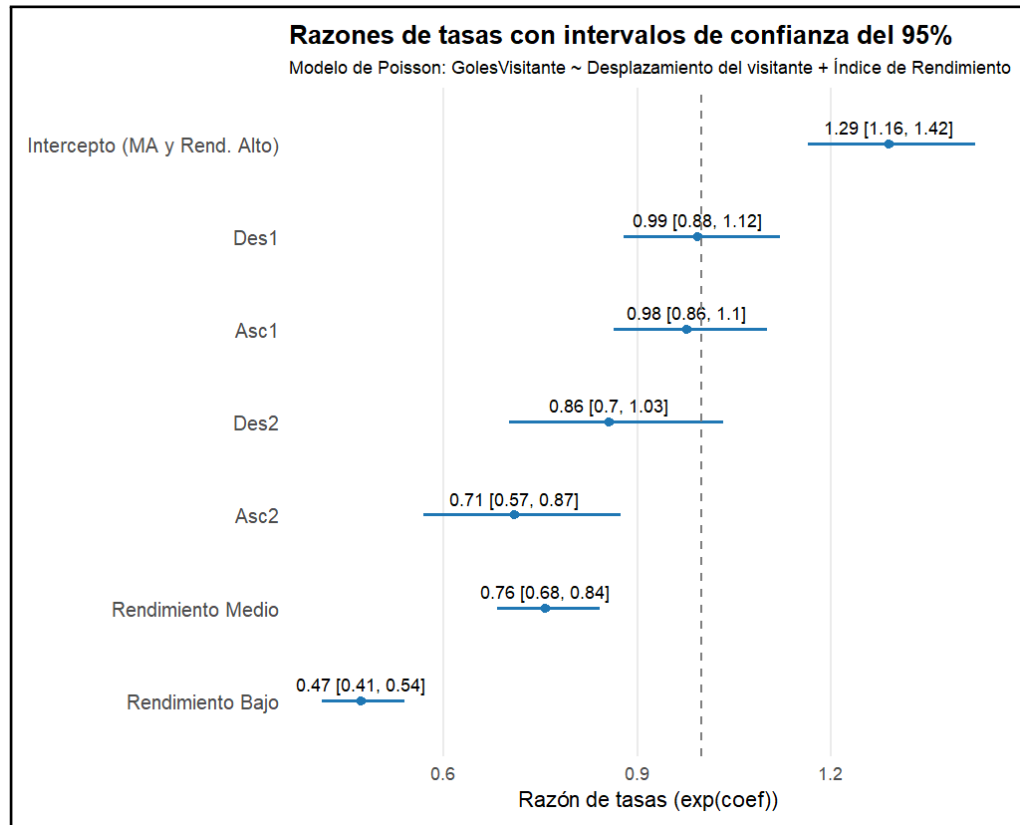


Figura 5.5: Razones de tasas con intervalos de confianza del 95% del Modelo 1.

Según la Tabla 5.12 y la Figura 5.5, los coeficientes estimados del modelo Poisson permiten calcular los cambios porcentuales esperados en el número de goles respecto al escenario base utilizando la fórmula $(e^{\beta_j} - 1) \times 100\%$. El escenario base corresponde a un equipo visitante con rendimiento alto al final del partido y sin desplazamiento (MA), para el cual se espera un promedio de 1.29 goles por partido ($\beta_0 = 0.2541$), lo que representa la situación más favorable para el visitante y es altamente significativo.

En particular, los equipos con rendimiento medio al final de partido ($\beta = -0.2774$) anotan en promedio un 24% menos de goles que el escenario base, mientras que los equipos con rendimiento bajo al final de partido ($\beta = -0.7525$) reducen su producción ofensiva en aproximadamente 53%, ambas diferencias altamente significativas.

En cuanto al desplazamiento altitudinal, solo el ascenso de dos categorías (Asc 2, correspondiente a un aumento en promedio de 2420 metros sobre el nivel del mar, $\beta = -0.3435$) presentó un efecto significativo, asociado con una disminución aproximada del 29% en la cantidad de goles

esperados respecto al escenario base; ascenso leve (Asc 1) o descensos (Desc 1 y Desc 2) no mostraron efectos estadísticamente significativos. La Figura 5.5 ilustra visualmente estas relaciones, mostrando las razones de tasas estimadas (e^{β_j}) y sus intervalos de confianza al 95%, confirmando que únicamente el rendimiento medio y bajo, así como el ascenso de dos categorías, presentan efectos significativos.

Este hallazgo indica que, en el Fútbol Profesional Colombiano, el rendimiento del visitante es el principal determinante del número de goles anotados, aunque un ascenso pronunciado en altitud representa una desventaja considerable.

Además, respalda la hipótesis planteada en el problema de investigación, según la cual los factores fisiológicos y los procesos de adaptación al entorno influyen de manera significativa en el desempeño ofensivo de los equipos cuando compiten bajo condiciones de altitud elevada, en concordancia con lo propuesto por Matijasevic (2017).

5.3.3 Modelo 2: Regresión Poisson para los goles del equipo visitante en función de los factores del índice de rendimiento

$$\begin{aligned}\log(\lambda_i) = & \beta_0 + \beta_1 CO_{\text{bajo},i} + \beta_2 CO_{\text{medio},i} \\ & + \beta_3 SD_{\text{bajo},i} + \beta_4 SD_{\text{medio},i} \\ & + \beta_5 EF_{\text{bajo},i} + \beta_6 EF_{\text{medio},i}\end{aligned}\quad (5.2)$$

donde:

- $CO_{\text{bajo},i}$, $CO_{\text{medio},i}$ variables indicadoras del nivel de rendimiento en el factor Control Ofensivo.
- $SD_{\text{bajo},i}$, $SD_{\text{medio},i}$ variables indicadoras del nivel de rendimiento en el factor Solidez Defensiva.
- $EF_{\text{bajo},i}$, $EF_{\text{medio},i}$ variables indicadoras del nivel de rendimiento en el factor Eficacia en Finalización.

