



---

**Universidad de Valladolid**

Facultad de Ciencias

TRABAJO FIN DE GRADO

GRADO EN ESTADÍSTICA

**Modelización estadística de resultados  
en fútbol a partir de indicadores de rotación de  
equipos**

Autor:

**Miguel García de Pedro**

Tutor:

**Teresa González Arteaga**



# Agradecimientos

*Quiero agradecer, en primer lugar, a mi tutora, Teresa González Arteaga, por su ayuda durante todo el desarrollo de este Trabajo de Fin de Grado.*

*También me gustaría dar las gracias a mis compañeros y amigos de la carrera, por todos estos años compartidos. Sin ellos, nada de esto habría sido posible.*

*Por último, a mi familia y a mi novia, gracias por estar siempre ahí. Por vuestro apoyo y por animarme en cada paso.*



# Resumen

Este Trabajo de Fin de Grado tiene como objetivo modelizar estadísticamente los resultados en el fútbol profesional a partir de indicadores de rotación de equipos, centrándose en cómo las decisiones relacionadas con las alineaciones y las sustituciones influyen en el desenlace de los partidos. Para ello, se ha recopilado y procesado un conjunto de datos detallado de la Primera División española entre las temporadas 2019 y 2024, excluyendo partidos de equipos con presupuestos extremos y primeras jornadas sin historial previo suficiente.

El estudio comienza con un análisis descriptivo de las variables objetivo (resultado del partido y goles de cada equipo) y de variables explicativas relacionadas con las rotaciones. A continuación, se implementan y comparan distintos modelos predictivos: regresión de Poisson, regresión logística multinomial (con y sin regularización mediante `glmnet`), y el algoritmo `XGBoost`, tanto en versiones de regresión como de clasificación directa.

Los resultados muestran que, los modelos más simples, como la regresión de Poisson, proporcionan un rendimiento aceptable, aunque con ciertas limitaciones. En cambio, al aplicar métodos como la regresión multinomial, la regresión regularizada con `glmnet` o la clasificación directa con `XGBoost`, se ha conseguido una mejora significativa en la precisión, alcanzando niveles de *accuracy* por encima del 48 %. Estos enfoques permiten capturar mejor la complejidad del problema y ofrecen resultados más competitivos. .

Este trabajo aporta evidencia empírica sobre el valor de incluir indicadores de rotación en los modelos estadísticos de predicción deportiva, y destaca el potencial de la estadística aplicada para mejorar la toma de decisiones en contextos competitivos.

**Palabras clave:** fútbol, modelos estadísticos, predicción de resultados, Poisson, `XGBoost`, regresión logística multinomial, rotación de equipos.



# Abstract

This Bachelor's Thesis aims to statistically model football match outcomes based on team rotation indicators, focusing on how lineup and substitution decisions may affect match results. To this end, a detailed dataset from the Spanish First Division (2019–2024) was compiled and processed, excluding matches involving financially extreme teams and early-season games lacking historical data.

The study begins with a descriptive analysis of both target variables (match result and goals scored) and explanatory variables related to player rotations. Subsequently, various predictive models are implemented and compared: Poisson regression, multinomial logistic regression (with and without regularization using `glmnet`), and XGBoost, both as regression and direct classification tools.

Results show that simpler models, such as Poisson regression, offer acceptable performance, although with certain limitations. In contrast, applying more advanced methods—such as multinomial regression, regularized regression with `glmnet`, or direct classification with XGBoost—has led to a significant improvement in precision, reaching *accuracy* levels above 48%. These more sophisticated approaches better capture the complexity of the problem and yield more competitive results.

This work provides empirical evidence on the value of including rotation indicators in football prediction models and highlights the potential of applied statistics to support decision-making in competitive settings.

**Key words:** football, statistical models, outcome prediction, Poisson, XGBoost, multinomial logistic regression, team rotation.



# Índice general

<b>1. Introducción</b>	<b>1</b>
1.1. Contexto y motivación . . . . .	1
1.2. Objetivo del trabajo . . . . .	2
1.3. Estructura del documento . . . . .	3
1.4. Tecnologías utilizadas . . . . .	3
<b>2. Conceptos Teóricos</b>	<b>5</b>
2.1. Introducción a la modelización estadística . . . . .	5
2.2. Regresión de Poisson . . . . .	6
2.3. Regresión logística multinomial . . . . .	7
2.4. Penalización y regresión regularizada (GLMNet) . . . . .	8
2.5. XGBoost . . . . .	10
<b>3. Conjunto de datos</b>	<b>13</b>
3.1. Tratamiento de los datos . . . . .	13
3.2. Estructura del conjunto de datos . . . . .	14
<b>4. Análisis Descriptivo</b>	<b>17</b>
4.1. Análisis de las variables respuesta . . . . .	17
4.2. Análisis descriptivo de las sustituciones . . . . .	23
4.3. Comparación entre normativas de sustituciones . . . . .	27
<b>5. Modelos predictivos</b>	<b>29</b>

5.1. Regresión de Poisson . . . . .	29
5.2. Regresión logística multinomial . . . . .	35
5.3. Regresión logística multinomial penalizada (glmnet) . . . . .	39
5.4. Modelo de regresión con XGBoost . . . . .	41
5.5. Clasificación directa del resultado con XGBoost . . . . .	44
<b>6. Conclusiones</b>	<b>47</b>
<b>Bibliografía</b>	<b>53</b>
<b>Repositorio del código</b>	<b>55</b>
<b>Listado completo de variables</b>	<b>57</b>
.1. Variables del conjunto de datos . . . . .	57
<b>Salidas de Modelos de Poisson</b>	<b>65</b>
.2. Modelo completo de goles del equipo local . . . . .	65
.3. Modelo reducido de goles del equipo local . . . . .	68
.4. Modelo completo de goles del equipo visitante . . . . .	68
.5. Modelo reducido de goles del equipo visitante . . . . .	71
<b>Salidas de Modelos Multinomiales</b>	<b>73</b>
.6. Modelo multinomial completo . . . . .	73
.7. Modelo multinomial reducido . . . . .	79

# Índice de figuras

2.1. Comparación visual entre regresión Lasso, Ridge y Elastic Net. Fuente: <i>Zou y Hastie (2005)</i> , adaptado desde <i>Medium</i> . . . . .	9
4.1. Distribución de goles del equipo local . . . . .	18
4.2. Distribución de goles del equipo visitante . . . . .	18
4.3. Boxplot: comparación de goles local vs visitante . . . . .	19
4.4. Media de goles por temporada . . . . .	19
4.5. Tendencia de goles del equipo local por jornada . . . . .	20
4.6. Tendencia de goles del equipo visitante por jornada . . . . .	20
4.7. Distribución de resultado del partido . . . . .	20
4.8. Distribución del resultado del partido por temporada . . . . .	21
4.9. Frecuencia de combinaciones de goles local vs visitante . . . . .	22
4.10. Comparación de proporciones de cambios por equipo y reglamento . . . . .	23
4.11. Comparación de cambios en alineación por reglamento y tipo de equipo . . . . .	25
4.12. Comparación del minuto promedio de cambios por reglamento y tipo de equipo . . . . .	26
5.1. Accuracy en test según alpha (glmnet multinomial, conjunto completo) . . . . .	40



# Índice de tablas

4.1. Distribución del resultado del partido (1, X, 2) . . . . .	20
4.2. Combinaciones de goles más frecuentes . . . . .	22
4.3. Combinaciones de goles más inusuales . . . . .	23
4.4. Estadísticas descriptivas de proporciones de cambios por equipo y reglamento . . . . .	24
4.5. Estadísticas descriptivas de cambios en la alineación por reglamento y tipo de equipo . . . . .	25
4.6. Estadísticas del minuto promedio de cambios por equipo y reglamento . . . . .	26
5.1. Comparación entre modelo completo y modelo reducido mediante test de razón de verosimilitud (modelo de Poisson para goles locales) . . . . .	31
5.2. Comparación entre modelo completo y modelo reducido mediante test de razón de verosimilitud (modelo de Poisson para goles visitantes) . . . . .	32
5.3. Matriz de confusión del modelo multinomial completo. . . . .	36
5.4. Matriz de confusión del modelo multinomial reducido. . . . .	37
5.5. Comparación entre modelo completo y reducido (regresión logística multinomial) . . . . .	38
5.6. Resultados del modelo glmnet multinomial según el parámetro $\alpha$ (conjunto completo) . . . . .	39
5.7. Matriz de confusión del resultado del partido (XGBoost, conjunto completo) . . . . .	42
5.8. Matriz de confusión del mejor modelo XGBoost (clasificación directa, todos los partidos) . . . . .	44
1. Resumen de coeficientes del modelo de Poisson para goles locales . . . . .	65
2. Resumen de coeficientes del modelo de Poisson para goles locales . . . . .	66
3. Resumen de coeficientes del modelo de Poisson para goles locales . . . . .	67
4. Resumen de coeficientes del modelo de Poisson reducido para goles locales . . . . .	68
5. Resumen de coeficientes del modelo de Poisson para goles locales . . . . .	68

6.	Resumen de coeficientes del modelo de Poisson para goles locales . . . . .	69
7.	Resumen de coeficientes del modelo de Poisson para goles visitantes . . . . .	70
8.	Resumen de coeficientes del modelo de Poisson para goles visitantes significativo . . . . .	71
9.	Resumen de coeficientes del modelo Multinomial completo . . . . .	73
10.	Resumen de coeficientes del modelo Multinomial completo . . . . .	74
11.	Resumen de coeficientes del modelo Multinomial completo . . . . .	75
12.	Resumen de coeficientes del modelo Multinomial completo . . . . .	76
13.	Resumen de coeficientes del modelo Multinomial completo . . . . .	77
14.	Resumen de coeficientes del modelo Multinomial completo . . . . .	78
15.	Resumen de coeficientes del modelo Multinomial reducido . . . . .	79
16.	Resumen de coeficientes del modelo Multinomial reducido . . . . .	80

# Capítulo 1

## Introducción

### 1.1. Contexto y motivación

En los últimos años, la estadística ha ido ganando protagonismo dentro del mundo del deporte, sobre todo gracias al enorme crecimiento en la disponibilidad de datos y al interés que despierta su análisis entre clubes, analistas e incluso aficionados. En el caso concreto del fútbol profesional, donde se mezclan factores económicos, estratégicos y emocionales, contar con modelos que ayuden a anticipar el resultado de los partidos se ha convertido en algo cada vez más útil, tanto para mejorar la toma de decisiones en el ámbito deportivo como para avanzar en el estudio académico del juego.

Los primeros intentos de aplicar la estadística al fútbol se remontan a modelos clásicos como el de Poisson, utilizado para predecir el número de goles en un partido [1]. A partir de ahí, se han ido incorporando enfoques más complejos, como los métodos bayesianos [2], las redes probabilísticas [3] y otras técnicas de modelización más sofisticadas que han permitido mejorar la capacidad predictiva de estos modelos. Aun así, el fútbol sigue siendo un deporte con un alto grado de aleatoriedad y está muy influido por factores situacionales, lo que supone todavía un desafío importante para cualquier modelo estadístico.

El fútbol ha vivido una evolución notable no solo como disciplina deportiva, sino también como fenómeno global. Hoy en día, no es solo el deporte más seguido del mundo, sino también una industria multimillonaria con una enorme repercusión social, mediática y económica. Las grandes competiciones generan audiencias masivas, los clubes funcionan como marcas internacionales, y las decisiones que se toman dentro y fuera del campo tienen un impacto que va mucho más allá del resultado deportivo. Esta profesionalización y expansión han incrementado la presión por optimizar el rendimiento y minimizar la incertidumbre, lo que ha favorecido el desarrollo de herramientas analíticas más sofisticadas, entre ellas los modelos estadísticos, como apoyo para la toma de decisiones estratégicas.

Uno de los factores que más influyen en el rendimiento de un equipo es la *rotación de jugadores*, es decir, los cambios en la alineación titular que se hacen de un partido a otro. En competiciones exigentes, como las grandes ligas o los torneos europeos, el calendario tan apretado obliga a los entrenadores a dosificar el esfuerzo de sus futbolistas, introduciendo rotaciones de forma regular. Según Chris Carling, Franck Le Gall y Gregory Dupont [4], la acumulación de minutos puede provocar fatiga y aumentar el riesgo de lesiones, lo que hace que estas rotaciones tengan sentido desde el punto de vista físico. Ahora bien, modificar la alineación también puede afectar a la cohesión del equipo o a su nivel competitivo, y eso, lógicamente, puede acabar influyendo en el resultado final del partido.

Desde un punto de vista estadístico, la rotación de jugadores puede considerarse una variable explicativa con potencial dentro de los modelos de predicción de resultados. Aunque existe bastante literatura centrada en los aspectos tácticos o físicos del juego, todavía son pocos los estudios que analizan de forma rigurosa cómo los cambios en la alineación influyen directamente en el resultado de un partido. En este trabajo se pretende precisamente cubrir ese hueco, planteando un modelo estadístico que tenga en cuenta no solo la rotación del equipo, sino también otros indicadores históricos de rendimiento que puedan ayudar a anticipar el desenlace de los encuentros.

La principal motivación de este trabajo es combinar técnicas estadísticas avanzadas con el análisis de un aspecto del juego que, hasta ahora, ha recibido poca atención: la dinámica de los cambios en la alineación. El objetivo es aportar evidencia empírica sobre la importancia real de este factor y comprobar hasta qué punto puede mejorar la capacidad predictiva de los modelos estadísticos aplicados al fútbol. Con ello, se busca contribuir a una línea de investigación aplicada que resulta útil tanto en el ámbito académico como en el deportivo.

## 1.2. Objetivo del trabajo

El objetivo principal de este trabajo es construir y analizar modelos estadísticos que ayuden a predecir el resultado de partidos de fútbol profesional, basándose en el comportamiento histórico de los equipos. Dentro de todas las variables posibles, el foco se pone especialmente en la rotación de jugadores: es decir, en cómo los cambios en la alineación y las sustituciones realizadas pueden estar relacionados con el desenlace del encuentro.

Para ello, se han recopilado datos de varias temporadas de la Primera División española, incluyendo información detallada sobre resultados, alineaciones, cambios y otros indicadores previos al partido. Con esta base, se han aplicado distintos modelos —como distribución de Poisson, distribución multinomial, modelos penalizados (ridge y lasso) o XGBoost— con el fin de evaluar qué capacidad tienen para anticipar el resultado de un partido en función de esas variables.

En conjunto, este trabajo pretende aportar una visión estadística clara sobre el papel que puede tener la gestión de la plantilla en el rendimiento deportivo, y explorar su utilidad práctica dentro

del análisis aplicado al fútbol.

## 1.3. Estructura del documento

A continuación se detalla la estructura del trabajo, que se compone de seis capítulos:

### ■ **Capítulo 1: Introducción**

Se contextualiza el problema de estudio y se expone la motivación del trabajo. Además, se definen los objetivos principales y se introduce la estructura general del documento.

### ■ **Capítulo 2: Conceptos Teóricos**

Se presentan los fundamentos estadísticos necesarios para comprender los métodos aplicados. Se explican los modelos utilizados, como la regresión de Poisson, la regresión logística multinomial, los modelos penalizados (GLMNet) y XGBoost, así como las métricas empleadas para evaluar su rendimiento.

### ■ **Capítulo 3: Conjunto de datos**

Se describe el origen y tratamiento del conjunto de datos utilizado, incluyendo los filtros aplicados, los indicadores prepartido y la segmentación según normativas de sustituciones. También se detalla la estructura final del dataset empleado en los modelos.

### ■ **Capítulo 4: Análisis Descriptivo**

Se realiza un análisis exploratorio de las variables clave, tanto de los resultados de los partidos como de las rotaciones y sustituciones. Se identifican patrones según la normativa vigente y se extraen conclusiones preliminares de interés.

### ■ **Capítulo 5: Modelos Predictivos**

Se aplican distintos modelos estadísticos para predecir el número de goles y el resultado final de los partidos. Se evalúa y compara su rendimiento tanto en tareas de regresión como de clasificación.

### ■ **Capítulo 6: Conclusiones**

Se resumen las principales aportaciones del trabajo, se discuten sus limitaciones y se proponen posibles líneas de mejora o extensión futura.

## 1.4. Tecnologías utilizadas

El desarrollo completo de este trabajo ha requerido el uso combinado de distintas tecnologías, todas ellas orientadas a garantizar la eficiencia, la reproducibilidad y la trazabilidad del análisis realizado.

La mayor parte del estudio, incluyendo la carga de datos, el tratamiento estadístico, la generación de gráficos y la implementación de modelos, se ha llevado a cabo en el lenguaje de programación R. Para ello se ha utilizado el entorno de desarrollo integrado RStudio, junto con un conjunto de librerías especializadas: `dplyr` y `tidyr` para manipulación de datos, `ggplot2` para la visualización, `glmnet` para la regresión penalizada y `xgboost` para el ajuste de modelos de boosting.

Adicionalmente, se ha empleado Python para tareas puntuales de depuración y filtrado del conjunto de datos, especialmente en lo relativo a la separación de partidos según la normativa de sustituciones (3 o 5 cambios) y la exclusión de equipos con presupuestos atípicos. Estas tareas se apoyaron en el uso de librerías como `pandas` y `numpy`.

Como sistema de control de versiones y almacenamiento, se ha utilizado GitHub, donde se ha mantenido un repositorio con todos los scripts, resultados intermedios y conjuntos de datos procesados. Esto ha permitido una gestión organizada del flujo de trabajo y ha facilitado la reproducibilidad de los análisis realizados.

# Capítulo 2

## Conceptos Teóricos

### 2.1. Introducción a la modelización estadística

#### 2.1.1. Problemas de clasificación y regresión

Cuando hablamos de modelos estadísticos predictivos, normalmente distinguimos entre dos grandes tipos de problemas: los modelos de regresión y los modelos de clasificación [5, 6]. La diferencia principal está en el tipo de variable que queremos predecir. En los modelos de regresión, lo que buscamos es modelizar una variable numérica (por ejemplo, cuántos goles marcará un equipo). En cambio, en los modelos de clasificación intentamos predecir una variable categórica (como si el equipo ganará, empatará o perderá).

En este trabajo se utilizan ambos enfoques. Por un lado, se ajustan modelos de regresión para estimar el número de goles que marcarán el equipo local y el visitante. Por otro, se han desarrollado modelos de clasificación para anticipar directamente el signo del resultado final: victoria local (1), empate (X) o victoria visitante (2). Esta combinación permite valorar qué tipo de modelo se ajusta mejor a este tipo de problema y comparar distintas formas de abordar la predicción.

#### 2.1.2. Variables predictoras y variable respuesta en el contexto del fútbol

Las variables predictoras que se han incluido en los modelos provienen de una recopilación de indicadores previos al partido, como goles marcados, encajados o resultados anteriores. Son datos que reflejan el rendimiento reciente de cada equipo y que, en conjunto, aportan una imagen bastante completa del estado de forma con el que llegan al encuentro.

Dependiendo del tipo de modelo, la variable que queremos predecir (variable respuesta) cambia. En el caso de los modelos de regresión, lo que se intenta estimar es el número de goles que marcará cada equipo. En los modelos de clasificación, en cambio, el objetivo es predecir directamente el

resultado del partido, simplificado en tres posibles categorías: 1, X o 2.

Se ha querido prestar atención al papel que puede jugar la rotación de jugadores y los cambios en la alineación para la selección de variables, algo que habitualmente se pasa por alto en modelos más básicos.

### 2.1.3. Métricas de evaluación

Una vez que se han entrenado los modelos, es importante poder medir su rendimiento de forma objetiva. Para ello, se han utilizado diferentes métricas dependiendo de si el modelo es de regresión o de clasificación.

En el caso de los modelos de clasificación, estas son las dos métricas utilizadas:

- **Accuracy:** mide qué porcentaje de predicciones han sido correctas.
- **Matriz de confusión:** permite ver con más detalle dónde se han producido los errores, es decir, si el modelo tiende a equivocarse más al predecir empates, victorias locales o visitantes.

En los modelos de regresión, se utiliza:

- **MAE (error absoluto medio):** calcula cuánto se desvía, de media, la predicción respecto al valor real, sin importar si el error es por exceso o por defecto.
- **RMSE (raíz del error cuadrático medio):** también mide la desviación, pero penalizando más los errores grandes.

Estas métricas permiten comparar de forma clara los distintos modelos ajustados.

## 2.2. Regresión de Poisson

La **regresión de Poisson** [7] se aplica cuando se quiere predecir un número de eventos que ocurren en un cierto intervalo (por ejemplo, goles en un partido). Este modelo parte de la hipótesis de que la variable que queremos predecir sigue una distribución de Poisson, es decir, una distribución de probabilidades que se utiliza habitualmente para modelar conteos.