El nivel de referencia para cada factor es el rendimiento alto.

5.3.4 Bonda de ajuste y calidad del modelo 2

Tabla 5.13: Parámetro de dispersión

Parámetro	Valor
Parámetro de dispersión ϕ	0.872

En la Tabla 5.13 se observa el parámetro de dispersión y este sugiere que no hay evidencia de sobredispersión en el modelo 2 ajustado.

Tabla 5.14: Evaluación del ajuste mediante desviancias y prueba de razón de verosimilitud del modelo 2

Medida	Valor	Grados de libertad
Desviación nula (D_{null})	2122.6	1897
Desviación residual (D_{res})	1898.5	1891
Diferencia de desviancias (D)	224.14	6
p -valor	< 0.001	–

Dado este resultado, el modelo con los predictores Control Ofensivo, Solidez Defensiva y Eficacia en Finalización mejora significativamente el ajuste respecto al modelo nulo.

Tabla 5.15: Estimación del modelo Poisson para los goles del visitante según los factores del índice de rendimiento

Variable	Coefficiente	e^{β_j}	Std. Error	z value	p-value
Intercepto	0.4383	1.55	0.0658	6.656	$p < 0.001$ ***
Control Ofensivo Bajo	-0.3703	0.69	0.0653	-5.665	$p < 0.001$ ***
Control Ofensivo Medio	-0.2327	0.79	0.0561	-4.144	$p < 0.001$ ***
Solidez Defensiva Bajo	0.0785	1.08	0.0650	1.208	0.227
Solidez Defensiva Medio	0.0158	1.01	0.0596	0.265	0.791
Eficacia en Finalización Bajo	-0.8751	0.41	0.0700	-12.491	$p < 0.001$ ***
Eficacia en Finalización Medio	-0.3219	0.72	0.0534	-6.028	$p < 0.001$ ***

Nota: *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$.

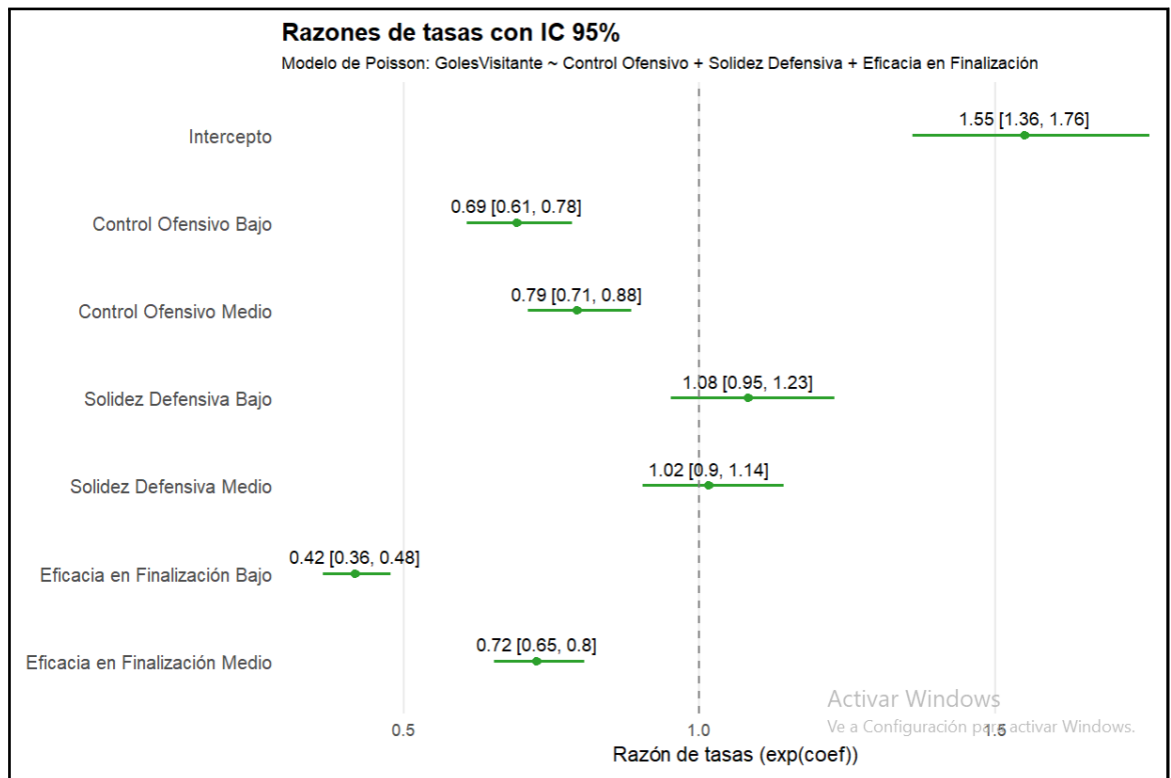


Figura 5.6: Razones de tasas con intervalos de confianza del 95% del Modelo 2.

Según la Tabla 5.15 y la Figura 5.6, los cambios porcentuales esperados en el número de goles del equipo visitante se calcularon tal como se hizo en el Modelo 1.

El escenario base corresponde a un equipo visitante con rendimiento alto en todos los factores del índice que mide el desempeño (Control Ofensivo, Solidez Defensiva y Eficacia en Finalización), para el cual se espera un promedio de 1.55 goles por partido ($\beta_0 = 0.4383$). Este escenario representa la situación más favorable para el equipo visitante y es significativo.

En particular, los equipos con Control Ofensivo medio al final de partido presentan una reducción de aproximadamente 21% en la cantidad de goles esperados, mientras que los que tienen Control Ofensivo bajo al final del partido disminuyen sus goles en un 31%, diferencias que resultan altamente significativas. Por otro lado, los niveles bajo y medio en Solidez Defensiva no muestran efectos significativos, lo que indica que este factor no afecta de manera clara la capacidad goleadora del visitante, lo cual tiene sentido en contexto, ya que la defensa impacta más en los goles recibidos que en los anotados.

El factor con mayor influencia sobre los goles esperados es la Eficacia en Finalización, donde los equipos con rendimiento medio al final del partido reducen sus goles en un 28% y aquellos con rendimiento bajo en aproximadamente 59%, ambas significativas, evidenciando que la capacidad de concretar oportunidades es determinante para la producción ofensiva de los visitantes.

Estos resultados muestran que, en el Fútbol Profesional Colombiano, el rendimiento del visitante en Control Ofensivo y Eficacia en Finalización son los factores más relevantes para anotar goles, mientras que la Solidez Defensiva no parece limitar de manera significativa la producción ofensiva de los equipos visitantes.

5.3.5 Comparación de indicadores entre el Modelo Poisson Mixto y el Modelo Poisson

Tabla 5.16: Comparación de AIC entre modelo lineal generalizado Poisson simple y modelo lineal generalizado Poisson mixto

Covariable (Índice de rendimiento y desplazamiento del visitante)		Covariable (Factores de rendimiento)	
Modelo	AIC	Modelo	AIC
Poisson simple (GLM)	4604.9	Poisson simple (GLM)	4523.5
Poisson mixto (GLMM)	4605.0	Poisson mixto (GLMM)	4524.0

En ambas comparaciones, el modelo Poisson simple presenta un AIC ligeramente menor (diferencias de 0.1 y 0.5 unidades, respectivamente), lo que indica que la inclusión del efecto aleatorio por equipo visitante no mejora de forma significativa el ajuste. Por tanto, los modelos Poisson simples resulta suficiente para describir el comportamiento de los goles anotados por el equipo visitante.

5.4 Resultado 2: Modelación de la relación entre la altura, el desempeño del equipo al final del partido y su influencia en el resultado

5.4.1 Modelo 3: Regresión ordinal del resultado del equipo visitante en función de la altura y el rendimiento del equipo al final del partido

$$\begin{aligned} \text{logit}(\Pr(Y_i \leq 0 \mid \mathbf{X}_i)) = & \alpha_0 - \beta_1 \text{Asc1}_i - \beta_2 \text{Asc2}_i - \beta_3 \text{Des1}_i \\ & - \beta_4 \text{Des2}_i - \beta_5 \text{RendMedio}_i - \beta_6 \text{RendBajo}_i \end{aligned} \quad (5.3)$$

$$\begin{aligned} \text{logit}(\Pr(Y_i \leq 1 \mid \mathbf{X}_i)) = & \alpha_1 - \beta_1 \text{Asc1}_i - \beta_2 \text{Asc2}_i - \beta_3 \text{Des1}_i \\ & - \beta_4 \text{Des2}_i - \beta_5 \text{RendMedio}_i - \beta_6 \text{RendBajo}_i \end{aligned} \quad (5.4)$$

donde $Y_i \leq 0$ corresponde a que el equipo visitante pierde el partido, y $Y_i \leq 1$ corresponde a que el equipo visitante pierde o empata el partido.

Evaluación del supuesto de Odds Proporcionales

En la Tabla 5.17 se observa que, bajo un nivel de significancia del 5%, no se rechaza la hipótesis nula asociada al cumplimiento del supuesto de odds proporcionales. Por lo tanto, se concluye que tanto el efecto de los grupos de altitud como el del índice de rendimiento se mantienen constantes a través de las diferentes categorías que describen el resultado del partido, lo que indica que el modelo ordinal satisface adecuadamente este supuesto.

Tabla 5.17: Prueba del supuesto de Odds Proporcionales para el modelo de regresión ordinal, utilizando la prueba de Brant

Modelo	χ^2	gl	Valor p
Full	7.91	6	0.24

5.4.2 Bondad de ajuste y calidad del Modelo Ordinal 1

Se realizaron las pruebas de bondad de ajuste del modelo. En primer lugar, la prueba de razón de verosimilitud presentada en la Tabla 5.18 muestra que la desviación del modelo completo es significativamente menor que la del modelo nulo.

Esto indica que las covariables incluidas en el Modelo 3 contribuyen de manera significativa a explicar la variabilidad del resultado del partido. El modelo que incorpora estas variables presenta un mejor ajuste que el modelo nulo, por lo que su inclusión resulta estadísticamente relevante.

Tabla 5.18: Prueba de razón de verosimilitud (LRT) para el modelo logit ordinal (Modelo 3)

Medida	Valor
Comparación entre modelo nulo y modelo completo	
Deviancia modelo nulo (D_{nulo})	3999.367
Deviancia modelo completo (D_{full})	3941.476
Diferencia de grados de libertad	6
Estadístico LR	57.8916
Valor crítico $\chi^2_{0.05, 6}$	12.5916
Valor-p	0

Por otra parte, al evaluar la bondad de ajuste mediante la Tabla 5.19, se observa que la desviación del modelo ajustado es superior al valor crítico de la distribución χ^2 , lo que indica que el Modelo 3 no está obteniendo el mejor ajuste posible, puede ser mejor.

Esta conclusión se complementa con la medida de calidad de ajuste Pseudo- R^2 de Nagelkerke, cuyo valor fue de 3.42%. Este resultado sugiere que el Modelo 3 presenta una capacidad global limitada para discriminar entre las categorías del resultado del partido. En consecuencia, aunque algunas covariables resultan estadísticamente significativas, su contribución al ajuste global del modelo es modesta.