Matemáticamente, se supone que la variable respuesta  $Y_i$  (en nuestro caso, el número de goles del equipo local o visitante) sigue una distribución de Poisson con parámetro  $\lambda_i$ :

$$Y_i \sim \text{Poisson}(\lambda_i)$$

donde  $\lambda_i$  representa el número esperado de goles para la observación  $i$ . Este valor esperado no es constante, sino que se modela como una función exponencial de las variables explicativas,  $x_i$ :

$$E(Y_i) = \lambda_i = \exp(\mathbf{x}_i^\top \boldsymbol{\beta})$$

o, de forma equivalente, aplicando logaritmo:

$$\log(\lambda_i) = \mathbf{x}_i^\top \boldsymbol{\beta}$$

De esta forma, el modelo estima cómo influyen las distintas variables predictoras (alineación, rendimiento anterior, cambios, etc.) sobre el número esperado de goles.

Este tipo de regresión es especialmente útil porque garantiza que las predicciones son siempre positivas (no se pueden predecir goles negativos) y porque tiene una interpretación clara: un coeficiente positivo indica que esa variable está asociada a un aumento del número de goles esperados, y uno negativo, a una disminución.

En este trabajo se ha utilizado la regresión de Poisson tanto para predecir los goles del equipo local como del visitante. Posteriormente, a partir de estas predicciones se infiere el posible resultado del partido (1, X o 2), comparando los goles estimados para cada equipo. Aunque es un modelo simple y relativamente fácil de interpretar, una de sus principales limitaciones es que asume que la varianza y la media de la distribución son iguales. En caso de que esto no sea así, se dice que existe sobredispersión y se podrían probar modelos de regresión binomial negativa. En este trabajo no se presenta esta circunstancia, como se comprueba en las secciones 5.1.2 y 5.1.3.

## 2.3. Regresión logística multinomial

Cuando la variable que queremos predecir es categórica y puede tomar más de dos valores, una de las herramientas estadísticas más utilizadas es la **regresión logística multinomial** [8]. Este modelo es una extensión natural de la regresión logística binaria, que permite trabajar con variables respuesta con tres o más categorías.

En este trabajo, se ha ajustado para predecir directamente el resultado del partido en términos de victoria local (1), empate (X) o victoria visitante (2), por lo que estamos ante un problema de clasificación multiclase. El modelo asume que la probabilidad de que una observación pertenezca a una determinada clase puede modelarse en función de las variables explicativas mediante una transformación logística.

Supongamos que  $Y$  es la variable respuesta, que puede tomar  $K$  valores distintos y  $p$  variables predictoras  $x_1, \dots, x_p$ . En este caso,  $K = 3$ . El modelo calcula la probabilidad de que una

observación pertenezca a cada clase  $k$  mediante la siguiente expresión:

$$P(Y_i = k) = \frac{\exp(\mathbf{x}_i^\top \boldsymbol{\beta}_{ik})}{1 + \sum_{j=1}^{K-1} \exp(\mathbf{x}_i^\top \boldsymbol{\beta}_{ji})} \quad \text{para } k = 1, \dots, K - 1$$

y

$$P(Y_i = K) = \frac{1}{1 + \sum_{j=1}^{K-1} \exp(\mathbf{x}_i^\top \boldsymbol{\beta}_{jk})}$$

donde  $\boldsymbol{\beta}_{kj}$  representa los coeficientes asociados a la clase  $k$  y la variable predictora  $x_j$ .

Este modelo tiene varias ventajas: es relativamente fácil de ajustar e interpretar, y permite estimar no solo la clase más probable, sino también la probabilidad asociada a cada una.

## 2.4. Penalización y regresión regularizada (GLMNet)

En problemas con muchas variables predictoras o cuando algunas de ellas están muy correlacionadas entre sí, los modelos estadísticos clásicos, como la regresión logística multinomial, pueden volverse inestables o tender al sobreajuste. Para evitar esto, una estrategia habitual consiste en aplicar **regresión regularizada**, que introduce una penalización en la función de pérdida del modelo con el objetivo de controlar el tamaño de los coeficientes.

En este trabajo se ha utilizado la librería `glmnet`, que permite ajustar modelos de *regresión logística multinomial regularizada*. Con este modelo se intenta predecir directamente el resultado del partido (1, X o 2), y para ello se minimiza una función de coste que combina la pérdida del modelo con una penalización:

$$\mathcal{L}(\boldsymbol{\beta}) = -\frac{1}{n} \sum_{i=1}^n \sum_{k=1}^K y_{ik} \log \hat{p}_{ik} + \lambda \cdot \mathcal{P}(\boldsymbol{\beta})$$

donde:

- $\lambda$  es el parámetro de regularización, que controla cuánto se penalizan los coeficientes grandes.
- La *Pérdida* está basada en la clasificación multiclase (log-loss).
- La *Penalización* puede adoptar diferentes formas según el tipo de regularización aplicada.

En concreto, `glmnet` permite trabajar con dos formas principales de penalización [9]:

- **Lasso** ( $\ell_1$ ): fuerza algunos coeficientes a ser exactamente cero, lo que equivale a hacer una selección automática de variables.

$$\mathcal{P}(\beta) = \|\beta\|_1 = \sum_j |\beta_j|$$

- **Ridge** ( $\ell_2$ ): reduce el tamaño de todos los coeficientes, sin llegar a anularlos.

$$\mathcal{P}(\beta) = \|\beta\|_2^2 = \sum_j \beta_j^2$$

Entre ambos extremos se encuentra el modelo conocido como **Elastic Net**, que combina las dos penalizaciones anteriores. El *balance* entre Lasso y Ridge se controla con el parámetro  $\alpha$ :

- $\alpha = 1 \rightarrow$  Lasso puro.
- $\alpha = 0 \rightarrow$  Ridge puro.
- $0 < \alpha < 1 \rightarrow$  Elastic Net.

En este TFG se han probado distintos valores de  $\alpha$  para evaluar qué tipo de penalización funciona mejor a la hora de predecir el resultado del partido. En cada caso, el valor óptimo de  $\lambda$  se ha determinado mediante validación cruzada.

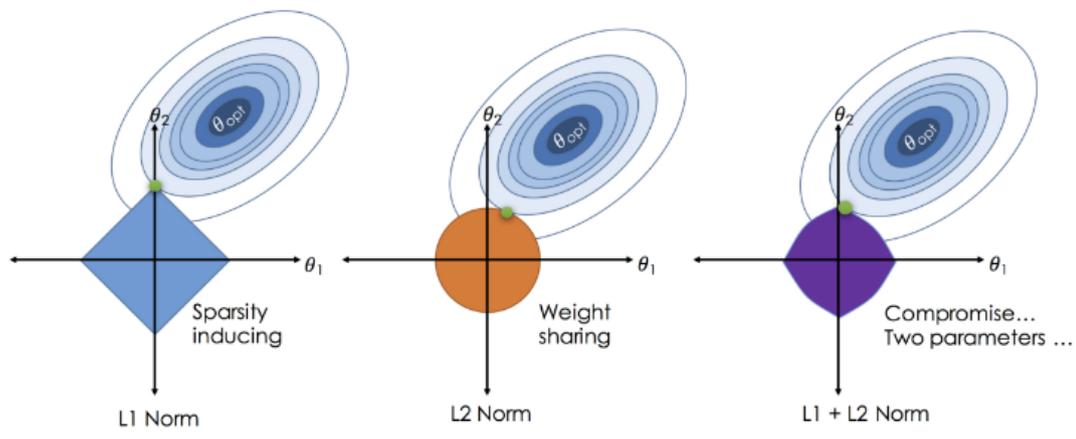


Figura 2.1: Comparación visual entre regresión Lasso, Ridge y Elastic Net. Fuente: *Zou y Hastie (2005)*, adaptado desde *Medium*.

La Figura 2.1 ilustra de forma intuitiva cómo actúan los tres tipos de regularización. En Lasso, la penalización tiene forma de rombo, lo que induce *sparsity*, es decir, fuerza a muchos coeficientes a ser exactamente cero. Ridge, en cambio, penaliza con una forma circular que tiende a “encoger” todos los coeficientes pero sin anularlos. Elastic Net combina ambas ideas, dando lugar a una forma intermedia. La intersección entre las curvas de nivel del error y la región penalizada determina la solución final del modelo.

Esta visualización ayuda a entender por qué Elastic Net puede ser una buena alternativa: permite reducir el sobreajuste como Ridge, pero también eliminar variables irrelevantes como Lasso.

En resumen, la regresión regularizada permite construir modelos más estables, más interpretables y con mejor capacidad de generalización, algo fundamental cuando se trabaja con predicciones deportivas donde las variables pueden estar muy correlacionadas o ser redundantes.

## 2.5. XGBoost

Uno de los algoritmos más potentes y populares en los últimos años para tareas de predicción es **XGBoost** (Extreme Gradient Boosting) [10]. Se trata de una implementación optimizada de la técnica de *boosting*, que ha demostrado excelentes resultados en múltiples competiciones y contextos reales, tanto en problemas de clasificación como de regresión.

El objetivo del *boosting* es construir un modelo fuerte a partir de la combinación de múltiples modelos débiles, que en este caso suelen ser árboles de decisión poco profundos (denominados *stumps*). La lógica detrás de este enfoque es iterativa: en cada paso, el algoritmo entrena un nuevo árbol que trata de corregir los errores cometidos por la combinación de árboles anteriores. En lugar de predecir directamente la variable respuesta, lo que aprende cada nuevo árbol es a modelar los *residuos* —es decir, las diferencias entre las predicciones acumuladas y los valores reales.

Matemáticamente, la predicción en la iteración  $t$  se actualiza como:

$$\hat{y}_i^{(t)} = \hat{y}_i^{(t-1)} + f_t(\mathbf{x}_i)$$

donde  $f_t$  es el nuevo árbol aprendido en la iteración  $t$ , y  $\hat{y}_i^{(t)}$  representa la predicción final acumulada para la observación  $i$ . Cada árbol aporta una pequeña corrección al error previo, y esta suma controlada permite construir modelos muy precisos sin caer fácilmente en el sobreajuste.

XGBoost refina esta idea con una serie de innovaciones clave que lo diferencian de otros algoritmos de *boosting* más tradicionales:

- **Regularización explícita:** introduce términos de penalización ( $\ell_1$  y  $\ell_2$ ) en la función objetivo, lo que permite controlar la complejidad del modelo y evitar el sobreajuste, algo especialmente importante cuando se trabaja con muchos predictores.
- **Optimización mediante derivadas de segundo orden:** en lugar de utilizar solo el gradiente (como en el *gradient boosting* clásico), XGBoost incorpora también la hessiana (segunda derivada), lo que mejora notablemente la estabilidad y la velocidad del proceso de aprendizaje.
- **Soporte para valores perdidos y sparsidad:** el algoritmo maneja automáticamente valores ausentes

o estructuras dispersas, lo que lo hace especialmente útil cuando se trabaja con conjuntos de datos reales e imperfectos.

- **Flexibilidad en la configuración:** permite ajustar numerosos hiperparámetros, como la profundidad máxima de los árboles (`max_depth`), el número de árboles a construir (`nrounds`), la tasa de aprendizaje ( $\eta$ ), la proporción de observaciones utilizadas por árbol (`subsample`), o el número de variables consideradas en cada división (`colsample_bytree`).

En el contexto de este TFG, XGBoost se ha utilizado con dos objetivos principales: por un lado, para predecir el número de goles marcados por cada equipo (empleando una función de pérdida cuadrática en tareas de regresión), y por otro, para predecir directamente el resultado del partido en términos de victoria local, empate o victoria visitante (tratando el problema como una clasificación multiclase). Esta doble aproximación ha permitido explorar tanto estrategias indirectas (inferir el signo del partido a partir de goles previstos) como enfoques directos de clasificación, con resultados notablemente consistentes.



# Capítulo 3

## Conjunto de datos

Los datos utilizados en este trabajo provienen del Trabajo de Fin de Máster en Ingeniería Informática desarrollado por José María Lozano Olmedo [11], en el que se abordaba la aplicación de técnicas de inteligencia artificial al análisis de rotación en equipos de fútbol. En este trabajo los datos se obtuvieron de la página web Resultados de Fútbol [12]. Este conjunto de datos fue la base inicial tanto para el Trabajo de Fin de Grado en Ingeniería Informática que estoy desarrollando en paralelo, como para el presente Trabajo de Fin de Grado en Estadística.

A partir de los datos brutos obtenidos del TFM mencionado, se realizó un proceso de limpieza y transformación que se explica a continuación:

### 3.1. Tratamiento de los datos

En primer lugar, se excluyeron los partidos correspondientes a las primeras jornadas de cada temporada, conservando solo los encuentros disputados a partir de la jornada 10. Esta decisión se tomó porque los indicadores utilizados en los modelos se calculan a partir del rendimiento histórico reciente, y en las primeras jornadas no hay suficiente información acumulada para obtener métricas fiables.

También se aplicaron filtros para eliminar equipos con presupuestos demasiado extremos (muy altos o muy bajos), con el objetivo de evitar distorsiones que pudieran afectar a los modelos. Esta información se obtuvo a partir del valor de mercado de las plantillas en cada temporada, según la página web Transfermarkt [13], y se eliminaron del análisis todos los partidos en los que participaron esos equipos.

El filtrado se realizó por separado para cada temporada, identificando sistemáticamente a los equipos más desequilibrantes. En concreto, se eliminaron siempre los dos equipos con mayor presupuesto (el Real Madrid y el Fútbol Club Barcelona), y en algunas temporadas también el Atlético de Madrid. En el extremo opuesto, se suprimieron entre uno y tres equipos con los

presupuestos más bajos, dependiendo del grado de asimetría en la distribución del valor de mercado en esa temporada.

Después, los datos se organizaron en función del número máximo de sustituciones permitidas en cada temporada (3 o 5 cambios), agrupando los partidos en bloques homogéneos para facilitar comparaciones. Esto permite analizar si los cambios en el reglamento han tenido algún efecto en el comportamiento de los equipo.

Finalmente, se eliminaron columnas no informativas (como identificadores) y se prepararon distintos conjuntos de datos que permitieran entrenar y evaluar modelos según el objetivo de cada análisis.

## 3.2. Estructura del conjunto de datos

El conjunto de datos utilizado en este trabajo está organizado siguiendo un modelo de *indicadores prepartido*. Esto significa que, para cada partido, se genera una serie de variables que resumen el estado de forma de los equipos justo antes del encuentro. Estas variables permiten capturar el contexto competitivo inmediato, incorporando tanto la dinámica reciente como aspectos estratégicos de la gestión de la plantilla. Su objetivo es proporcionar una base informativa sólida para alimentar los modelos estadísticos y facilitar la predicción de resultados.

En total, se han considerado **1139 partidos** disputados en la **Primera División española** durante el periodo comprendido entre las temporadas **2019 y 2024**.

Cada fila del conjunto de datos representa un partido concreto, con información tanto del equipo local como del visitante. Se incluyen variables contextuales (como la jornada, la temporada o el número de cambios permitidos según el reglamento vigente) y un amplio conjunto de variables explicativas generadas a partir del comportamiento reciente de los equipos (el listado completo con las 176 variables puede verse en el Anexo .1).

Los indicadores utilizados se construyen a partir de un historial de partidos anteriores, y pueden calcularse de dos formas diferenciadas:

- **En general:** considerando todos los partidos previos del equipo, independientemente de si fueron jugados como local o visitante.
- **En contexto:** filtrando solo aquellos encuentros disputados en la misma condición que el partido analizado (es decir, solo partidos como local si el equipo juega en casa, o como visitante si juega fuera).

Esta distinción es relevante, ya que muchos equipos presentan comportamientos significativamente distintos según el lugar en el que juegan, lo cual puede afectar de manera notable a las predicciones.

Las variables explicativas incluidas en el conjunto de datos se agrupan en las siguientes categorías principales:

- **Rendimiento previo:** porcentaje de victorias, empates y derrotas; puntos acumulados; goles a favor y en contra; media de goles por partido.
- **Comportamiento ofensivo y defensivo:** número medio de tiros por partido; porcentaje de posesión del balón; saques de esquina a favor y en contra; frecuencia de porterías a cero.
- **Aspectos disciplinarios:** tarjetas amarillas y rojas recibidas; número medio de faltas cometidas.
- **Rotación y sustituciones:** número de cambios realizados en la alineación titular respecto al partido anterior; proporción de sustituciones efectuadas durante el partido; minutos promedio de los cambios; desglose por tipo de cambio (táctico, posicional, por línea).

Además de estas variables explicativas, cada registro incorpora también las **variables respuesta**, que constituyen el objetivo de los modelos de predicción. Concretamente, se incluyen:

- El **resultado del partido** codificado en tres categorías: victoria local (1), empate (X) y victoria visitante (2).
- El **número de goles** marcados por el equipo local y el visitante.

Esta estructura permite abordar tanto problemas de clasificación (predicción directa del signo del resultado) como problemas de regresión (estimación de goles), ofreciendo así una visión completa y versátil del rendimiento de los equipos a partir de sus indicadores previos.



# Capítulo 4

## Análisis Descriptivo

Antes de aplicar modelos estadísticos, es fundamental comprender cómo se comportan las variables principales del estudio. En este capítulo se presenta un análisis exploratorio de los datos, tanto de las variables respuesta (goles y resultado del partido), como de algunas variables relacionadas con las sustituciones, que tienen un papel clave en el enfoque de este trabajo.

El objetivo de este análisis es observar patrones generales, detectar posibles diferencias entre contextos (como las distintas normativas de cambios) y entender mejor cómo se distribuyen y relacionan algunas de las variables que más adelante se utilizan en los modelos predictivos.

### 4.1. Análisis de las variables respuesta

#### 4.1.1. Distribución de goles por equipo (local y visitante)

El primer paso del análisis descriptivo consiste en estudiar cómo se distribuyen los goles anotados por los equipos locales y visitantes. En la Figura 4.1 y Figura 4.2 se muestran los gráficos de barras de ambos casos. Como se puede observar, la mayoría de partidos terminan con 1 o 2 goles anotados por cada equipo, aunque existen casos más extremos, con marcadores de hasta 6 goles.

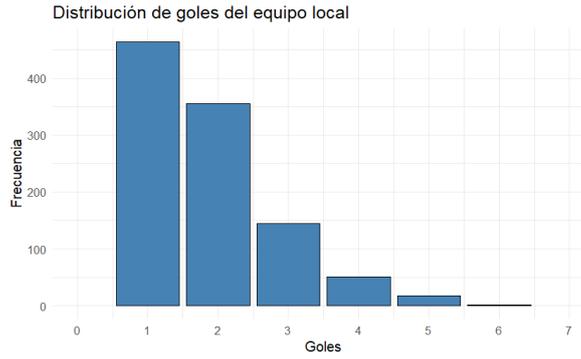


Figura 4.1: Distribución de goles del equipo local

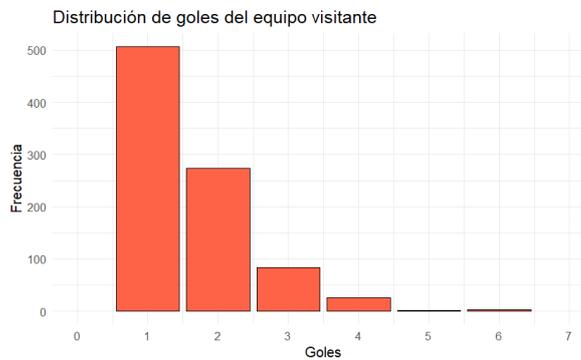


Figura 4.2: Distribución de goles del equipo visitante

A nivel descriptivo, en el conjunto de partidos estudiados, los equipos locales presentan una media de **1,42** goles por partido, mientras que los visitantes marcan de media **1,07** goles. La dispersión también es algo mayor en el caso de los locales (desviación estándar de 1,15 frente a 1,02), y el coeficiente de asimetría de ambas distribuciones es positiva, lo que indica una mayor frecuencia de resultados bajos y algunos casos con marcadores altos poco frecuentes. La curtosis es algo más elevada en los visitantes (4,25), lo que indica una mayor concentración en torno a la moda y colas más pesadas.

Se puede observar que los equipos locales tienden a marcar más goles que los visitantes en el boxplot de la Figura 4.3, donde se ve que la mediana de goles para ambos es 1, pero los equipos locales tienen un rango intercuartílico más elevado y más valores extremos.