Resultados similares fueron reportados en el análisis de partidos de la Copa Conmebol Cabrera-Hernández, García-Corzo, Povea-Combariza et al. (2023), donde también se encontró que los modelos presentaban una capacidad limitada de ajuste y discriminación para explicar el resultado final del partido. Este hallazgo es consistente con el hecho de que el resultado de un partido de fútbol depende de numerosos factores técnicos, tácticos, físicos, psicológicos y contextuales, de los cuales en este estudio

solo se consideran algunos, específicamente los relacionados con la altura y el rendimiento del equipo.

Tabla 5.19: Bondad de ajuste del modelo logit ordinal (Modelo 3)

Medida	Valor
Desviación del modelo ajustado (D_{full})	3941.476
Grados de libertad residuales (gl)	1890
Valor crítico $\chi^2_{0.05, gl}$	1992.253

Tabla 5.20: Estimación del modelo ordinal para el resultado del partido según la categoría de desplazamiento y el rendimiento

Variable	Coef. (β)	e^{β}	Std. Error	z value	p-value
Asc 1	0.0409	1.042	0.1129	0.3628	0.7167
Asc 2	0.5163	1.676	0.1776	2.9079	0.0036 **
Desc 1	0.0539	1.055	0.1140	0.4727	0.6364
Desc 2	0.1966	1.217	0.1701	1.1559	0.2477
Rendimiento Medio	0.3181	1.374	0.1053	3.0195	0.0025 **
Rendimiento Bajo	0.7914	2.206	0.1182	6.6978	$p < 0.001$ ***
Interceptos	Valor		Std. Error	z value	p-value
Ganó Empató	-0.7778		0.1057	-7.3569	$p < 0.001$ ***
Empató Perdió	0.5719		0.1048	5.4570	$p < 0.001$ ***

Nota: *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$.

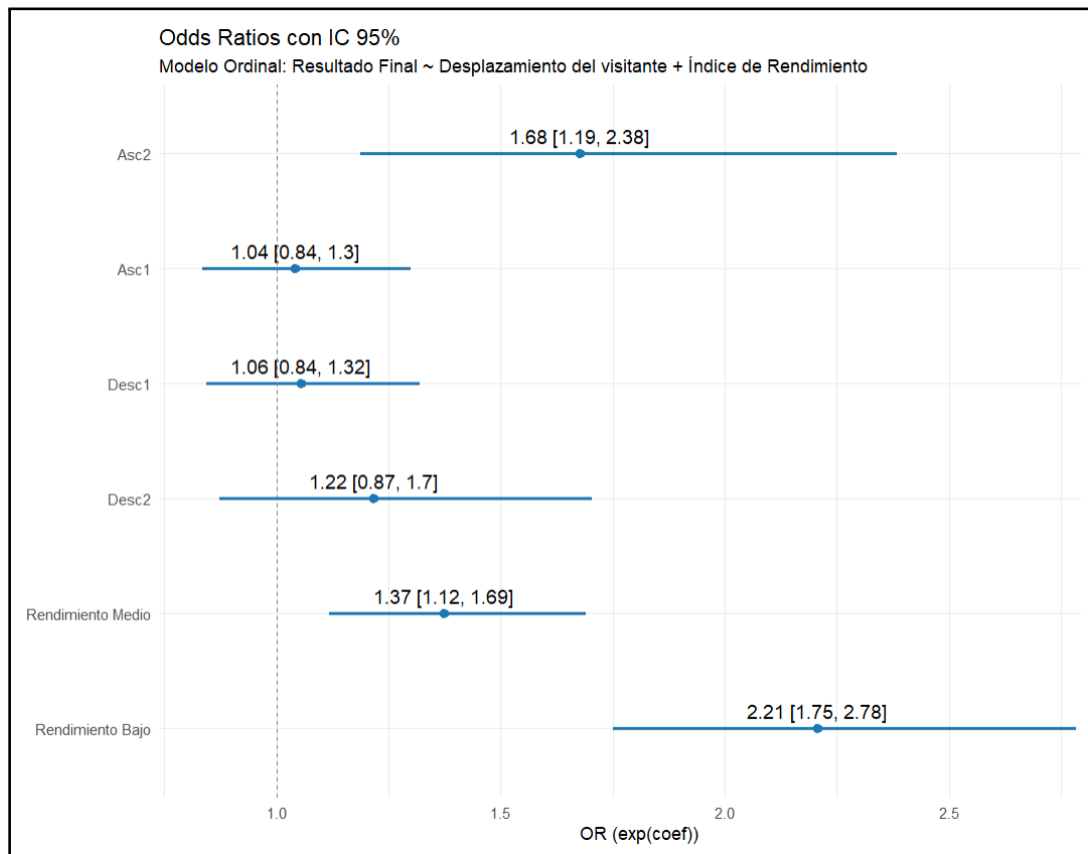


Figura 5.7: Razones de tasas con intervalos de confianza del 95% del Modelo 3.

Después de evaluar la significancia individual de los parámetros del Modelo 3, se observa en la Tabla 5.20 que únicamente algunas condiciones asociadas al desplazamiento y al rendimiento del equipo visitante al final del partido presentan efectos estadísticamente significativos sobre el resultado.

En particular, la categoría Asc 2 resulta estadísticamente significativa. Este resultado indica que, en comparación con la condición de referencia de misma altitud (MA), cuando el equipo visitante asciende dos categorías de altitud, sus odds acumuladas de obtener un resultado menos favorable (perder frente a empatar o ganar, o empatar frente a ganar) aumentan en un 67.6%. En consecuencia, este tipo de desplazamiento se asocia con una mayor probabilidad de que el equipo visitante obtenga un resultado desfavorable.

Por el contrario, los desplazamientos de menor magnitud (Asc 1, Desc 1 y Desc 2) no presentan efectos estadísticamente significativos, lo que sugiere que estos cambios de altitud no generan un impacto concluyente sobre el resultado del partido del equipo visitante.

En relación con el rendimiento del equipo visitante, tanto el rendimiento medio como el rendimiento bajo al final del partido muestran efectos estadísticamente significativos. Tomando como referencia un rendimiento alto al final del partido, un rendimiento medio incrementa las odds acumuladas de un resultado menos favorable en un 37.4%. Este efecto se intensifica cuando el equipo visitante presenta un rendimiento bajo, caso en el cual las odds aumentan en un 120.6%, constituyéndose como el efecto de mayor magnitud dentro del modelo.

La Figura 5.7 complementa la interpretación del Modelo 4 al presentar de forma gráfica las razones de odds estimadas junto con sus intervalos de confianza al 95%.

5.4.3 Modelo 4: Regresión ordinal del resultado del equipo visitante en función de los factores del índice de rendimiento

$$\begin{aligned} \text{logit}(\Pr(Y_i \leq 0 \mid \mathbf{X}_i)) = & \alpha_0 - \beta_1 CO_{\text{bajo},i} - \beta_2 CO_{\text{medio},i} \\ & - \beta_3 SD_{\text{bajo},i} - \beta_4 SD_{\text{medio},i} \\ & - \beta_5 EF_{\text{bajo},i} - \beta_6 EF_{\text{medio},i} \quad (5.5) \end{aligned}$$

$$\begin{aligned} \text{logit}(\Pr(Y_i \leq 1 \mid \mathbf{X}_i)) = & \alpha_1 - \beta_1 CO_{\text{bajo},i} - \beta_2 CO_{\text{medio},i} \\ & - \beta_3 SD_{\text{bajo},i} - \beta_4 SD_{\text{medio},i} \\ & - \beta_5 EF_{\text{bajo},i} - \beta_6 EF_{\text{medio},i} \quad (5.6) \end{aligned}$$

Evaluación del supuesto de Odds Proporcionales

Se realizó la prueba del supuesto de odds proporcional. En la Tabla 5-12 se observa que, bajo un nivel de significancia del 5%, no se rechaza la hipótesis nula del cumplimiento del supuesto. Por lo tanto, se infiere que el efecto de los grupos de altitud es el mismo sobre las categorías que describen el resultado del partido.

Tabla 5.21: Prueba del supuesto de Odds Proporcionales para el modelo de regresión ordinal, utilizando la prueba de Brant

Modelo	χ^2	gl	Valor p
Full	5.80	6	0.45

5.4.4 Bondad de ajuste y calidad del Modelo Ordinal 2

La prueba de razón de verosimilitud presentada en la Tabla 5.22 indica que la desviancia del modelo completo es significativamente menor que la del modelo nulo. Esto confirma que las covariables incluidas en el Modelo 4 contribuyen de manera estadísticamente significativa a explicar el resultado del partido.

Tabla 5.22: Prueba de razón de verosimilitud (LRT) para el modelo logit ordinal (Modelo 4)

Medida	Valor
Comparación entre modelo nulo y modelo completo	
Deviancia modelo nulo (D_{nulo})	3999.367
Deviancia modelo completo (D_{full})	3916.703
Diferencia de grados de libertad	6
Estadístico LR	82.6639
Valor crítico $\chi^2_{0.05, 6}$	12.5916
Valor-p	0

Sin embargo, la Tabla 5.23 muestra que la desviación del modelo ajustado es mayor que el valor crítico de la distribución χ^2 , lo que sugiere que el Modelo 4 no presenta un ajuste adecuado a los datos.

Tabla 5.23: Bondad de ajuste del modelo logit ordinal (Modelo 4)

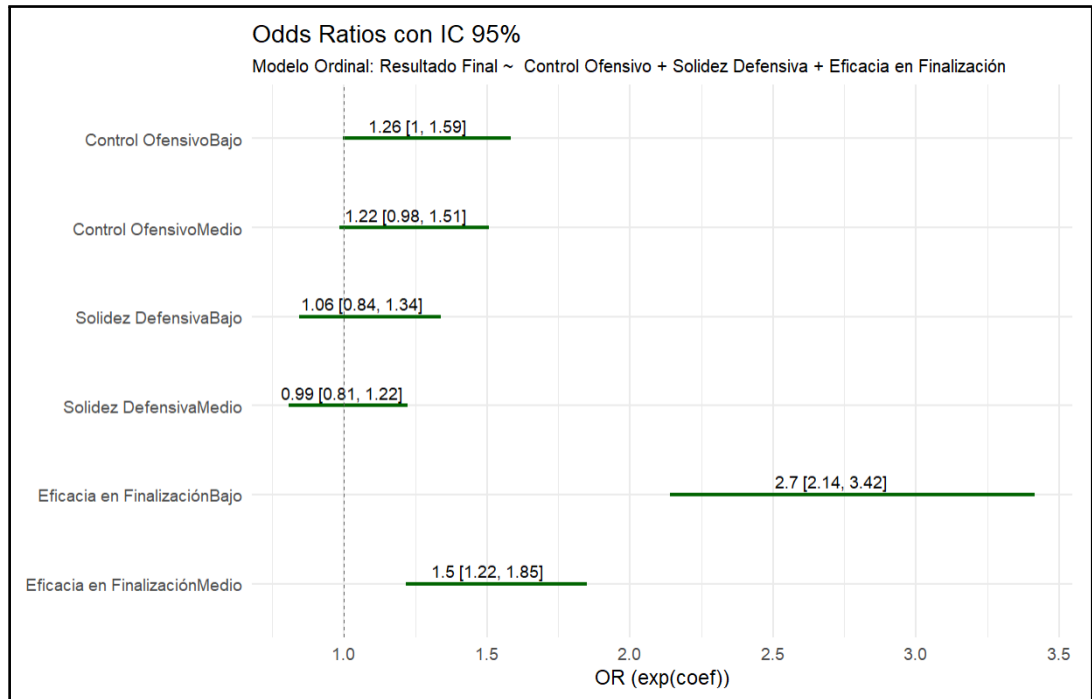
Medida	Valor
Bondad de ajuste del modelo completo	
Desviación del modelo ajustado (D_{full})	3916.703
Grados de libertad residuales (gl)	1890
Valor crítico $\chi^2_{0.05, gl}$	1992.253

En términos de calidad de ajuste, el Pseudo- R^2 de Nagelkerke fue de 4.85%, lo cual sugiere que el Modelo 4, aunque incorpora covariables relevantes, presenta una capacidad global limitada para discriminar entre los resultados del partido.