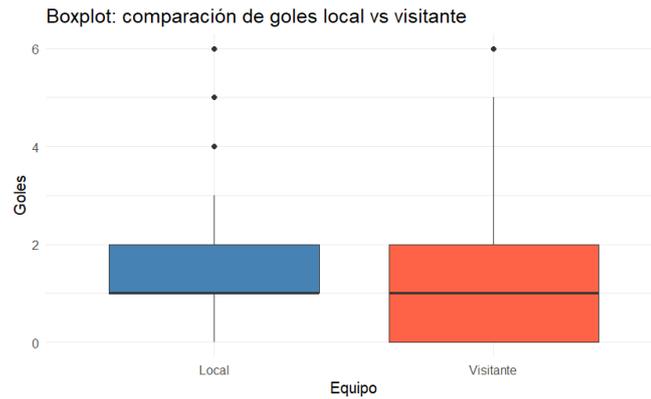


Figura 4.3: Boxplot: comparación de goles local vs visitante

En cuanto a la evolución temporal, la Figura 4.4 muestra la media de goles anotados por temporada. Se observa que los goles de los equipos locales se han mantenido relativamente estables, con una ligera disminución en las últimas temporadas. En cambio, los goles visitantes presentan más oscilación, con un pico en la temporada 2021 y mínimos en 2020 y 2023.

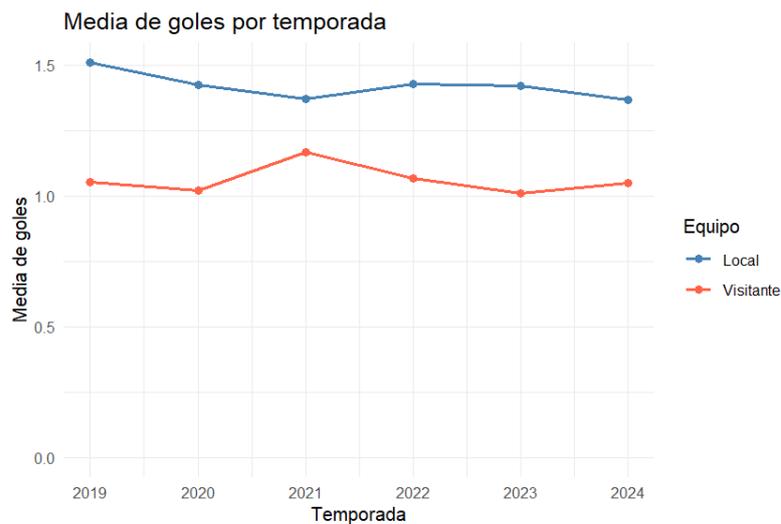


Figura 4.4: Media de goles por temporada

Finalmente, el análisis por jornada también ofrece información interesante. Tal como se muestra en las Figuras 4.5 y 4.6, la media de goles de los equipos locales tiende a disminuir ligeramente conforme avanza la temporada, mientras que los goles visitantes muestran una tendencia creciente, aunque en ambos casos la variación es mínima.

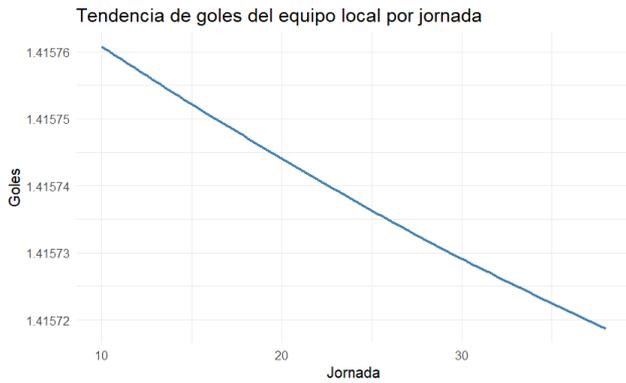


Figura 4.5: Tendencia de goles del equipo local por jornada

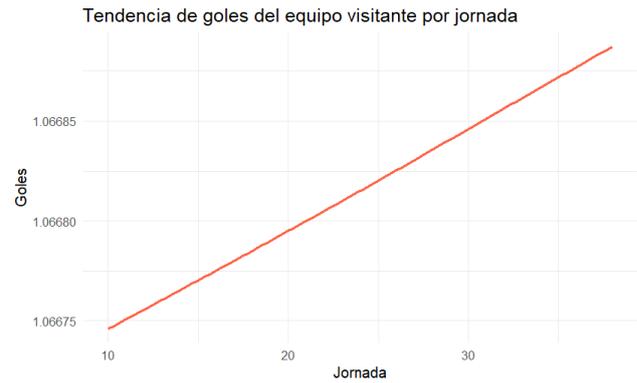


Figura 4.6: Tendencia de goles del equipo visitante por jornada

### 4.1.2. Distribución del resultado del partido (1, X, 2)

Además de los goles individuales de cada equipo, una de las variables más relevantes para el estudio es el resultado global del partido, que puede ser victoria local (1), empate (X) o victoria visitante (2). La Figura 4.7 muestra la distribución de estas tres categorías en el conjunto de partidos analizado.



Figura 4.7: Distribución de resultado del partido

Resultado del partido	Frecuencia	Proporción (%)
Victoria local (1)	601	44.6 %
Empate (X)	397	29.5 %
Victoria visitante (2)	349	25.9 %

Tabla 4.1: Distribución del resultado del partido (1, X, 2)

Tal como se muestra en la Tabla 4.1, la victoria local fue el resultado más frecuente, seguida de

los empates y, finalmente, las victorias visitantes.

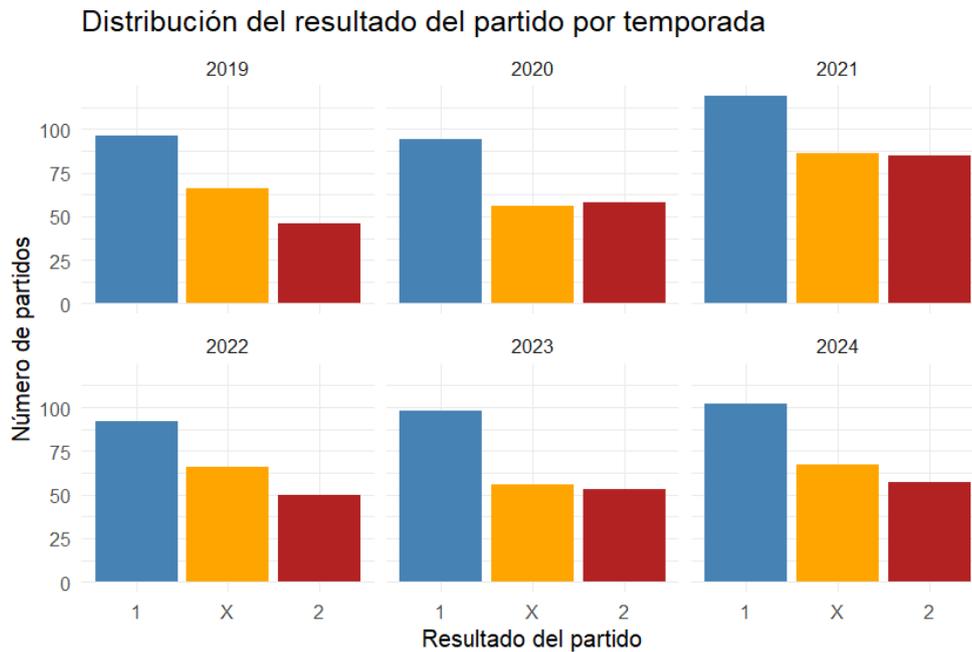


Figura 4.8: Distribución del resultado del partido por temporada

Como se observa en la Figura 4.8, existe una clara tendencia en todas las temporadas a que la victoria del equipo local (1) sea el resultado más frecuente. Este patrón es especialmente acusado en las temporadas 2021 y 2024. Por el contrario, las victorias visitantes (2) son sistemáticamente el resultado menos frecuente, aunque con variaciones interanuales. El empate (X) mantiene una frecuencia intermedia y relativamente estable entre temporadas.

Más allá de la clasificación general del partido en términos de 1, X o 2, es interesante observar qué combinaciones concretas de goles son las más habituales y cuáles apenas se producen. La Figura 4.9 muestra un mapa de calor con la frecuencia de aparición de cada pareja de goles anotados por el equipo local y el visitante. Como se aprecia, la mayor concentración de partidos se encuentra en marcadores con pocos goles, y especialmente cuando el equipo local anota uno o dos tantos.

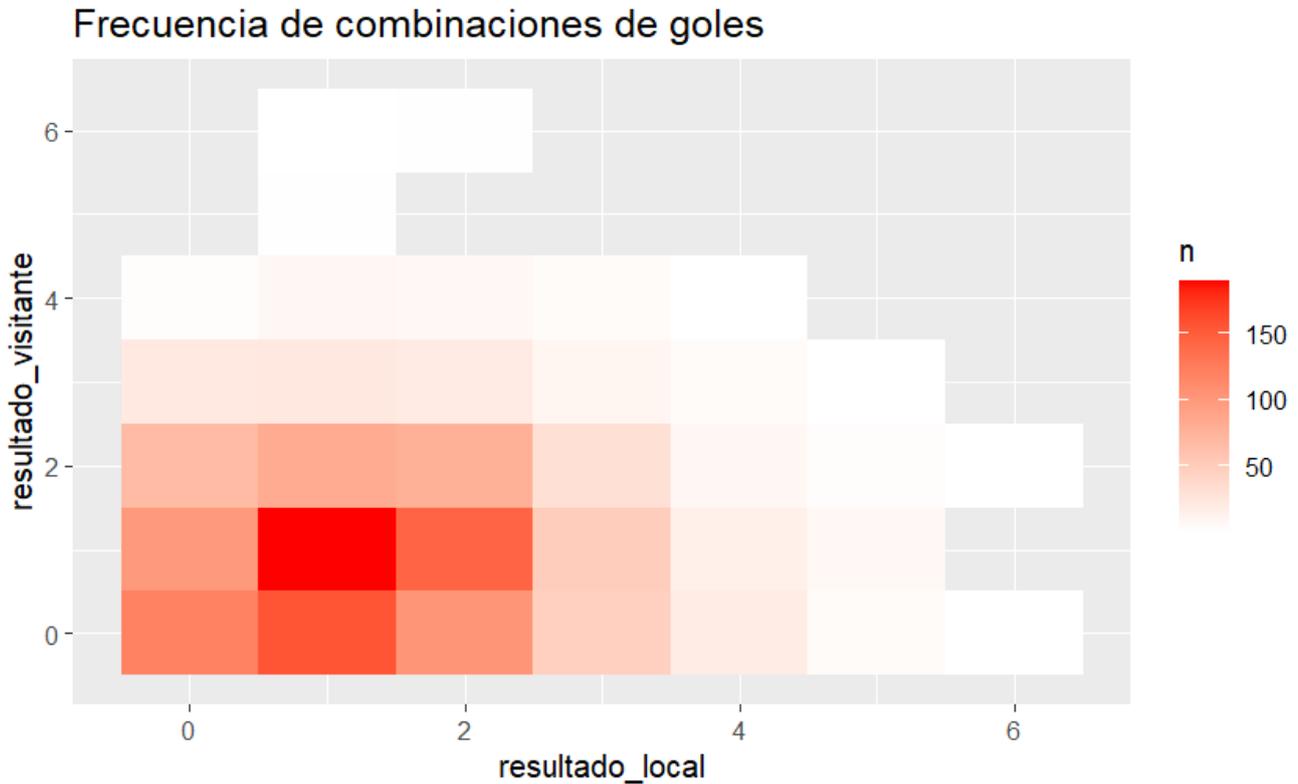


Figura 4.9: Frecuencia de combinaciones de goles local vs visitante

Para completar esta visualización, en la Tabla 4.2 se recogen las combinaciones de goles más frecuentes. El empate 1–1 y las victorias locales por 1–0 o 2–1 ocupan los primeros puestos. También destaca la frecuencia de empates sin goles, algo habitual en partidos muy igualados.

Goles Local	Goles Visitante	Frecuencia
1	1	189
1	0	156
2	1	144
0	0	120
2	0	103

Tabla 4.2: Combinaciones de goles más frecuentes

En el extremo opuesto, la Tabla 4.3 muestra algunas de las combinaciones más inusuales. Se trata de resultados que apenas se han registrado una única vez, como encuentros con 6 goles de un solo equipo o marcadores simétricos poco comunes (como el 4–4). Aunque estas situaciones son excepcionales, es importante identificarlas porque pueden tener un efecto desproporcionado en algunos modelos si no se controlan adecuadamente.

Goles Local	Goles Visitante	Frecuencia
1	6	1
4	4	1
5	3	1
6	0	1
6	2	1

Tabla 4.3: Combinaciones de goles más inusuales

## 4.2. Análisis descriptivo de las sustituciones

### 4.2.1. Proporción de cambios por equipo y reglamento

Uno de los primeros aspectos a explorar en relación con las sustituciones es la proporción de cambios realizados por los equipos en función del reglamento vigente. Antes de la pandemia de COVID-19, el reglamento establecía un máximo de tres sustituciones por equipo. Sin embargo, a raíz de esta situación excepcional, se introdujo una modificación que permitió ampliar ese límite a cinco cambios, medida que se ha mantenido hasta la actualidad [14].

En la Figura 4.10, se muestra un boxplot comparativo de la proporción de cambios realizados por los equipos locales y visitantes bajo ambos reglamentos. Como es esperable, los valores se agrupan en torno a 3 cuando la normativa solo permitía 3 cambios, y aumentan claramente en el contexto de 5 sustituciones, donde la mayoría de los equipos se acercan al máximo permitido.

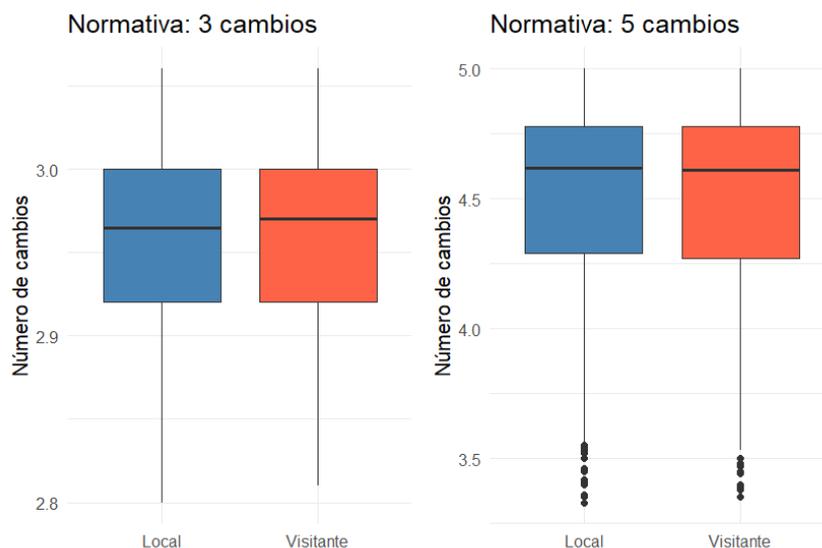


Figura 4.10: Comparación de proporciones de cambios por equipo y reglamento

La Tabla 4.4 resume los principales estadísticos por tipo de equipo y reglamento. En ambos contextos, no se aprecian grandes diferencias entre equipos locales y visitantes. Con 3 cambios, la media de sustituciones ronda los 2,95; mientras que con 5 cambios, se eleva hasta 4,5.

Reglamento	Equipo	Media	Mediana	Q1	Q3	SD
3 cambios	Local	2.95	2.96	2.92	3.00	0.049
3 cambios	Visitante	2.96	2.97	2.92	3.00	0.049
5 cambios	Local	4.50	4.62	4.29	4.78	0.373
5 cambios	Visitante	4.50	4.61	4.27	4.78	0.371

Tabla 4.4: Estadísticas descriptivas de proporciones de cambios por equipo y reglamento

También se observa que, bajo la normativa de 5 cambios, la dispersión de las proporciones es mayor, lo cual tiene sentido dado que los equipos cuentan con más margen estratégico para decidir cuándo y cuántos cambios realizar. La asimetría negativa y curtosis algo elevadas en ambos contextos sugieren una distribución ligeramente sesgada hacia valores cercanos al límite permitido.

Este análisis confirma que los equipos tienden a agotar las sustituciones disponibles en función del reglamento, sin grandes diferencias entre locales y visitantes.

#### 4.2.2. Cambios en la alineación: defensas, medios y delanteros

Además de las sustituciones realizadas durante el partido, resulta muy interesante analizar cómo varían las alineaciones titulares de un encuentro a otro. Este tipo de cambios, que reflejan decisiones estratégicas del entrenador, pueden estar influenciados por el reglamento, el rendimiento reciente o la rotación de la plantilla.

En este apartado se estudia la proporción total de cambios en la alineación de cada equipo (local y visitante). La Figura 4.11 muestra cómo se distribuyen estas proporciones en función del reglamento vigente.

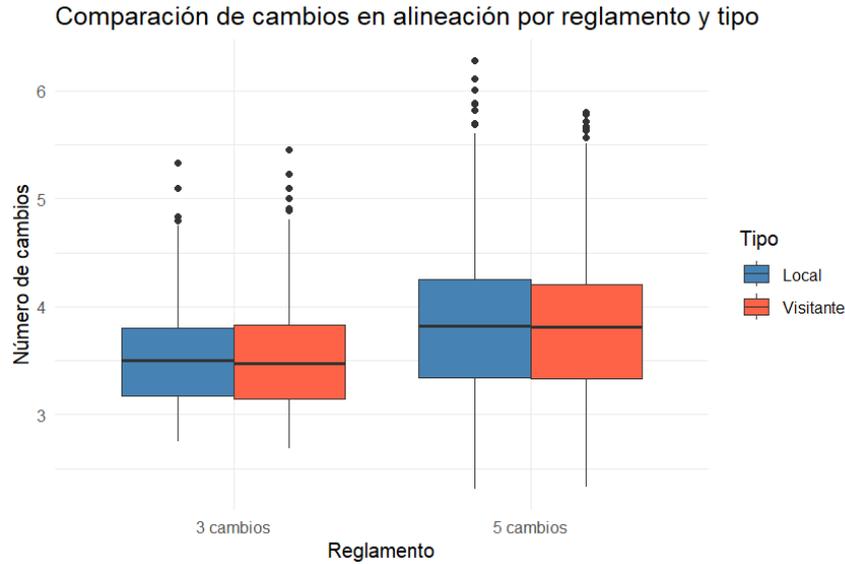


Figura 4.11: Comparación de cambios en alineación por reglamento y tipo de equipo

Como puede observarse, los equipos tienden a modificar algo más sus alineaciones bajo la normativa de 5 cambios. No obstante, las diferencias no son tan marcadas como en el caso de las sustituciones durante el propio partido. En ambos reglamentos, tanto equipos locales como visitantes presentan medias similares, ligeramente por encima de 3,5 cambios de una jornada a la siguiente.

La Tabla 4.5 recoge los principales estadísticos descriptivos. Se aprecia una ligera mayor dispersión en el caso del reglamento de 5 cambios, lo que puede indicar una mayor flexibilidad táctica.

Reglamento	Equipo	Media	Mediana	Q1	Q3	SD
3 cambios	Local	3.54	3.50	3.17	3.80	0.515
3 cambios	Visitante	3.55	3.47	3.14	3.83	0.546
5 cambios	Local	3.82	3.82	3.34	4.25	0.663
5 cambios	Visitante	3.81	3.81	3.33	4.21	0.653

Tabla 4.5: Estadísticas descriptivas de cambios en la alineación por reglamento y tipo de equipo

En general, los datos sugieren que el número de sustituciones permitido durante el partido también puede influir indirectamente en las decisiones sobre la alineación inicial, aunque este efecto es más suave. La asimetría positiva en las distribuciones indica que, aunque la mayoría de los equipos hacen entre 3 y 4 cambios, existen algunos que varían sus alineaciones de forma más agresiva entre jornadas.

### 4.2.3. Minuto promedio de los cambios

Otra dimensión interesante del análisis de sustituciones es el momento del partido en que se producen. El minuto en que se realiza un cambio puede tener implicaciones estratégicas importantes: cambios tempranos pueden responder a lesiones o ajustes tácticos urgentes, mientras que cambios tardíos suelen reflejar una gestión del resultado o de la carga física.

La Figura 4.12 muestra la comparación del minuto promedio de las sustituciones realizadas por equipos locales y visitantes bajo cada uno de los reglamentos (3 y 5 cambios).