Tabla 5.24: Estimación del modelo ordinal para el resultado del partido según los factores del índice de rendimiento

Variable	Coef. (β)	e^β	Std. Error	z value	p-value
Control Ofensivo Bajo	0.2280	1.256	0.1188	1.9194	0.0550
Control Ofensivo Medio	0.1967	1.217	0.1087	1.8094	0.0703
Solidez Defensiva Bajo	0.0612	1.063	0.1179	0.5194	0.6034
Solidez Defensiva Medio	-0.0055	0.995	0.1060	-0.0515	0.9590
Eficacia en Finalización Bajo	0.9943	2.703	0.1191	8.3455	$p < 0.001$ ***
Eficacia en Finalización Medio	0.4058	1.501	0.1068	3.8011	$p < 0.001$ ***
Interceptos	Valor		Std. Error	z value	p-value
Ganó Empató	-0.5850		0.1339	-4.3704	$p < 0.001$ ***
Empató Perdió	0.7787		0.1344	5.7960	$p < 0.001$ ***

Nota: *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$.

**Figura 5.8:** Razones de tasas con intervalos de confianza del 95% del Modelo 4.

Después de evaluar la significancia individual de los parámetros del Modelo 4, se observa en la Tabla 5.24 que únicamente el factor Eficacia en Finalización del equipo visitante presenta efectos estadísticamente significativos sobre el resultado final del partido. Tanto el nivel bajo como el nivel medio de esta dimensión del rendimiento muestran una asociación clara con un aumento en las odds acumuladas de obtener un resultado menos favorable para el equipo visitante.

En particular, cuando el equipo visitante presenta una Eficacia en Finalización baja al final del partido, las odds acumuladas de obtener un resultado desfavorable (perder frente a empatar o ganar, o empatar frente a ganar) aumentan en un 170.3% ($OR = 2.703$), en comparación con equipos que presentan una eficacia alta al final del partido. De manera similar, si el equipo visitante presenta una Eficacia en Finalización media al final del partido, dichas odds aumentan en un 50.1% ($OR = 1.501$).

En contraste, los factores Control Ofensivo y Solidez Defensiva del equipo visitante, tanto en sus niveles bajo como medio, no presentan efectos estadísticamente significativos. Esto sugiere que, dentro del Modelo 4, estas dimensiones del rendimiento no generan cambios concluyentes en la probabilidad del resultado final del partido del equipo visitante.

En este contexto, los resultados indican que, aun cuando un equipo visitante pueda mostrar un adecuado Control Ofensivo o Solidez Defensiva, la falta de Eficacia en la Finalización de las jugadas ofensivas incrementa de manera significativa la probabilidad de obtener un resultado desfavorable al final del partido.

La Figura 5.8 complementa la interpretación del Modelo 4 al presentar de forma gráfica las razones de odds estimadas junto con sus intervalos de confianza al 95%.

Capítulo 6

Conclusiones

En el contexto del Fútbol Profesional Colombiano (FPC) no se cuenta con variedad de estudios que analicen con qué variables se relacionan el resultado y el número de goles anotados por los equipos visitantes. En este trabajo se parte de examinar esa relación del resultado y el número de goles anotados por los equipos visitantes con un índice que mide el desempeño de los equipo al final de cada partido creado por Chávez-Avendaño (2025). Además, se incorpora como variable contextual el desplazamiento realizado por el equipo visitante para disputar el encuentro, medido a través de la altitud sobre el nivel del mar, bajo la hipótesis de que la altura sobre el nivel del mar puede influir en el resultado y en la capacidad anotadora de los equipos visitantes.

De acuerdo a las definiciones propuesta por Chávez-Avendaño (2025) para cada factor (Control Ofensivo, Solidez Defensiva y Eficacia en la Finalización), y según los resultados obtenidos en el presente estudio, se concluye que los factores del índice que mide el rendimiento en los partidos de los equipos, se encuentran bien contruidos y muestran coherencia con la dinámica real del fútbol profesional, ya que los factores Control Ofensivo y Eficacia en Finalización sí deberían tener una influencia directa sobre el número de goles anotados, y la Solidez Defensiva no, esto dado los indicadores que lo componen y esto efectivamente se encontró.

El factor Solidez Defensiva se concluye que siguiendo la lógica del juego debería ser más determinante para analizar el comportamiento del equipo cuando recibe goles.

Al estudiar la relación entre el número de goles anotados y el desplazamiento que tuvo que hacer el equipo visitante para disputar el encuentro, se evidenció que cuando el equipo visitante debe ascender dos categorías (en promedio, 2420 m.s.n.m.), su capacidad para anotar goles se ve afectada.

Además, se evidenció que niveles bajos o medios de rendimiento global al final del partido se asocian con una menor capacidad anotadora, lo que permite concluir que el rendimiento del equipo, medido a través del índice utilizado en este estudio, se relaciona de manera significativa con la cantidad de goles que anota.

Cuando el equipo visitante asciende dos categorías de altitud (Asc 2), en comparación con permanecer en misma altitud (MA), aumenta en un 67.6% sus odds de obtener un resultado más desfavorable, lo cual es coherente con estudios previos sobre partidos de la Copa Conmebol Cabrera-Hernández, García-Corzo, Povea-Combariza et al. (2023)

Respecto al rendimiento, tanto el rendimiento medio como bajo al final de partido incrementan significativamente las odds de un resultado desfavorable para el visitante (perder frente a empatar o ganar, o empatar frente a ganar), lo que confirma que un rendimiento inferior con respecto al alto al final del partido impacta negativamente en el resultado.