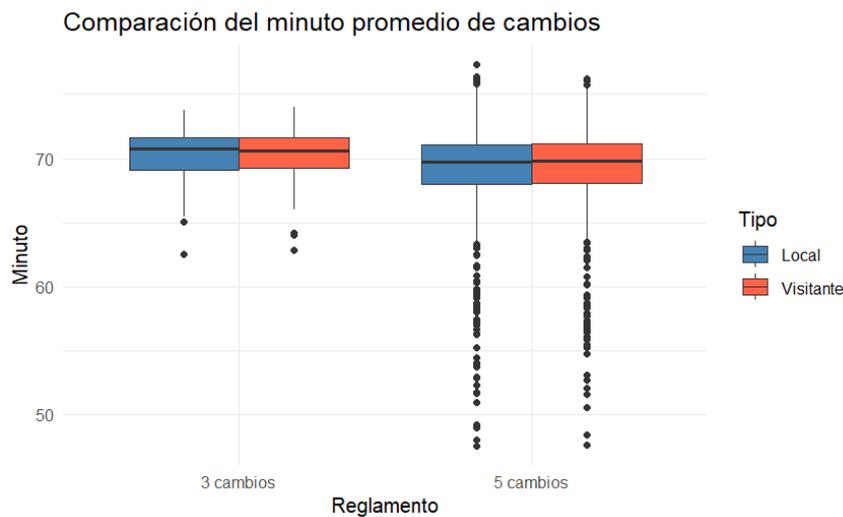


Figura 4.12: Comparación del minuto promedio de cambios por reglamento y tipo de equipo

En la Tabla 4.6 se recogen los principales estadísticos. Bajo la normativa de 3 cambios, los equipos realizaban las sustituciones en torno al minuto 70, con muy poca dispersión. En cambio, cuando se permite realizar 5 cambios, el momento medio en que se producen las sustituciones se adelanta ligeramente, y se observa una mayor variabilidad.

Reglamento	Equipo	Media	Mediana	Q1	Q3	SD	Asimetría	Curtosis
3 cambios	Local	70.3	70.8	69.1	71.6	1.96	-0.802	3.61
3 cambios	Visitante	70.3	70.6	69.3	71.7	1.97	-0.783	3.67
5 cambios	Local	69.1	69.7	68.0	71.1	3.69	-2.29	11.2
5 cambios	Visitante	69.2	69.8	68.1	71.2	3.49	-2.12	10.7

Tabla 4.6: Estadísticas del minuto promedio de cambios por equipo y reglamento

Los valores de asimetría y curtosis sugieren que, en el reglamento de 5 cambios, algunos equipos realizan sustituciones en momentos bastante alejados de la mediana, lo que justifica la mayor dispersión observada. La tendencia general, sin embargo, indica que tanto locales como

visitantes tienden a hacer cambios entre los minutos 68 y 72, independientemente del reglamento, aunque con mayor flexibilidad táctica cuando se permiten más sustituciones.

### **4.3. Comparación entre normativas de sustituciones**

El cambio reglamentario que permitió ampliar el número máximo de sustituciones por equipo de 3 a 5 ha tenido un impacto claro en el comportamiento de los equipos, tanto durante el partido como entre jornadas. Los análisis realizados en este capítulo permiten identificar varios patrones diferenciados en función del reglamento vigente.

En primer lugar, la proporción de sustituciones realizadas dentro del partido se ajusta prácticamente al límite permitido en ambos casos. Bajo la normativa de 3 cambios, tanto los equipos locales como los visitantes utilizan una media cercana a 3 sustituciones por encuentro. Con 5 cambios, esta media aumenta hasta los 4,5 cambios por equipo, con una ligera mayor variabilidad, lo que sugiere un mayor margen para la toma de decisiones tácticas o físicas durante el encuentro.

En segundo lugar, el número de cambios en la alineación inicial de un partido a otro también se ve afectado, aunque en menor medida. Los equipos modifican sus onces iniciales ligeramente más en el contexto de 5 sustituciones, con un aumento medio de 0.3 cambios. Esto indica que disponer de más cambios durante el encuentro no solo afecta a la gestión del partido en sí, sino que también se asocia a una mayor rotación entre jornadas. Es probable que esta flexibilidad adicional anime a los entrenadores a introducir más variaciones en la alineación inicial, permitiendo una planificación más dinámica del equipo.

Por último, el análisis del minuto promedio en el que se realizan las sustituciones refleja que, aunque los cambios siguen produciéndose mayoritariamente entre los minutos 68 y 72, existe una mayor dispersión temporal bajo el reglamento de 5 cambios. Esto sugiere que algunos equipos aprovechan esta mayor flexibilidad para intervenir en momentos más variados del encuentro, ya sea para proteger el resultado, introducir piernas frescas o responder a situaciones imprevistas.

En conjunto, estos resultados muestran que el reglamento de sustituciones no solo afecta al número de cambios realizados, sino también al modo en que los equipos gestionan sus plantillas y ajustan sus decisiones tácticas. Esta diferencia debe tenerse en cuenta en los modelos predictivos, especialmente si se entrenan sobre datos que combinan ambos tipos de normativa.



# Capítulo 5

## Modelos predictivos

Tras el análisis descriptivo del conjunto de datos, el siguiente paso es aplicar distintos modelos estadísticos con el objetivo de predecir el resultado del partido (1, X o 2) o el número de goles anotados por cada equipo. En este capítulo se describen los modelos utilizados, se explica cómo se han implementado y se justifica su elección en función del tipo de problema.

Se utilizan modelos de regresión para variables numéricas (como el número de goles) y modelos de clasificación multiclase para predecir directamente el resultado del encuentro. Todos los modelos se han implementado en R, utilizando paquetes específicos y estructuras adaptadas al problema.

### 5.1. Regresión de Poisson

Uno de los modelos más clásicos para abordar problemas de conteo es la regresión de Poisson. En este trabajo se ha utilizado para predecir de forma separada el número de goles que marcarán el equipo local y el equipo visitante, a partir de un amplio conjunto de variables explicativas derivadas de los indicadores prepartido.

#### 5.1.1. Preparación del conjunto de datos

Para el ajuste de los modelos de regresión de Poisson se empleó el conjunto de datos completo correspondiente a la Primera División española, que incluye múltiples temporadas. Como paso inicial, se eliminó la temporada 2020, debido a que en ella se produjo una transición normativa con respecto al número de sustituciones permitidas por equipo debido a la pandemia del COVID-19, lo que generaba heterogeneidad difícil de modelar adecuadamente.

Posteriormente, se incorporó una nueva variable dicotómica denominada `cinco_cambios`, cuyo objetivo es identificar si en el partido estaba vigente la normativa de cinco sustituciones

(valor 1) o la anterior de tres sustituciones (valor 0). De acuerdo con la evolución reglamentaria de la competición, los partidos correspondientes a la temporada 2019 fueron codificados con un 0, y aquellos a partir de la temporada 2021 con un 1.

Una vez creada esta variable, se eliminaron las columnas no predictivas, como los identificadores de partido y equipo, la variable con el resultado categórico (`resultado_partido`), la temporada y el nombre de la liga, ya que solo se trabaja con una competición. Además, se excluyeron las variables `resultado_local` y `resultado_visitante` en el modelo de clasificación multinomial, aunque en el modelo de Poisson se mantuvieron como variables respuesta.

Finalmente, los datos fueron divididos en dos subconjuntos: entrenamiento (80 %) y test (20 %), manteniendo la proporción de clases mediante partición estratificada. Esta división se realizó por separado para las variables objetivo: goles del equipo local y goles del equipo visitante, permitiendo así ajustar modelos independientes para ambas situaciones.

### 5.1.2. Modelo de regresión de Poisson para goles del equipo local

Con el objetivo de modelizar el número de goles anotados por el equipo local en un partido, se ajustó un modelo de regresión de Poisson utilizando como variable respuesta `resultado_local`. Inicialmente se emplearon todas las variables disponibles como predictores (171), incluyendo indicadores de rendimiento, posesión, goles esperados, disciplina, sustituciones y características del rival.

El modelo completo alcanzó una deviance de 800.51, con un cociente deviance/grados de libertad residuales de 1.0818, lo que indica una ligera sobredispersión, pero aceptable para mantener el uso del modelo de Poisson estándar. La evaluación del modelo sobre el conjunto de test mostró un *MAE* (error absoluto medio) de 0.9797 y una *RMSE* (raíz del error cuadrático medio) de 1.2495. La salida completa del modelo, incluyendo los coeficientes estimados, errores estándar y significancias individuales, se incluye en el Anexo .2.

En particular, se examinó el efecto de la variable `cinco_cambios`, obteniéndose un coeficiente estimado de 0.205 (p-valor = 0.6185), lo cual indica que su influencia no fue estadísticamente significativa dentro del modelo completo.

Posteriormente, se construyó un modelo reducido incluyendo únicamente aquellas variables que resultaron significativas al nivel  $\alpha = 0.05$ . Las variables seleccionadas fueron:

- porcentaje (de goles) local más 1,5 en sitio
- porcentaje (de goles) local más 3,5 en sitio
- porcentaje (de goles) local más 4,5 en sitio
- proporción visitante cambios en general

- proporción local posesión en general
- media visitante cambios minutos en general
- proporción local cambios alineación delantero en general
- resultado\_visitante

La ecuación del modelo reducido es:

$$\begin{aligned} \log \left( E(\widehat{\text{resultado\_local}}) \right) = & -1.113 + 0.004 \cdot \text{'porcentaje local más 1,5 en sitio'} + 0.003 \cdot \text{'porcentaje local más 3,5 en sitio'} \\ & + 0.006 \cdot \text{'porcentaje local más 4,5 en sitio'} - 0.012 \cdot \text{'proporcion visitante cambios en general'} \\ & + 0.028 \cdot \text{'proporcion local posesión en general'} + 0.010 \cdot \text{'media visitante cambios minutos en general'} \\ & - 0.181 \cdot \text{'proporcion local cambios alineación delantero en general'} + 0.042 \cdot \text{resultado\_visitante} \end{aligned} \quad (5.1)$$

El modelo reducido obtuvo una deviance de 984.02, con un indicador de sobredispersión de 1.0897. La evaluación sobre el conjunto de test mostró una mejora clara respecto al modelo completo, con *MAE* de 0.9206 y *RMSE* de 1.1587.

Finalmente, se realizó una comparación formal entre ambos modelos mediante un test de razón de verosimilitud (Chi-cuadrado), obteniéndose un valor  $G^2 = 183.5$  con 163 grados de libertad y un p-valor de 0.1297. Esto indica que la reducción del modelo no implica una pérdida significativa de capacidad predictiva, por lo que el modelo reducido puede considerarse preferible al completo por su mayor interpretabilidad y menor complejidad (véase Tabla 5.1). El resumen del ajuste del modelo reducido se encuentra disponible en el Anexo 3.

Resid. Df	Resid. Dev	Df	Deviance	Pr(¿Chi)
903	984.02			
740	800.51	163	183.50	0.1297

Tabla 5.1: Comparación entre modelo completo y modelo reducido mediante test de razón de verosimilitud (modelo de Poisson para goles locales)

### 5.1.3. Modelo de regresión de Poisson para goles del equipo visitante

Se procedió del mismo modo que en el caso de los goles locales, ajustando un modelo de regresión de Poisson con la variable respuesta `resultado_visitante`. El modelo completo incluyó todas las variables predictoras disponibles (171), obteniendo una deviance de 785.99 y un cociente deviance/grados de libertad residuales de 1.0621, lo cual indica un ajuste adecuado sin evidencia significativa de sobredispersión.

La evaluación del modelo completo sobre el conjunto de test arrojó un *MAE* de 0.9147 y un *RMSE* de 1.1015. En cuanto a la variable `cinco_cambios`, su coeficiente estimado fue 0.144

con un p-valor de 0.7508, por lo que no se consideró significativa en este modelo. El resumen completo del ajuste se encuentra disponible en el Anexo [4](#).

A continuación, se construyó un modelo reducido que incluía únicamente aquellas variables que resultaron significativas al nivel  $\alpha = 0.05$ . Las variables seleccionadas fueron:

- proporción local puntos en general
- porcentaje local más 2,5 marcados en sitio
- porcentaje local más 0,5 encajados en general
- proporción visitante cambios amarillas en general
- media visitante cambios minutos en general
- proporción visitante cambios 76 a final en general
- resultado\_local

La ecuación del modelo reducido es:

$$\begin{aligned} \log(E(\widehat{\text{resultado\_visitante}})) = & 1.273 - 0.125 \cdot \text{'proporcion local puntos en general'} \\ & + 0.003 \cdot \text{'porcentaje local más 2,5 marcados en sitio'} + 0.001 \cdot \text{'porcentaje local más 0,5 encajados en general'} \\ & - 0.484 \cdot \text{'proporcion visitante cambios amarillas en general'} - 0.016 \cdot \text{'media visitante cambios minutos en general'} \\ & + 0.018 \cdot \text{'proporcion visitante cambios 76 a final en general'} + 0.052 \cdot \text{resultado\_local} \end{aligned} \quad (5.2)$$

Este modelo reducido obtuvo una deviance de 973.63, con un indicador de sobredispersión de 1.077. A pesar de la deviance algo mayor (esperable al reducir el número de predictores), se observó una mejora en la capacidad predictiva, con un *MAE* de 0.8326 y un *RMSE* de 1.0376 en el conjunto de test.

La comparación formal entre el modelo completo y el reducido mediante un test de razón de verosimilitud arrojó un valor  $G^2 = 187.64$  con 164 grados de libertad y un p-valor de 0.0996. Dado que el p-valor supera el umbral del 5 %, no se observa una pérdida significativa de información al reducir el modelo. En consecuencia, el modelo reducido es preferible por razones de interpretabilidad y parquedad (véase Tabla [5.2](#)). El resumen del ajuste se incluye en el Anexo [5](#).

Resid. Df	Resid. Dev	Df	Deviance	Pvalor
904	973.63			
740	785.99	164	187.64	0.0996

Tabla 5.2: Comparación entre modelo completo y modelo reducido mediante test de razón de verosimilitud (modelo de Poisson para goles visitantes)

#### 5.1.4. Inferencia del resultado del partido con modelos reducidos

A partir de las predicciones generadas por los modelos de Poisson reducidos para los goles del equipo local y visitante, se procedió a **inferir el resultado del partido** (victoria local, empate o victoria visitante) a partir del número de goles estimado.

Para ello, se redondearon las predicciones a un decimal, con el objetivo de suavizar pequeñas diferencias que no resultan significativas en el contexto futbolístico. A continuación, se aplicó una regla de decisión:

- Se predice un empate (X) cuando la diferencia absoluta entre goles predichos de locales y visitantes es inferior a 0.05.
- Se predice victoria local (1) si el equipo local supera al visitante en goles con una diferencia mayor de 0.05.
- Se predice victoria visitante (2) si el equipo visitante supera al local en goles con una diferencia mayor de 0.05.

A partir del resultado del partido predicho de esta forma, se construye la siguiente matriz de confusión, comparando el resultado predicho con el resultado real:

Predicho \ Real	1	2	X
1	91	46	52
2	8	9	10
X	7	3	1

A pesar de la simplicidad de la regla, se obtiene una precisión del **44.5 %** sobre el conjunto de test:

$$\text{Accuracy (resultado partido)} = 0.4449$$

Este resultado no es del todo insatisfactorio, pero revela una clara tendencia a favorecer la victoria del equipo local. Esto puede observarse en la distribución de predicciones dentro de la matriz de confusión, donde la categoría 1 (victoria local) es la más frecuente con diferencia. Este sesgo puede deberse tanto al comportamiento del modelo como a la propia estructura de los datos, donde la victoria local es estadísticamente más habitual. En cualquier caso, el análisis permite comprobar la coherencia general de los modelos de Poisson y su aplicabilidad como herramienta auxiliar para la clasificación del resultado final del partido.

## Resumen

Los modelos de regresión de Poisson han demostrado ser una herramienta útil y coherente para predecir el número de goles en un partido de fútbol, tanto para el equipo local como para el visitante. Este tipo de modelo, clásico en problemas de conteo, se basa en estimar la media del número de goles en función de un conjunto de variables explicativas, permitiendo interpretar cómo influyen distintos factores en el resultado numérico.

En este trabajo, se compararon modelos completos (con todas las variables disponibles) frente a modelos reducidos (con solo las variables estadísticamente significativas). Los resultados mostraron que los modelos reducidos no solo son más simples e interpretables, sino que además obtuvieron mejores métricas predictivas, como menor error medio.

Por otro lado, se analizó si la normativa de tres o cinco sustituciones por equipo (variable `cinco_cambios`) tenía un impacto relevante en la predicción de goles. En ambos modelos (local y visitante), esta variable no resultó significativa, lo que sugiere que la introducción de la nueva normativa no tuvo un efecto claro sobre el número total de goles en los partidos.

Cabe destacar que, aunque la variable `cinco_cambios` no fue significativa, sí se identificaron algunas variables relacionadas con las sustituciones que resultaron estadísticamente relevantes en los modelos reducidos. Esto indica que ciertos patrones en la gestión de los cambios por parte de los equipos pueden estar asociados al comportamiento goleador. Además, se observa una relación significativa entre los goles del equipo contrario: el número de goles del visitante es una variable relevante para predecir los goles del local, y viceversa. Este hallazgo sugiere una cierta interdependencia entre los resultados de ambos equipos durante el desarrollo del partido.

En conjunto, la regresión de Poisson ha permitido capturar de forma efectiva la dinámica goleadora de los encuentros y, aunque no está diseñada para predecir directamente el resultado del partido, sus predicciones pueden combinarse para inferirlo con una precisión razonable.

## 5.2. Regresión logística multinomial

A diferencia de los modelos de regresión de Poisson, que predicen los goles de cada equipo por separado, la regresión logística multinomial permite modelar directamente la variable categórica `resultado_partido`, abordando así el problema como una clasificación multiclase. Este enfoque evita la necesidad de transformar predicciones numéricas en un signo del resultado, ofreciendo una solución más directa y coherente con el objetivo final del análisis.

### 5.2.1. Preparación de los datos

Se ha utilizado el conjunto de datos completo, excluyendo la temporada 2020 por considerarse una temporada atípica con interrupciones y condiciones excepcionales debido a la pandemia COVID-19. Se ha añadido la variable binaria `cinco_cambios` para identificar si el partido se jugó bajo la normativa de cinco sustituciones (temporadas posteriores a 2019).

A continuación, se eliminaron columnas irrelevantes como identificadores, temporada, jornada y las variables de goles individuales (`resultado_local` y `resultado_visitante`).

### 5.2.2. Ajuste del modelo completo

La variable respuesta `resultado_partido` se transformó en un factor con niveles “1”, “X” y “2”.

El modelo de regresión logística multinomial se ajusta sobre el conjunto de entrenamiento empleando todas las variables disponibles como predictores (171). El proceso de optimización se detuvo tras 100 iteraciones, alcanzando una **deviance final de 1641.66** y un **AIC de 2321.66**, lo que proporciona una primera medida de la calidad del ajuste teniendo en cuenta el equilibrio entre bondad de ajuste y complejidad del modelo.

La salida detallada del modelo completo, con los coeficientes estimados para cada categoría de resultado del partido y sus errores estándar, se presenta en el **Anexo .6**.

En cuanto a la capacidad predictiva, se evaluó el rendimiento del modelo sobre el conjunto de test. La matriz de confusión obtenida, que se muestra en la Tabla 5.3, permite visualizar cómo se distribuyen las predicciones entre las diferentes clases reales del resultado del partido.

Tabla 5.3: Matriz de confusión del modelo multinomial completo.

Predicción	Real = 1	Real = X	Real = 2
1	59	37	24
X	23	17	15
2	19	14	19

La **precisión global (accuracy)** alcanzada fue del **41.85 %**.

### 5.2.3. Modelo reducido: selección de variables significativas

A partir del modelo completo, se calculan los valores  $z$  y los correspondientes  $p$ -valores para cada variable y categoría del resultado. Se seleccionan aquellas variables que presentan significación estadística a un nivel del 5% en al menos una de las categorías. El código empleado para este filtrado es el que se observa a continuación en el Código 5.1.

```

1 coefs <- summary(modelo_multinom)$coefficients
2 se <- summary(modelo_multinom)$standard.errors
3 z <- coefs / se
4 p_values <- 2 * (1 - pnorm(abs(z)))
5
6 # Transponer p_values para que las variables est n como filas
7 p_valores_t <- t(p_values)
8
9 # Obtener nombres de variables con alg n p-valor < 0.05 en cualquier clase
10 variables_significativas <- rownames(p_valores_t)[apply(p_valores_t, 1,
  function(x) any(x < 0.05))]

```

Código 5.1: Cálculo de p-valores y selección de variables significativas

La variable `cinco_cambios` no fue incluida en el modelo reducido, al no alcanzar el umbral de significación.

El modelo reducido fue ajustado únicamente con las 30 variables seleccionadas (el resumen completo del modelo puede consultarse en el Anexo .7). El proceso de optimización del log-likelihood convergió tras 50 iteraciones.

La fórmula del modelo reducido es:

$$\begin{aligned}
\log\left(\frac{P(Y = \text{empate})}{P(Y = 1)}\right) &= -5.481 - 0.023 \cdot x_1 + 0.006 \cdot x_2 + 0.764 \cdot x_3 + 0.103 \cdot x_4 + 0.729 \cdot x_5 + 0.019 \cdot x_6 \\
&\quad - 0.003 \cdot x_7 + 0.014 \cdot x_8 - 0.021 \cdot x_9 - 0.633 \cdot x_{10} - 0.462 \cdot x_{11} + 0.025 \cdot x_{12} \\
&\quad + 0.150 \cdot x_{13} + 0.177 \cdot x_{14} + 0.161 \cdot x_{15} - 0.017 \cdot x_{16} - 0.019 \cdot x_{17} + 0.037 \cdot x_{18} \\
&\quad - 0.004 \cdot x_{19} + 0.266 \cdot x_{20} - 1.289 \cdot x_{21} + 0.819 \cdot x_{22} + 0.049 \cdot x_{23} \\
&\quad - 0.326 \cdot x_{24} - 0.978 \cdot x_{25} + 0.125 \cdot x_{26} + 0.401 \cdot x_{27} + 0.207 \cdot x_{28} - 0.232 \cdot x_{29} \\
\log\left(\frac{P(Y = 2)}{P(Y = 1)}\right) &= 1.276 - 0.041 \cdot x_1 + 0.003 \cdot x_2 + 1.248 \cdot x_3 + 0.143 \cdot x_4 + 0.671 \cdot x_5 + 0.015 \cdot x_6 \\
&\quad + 0.018 \cdot x_7 + 0.026 \cdot x_8 + 0.027 \cdot x_9 - 0.753 \cdot x_{10} - 0.440 \cdot x_{11} - 1.007 \cdot x_{12} \\
&\quad - 0.814 \cdot x_{13} + 1.341 \cdot x_{14} + 0.725 \cdot x_{15} - 0.017 \cdot x_{16} - 0.028 \cdot x_{17} + 0.044 \cdot x_{18} \\
&\quad - 0.005 \cdot x_{19} - 0.077 \cdot x_{20} - 0.856 \cdot x_{21} + 0.777 \cdot x_{22} - 0.015 \cdot x_{23} \\
&\quad - 0.037 \cdot x_{24} - 2.132 \cdot x_{25} + 2.090 \cdot x_{26} - 0.363 \cdot x_{27} + 0.129 \cdot x_{28} - 0.096 \cdot x_{29}
\end{aligned} \tag{5.3}$$

En cuanto a las métricas, este modelo alcanzó una **deviance de 1832.388** y un **AIC de 1952.388**, notablemente inferiores a los del modelo completo, lo que indica una mejor relación entre ajuste y complejidad.

Respecto al rendimiento predictivo, la matriz de confusión correspondiente se muestra en la Tabla 5.4. El modelo alcanzó una **accuracy del 48.9 %**, mejorando la capacidad de clasificación respecto al modelo inicial.

Tabla 5.4: Matriz de confusión del modelo multinomial reducido.

Predicción	Real = 1	Real = X	Real = 2
1	77	42	30
X	16	18	12
2	8	8	16

Para comprobar si la simplificación del modelo implica una pérdida significativa de ajuste, se realizó una comparación formal entre el modelo completo y el modelo reducido mediante un test de razón de verosimilitud. El valor de la estadística fue  $G^2 = 190,73$  con 280 grados de libertad y un p-valor de 0.9999, lo que indica que la reducción del número de variables no compromete la calidad del ajuste (véase Tabla 5.5). Por tanto, el modelo reducido es preferible por su mayor interpretabilidad y menor complejidad.

Resid. df	Resid. Dev	Test	Df	LR stat.	Pvalor
1764	1832.388				
1484	1641.661	1 vs 2	280	190.73	0.9999

Tabla 5.5: Comparación entre modelo completo y reducido (regresión logística multinomial)

## Resumen

La regresión logística multinomial ha permitido abordar de forma directa la predicción del resultado categórico del partido (victoria local, empate o victoria visitante) sin necesidad de modelar goles por separado. Este enfoque ha demostrado ser efectivo y coherente con el objetivo del análisis.

En el estudio se compararon dos modelos: uno completo, que incluía todas las variables disponibles, y otro reducido, construido únicamente con aquellas que resultaron significativas. Los resultados evidencian que el modelo reducido no solo mejora la precisión de clasificación (alcanzando un **48.9 %** de acierto frente al **41.85 %** del modelo completo), sino que también ofrece una estructura más sencilla y mejor ajustada.

Por otro lado, se examinó el papel de la variable `cinco_cambios`, diseñada para capturar el efecto de la normativa de cinco sustituciones. Esta variable no fue seleccionada en el modelo reducido, lo que indica que su influencia en el resultado final del partido no es significativa desde el punto de vista estadístico.

En definitiva, la regresión logística multinomial se consolida como un enfoque adecuado para clasificar el desenlace de los partidos, y los resultados obtenidos refuerzan la importancia de aplicar técnicas de selección de variables para mejorar la interpretabilidad sin sacrificar rendimiento.

## 5.3. Regresión logística multinomial penalizada (glmnet)

Se ha explorado el uso de regresión logística multinomial penalizada mediante la librería `glmnet`. Este enfoque introduce regularización sobre los coeficientes del modelo, lo que permite manejar de forma más eficiente situaciones con muchas variables explicativas, correlaciones entre ellas o riesgo de sobreajuste.

### 5.3.1. Ajuste del modelo y validación cruzada

El ajuste se ha realizado con la función `cv.glmnet()`, que incorpora validación cruzada para seleccionar automáticamente el valor óptimo del parámetro de penalización `lambda`. Además, se ha explorado el efecto del parámetro `alpha`, que determina el tipo de regularización:

- $\alpha = 0$ : penalización Ridge (norma L2),
- $\alpha = 1$ : penalización Lasso (norma L1),
- $0 < \alpha < 1$ : combinación de ambas (Elastic Net).

Para cada  $\alpha$  (0, 0.25, 0.5, 0.75, 1), se seleccionó el mejor  $\lambda$  y se evaluó la *accuracy* sobre el conjunto de prueba. El conjunto de datos se dividió en entrenamiento (80%) y test (20%), y las matrices de predictores se prepararon con `model.matrix()` para ser compatibles con `glmnet`.

### 5.3.2. Resultados

En primer lugar, se aplicó el modelo a la totalidad de los partidos disponibles. Los resultados fueron los siguientes:

Alpha	Lambda óptimo	Accuracy en test
0.00	4.0626	46,64 %
0.25	0.0867	<b>47,01 %</b>
0.50	0.0434	46,64 %
0.75	0.0289	46,27 %
1.00	0.0217	45,52 %

Tabla 5.6: Resultados del modelo `glmnet` multinomial según el parámetro  $\alpha$  (conjunto completo)

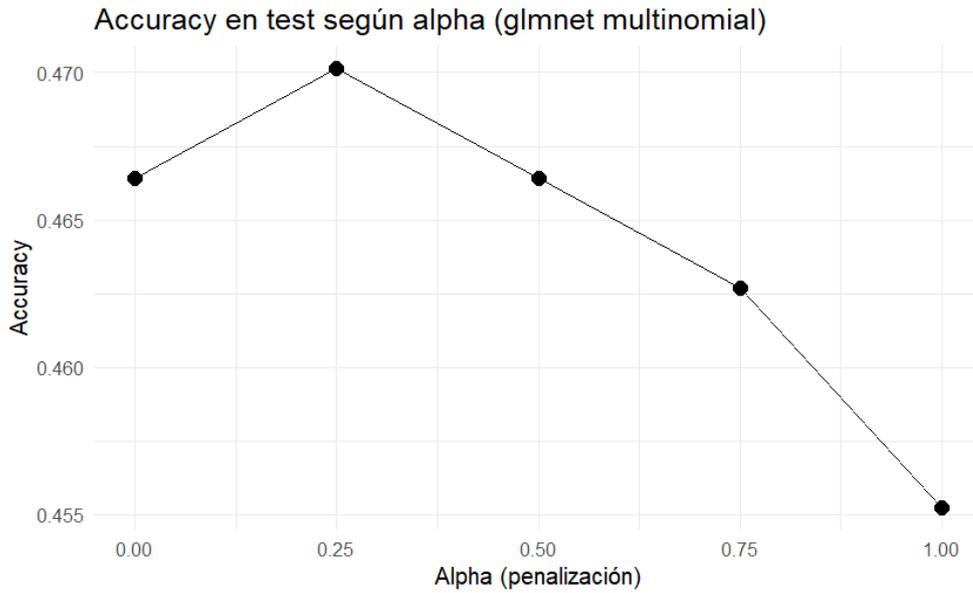


Figura 5.1: Accuracy en test según alpha (glmnet multinomial, conjunto completo)

El mejor resultado se obtuvo con  $\alpha = 0,25$ , (Figura 5.1) lo que corresponde a una penalización tipo Elastic Net. Este modelo alcanzó una **accuracy del 47,01 %** (Tabla 5.6).

### 5.3.3. Conclusión

La regresión logística multinomial penalizada mediante `glmnet` ha mostrado un rendimiento notable, especialmente cuando se ajusta con una penalización tipo Elastic Net. Este ha sido competitivo frente a modelos clásicos como Poisson y Multinomial.

En conjunto, este modelo aporta una solución sólida, flexible y estadísticamente robusta para la clasificación directa del resultado del partido.

## 5.4. Modelo de regresión con XGBoost

Se ha explorado el uso del algoritmo XGBoost (*Extreme Gradient Boosting*) para predecir el número de goles marcados por cada equipo. XGBoost es un método basado en árboles de decisión que optimiza la función de pérdida mediante gradiente, incorporando técnicas como regularización, subsampling y paralelización. Su flexibilidad y eficiencia lo convierten en una herramienta potente para tareas de predicción numérica en contextos complejos como el deportivo.

### 5.4.1. Ajuste del modelo y búsqueda de hiperparámetros

Se ajustaron dos modelos independientes: uno para predecir los goles del equipo local y otro para los del visitante. En ambos casos se llevó a cabo una búsqueda en cuadrícula sobre los principales hiperparámetros del algoritmo:

- `max_depth`: profundidad máxima de los árboles (3, 5, 7)
- `eta`: tasa de aprendizaje (0.01, 0.05, 0.1)
- `subsample`: proporción de muestras utilizadas por iteración (0.8, 1)
- `colsample_bytree`: proporción de predictores considerados por árbol (0.8, 1)

Cada configuración se evaluó con 100 iteraciones (`nrounds = 100`) sobre un conjunto de prueba (20 % del total), utilizando el RMSE y el MAE como métricas de rendimiento.

La predicción del signo del partido (1, X o 2) se derivó a partir de los goles estimados, aplicando la siguiente regla:

- Se predijo un empate (X) cuando la diferencia absoluta entre goles predichos de equipo local y visitante fue inferior a 0.05.
- Se predijo victoria local (1) si el equipo local superaba al visitante en goles con una diferencia mayor de 0.05.
- Se predijo victoria visitante (2) si el equipo visitante superaba al local en goles con una diferencia mayor de 0.05.

### 5.4.2. Resultados

Los mejores modelos XGBoost aplicados sobre el conjunto completo se ajustaron con los siguientes hiperparámetros optimizados: `max_depth = 3`, `eta = 0.05`, `subsample =`

1.00 y `colsample_bytree = 1.00`. Estos valores se mantuvieron constantes tanto para la predicción de goles locales como visitantes.

Los resultados obtenidos fueron los siguientes:

- **Goles locales:** MAE = 0,8647; RMSE = 1,1189
- **Goles visitantes:** MAE = 0,7514; RMSE = 0,9898

A partir de las predicciones de goles, se infirió el resultado del partido (1, X o 2), obteniéndose la matriz de confusión mostrada en la Tabla 5.7.

<b>Predicho \ Real</b>	1	2	X
1	95	61	57
2	13	16	8
X	7	6	5

Tabla 5.7: Matriz de confusión del resultado del partido (XGBoost, conjunto completo)

Como se observa, el modelo tiende a sobrepredicir victorias locales, mientras que presenta más dificultades para identificar empates y victorias visitantes. El **accuracy global** fue del **43,28 %**, un rendimiento comparable al de los modelos de Poisson, aunque con mayor flexibilidad y capacidad no lineal. A pesar de que XGBoost no está específicamente diseñado para clasificación multiclase, su uso en la predicción de goles ofrece una vía eficaz para derivar el resultado final con un nivel de acierto aceptable.

## Conclusión

El modelo XGBoost ha demostrado ser una herramienta robusta para predecir el número de goles en un partido de fútbol, permitiendo posteriormente inferir el resultado del encuentro. A pesar de que este algoritmo no está específicamente diseñado para clasificación multiclase, su capacidad para modelar relaciones no lineales y complejas entre variables predictoras le permite ofrecer un rendimiento competitivo.

Los errores obtenidos en la predicción de goles (MAE y RMSE) se situaron en niveles aceptables tanto para el equipo local como para el visitante. A partir de estas predicciones, se derivó el resultado del partido, obteniéndose una precisión global del 43,28 %. Este valor es comparable al alcanzado por modelos clásicos como la regresión de Poisson, pero con la ventaja adicional de capturar patrones más complejos gracias a la estructura flexible del modelo.

Sin embargo, la matriz de confusión revela una tendencia del modelo a favorecer la predicción de victorias locales, lo que sugiere un cierto sesgo hacia la clase mayoritaria. Este comportamiento

es común en contextos con clases desbalanceadas y podría mitigarse con estrategias como el ajuste de pesos por clase o técnicas de balanceo en el conjunto de entrenamiento.

## 5.5. Clasificación directa del resultado con XGBoost

Además de los modelos de regresión, se ha explorado un enfoque de clasificación directa del signo del partido (1, X o 2) mediante XGBoost, utilizando su configuración multiclase. Este modelo permite abordar el objetivo final sin depender de predicciones intermedias de goles.

### 5.5.1. Ajuste y validación del modelo

La variable `resultado_partido` se recodificó como un factor con niveles ordinales ("1", "X", "2"), transformados internamente en etiquetas numéricas (0, 1, 2). Se eliminaron columnas no predictivas como los goles reales y los identificadores, y se dividió el conjunto en entrenamiento (80 %) y prueba (20 %).

Se llevó a cabo una búsqueda en cuadrícula sobre los siguientes hiperparámetros:

- `max_depth`: 3, 5, 7
- `eta` (tasa de aprendizaje): 0.01, 0.05, 0.1
- `subsample`: 0.8, 1.0
- `colsample_bytree`: 0.8, 1.0

Cada configuración se evaluó utilizando la métrica `error` (error de clasificación) y se seleccionó el modelo con mayor *accuracy* sobre el conjunto de prueba.

### 5.5.2. Resultados

El mejor modelo XGBoost entrenado para predecir directamente el resultado del partido (victoria local, empate o victoria visitante) utilizó los siguientes hiperparámetros: `max_depth = 3`, `eta = 0.01`, `subsample = 0.80` y `colsample_bytree = 1.00`.

La Tabla 5.8 presenta la matriz de confusión correspondiente a este modelo, evaluado sobre el conjunto completo de partidos.

Predicho \ Real	1	X	2
1	103	58	41
X	5	7	8
2	14	12	20

Tabla 5.8: Matriz de confusión del mejor modelo XGBoost (clasificación directa, todos los partidos)

El modelo mostró una elevada capacidad para identificar correctamente victorias locales, aunque continúa presentando dificultades en la detección de empates y victorias visitantes, posiblemente debido al desbalance en la distribución de clases. La **accuracy alcanzada fue del 48,51 %**, superando al modelo basado en regresión sobre goles y también a los modelos clásicos, lo que refuerza el interés de abordar el problema como una tarea de clasificación multiclase directa.

Estos resultados ponen de manifiesto la ventaja de utilizar algoritmos modernos de aprendizaje automático cuando el objetivo es predecir el resultado del partido directamente, sin necesidad de inferencias intermedias.

## Conclusión

El enfoque de clasificación directa mediante XGBoost ha ofrecido resultados prometedores en la predicción del resultado del partido, alcanzando una precisión del 48,51 % en el conjunto completo de datos. Este valor supera al obtenido mediante la predicción indirecta basada en goles, lo que demuestra el potencial de los algoritmos de aprendizaje automático para abordar directamente tareas de clasificación multiclase en contextos deportivos.

El modelo se mostró especialmente eficaz en la identificación de victorias locales, aunque su rendimiento fue menor en la predicción de empates y victorias visitantes, una limitación probablemente relacionada con el desequilibrio en la distribución de clases. A pesar de ello, la mejora global en precisión y la simplicidad del enfoque —al evitar el paso intermedio de estimar los goles— justifican el interés por esta estrategia.

En conjunto, los resultados obtenidos confirman que tratar el problema como una tarea de clasificación directa, en lugar de derivarlo a partir de una regresión sobre goles, permite obtener modelos más eficaces y adaptados a la naturaleza discreta del resultado final. Futuras extensiones podrían considerar técnicas de balanceo, calibración de probabilidades o ensamblado de modelos para mejorar aún más el rendimiento.



# Capítulo 6

## Conclusiones

En este trabajo se ha explorado la capacidad de la estadística para modelizar y anticipar resultados en el fútbol profesional, centrándose en un aspecto poco tratado hasta ahora: la rotación de los equipos. A partir de una base de datos cuidadosamente tratada y segmentada según la normativa de sustituciones, se han aplicado distintos modelos predictivos con el objetivo de evaluar hasta qué punto las decisiones relacionadas con cambios en la alineación o en el transcurso del partido pueden estar asociadas al resultado final.

Desde el punto de vista predictivo, los resultados muestran que los modelos de regresión de Poisson, ofrecen un rendimiento razonable, pero limitado. En cambio, modelos como la regresión Multinomial, la regresión penalizada con `glmnet` o la clasificación directa mediante XGBoost han logrado niveles de precisión muy competitivos, donde se ha alcanzado una *accuracy* superior al 48 %.

A pesar de estos resultados positivos, también deben señalarse algunas limitaciones. El estudio se ha centrado exclusivamente en la Primera División española, lo que puede condicionar la generalización de los hallazgos. Además, aunque se ha intentado reflejar la dinámica de los equipos a través de indicadores prepartido, no se ha abordado una perspectiva temporal o secuencial más detallada, que podría enriquecer aún más los modelos.

Como posibles líneas futuras, se propone ampliar el conjunto de datos a otras ligas o competiciones, lo que permitiría evaluar la robustez de los modelos frente a estilos de juego y contextos tácticos diferentes. También sería interesante incorporar más temporadas, para mejorar la representatividad y el poder explicativo de las variables consideradas. Otra vía de mejora pasa por optimizar los modelos actuales ajustando con mayor precisión sus hiperparámetros, así como explorar validaciones temporales más finas o segmentaciones alternativas del conjunto de datos, que puedan capturar matices específicos del calendario competitivo o de la evolución de los equipos. Por último, una mejor caracterización de la rotación, por ejemplo diferenciando entre cambios obligados y estratégicos, podría aportar información adicional de gran valor.

En definitiva, este trabajo pone de manifiesto el potencial de la estadística aplicada al deporte para generar conocimiento útil, tanto en el ámbito académico como en el profesional, y abre la puerta a nuevas investigaciones centradas en la gestión táctica y estratégica de los equipos desde una perspectiva cuantitativa.





# **Anexos**



# Bibliografía

- [1] Michael J. Maher. Modelling association football scores. *Statistica Neerlandica*, 36(3):109–118, 1982.
- [2] Håvard Rue and Øyvind Salvesen. Prediction and retrospective analysis of soccer matches in a league. *The Statistician*, 49(3):399–418, 2000.
- [3] Amir Joseph, Norman Fenton, and Martin Neil. Predicting football results using bayesian networks. *Knowledge-Based Systems*, 19(7):544–553, 2006.
- [4] Chris Carling, Franck Le Gall, and Gregory Dupont. Are physical performance and injury risk in a professional soccer team in match-play affected over a prolonged period of fixture congestion? *International Journal of Sports Medicine*, 33(01):36–42, 2012.
- [5] Gareth James, Daniela Witten, Trevor Hastie, and Robert Tibshirani. *An Introduction to Statistical Learning: With Applications in R*. Springer, second edition, 2021.
- [6] Jerome Friedman, Trevor Hastie, and Robert Tibshirani. *The Elements of Statistical Learning*. Springer, 2001.
- [7] Peter McCullagh and John A. Nelder. *Generalized Linear Models*. Chapman and Hall/CRC, second edition, 1989.
- [8] V. Pando Fernández and R. San Martín Fernández. Regresión logística multinomial. *Cuadernos de la Sociedad Española de Ciencias Forestales*, (18), 2004. Actas de la Reunión de Modelización Forestal.
- [9] Data Science Team. From linear regression to ridge regression, the lasso and the elastic net. <https://medium.com/data-science/from-linear-regression-to-ridge-regression-the-lasso-and-the-elastic-net-4eaecaf5f7e6>, 2020. Consultado en mayo de 2025.
- [10] Tianqi Chen and Carlos Guestrin. Xgboost: A scalable tree boosting system. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 785–794, 2016.