Entre los factores del índice de rendimiento, únicamente la Eficacia en Finalización presenta efectos estadísticamente significativos: sus niveles bajo y medio aumentan la probabilidad de que el equipo visitante obtenga un resultado menos favorable.

Los modelos desarrollados en este estudio constituyen herramientas útiles para el análisis del Fútbol Profesional Colombiano y para apoyar la toma de decisiones tácticas. En particular, los resultados sugieren que el desplazamiento a altitudes elevadas, como 2420 m s. n. m., afecta tanto la capacidad anotadora como los resultados finales de los equipos visitantes. En este sentido, los hallazgos resaltan que es indispensable realizar una adecuada planificación de los procesos de aclimatación y preparación física por parte de los cuerpos técnicos.

Además, la Eficacia en Finalización, cuyo factor está compuesto por indicadores como tiros a portería, pases de finalización efectivos, centros acertados, penetraciones al área e interceptaciones, se consolida como el factor más determinante tanto para obtener resultados favorables como para incrementar el número de goles anotados. Este resultado orienta el trabajo táctico y técnico hacia el fortalecimiento de los procesos de creación y concreción de oportunidades de gol, aspectos clave para incrementar la probabilidad de victoria y el número de goles anotados en los partidos disputados.

Capítulo 7

Recomendaciones

Se sugiere perfeccionar el índice que mide el rendimiento al final del partido, dado que en su versión actual solo explica aproximadamente el 33% de la variabilidad. En una versión mejorada, podría incorporarse la variable anotar el primer gol, la cual ha sido destacada en la literatura como un factor fundamental en el análisis de los resultados en el fútbol. Asimismo, sería conveniente explorar otros métodos de construcción del índice, como el Análisis Factorial Múltiple (AFM), con el fin de obtener una medida más robusta del rendimiento.

Adicionalmente, se recomienda replicar los análisis considerando el rendimiento de los equipos en condición de local. En este caso, la altitud no representa un factor determinante, dado que el equipo no experimenta desplazamiento. No obstante, resulta pertinente evaluar su rendimiento en casa, ya que las dinámicas propias del juego local pueden ofrecer información complementaria a los hallazgos del visitante.

Es importante señalar que para los equipos locales el supuesto de independencia podría verse comprometido, debido a que tienden a mantener un estilo de juego "constante" en su estadio. En este escenario, la utilización de modelos generalizados mixtos con efectos aleatorios a nivel de equipo podría resultar más adecuada para capturar dicha dependencia.

Por otra parte, los resultados mostraron que la Solidez Defensiva es independiente del número de goles anotados y el resultado final, lo cual indica que no constituye un factor determinante en la ofensiva visitante. Sin embargo, este factor podría cobrar relevancia en otro tipo de estudios, como por ejemplo en el análisis de goles recibidos, tanto para equipos locales como visitantes.

Capítulo 8

Anexos

Estadio	Ciudad	Departamento
Alberto Grisales	Rionegro	Antioquia
Daniel Villa Zapata	Barrancabermeja	Santander
Pascual Guerrero	Cali	Valle del Cauca
La Independencia	Tunja	Boyacá
Alfonso López	Bucaramanga	Santander
Palogrande	Manizales	Caldas
Deportivo Cali	Palmira	Valle del Cauca
Doce de Octubre	Tuluá	Valle del Cauca
General Santander	Cúcuta	Norte de Santander
Polideportivo Sur	Envigado	Antioquia
Metropolitano de Techo	Bogotá	Cundinamarca
Guillermo Plazas Alcid	Neiva	Huila
Jaraguay	Montería	Córdoba
Metropolitano	Barranquilla	Atlántico
Metropolitano Itagüí	Itagüí	Antioquia
Sierra Nevada	Santa Marta	Magdalena
Atanasio Girardot	Medellín	Antioquia
El Campín	Bogotá	Cundinamarca
Departamental Libertad	Pasto	Nariño
Hernán Ramírez Villegas	Pereira	Risaralda
Manuel Murillo Toro	Ibagué	Tolima

Tabla 8.1: Estadios, ciudades y departamentos del fútbol profesional colombiano.

Tabla 8.2: Distribución de goles por equipo visitante.

Equipo	Sin goles	Un gol	Dos goles	Tres o más goles	Partidos
Águilas	32	31	20	5	88
Alianza	33	38	15	2	88
América	29	29	17	6	81
Boyacá	28	15	5	0	48
Bucaramanga	40	35	12	4	91
Caldas	35	32	17	7	91
Cali	36	41	15	8	100
Cortuluá	11	20	7	2	40
Cúcuta	11	14	4	3	32
Envigado	36	33	12	5	86
Equidad	33	39	13	4	89
Fortaleza	10	3	5	0	18
Huila	49	24	3	3	79
Jaguars	45	28	16	3	92
Junior	38	31	27	8	104
Leones	8	8	2	1	19
Magdalena	12	10	0	0	22
Medellín	29	37	28	7	101
Millonarios	30	39	24	5	98
Nacional	33	37	20	16	106
Pasto	39	37	12	6	94
Patriotas	49	30	9	3	91
Pereira	6	3	3	1	13
Santa Fe	40	40	12	11	103
Tigres	11	8	1	0	20
Tolima	31	41	23	9	104