- [11] José María Lozano Olmedo. Comparación y evaluación de diferentes técnicas de ia para un modelo de rotación de empleados en empresas aplicado a equipos de fútbol. Master's thesis, Universidad de Valladolid, 2024. Trabajo Fin de Máster. Consultado en mayo de 2025.
- [12] Resultados Fútbol. Resultados de fútbol, 2025. Consulta realizada el 15 de junio de 2025.
- [13] Transfermarkt. Transfermarkt - valores de mercado de equipos y jugadores, 2024. Accedido en mayo de 2025.
- [14] AS. La ifab aprueba que los cinco cambios se queden para siempre, 2021. Consultado en mayo de 2025.

# Repositorio del código

Todo el código fuente desarrollado para este Trabajo de Fin de Grado, tanto en R como en Python, se encuentra disponible en el siguiente repositorio de GitHub:

[https://github.com/miguelgp02/TFG\\_Estadistica.git](https://github.com/miguelgp02/TFG_Estadistica.git)

Este repositorio incluye:

- Los scripts de análisis descriptivo y modelado estadístico en R.
- Los scripts de tratamiento previo de datos en Python.
- Los conjuntos de datos utilizados en el estudio.



# Listado completo de variables

## .1. Variables del conjunto de datos

A continuación se presenta el listado completo de las 175 variables utilizadas en el estudio:

1. id\_indicadores\_equipo\_prepartido
2. id\_partido
3. porcentaje local ganados en sitio
4. porcentaje local ganados en general
5. porcentaje local empatados en sitio
6. porcentaje local empatados en general
7. porcentaje local perdidos en sitio
8. porcentaje local perdidos en general
9. porcentaje visitante ganados en sitio
10. porcentaje visitante ganados en general
11. porcentaje visitante empatados en sitio
12. porcentaje visitante empatados en general
13. porcentaje visitante perdidos en sitio
14. porcentaje visitante perdidos en general
15. proporcion local puntos en sitio
16. proporcion local puntos en general
17. proporcion visitante puntos en sitio

18. proporcion visitante puntos en general
19. porcentaje local mas 1,5 en sitio
20. porcentaje local mas 1,5 en general
21. porcentaje visitante mas 1,5 en sitio
22. porcentaje visitante mas 1,5 en general
23. porcentaje local mas 2,5 en sitio
24. porcentaje local mas 2,5 en general
25. porcentaje visitante mas 2,5 en sitio
26. porcentaje visitante mas 2,5 en general
27. porcentaje local mas 3,5 en sitio
28. porcentaje local mas 3,5 en general
29. porcentaje visitante mas 3,5 en sitio
30. porcentaje visitante mas 3,5 en general
31. porcentaje local mas 4,5 en sitio
32. porcentaje local mas 4,5 en general
33. porcentaje visitante mas 4,5 en sitio
34. porcentaje visitante mas 4,5 en general
35. proporcion local goles totales en sitio
36. proporcion local goles totales en general
37. proporcion local goles marcados en sitio
38. proporcion local goles marcados en general
39. proporcion local goles encajados en sitio
40. proporcion local goles encajados en general
41. proporcion visitante goles totales en sitio
42. proporcion visitante goles totales en general
43. proporcion visitante goles marcados en sitio

44. proporcion visitante goles marcados en general
45. proporcion visitante goles encajados en sitio
46. proporcion visitante goles encajados en general
47. porcentaje local mas 0,5 marcados en sitio
48. porcentaje local mas 0,5 marcados en general
49. porcentaje local mas 1,5 marcados en sitio
50. porcentaje local mas 1,5 marcados en general
51. porcentaje local mas 2,5 marcados en sitio
52. porcentaje local mas 2,5 marcados en general
53. porcentaje local mas 0,5 encajados en sitio
54. porcentaje local mas 0,5 encajados en general
55. porcentaje local mas 1,5 encajados en sitio
56. porcentaje local mas 1,5 encajados en general
57. porcentaje local mas 2,5 encajados en sitio
58. porcentaje local mas 2,5 encajados en general
59. porcentaje visitante mas 0,5 marcados en sitio
60. porcentaje visitante mas 0,5 marcados en general
61. porcentaje visitante mas 1,5 marcados en sitio
62. porcentaje visitante mas 1,5 marcados en general
63. porcentaje visitante mas 2,5 marcados en sitio
64. porcentaje visitante mas 2,5 marcados en general
65. porcentaje visitante mas 0,5 encajados en sitio
66. porcentaje visitante mas 0,5 encajados en general
67. porcentaje visitante mas 1,5 encajados en sitio
68. porcentaje visitante mas 1,5 encajados en general
69. porcentaje visitante mas 2,5 encajados en sitio

70. porcentaje visitante mas 2,5 encajados en general
71. proporcion local amarillas en sitio
72. proporcion local amarillas en general
73. proporcion visitante amarillas en sitio
74. proporcion visitante amarillas en general
75. proporcion local rojas en sitio
76. proporcion local rojas en general
77. proporcion visitante rojas en sitio
78. proporcion visitante rojas en general
79. proporcion local cambios en sitio
80. proporcion local cambios en general
81. proporcion visitante cambios en sitio
82. proporcion visitante cambios en general
83. proporcion local posesion en sitio
84. proporcion local posesion en general
85. proporcion visitante posesion en sitio
86. proporcion visitante posesion en general
87. proporcion local total tiros en sitio
88. proporcion local total tiros en general
89. proporcion visitante total tiros en sitio
90. proporcion visitante total tiros en general
91. proporcion local corners a favor en sitio
92. proporcion local corners a favor en general
93. proporcion visitante corners a favor en sitio
94. proporcion visitante corners a favor en general
95. proporcion local corners en contra en sitio

96. proporcion local corners en contra en general
97. proporcion visitante corners en contra en sitio
98. proporcion visitante corners en contra en general
99. proporcion local cambios lesionados sitio
100. proporcion local cambios lesionados en general
101. proporcion visitante cambios lesionados en sitio
102. proporcion visitante cambios lesionados en general
103. proporcion local cambios amarillas sitio
104. proporcion local cambios amarillas en general
105. proporcion visitante cambios amarillas en sitio
106. proporcion visitante cambios amarillas en general
107. proporcion local cambios goleadores sitio
108. proporcion local cambios goleadores en general
109. proporcion visitante cambios goleadores en sitio
110. proporcion visitante cambios goleadores en general
111. proporcion local cambios asistentes sitio
112. proporcion local cambios asistentes en general
113. proporcion visitante cambios asistentes en sitio
114. proporcion visitante cambios asistentes en general
115. media local cambios minutos sitio
116. media local cambios minutos en general
117. media visitante cambios minutos sitio
118. media visitante cambios minutos en general
119. proporcion local cambios delanteros a centrocampistas sitio
120. proporcion local cambios delanteros a centrocampistas en general
121. proporcion visitante cambios delanteros a centrocampistas en sitio

122. proporcion visitante cambios delanteros a centrocampistas en general
123. proporcion local cambios delanteros a defensas sitio
124. proporcion local cambios delanteros a defensas en general
125. proporcion visitante cambios delanteros a defensas en sitio
126. proporcion visitante cambios delanteros a defensas en general
127. proporcion local cambios centrocampistas a delanteros sitio
128. proporcion local cambios centrocampistas a delanteros en general
129. proporcion visitante cambios centrocampistas a delanteros en sitio
130. proporcion visitante cambios centrocampistas a delanteros en general
131. proporcion local cambios centrocampistas a defensas sitio
132. proporcion local cambios centrocampistas a defensas en general
133. proporcion visitante cambios centrocampistas a defensas en sitio
134. proporcion visitante cambios centrocampistas a defensas en general
135. proporcion local cambios defensas a delanteros sitio
136. proporcion local cambios defensas a delanteros en general
137. proporcion visitante cambios defensas a delanteros en sitio
138. proporcion visitante cambios defensas a delanteros en general
139. proporcion local cambios defensas a centrocampistas sitio
140. proporcion local cambios defensas a centrocampistas en general
141. proporcion visitante cambios defensas a centrocampistas en sitio
142. proporcion visitante cambios defensas a centrocampistas en general
143. proporcion local cambios antes descanso sitio
144. proporcion local cambios antes descanso en general
145. proporcion visitante cambios antes descanso en sitio
146. proporcion visitante cambios antes descanso en general
147. proporcion local cambios 45 a 60 sitio

- 148. proporcion local cambios 45 a 60 en general
- 149. proporcion visitante cambios 45 a 60 en sitio
- 150. proporcion visitante cambios 45 a 60 en general
- 151. proporcion local cambios 61 a 75 sitio
- 152. proporcion local cambios 61 a 75 en general
- 153. proporcion visitante cambios 61 a 75 en sitio
- 154. proporcion visitante cambios 61 a 75 en general
- 155. proporcion local cambios 76 a final sitio
- 156. proporcion local cambios 76 a final en general
- 157. proporcion visitante cambios 76 a final en sitio
- 158. proporcion visitante cambios 76 a final en general
- 159. proporcion local cambios alineacion defensa sitio
- 160. proporcion local cambios alineacion defensa en general
- 161. proporcion visitante cambios alineacion defensa en sitio
- 162. proporcion visitante cambios alineacion defensa en general
- 163. proporcion local cambios alineacion centrocampista sitio
- 164. proporcion local cambios alineacion centrocampista en general
- 165. proporcion visitante cambios alineacion centrocampista en sitio
- 166. proporcion visitante cambios alineacion centrocampista en general
- 167. proporcion local cambios alineacion delantero sitio
- 168. proporcion local cambios alineacion delantero en general
- 169. proporcion visitante cambios alineacion delantero en sitio
- 170. proporcion visitante cambios alineacion delantero en general
- 171. resultado\_local
- 172. resultado\_visitante
- 173. resultado\_partido

174. jornada

175. temporada

# Salidas de Modelos de Poisson

## .2. Modelo completo de goles del equipo local

	Estimate	Std. Error	z value	Pr(> z )
(Intercept)	1846.84	1343.89	1.37	0.17
'porcentaje local ganados en sitio'	-3.96	7.14	-0.55	0.58
'porcentaje local ganados en general'	-0.25	6.10	-0.04	0.97
'porcentaje local empatados en sitio'	-3.98	7.12	-0.56	0.58
'porcentaje local empatados en general'	0.10	6.08	0.02	0.99
'porcentaje local perdidos en sitio'	-4.00	7.12	-0.56	0.57
'porcentaje local perdidos en general'	0.29	6.07	0.05	0.96
'porcentaje visitante ganados en sitio'	-10.45	7.70	-1.36	0.17
'porcentaje visitante ganados en general'	-4.83	6.52	-0.74	0.46
'porcentaje visitante empatados en sitio'	-10.23	7.66	-1.34	0.18
'porcentaje visitante empatados en general'	-4.72	6.50	-0.73	0.47
'porcentaje visitante perdidos en sitio'	-10.11	7.65	-1.32	0.19
'porcentaje visitante perdidos en general'	-4.67	6.49	-0.72	0.47
'proporcion local puntos en sitio'	-1.05	11.96	-0.09	0.93
'proporcion local puntos en general'	17.79	12.03	1.48	0.14
'proporcion visitante puntos en sitio'	11.20	11.99	0.93	0.35
'proporcion visitante puntos en general'	4.81	11.87	0.41	0.69
'porcentaje local mas 1,5 en sitio'	-0.02	0.01	-2.05	0.04
'porcentaje local mas 1,5 en general'	0.02	0.01	1.34	0.18
'porcentaje visitante mas 1,5 en sitio'	0.00	0.01	0.26	0.80
'porcentaje visitante mas 1,5 en general'	-0.00	0.01	-0.11	0.91
'porcentaje local mas 2,5 en sitio'	-0.01	0.01	-1.75	0.08
'porcentaje local mas 2,5 en general'	-0.00	0.01	-0.05	0.96
'porcentaje visitante mas 2,5 en sitio'	-0.01	0.01	-1.71	0.09
'porcentaje visitante mas 2,5 en general'	0.01	0.01	0.93	0.35
'porcentaje local mas 3,5 en sitio'	-0.02	0.01	-2.62	0.01
'porcentaje local mas 3,5 en general'	0.02	0.01	1.19	0.23
'porcentaje visitante mas 3,5 en sitio'	0.00	0.01	0.05	0.96
'porcentaje visitante mas 3,5 en general'	-0.00	0.01	-0.38	0.70
'porcentaje local mas 4,5 en sitio'	-0.02	0.01	-2.13	0.03
'porcentaje local mas 4,5 en general'	0.01	0.02	0.69	0.49
'porcentaje visitante mas 4,5 en sitio'	-0.01	0.01	-1.19	0.23
'porcentaje visitante mas 4,5 en general'	0.00	0.01	0.06	0.95

Tabla 1: Resumen de coeficientes del modelo de Poisson para goles locales

	Estimate	Std. Error	z value	Pr(> z )
'proporcion local goles totales en sitio'	4.81	6.98	0.69	0.49
'proporcion local goles totales en general'	-2.46	6.70	-0.37	0.71
'proporcion local goles marcados en sitio'	-4.95	7.00	-0.71	0.48
'proporcion local goles marcados en general'	3.54	6.73	0.53	0.60
'proporcion local goles encajados en sitio'	-2.75	6.95	-0.40	0.69
'proporcion local goles encajados en general'	1.16	6.75	0.17	0.86
'proporcion visitante goles totales en sitio'	0.80	7.08	0.11	0.91
'proporcion visitante goles totales en general'	-5.83	6.91	-0.84	0.40
'proporcion visitante goles marcados en sitio'	0.37	7.02	0.05	0.96
'proporcion visitante goles marcados en general'	5.14	6.88	0.75	0.46
'proporcion visitante goles encajados en sitio'	-0.93	7.12	-0.13	0.90
'proporcion visitante goles encajados en general'	6.21	6.94	0.90	0.37
'porcentaje local mas 0,5 marcados en sitio'	-0.00	0.01	-0.14	0.89
'porcentaje local mas 0,5 marcados en general'	-0.01	0.01	-0.72	0.47
'porcentaje local mas 1,5 marcados en sitio'	0.02	0.01	1.67	0.10
'porcentaje local mas 1,5 marcados en general'	-0.02	0.01	-1.55	0.12
'porcentaje local mas 2,5 marcados en sitio'	0.02	0.01	1.44	0.15
'porcentaje local mas 2,5 marcados en general'	-0.02	0.02	-0.76	0.45
'porcentaje local mas 0,5 encajados en sitio'	-0.00	0.01	-0.58	0.56
'porcentaje local mas 0,5 encajados en general'	0.01	0.01	0.63	0.53
'porcentaje local mas 1,5 encajados en sitio'	0.00	0.01	0.38	0.70
'porcentaje local mas 1,5 encajados en general'	-0.00	0.01	-0.32	0.75
'porcentaje local mas 2,5 encajados en sitio'	-0.01	0.01	-0.48	0.63
'porcentaje local mas 2,5 encajados en general'	0.01	0.02	0.43	0.67
'porcentaje visitante mas 0,5 marcados en sitio'	-0.01	0.01	-0.64	0.52
'porcentaje visitante mas 0,5 marcados en general'	0.01	0.01	1.05	0.29
'porcentaje visitante mas 1,5 marcados en sitio'	-0.01	0.01	-0.59	0.56
'porcentaje visitante mas 1,5 marcados en general'	0.01	0.01	0.75	0.45
'porcentaje visitante mas 2,5 marcados en sitio'	-0.02	0.01	-1.59	0.11
'porcentaje visitante mas 2,5 marcados en general'	0.02	0.02	1.26	0.21
'porcentaje visitante mas 0,5 encajados en sitio'	0.00	0.01	0.34	0.73
'porcentaje visitante mas 0,5 encajados en general'	-0.01	0.01	-0.68	0.49
'porcentaje visitante mas 1,5 encajados en sitio'	0.00	0.01	0.47	0.64
'porcentaje visitante mas 1,5 encajados en general'	-0.01	0.01	-0.70	0.48
'porcentaje visitante mas 2,5 encajados en sitio'	0.00	0.01	0.00	1.00
'porcentaje visitante mas 2,5 encajados en general'	0.01	0.01	0.45	0.65
'proporcion local amarillas en sitio'	-0.05	0.18	-0.28	0.78
'proporcion local amarillas en general'	0.13	0.24	0.57	0.57
'proporcion visitante amarillas en sitio'	-0.23	0.17	-1.33	0.18
'proporcion visitante amarillas en general'	0.17	0.20	0.86	0.39
'proporcion local rojas en sitio'	0.34	0.51	0.66	0.51
'proporcion local rojas en general'	0.23	0.76	0.30	0.76
'proporcion visitante rojas en sitio'	-0.42	0.51	-0.83	0.41
'proporcion visitante rojas en general'	-0.12	0.82	-0.15	0.88
'proporcion local cambios en sitio'	-1.10	1.37	-0.80	0.42
'proporcion local cambios en general'	-1.17	1.64	-0.72	0.47
'proporcion visitante cambios en sitio'	-1.69	1.32	-1.28	0.20
'proporcion visitante cambios en general'	3.24	1.60	2.02	0.04
'proporcion local posesion en sitio'	-0.03	0.02	-1.32	0.19
'proporcion local posesion en general'	0.05	0.03	2.08	0.04
'proporcion visitante posesion en sitio'	0.00	0.02	0.03	0.97
'proporcion visitante posesion en general'	0.00	0.02	0.13	0.89
'proporcion local total tiros en sitio'	0.07	0.05	1.31	0.19
'proporcion local total tiros en general'	-0.07	0.07	-0.98	0.33
'proporcion visitante total tiros en sitio'	0.01	0.05	0.15	0.88
'proporcion visitante total tiros en general'	-0.05	0.06	-0.82	0.41
'proporcion local corners a favor en sitio'	0.06	0.08	0.79	0.43
'proporcion local corners a favor en general'	0.03	0.11	0.25	0.80
'proporcion visitante corners a favor en sitio'	-0.01	0.08	-0.11	0.91
'proporcion visitante corners a favor en general'	0.04	0.11	0.33	0.74
'proporcion local corners en contra en sitio'	0.03	0.08	0.45	0.66
'proporcion local corners en contra en general'	0.06	0.10	0.60	0.55
'proporcion visitante corners en contra en sitio'	-0.00	0.07	-0.01	1.00
'proporcion visitante corners en contra en general'	0.08	0.10	0.78	0.44
'proporcion local cambios lesionados sitio'	-0.48	0.36	-1.33	0.18
'proporcion local cambios lesionados en general'	0.62	0.49	1.27	0.21
'proporcion visitante cambios lesionados en sitio'	0.35	0.33	1.06	0.29
'proporcion visitante cambios lesionados en general'	0.12	0.43	0.28	0.78

Tabla 2: Resumen de coeficientes del modelo de Poisson para goles locales

	Estimate	Std. Error	z value	Pr(> z )
'proporcion local cambios amarillas sitio'	-0.08	0.30	-0.26	0.80
'proporcion local cambios amarillas en general'	-0.05	0.40	-0.12	0.91
'proporcion visitante cambios amarillas en sitio'	0.27	0.28	0.94	0.35
'proporcion visitante cambios amarillas en general'	-0.25	0.39	-0.64	0.52
'proporcion local cambios goleadores sitio'	0.31	0.39	0.79	0.43
'proporcion local cambios goleadores en general'	-0.20	0.53	-0.38	0.71
'proporcion visitante cambios goleadores en sitio'	-0.47	0.41	-1.16	0.24
'proporcion visitante cambios goleadores en general'	-0.04	0.57	-0.07	0.94
'proporcion local cambios asistentes sitio'	-0.26	0.44	-0.59	0.56
'proporcion local cambios asistentes en general'	0.35	0.62	0.57	0.57
'proporcion visitante cambios asistentes en sitio'	0.77	0.44	1.73	0.08
'proporcion visitante cambios asistentes en general'	-0.68	0.57	-1.20	0.23
'media local cambios minutos sitio'	-0.01	0.06	-0.10	0.92
'media local cambios minutos en general'	-0.08	0.08	-1.03	0.30
'media visitante cambios minutos en general'	0.17	0.08	2.29	0.02
'proporcion local cambios delanteros a centrocampistas sitio'	0.01	0.34	0.03	0.97
'proporcion local cambios delanteros a centrocampistas en general'	0.11	0.41	0.27	0.79
'proporcion visitante cambios delanteros a centrocampistas en sitio'	0.05	0.34	0.15	0.88
'proporcion visitante cambios delanteros a centrocampistas en general'	0.04	0.39	0.10	0.92
'proporcion local cambios delanteros a defensas sitio'	-0.07	0.42	-0.16	0.87
'proporcion local cambios delanteros a defensas en general'	-0.03	0.54	-0.05	0.96
'proporcion visitante cambios delanteros a defensas en sitio'	0.40	0.43	0.92	0.36
'proporcion visitante cambios delanteros a defensas en general'	-0.65	0.51	-1.29	0.20
'proporcion local cambios centrocampistas a delanteros sitio'	-0.01	0.29	-0.02	0.99
'proporcion local cambios centrocampistas a delanteros en general'	0.09	0.33	0.29	0.78
'proporcion visitante cambios centrocampistas a delanteros en sitio'	-0.36	0.30	-1.20	0.23
'proporcion visitante cambios centrocampistas a delanteros en general'	0.67	0.43	1.54	0.12
'proporcion local cambios centrocampistas a defensas sitio'	0.62	0.51	1.21	0.22
'proporcion local cambios centrocampistas a defensas en general'	-0.31	0.69	-0.45	0.65
'proporcion visitante cambios centrocampistas a defensas en sitio'	-0.50	0.52	-0.95	0.34
'proporcion visitante cambios centrocampistas a defensas en general'	0.27	0.70	0.39	0.70
'proporcion local cambios defensas a delanteros sitio'	-0.40	0.43	-0.92	0.36
'proporcion local cambios defensas a delanteros en general'	-0.05	0.47	-0.10	0.92
'proporcion visitante cambios defensas a delanteros en sitio'	0.13	0.43	0.31	0.75
'proporcion visitante cambios defensas a delanteros en general'	-0.13	0.64	-0.20	0.84
'proporcion local cambios defensas a centrocampistas sitio'	-0.65	0.59	-1.10	0.27
'proporcion local cambios defensas a centrocampistas en general'	0.55	0.71	0.77	0.44
'proporcion visitante cambios defensas a centrocampistas en sitio'	0.20	0.61	0.33	0.74
'proporcion visitante cambios defensas a centrocampistas en general'	-0.40	0.84	-0.47	0.64
'proporcion local cambios antes descanso sitio'	1.82	1.35	1.35	0.18
'proporcion local cambios antes descanso en general'	0.09	1.52	0.06	0.95
'proporcion visitante cambios antes descanso en sitio'	0.88	1.29	0.68	0.50
'proporcion visitante cambios antes descanso en general'	-3.08	1.57	-1.96	0.05
'proporcion local cambios 45 a 60 sitio'	0.98	1.31	0.75	0.45
'proporcion local cambios 45 a 60 en general'	1.02	1.48	0.69	0.49
'proporcion visitante cambios 45 a 60 en sitio'	0.78	1.23	0.64	0.53
'proporcion visitante cambios 45 a 60 en general'	-1.95	1.45	-1.34	0.18
'proporcion local cambios 61 a 75 sitio'	0.78	1.36	0.57	0.57
'proporcion local cambios 61 a 75 en general'	1.47	1.59	0.92	0.36
'proporcion visitante cambios 61 a 75 en sitio'	1.26	1.30	0.97	0.33
'proporcion visitante cambios 61 a 75 en general'	-2.92	1.56	-1.87	0.06
'proporcion local cambios 76 a final sitio'	0.72	1.46	0.49	0.62
'proporcion local cambios 76 a final en general'	1.51	1.83	0.83	0.41
'proporcion visitante cambios 76 a final en sitio'	1.50	1.40	1.07	0.28
'proporcion visitante cambios 76 a final en general'	-3.15	1.68	-1.87	0.06
'proporcion local cambios alineacion defensa sitio'	-0.06	0.25	-0.22	0.83
'proporcion local cambios alineacion defensa en general'	0.12	0.31	0.39	0.70
'proporcion visitante cambios alineacion defensa en sitio'	0.28	0.26	1.07	0.28
'proporcion visitante cambios alineacion defensa en general'	-0.23	0.30	-0.75	0.45
'proporcion local cambios alineacion centrocampista sitio'	-0.03	0.29	-0.12	0.91
'proporcion local cambios alineacion centrocampista en general'	-0.05	0.31	-0.16	0.87
'proporcion visitante cambios alineacion centrocampista en sitio'	0.15	0.29	0.52	0.60
'proporcion visitante cambios alineacion centrocampista en general'	-0.23	0.39	-0.59	0.56
'proporcion local cambios alineacion delantero sitio'	0.41	0.29	1.42	0.16
'proporcion local cambios alineacion delantero en general'	-0.80	0.32	-2.49	0.01
'proporcion visitante cambios alineacion delantero en sitio'	0.00	0.27	0.01	0.99
'proporcion visitante cambios alineacion delantero en general'	0.04	0.35	0.11	0.91
resultado_visitante	0.06	0.03	2.01	0.04
jornada	0.00	0.01	0.60	0.55
cinco_cambios	0.20	0.41	0.50	0.62

Tabla 3: Resumen de coeficientes del modelo de Poisson para goles locales

### .3. Modelo reducido de goles del equipo local

	Estimate	Std. Error	z value	Pr(> z )
(Intercept)	-1.11	0.68	-1.63	0.10
'porcentaje local mas 1,5 en sitio'	0.00	0.00	1.43	0.15
'porcentaje local mas 3,5 en sitio'	-0.00	0.00	-0.96	0.34
'porcentaje local mas 4,5 en sitio'	0.00	0.00	0.07	0.95
'proporcion visitante cambios en general'	-0.01	0.04	-0.25	0.80
'proporcion local posesion en general'	0.02	0.00	4.88	0.00
'media visitante cambios minutos en general'	0.01	0.01	0.76	0.45
'proporcion local cambios alineacion delantero en general'	-0.18	0.09	-1.93	0.05
resultado_visitante	0.04	0.03	1.48	0.14

Tabla 4: Resumen de coeficientes del modelo de Poisson reducido para goles locales

### .4. Modelo completo de goles del equipo visitante

	Estimate	Std. Error	z value	Pr(> z )
(Intercept)	1584.22	1573.82	1.01	0.31
'porcentaje local ganados en sitio'	-8.78	7.88	-1.11	0.27
'porcentaje local ganados en general'	-7.49	7.01	-1.07	0.28
'porcentaje local empatados en sitio'	-9.05	7.87	-1.15	0.25
'porcentaje local empatados en general'	-6.91	6.98	-0.99	0.32
'porcentaje local perdidos en sitio'	-9.19	7.86	-1.17	0.24
'porcentaje local perdidos en general'	-6.60	6.97	-0.95	0.34
'porcentaje visitante ganados en sitio'	9.93	8.80	1.13	0.26
'porcentaje visitante ganados en general'	-10.25	7.51	-1.37	0.17
'porcentaje visitante empatados en sitio'	9.99	8.76	1.14	0.25
'porcentaje visitante empatados en general'	-10.10	7.47	-1.35	0.18
'porcentaje visitante perdidos en sitio'	10.04	8.74	1.15	0.25
'porcentaje visitante perdidos en general'	-10.03	7.46	-1.34	0.18
'proporcion local puntos en sitio'	-12.68	13.57	-0.93	0.35
'proporcion local puntos en general'	28.04	13.70	2.05	0.04
'proporcion visitante puntos en sitio'	3.71	13.55	0.27	0.78
'proporcion visitante puntos en general'	7.32	13.02	0.56	0.57
'porcentaje local mas 1,5 en sitio'	-0.00	0.01	-0.13	0.89
'porcentaje local mas 1,5 en general'	0.01	0.02	0.55	0.58
'porcentaje visitante mas 1,5 en sitio'	0.01	0.01	0.99	0.32
'porcentaje visitante mas 1,5 en general'	-0.02	0.02	-1.02	0.31
'porcentaje local mas 2,5 en sitio'	0.00	0.01	0.21	0.84
'porcentaje local mas 2,5 en general'	-0.00	0.01	-0.28	0.78
'porcentaje visitante mas 2,5 en sitio'	0.01	0.01	1.19	0.23
'porcentaje visitante mas 2,5 en general'	-0.01	0.01	-0.67	0.50
'porcentaje local mas 3,5 en sitio'	-0.00	0.01	-0.08	0.94
'porcentaje local mas 3,5 en general'	-0.01	0.02	-0.79	0.43
'porcentaje visitante mas 3,5 en sitio'	-0.00	0.01	-0.33	0.74
'porcentaje visitante mas 3,5 en general'	-0.00	0.01	-0.20	0.84
'porcentaje local mas 4,5 en sitio'	0.00	0.01	0.04	0.97
'porcentaje local mas 4,5 en general'	-0.00	0.02	-0.10	0.92
'porcentaje visitante mas 4,5 en sitio'	-0.00	0.01	-0.30	0.76
'porcentaje visitante mas 4,5 en general'	0.01	0.02	0.46	0.65

Tabla 5: Resumen de coeficientes del modelo de Poisson para goles locales

	Estimate	Std. Error	z value	Pr(> z )
'proporcion local goles totales en sitio'	8.25	8.04	1.03	0.31
'proporcion local goles totales en general'	-2.11	7.23	-0.29	0.77
'proporcion local goles marcados en sitio'	-7.87	8.05	-0.98	0.33
'proporcion local goles marcados en general'	0.79	7.21	0.11	0.91
'proporcion local goles encajados en sitio'	-8.65	8.03	-1.08	0.28
'proporcion local goles encajados en general'	3.04	7.24	0.42	0.68
'proporcion visitante goles totales en sitio'	-12.15	8.42	-1.44	0.15
'proporcion visitante goles totales en general'	-6.45	8.16	-0.79	0.43
'proporcion visitante goles marcados en sitio'	11.63	8.39	1.39	0.17
'proporcion visitante goles marcados en general'	7.39	8.12	0.91	0.36
'proporcion visitante goles encajados en sitio'	10.72	8.46	1.27	0.20
'proporcion visitante goles encajados en general'	8.07	8.19	0.99	0.32
'porcentaje local mas 0,5 marcados en sitio'	-0.01	0.01	-0.99	0.32
'porcentaje local mas 0,5 marcados en general'	0.03	0.02	1.84	0.07
'porcentaje local mas 1,5 marcados en sitio'	-0.01	0.01	-0.53	0.59
'porcentaje local mas 1,5 marcados en general'	0.02	0.02	1.21	0.23
'porcentaje local mas 2,5 marcados en sitio'	-0.03	0.02	-2.02	0.04
'porcentaje local mas 2,5 marcados en general'	0.05	0.02	1.96	0.05
'porcentaje local mas 0,5 encajados en sitio'	0.01	0.01	1.17	0.24
'porcentaje local mas 0,5 encajados en general'	-0.03	0.01	-2.20	0.03
'porcentaje local mas 1,5 encajados en sitio'	0.01	0.01	0.55	0.58
'porcentaje local mas 1,5 encajados en general'	-0.01	0.02	-0.63	0.53
'porcentaje local mas 2,5 encajados en sitio'	0.01	0.01	0.97	0.33
'porcentaje local mas 2,5 encajados en general'	-0.02	0.02	-0.78	0.44
'porcentaje visitante mas 0,5 marcados en sitio'	-0.00	0.01	-0.47	0.64
'porcentaje visitante mas 0,5 marcados en general'	-0.00	0.01	-0.35	0.72
'porcentaje visitante mas 1,5 marcados en sitio'	-0.01	0.01	-1.22	0.22
'porcentaje visitante mas 1,5 marcados en general'	0.02	0.01	1.03	0.30
'porcentaje visitante mas 2,5 marcados en sitio'	-0.01	0.01	-0.68	0.50
'porcentaje visitante mas 2,5 marcados en general'	-0.00	0.02	-0.15	0.88
'porcentaje visitante mas 0,5 encajados en sitio'	0.00	0.01	0.45	0.65
'porcentaje visitante mas 0,5 encajados en general'	0.00	0.02	0.29	0.77
'porcentaje visitante mas 1,5 encajados en sitio'	0.01	0.01	1.34	0.18
'porcentaje visitante mas 1,5 encajados en general'	-0.03	0.02	-1.55	0.12
'porcentaje visitante mas 2,5 encajados en sitio'	0.02	0.01	1.09	0.28
'porcentaje visitante mas 2,5 encajados en general'	-0.00	0.02	-0.08	0.94
'proporcion local amarillas en sitio'	-0.13	0.20	-0.69	0.49
'proporcion local amarillas en general'	0.10	0.26	0.38	0.70
'proporcion visitante amarillas en sitio'	-0.13	0.19	-0.66	0.51
'proporcion visitante amarillas en general'	0.08	0.22	0.34	0.73
'proporcion local rojas en sitio'	-0.47	0.61	-0.77	0.44
'proporcion local rojas en general'	0.13	0.84	0.15	0.88
'proporcion visitante rojas en sitio'	-0.64	0.57	-1.13	0.26
'proporcion visitante rojas en general'	0.76	0.88	0.87	0.38
'proporcion local cambios en sitio'	1.25	1.66	0.75	0.45
'proporcion local cambios en general'	-0.84	1.98	-0.42	0.67
'proporcion visitante cambios en sitio'	0.89	1.67	0.53	0.59
'proporcion visitante cambios en general'	-3.63	1.92	-1.89	0.06
'proporcion local posesion en sitio'	-0.01	0.03	-0.38	0.70
'proporcion local posesion en general'	-0.00	0.03	-0.05	0.96
'proporcion visitante posesion en sitio'	0.03	0.03	1.01	0.31
'proporcion visitante posesion en general'	-0.01	0.03	-0.42	0.68
'proporcion local total tiros en sitio'	0.00	0.06	0.02	0.98
'proporcion local total tiros en general'	-0.01	0.08	-0.14	0.89
'proporcion visitante total tiros en sitio'	-0.02	0.06	-0.26	0.80
'proporcion visitante total tiros en general'	0.00	0.07	0.02	0.98
'proporcion local corners a favor en sitio'	-0.12	0.09	-1.37	0.17
'proporcion local corners a favor en general'	0.23	0.13	1.79	0.07
'proporcion visitante corners a favor en sitio'	-0.10	0.09	-1.06	0.29
'proporcion visitante corners a favor en general'	0.17	0.13	1.31	0.19
'proporcion local corners en contra en sitio'	-0.02	0.09	-0.25	0.80
'proporcion local corners en contra en general'	0.18	0.11	1.72	0.09
'proporcion visitante corners en contra en sitio'	-0.01	0.09	-0.14	0.89
'proporcion visitante corners en contra en general'	-0.06	0.12	-0.50	0.62
'proporcion local cambios lesionados sitio'	-0.04	0.40	-0.11	0.91
'proporcion local cambios lesionados en general'	-0.21	0.57	-0.37	0.71
'proporcion visitante cambios lesionados en sitio'	0.65	0.39	1.67	0.09
'proporcion visitante cambios lesionados en general'	-0.65	0.50	-1.30	0.19
'proporcion local cambios amarillas sitio'	0.26	0.34	0.78	0.43
'proporcion local cambios amarillas en general'	-0.14	0.46	-0.32	0.75
'proporcion visitante cambios amarillas en sitio'	0.56	0.32	1.73	0.08
'proporcion visitante cambios amarillas en general'	-1.08	0.44	-2.47	0.01

Tabla 6: Resumen de coeficientes del modelo de Poisson para goles locales

	Estimate	Std. Error	z value	Pr(> z )
'proporcion local cambios goleadores sitio'	-0.29	0.46	-0.63	0.53
'proporcion local cambios goleadores en general'	1.05	0.63	1.67	0.10
'proporcion visitante cambios goleadores en sitio'	-0.18	0.45	-0.41	0.68
'proporcion visitante cambios goleadores en general'	0.15	0.62	0.24	0.81
'proporcion local cambios asistentes sitio'	0.72	0.51	1.41	0.16
'proporcion local cambios asistentes en general'	-1.11	0.71	-1.55	0.12
'proporcion visitante cambios asistentes en sitio'	-0.09	0.51	-0.17	0.87
'proporcion visitante cambios asistentes en general'	0.06	0.65	0.09	0.93
'media local cambios minutos sitio'	-0.01	0.07	-0.22	0.83
'media local cambios minutos en general'	0.07	0.09	0.71	0.48
'media visitante cambios minutos sitio'	0.09	0.06	1.34	0.18
'media visitante cambios minutos en general'	-0.21	0.09	-2.50	0.01
'proporcion local cambios delanteros a centrocampistas sitio'	-0.07	0.40	-0.18	0.86
'proporcion local cambios delanteros a centrocampistas en general'	0.17	0.49	0.34	0.73
'proporcion visitante cambios delanteros a centrocampistas en sitio'	0.35	0.39	0.88	0.38
'proporcion visitante cambios delanteros a centrocampistas en general'	-0.41	0.43	-0.95	0.34
'proporcion local cambios delanteros a defensas sitio'	0.43	0.49	0.87	0.38
'proporcion local cambios delanteros a defensas en general'	-0.12	0.63	-0.19	0.85
'proporcion visitante cambios delanteros a defensas en sitio'	-0.27	0.51	-0.53	0.60
'proporcion visitante cambios delanteros a defensas en general'	0.02	0.59	0.04	0.97
'proporcion local cambios centrocampistas a delanteros sitio'	0.07	0.34	0.21	0.83
'proporcion local cambios centrocampistas a delanteros en general'	0.04	0.39	0.10	0.92
'proporcion visitante cambios centrocampistas a delanteros en sitio'	-0.41	0.35	-1.17	0.24
'proporcion visitante cambios centrocampistas a delanteros en general'	0.42	0.51	0.82	0.41
'proporcion local cambios centrocampistas a defensas sitio'	-0.32	0.56	-0.58	0.56
'proporcion local cambios centrocampistas a defensas en general'	0.88	0.79	1.11	0.27
'proporcion visitante cambios centrocampistas a defensas en sitio'	0.22	0.60	0.37	0.71
'proporcion visitante cambios centrocampistas a defensas en general'	0.23	0.83	0.28	0.78
'proporcion local cambios defensas a delanteros sitio'	0.79	0.50	1.57	0.12
'proporcion local cambios defensas a delanteros en general'	-0.32	0.55	-0.59	0.56
'proporcion visitante cambios defensas a delanteros en sitio'	-0.21	0.51	-0.41	0.68
'proporcion visitante cambios defensas a delanteros en general'	-0.38	0.74	-0.52	0.61
'proporcion local cambios defensas a centrocampistas sitio'	0.09	0.71	0.13	0.90
'proporcion local cambios defensas a centrocampistas en general'	-0.48	0.80	-0.59	0.55
'proporcion visitante cambios defensas a centrocampistas en sitio'	-1.13	0.74	-1.53	0.13
'proporcion visitante cambios defensas a centrocampistas en general'	1.23	0.96	1.29	0.20
'proporcion local cambios antes descanso sitio'	-1.61	1.59	-1.01	0.31
'proporcion local cambios antes descanso en general'	2.94	1.88	1.56	0.12
'proporcion visitante cambios antes descanso en sitio'	-0.22	1.63	-0.13	0.89
'proporcion visitante cambios antes descanso en general'	0.91	1.91	0.48	0.63
'proporcion local cambios 45 a 60 sitio'	-1.55	1.59	-0.97	0.33
'proporcion local cambios 45 a 60 en general'	1.28	1.82	0.70	0.48
'proporcion visitante cambios 45 a 60 en sitio'	-1.04	1.59	-0.66	0.51
'proporcion visitante cambios 45 a 60 en general'	2.98	1.79	1.67	0.09
'proporcion local cambios 61 a 75 sitio'	-0.83	1.65	-0.50	0.61
'proporcion local cambios 61 a 75 en general'	0.57	1.93	0.30	0.77
'proporcion visitante cambios 61 a 75 en sitio'	-0.87	1.65	-0.53	0.60
'proporcion visitante cambios 61 a 75 en general'	3.15	1.90	1.66	0.10
'proporcion local cambios 76 a final sitio'	-0.59	1.77	-0.33	0.74
'proporcion local cambios 76 a final en general'	0.01	2.19	0.00	1.00
'proporcion visitante cambios 76 a final en sitio'	-1.65	1.76	-0.94	0.35
'proporcion visitante cambios 76 a final en general'	4.46	2.00	2.23	0.03
'proporcion local cambios alineacion defensa sitio'	-0.08	0.30	-0.26	0.80
'proporcion local cambios alineacion defensa en general'	0.11	0.35	0.32	0.75
'proporcion visitante cambios alineacion defensa en sitio'	0.31	0.29	1.09	0.28
'proporcion visitante cambios alineacion defensa en general'	0.22	0.35	0.64	0.52
'proporcion local cambios alineacion centrocampista sitio'	0.00	0.35	0.00	1.00
'proporcion local cambios alineacion centrocampista en general'	-0.17	0.38	-0.46	0.64
'proporcion visitante cambios alineacion centrocampista en sitio'	-0.16	0.33	-0.46	0.64
'proporcion visitante cambios alineacion centrocampista en general'	0.04	0.44	0.08	0.93
'proporcion local cambios alineacion delantero sitio'	0.31	0.33	0.94	0.35
'proporcion local cambios alineacion delantero en general'	-0.37	0.35	-1.04	0.30
'proporcion visitante cambios alineacion delantero en sitio'	0.51	0.33	1.56	0.12
'proporcion visitante cambios alineacion delantero en general'	-0.79	0.41	-1.94	0.05
resultado_local	0.07	0.03	2.08	0.04
jornada	0.01	0.01	1.19	0.24
cinco_cambios	0.14	0.45	0.32	0.75