Tabla 8.3: Resultado final según el equipo visitante

Equipo	Ganó	Empató	Perdió	Partidos
Aguilas	17	36	35	88
Alianza	14	25	49	88
América	21	26	34	81
Boyacá	3	5	40	48
Bucaramanga	23	28	40	91
Caldas	23	27	41	91
Cali	20	32	48	100
Cortuluá	9	9	22	40
Cúcuta	9	5	18	32
Envigado	16	24	46	86
Equidad	20	30	39	89
Fortaleza	1	3	14	18
Huila	9	26	44	79
Jaguars	17	20	55	92
Junior	26	42	36	104
Leones	1	5	13	19
Magdalena	2	5	15	22
Medellín	32	25	44	101
Millonarios	31	30	37	98
Nacional	41	34	31	106
Pasto	18	27	49	94
Patriotas	11	30	50	91
Pereira	3	4	6	13
Santa Fe	34	32	37	103
Tigres	3	5	12	20
Tolima	34	30	40	104

Tabla 8.4: Nivel de rendimiento según el equipo visitante

Equipo	Rend- Alto	Rend - Medio	Rend - Bajo	Partidos
Aguilas	20	44	24	88
Alianza	16	41	31	88
América	23	36	22	81
Boyacá	7	20	21	48
Bucaramanga	13	42	36	91
Caldas	29	39	23	91
Cali	25	53	22	100
Cortuluá	4	23	13	40
Cúcuta	5	18	9	32
Envigado	18	32	36	86
Equidad	26	37	26	89
Fortaleza	3	7	8	18
Huila	14	33	32	79
Jaguars	18	42	32	92
Junior	33	54	17	104
Leones	3	4	12	19
Magdalena	4	8	10	22
Medellín	32	44	25	101
Millonarios	32	53	13	98
Nacional	54	39	13	106
Pasto	19	48	27	94
Patriotas	17	39	35	91
Pereira	2	7	4	13
Santa Fe	32	52	19	103
Tigres	2	7	11	20
Tolima	37	45	22	104

Tabla 8.5: Resumen descriptivo de goles por nivel del índice de rendimiento.

Índice de Rendimiento	Media de goles	SD goles	Mín. goles	Máx. goles	n
Rendimiento Alto	1.240	1.040	0	6	488
Rendimiento Bajo	0.578	0.696	0	3	543
Rendimiento Medio	0.935	0.949	0	5	867

Bibliografía

Silva, L. C. (1997). *Cultura Estadística e Investigación Científica en el Campo de la Salud: Una Mirada Crítica*. Díaz de Santos.

Bland, J. M. e D. G. Altman (2000). «Statistics Notes: The Odds Ratio». In: *BMJ* 320.7247, p. 1468.

Lee, A. J. (2005). «Regression models for forecasting goals and match results in association football». In: *International Journal of Forecasting* 21.2, pp. 331–340.

Lago-Peñas, Carlos (2008). «El análisis del rendimiento en los deportes de equipo. Algunas consideraciones metodológicas». In: *Acción Motriz* 1.1, pp. 41–58.

Coxe, Stefany, Stephen G. West e Leona S. Aiken (2009). «The Analysis of Count Data: A Gentle Introduction to Poisson Regression and Its Alternatives». In: *Journal of Personality Assessment* 91.2, pp. 121–136.

Lago-Peñas, Carlos, Luis Casáis, Alexandre Dellal, Enrique Rey e Eduardo Domínguez (2010). «La influencia de la localización del partido, el nivel del oponente y el marcador en la posesión del balón en el fútbol de alto nivel». In: *Apunts: Educación Física y Deportes* 102, pp. 78–86.

Montgomery, Douglas C., Elizabeth A. Peck e G. Geoffrey Vining (2012). *Introduction to Linear Regression Analysis*. 5ª ed. Wiley.

- García-Artiles, María-Dolores, Emilio Gómez-Déniz e Nancy Dávila Cárdenes (2014). «El modelo Poisson generalizado inflado de ceros: una aplicación en el entorno educativo universitario». In: *Revista Electrónica de Comunicaciones y Trabajos de ASEPUMA. Rect@* 15, pp. 159–171.
- Agresti, Alan (2015). *Foundations of Linear and Generalized Linear Models*. Hoboken, NJ: Wiley.
- Matijasevic, Eduardo (2017). «Campeonato Mundial de Fútbol de las Llanuras». In: *Acta Médica Colombiana* 42.3, pp. 149–162.
- Castellano, Julián (2018). «Relación entre indicadores de rendimiento y el éxito en el fútbol profesional». In: *Revista Iberoamericana de Psicología del Ejercicio y el Deporte* 13.1, pp. 41–47.
- Gómez Gallego, Antonio (2019). «Análisis del número de bares, restaurantes y cafeterías en los municipios de la provincia de Jaén mediante un modelo de regresión de Poisson».
- Bilek, Günal e Betül Aygün (2022). «Factors Associated with Match Result and Number of Goals Scored and Conceded in the English Premier League». In: *Bitlis Eren University Journal of Science* 11.1, pp. 227–236.
- Infobae (2022). *Bolivia recibe a Brasil en La Paz por la última fecha de las Eliminatorias Sudamericanas*. URL: <https://www.infobae.com/america/deportes/2022/03/29/bolivia-recibe-a-brasil-en-la-paz-por-la-ultima-fecha-de-las-eliminatorias-sudamericanas-hora-tv-y-formaciones/> (visitato il giorno 29/03/2022).
- Cabrera-Hernández, Moisés Arturo, Sergio Andrés García-Corzo, Camilo Ernesto Povea-Combariza et al. (2023). «CONMEBOL Libertadores Cup: Altitude Impact on Goals and Results in 16 Years of Soccer Matches». In: *Journal of Human Sport and Exercise* 18.1, pp. 34–45.

Chávez-Avendaño, N. A. (2025). «Construcción de un indicador para evaluar el rendimiento de equipos de fútbol usando variables latentes». Tesis de pregrado. Santiago de Cali, Colombia: Universidad del Valle, Facultad de Ingeniería, Escuela de Estadística.