Tabla 7: Resumen de coeficientes del modelo de Poisson para goles visitantes

**.5. Modelo reducido de goles del equipo visitante**

	Estimate	Std. Error	z value	Pr(> z )
(Intercept)	1.27	0.84	1.51	0.13
‘proporcion local puntos en general‘	-0.12	0.13	-0.96	0.34
‘porcentaje local mas 2,5 marcados en sitio‘	-0.00	0.00	-1.12	0.26
‘porcentaje local mas 0,5 encajados en general‘	0.00	0.00	0.77	0.44
‘proporcion visitante cambios amarillas en general‘	-0.48	0.15	-3.24	0.00
‘media visitante cambios minutos en general‘	-0.01	0.01	-1.18	0.24
‘proporcion visitante cambios 76 a final en general‘	0.01	0.10	0.06	0.95
resultado_local	0.05	0.03	1.91	0.06

Tabla 8: Resumen de coeficientes del modelo de Poisson para goles visitantes significativo



# Salidas de Modelos Multinomiales

## .6. Modelo multinomial completo

	Clase	Variable	Estimate	Std..Error	z.value	Pr...z..
1	X	(Intercept)	-0.00	0.00	-1.04	0.30
2	X	'porcentaje local ganados en sitio'	0.01	0.03	0.19	0.85
3	X	'porcentaje local ganados en general'	-0.07	0.03	-2.34	0.02
4	X	'porcentaje local empatados en sitio'	-0.02	0.02	-1.08	0.28
5	X	'porcentaje local empatados en general'	-0.03	0.03	-1.11	0.27
6	X	'porcentaje local perdidos en sitio'	0.01	0.03	0.39	0.70
7	X	'porcentaje local perdidos en general'	-0.04	0.03	-1.20	0.23
8	X	'porcentaje visitante ganados en sitio'	-0.03	0.03	-1.06	0.29
9	X	'porcentaje visitante ganados en general'	0.02	0.03	0.58	0.56
10	X	'porcentaje visitante empatados en sitio'	-0.03	0.02	-1.21	0.23
11	X	'porcentaje visitante empatados en general'	-0.02	0.02	-0.97	0.33
12	X	'porcentaje visitante perdidos en sitio'	0.05	0.03	1.72	0.09
13	X	'porcentaje visitante perdidos en general'	-0.05	0.03	-1.57	0.12
14	X	'proporcion local puntos en sitio'	0.01	0.01	1.48	0.14
15	X	'proporcion local puntos en general'	-0.02	0.01	-2.51	0.01
16	X	'proporcion visitante puntos en sitio'	0.02	0.01	2.01	0.04
17	X	'proporcion visitante puntos en general'	0.04	0.01	5.20	0.00
18	X	'porcentaje local mas 1,5 en sitio'	0.03	0.03	1.13	0.26
19	X	'porcentaje local mas 1,5 en general'	-0.05	0.04	-1.19	0.24
20	X	'porcentaje visitante mas 1,5 en sitio'	0.03	0.03	1.14	0.26
21	X	'porcentaje visitante mas 1,5 en general'	0.04	0.04	1.04	0.30
22	X	'porcentaje local mas 2,5 en sitio'	0.07	0.02	2.75	0.01
23	X	'porcentaje local mas 2,5 en general'	-0.05	0.03	-1.53	0.13
24	X	'porcentaje visitante mas 2,5 en sitio'	0.03	0.02	1.41	0.16
25	X	'porcentaje visitante mas 2,5 en general'	-0.02	0.03	-0.75	0.45
26	X	'porcentaje local mas 3,5 en sitio'	0.03	0.03	1.24	0.22
27	X	'porcentaje local mas 3,5 en general'	-0.07	0.04	-1.64	0.10
28	X	'porcentaje visitante mas 3,5 en sitio'	0.03	0.02	1.26	0.21
29	X	'porcentaje visitante mas 3,5 en general'	-0.03	0.03	-0.76	0.45
30	X	'porcentaje local mas 4,5 en sitio'	0.03	0.03	1.08	0.28
31	X	'porcentaje local mas 4,5 en general'	-0.02	0.04	-0.43	0.67
32	X	'porcentaje visitante mas 4,5 en sitio'	0.02	0.03	0.77	0.44
33	X	'porcentaje visitante mas 4,5 en general'	0.03	0.04	0.72	0.47

Tabla 9: Resumen de coeficientes del modelo Multinomial completo

	Clase	Variable	Estimate	Std..Error	z.value	Pr...z.
34	X	'proporcion local goles totales en sitio'	0.03	0.65	0.04	0.97
35	X	'proporcion local goles totales en general'	0.27	0.58	0.46	0.65
36	X	'proporcion local goles marcados en sitio'	-0.12	0.69	-0.18	0.86
37	X	'proporcion local goles marcados en general'	-0.29	0.39	-0.75	0.46
38	X	'proporcion local goles encajados en sitio'	0.08	0.73	0.11	0.91
39	X	'proporcion local goles encajados en general'	0.54	0.52	1.04	0.30
40	X	'proporcion visitante goles totales en sitio'	-1.51	0.64	-2.35	0.02
41	X	'proporcion visitante goles totales en general'	-1.40	0.39	-3.55	0.00
42	X	'proporcion visitante goles marcados en sitio'	-1.25	0.73	-1.70	0.09
43	X	'proporcion visitante goles marcados en general'	-0.93	0.45	-2.09	0.04
44	X	'proporcion visitante goles encajados en sitio'	-0.25	0.62	-0.39	0.69
45	X	'proporcion visitante goles encajados en general'	-0.39	0.43	-0.91	0.37
46	X	'porcentaje local mas 0,5 marcados en sitio'	-0.03	0.02	-1.32	0.19
47	X	'porcentaje local mas 0,5 marcados en general'	0.05	0.03	1.69	0.09
48	X	'porcentaje local mas 1,5 marcados en sitio'	-0.04	0.02	-1.46	0.14
49	X	'porcentaje local mas 1,5 marcados en general'	0.05	0.03	1.38	0.17
50	X	'porcentaje local mas 2,5 marcados en sitio'	-0.07	0.03	-2.49	0.01
51	X	'porcentaje local mas 2,5 marcados en general'	0.08	0.04	1.93	0.05
52	X	'porcentaje local mas 0,5 encajados en sitio'	-0.01	0.02	-0.56	0.57
53	X	'porcentaje local mas 0,5 encajados en general'	-0.01	0.03	-0.42	0.67
54	X	'porcentaje local mas 1,5 encajados en sitio'	-0.06	0.03	-2.49	0.01
55	X	'porcentaje local mas 1,5 encajados en general'	0.10	0.04	2.58	0.01
56	X	'porcentaje local mas 2,5 encajados en sitio'	-0.04	0.03	-1.22	0.22
57	X	'porcentaje local mas 2,5 encajados en general'	0.01	0.04	0.20	0.84
58	X	'porcentaje visitante mas 0,5 marcados en sitio'	0.02	0.02	0.90	0.37
59	X	'porcentaje visitante mas 0,5 marcados en general'	-0.00	0.03	-0.01	0.99
60	X	'porcentaje visitante mas 1,5 marcados en sitio'	0.02	0.02	0.91	0.36
61	X	'porcentaje visitante mas 1,5 marcados en general'	0.02	0.03	0.43	0.67
62	X	'porcentaje visitante mas 2,5 marcados en sitio'	-0.01	0.03	-0.43	0.67
63	X	'porcentaje visitante mas 2,5 marcados en general'	0.02	0.04	0.61	0.54
64	X	'porcentaje visitante mas 0,5 encajados en sitio'	-0.04	0.02	-2.02	0.04
65	X	'porcentaje visitante mas 0,5 encajados en general'	0.05	0.03	1.82	0.07
66	X	'porcentaje visitante mas 1,5 encajados en sitio'	-0.03	0.02	-1.30	0.19
67	X	'porcentaje visitante mas 1,5 encajados en general'	0.02	0.04	0.63	0.53
68	X	'porcentaje visitante mas 2,5 encajados en sitio'	-0.01	0.03	-0.39	0.69
69	X	'porcentaje visitante mas 2,5 encajados en general'	0.03	0.04	0.89	0.37
70	X	'proporcion local amarillas en sitio'	-0.60	0.41	-1.47	0.14
71	X	'proporcion local amarillas en general'	0.24	0.51	0.46	0.64
72	X	'proporcion visitante amarillas en sitio'	1.25	0.45	2.76	0.01
73	X	'proporcion visitante amarillas en general'	-0.55	0.51	-1.09	0.27
74	X	'proporcion local rojas en sitio'	-1.32	0.79	-1.67	0.10
75	X	'proporcion local rojas en general'	-0.96	0.49	-1.97	0.05
76	X	'proporcion visitante rojas en sitio'	0.35	0.81	0.43	0.67
77	X	'proporcion visitante rojas en general'	0.43	0.50	0.87	0.38
78	X	'proporcion local cambios en sitio'	0.05	0.59	0.09	0.93
79	X	'proporcion local cambios en general'	-0.47	0.60	-0.78	0.43
80	X	'proporcion visitante cambios en sitio'	-0.21	0.56	-0.37	0.71
81	X	'proporcion visitante cambios en general'	0.60	0.63	0.95	0.34
82	X	'proporcion local posesion en sitio'	0.05	0.07	0.71	0.48
83	X	'proporcion local posesion en general'	-0.09	0.07	-1.19	0.24
84	X	'proporcion visitante posesion en sitio'	0.07	0.07	1.04	0.30
85	X	'proporcion visitante posesion en general'	-0.07	0.07	-1.04	0.30
86	X	'proporcion local total tiros en sitio'	-0.28	0.15	-1.91	0.06
87	X	'proporcion local total tiros en general'	0.27	0.19	1.39	0.16
88	X	'proporcion visitante total tiros en sitio'	0.01	0.16	0.05	0.96
89	X	'proporcion visitante total tiros en general'	0.14	0.18	0.79	0.43
90	X	'proporcion local corners a favor en sitio'	0.10	0.22	0.45	0.65
91	X	'proporcion local corners a favor en general'	-0.15	0.29	-0.51	0.61
92	X	'proporcion visitante corners a favor en sitio'	-0.08	0.22	-0.35	0.73
93	X	'proporcion visitante corners a favor en general'	0.30	0.29	1.03	0.30
94	X	'proporcion local corners en contra en sitio'	-0.03	0.22	-0.14	0.89
95	X	'proporcion local corners en contra en general'	-0.49	0.26	-1.84	0.07
96	X	'proporcion visitante corners en contra en sitio'	-0.06	0.21	-0.27	0.79
97	X	'proporcion visitante corners en contra en general'	0.05	0.29	0.15	0.88
98	X	'proporcion local cambios lesionados sitio'	1.22	0.70	1.75	0.08
99	X	'proporcion local cambios lesionados en general'	-1.00	0.82	-1.22	0.22
100	X	'proporcion visitante cambios lesionados en sitio'	0.66	0.71	0.93	0.35
101	X	'proporcion visitante cambios lesionados en general'	0.22	0.84	0.26	0.80
102	X	'proporcion local cambios amarillas sitio'	0.19	0.66	0.30	0.77
103	X	'proporcion local cambios amarillas en general'	1.06	0.83	1.28	0.20
104	X	'proporcion visitante cambios amarillas en sitio'	-0.79	0.63	-1.25	0.21
105	X	'proporcion visitante cambios amarillas en general'	-0.71	0.76	-0.92	0.36

Tabla 10: Resumen de coeficientes del modelo Multinomial completo

	Clase	Variable	Estimate	Std. Error	z. value	Pr...z..
106	X	'proporcion local cambios goleadores sitio'	-0.38	0.74	-0.52	0.60
107	X	'proporcion local cambios goleadores en general'	-0.60	0.79	-0.76	0.45
108	X	'proporcion visitante cambios goleadores en sitio'	1.17	0.64	1.83	0.07
109	X	'proporcion visitante cambios goleadores en general'	0.15	0.65	0.22	0.82
110	X	'proporcion local cambios asistentes sitio'	1.46	0.69	2.12	0.03
111	X	'proporcion local cambios asistentes en general'	-0.17	0.67	-0.25	0.80
112	X	'proporcion visitante cambios asistentes en sitio'	-0.14	0.74	-0.18	0.85
113	X	'proporcion visitante cambios asistentes en general'	0.49	0.72	0.69	0.49
114	X	'media local cambios minutos sitio'	0.01	0.10	0.13	0.90
115	X	'media local cambios minutos en general'	-0.05	0.10	-0.52	0.60
116	X	'media visitante cambios minutos sitio'	0.24	0.10	2.36	0.02
117	X	'media visitante cambios minutos en general'	-0.09	0.11	-0.78	0.44
118	X	'proporcion local cambios delanteros a centrocampistas sitio'	-0.56	0.69	-0.81	0.42
119	X	'proporcion local cambios delanteros a centrocampistas en general'	-0.19	0.79	-0.24	0.81
120	X	'proporcion visitante cambios delanteros a centrocampistas en sitio'	0.53	0.70	0.77	0.44
121	X	'proporcion visitante cambios delanteros a centrocampistas en general'	0.01	0.73	0.02	0.99
122	X	'proporcion local cambios delanteros a defensas sitio'	0.90	0.57	1.59	0.11
123	X	'proporcion local cambios delanteros a defensas en general'	0.26	0.58	0.45	0.65
124	X	'proporcion visitante cambios delanteros a defensas en sitio'	0.04	0.67	0.06	0.95
125	X	'proporcion visitante cambios delanteros a defensas en general'	-0.30	0.75	-0.40	0.69
126	X	'proporcion local cambios centrocampistas a delanteros sitio'	-1.96	0.67	-2.91	0.00
127	X	'proporcion local cambios centrocampistas a delanteros en general'	1.44	0.77	1.86	0.06
128	X	'proporcion visitante cambios centrocampistas a delanteros en sitio'	0.66	0.62	1.06	0.29
129	X	'proporcion visitante cambios centrocampistas a delanteros en general'	-0.50	0.82	-0.60	0.55
130	X	'proporcion local cambios centrocampistas a defensas sitio'	-1.55	0.67	-2.31	0.02
131	X	'proporcion local cambios centrocampistas a defensas en general'	-1.13	0.57	-1.98	0.05
132	X	'proporcion visitante cambios centrocampistas a defensas en sitio'	-0.47	0.67	-0.70	0.48
133	X	'proporcion visitante cambios centrocampistas a defensas en general'	-0.21	0.50	-0.41	0.68
134	X	'proporcion local cambios defensas a delanteros sitio'	-0.74	0.78	-0.95	0.34
135	X	'proporcion local cambios defensas a delanteros en general'	-0.42	0.80	-0.52	0.60
136	X	'proporcion visitante cambios defensas a delanteros en sitio'	0.20	0.56	0.35	0.73
137	X	'proporcion visitante cambios defensas a delanteros en general'	0.06	0.63	0.10	0.92
138	X	'proporcion local cambios defensas a centrocampistas sitio'	-1.24	0.70	-1.77	0.08
139	X	'proporcion local cambios defensas a centrocampistas en general'	1.08	0.53	2.04	0.04
140	X	'proporcion visitante cambios defensas a centrocampistas en sitio'	-0.08	0.64	-0.13	0.90
141	X	'proporcion visitante cambios defensas a centrocampistas en general'	-0.14	0.51	-0.28	0.78
142	X	'proporcion local cambios antes descanso sitio'	0.11	0.77	0.14	0.89
143	X	'proporcion local cambios antes descanso en general'	0.20	0.59	0.34	0.73
144	X	'proporcion visitante cambios antes descanso en sitio'	0.26	0.78	0.33	0.74
145	X	'proporcion visitante cambios antes descanso en general'	0.06	0.59	0.10	0.92
146	X	'proporcion local cambios 45 a 60 sitio'	-0.40	0.52	-0.76	0.45
147	X	'proporcion local cambios 45 a 60 en general'	-0.53	0.66	-0.80	0.42
148	X	'proporcion visitante cambios 45 a 60 en sitio'	-0.03	0.52	-0.06	0.95
149	X	'proporcion visitante cambios 45 a 60 en general'	1.19	0.72	1.67	0.10
150	X	'proporcion local cambios 61 a 75 sitio'	-0.67	0.42	-1.59	0.11
151	X	'proporcion local cambios 61 a 75 en general'	-0.12	0.49	-0.26	0.80
152	X	'proporcion visitante cambios 61 a 75 en sitio'	0.69	0.40	1.75	0.08
153	X	'proporcion visitante cambios 61 a 75 en general'	-1.08	0.53	-2.02	0.04
154	X	'proporcion local cambios 76 a final sitio'	0.37	0.56	0.66	0.51
155	X	'proporcion local cambios 76 a final en general'	-0.61	0.55	-1.12	0.26
156	X	'proporcion visitante cambios 76 a final en sitio'	-0.77	0.55	-1.39	0.16
157	X	'proporcion visitante cambios 76 a final en general'	0.58	0.59	0.99	0.32
158	X	'proporcion local cambios alineacion defensa sitio'	-0.02	0.63	-0.03	0.98
159	X	'proporcion local cambios alineacion defensa en general'	0.21	0.77	0.27	0.79
160	X	'proporcion visitante cambios alineacion defensa en sitio'	0.17	0.64	0.27	0.79
161	X	'proporcion visitante cambios alineacion defensa en general'	0.74	0.70	1.05	0.29
162	X	'proporcion local cambios alineacion centrocampista sitio'	-0.37	0.71	-0.53	0.60
163	X	'proporcion local cambios alineacion centrocampista en general'	1.26	0.76	1.66	0.10
164	X	'proporcion visitante cambios alineacion centrocampista en sitio'	-1.27	0.62	-2.06	0.04
165	X	'proporcion visitante cambios alineacion centrocampista en general'	0.50	0.79	0.63	0.53
166	X	'proporcion local cambios alineacion delantero sitio'	-0.55	0.68	-0.82	0.41
167	X	'proporcion local cambios alineacion delantero en general'	0.89	0.74	1.19	0.23
168	X	'proporcion visitante cambios alineacion delantero en sitio'	0.39	0.66	0.59	0.56
169	X	'proporcion visitante cambios alineacion delantero en general'	-0.08	0.79	-0.10	0.92
170	X	cinco_cambios	0.72	0.89	0.82	0.41
171	2	(Intercept)	0.00	0.00	1.17	0.24
172	2	'porcentaje local ganados en sitio'	0.01	0.03	0.40	0.69
173	2	'porcentaje local ganados en general'	-0.02	0.03	-0.76	0.45
174	2	'porcentaje local empatados en sitio'	0.02	0.02	0.97	0.33
175	2	'porcentaje local empatados en general'	-0.00	0.03	-0.07	0.94
176	2	'porcentaje local perdidos en sitio'	0.01	0.03	0.55	0.58
177	2	'porcentaje local perdidos en general'	-0.00	0.03	-0.02	0.99

Tabla 11: Resumen de coeficientes del modelo Multinomial completo

	Clase	Variable	Estimate	Std..Error	z.value	Pr...z..
178	2	'porcentaje visitante ganados en sitio'	0.04	0.03	1.38	0.17
179	2	'porcentaje visitante ganados en general'	0.04	0.03	1.25	0.21
180	2	'porcentaje visitante empatados en sitio'	0.03	0.02	1.18	0.24
181	2	'porcentaje visitante empatados en general'	0.03	0.03	0.93	0.35
182	2	'porcentaje visitante perdidos en sitio'	0.07	0.03	2.40	0.02
183	2	'porcentaje visitante perdidos en general'	-0.02	0.03	-0.60	0.55
184	2	'proporción local puntos en sitio'	-0.00	0.01	-0.65	0.51
185	2	'proporción local puntos en general'	0.04	0.01	5.15	0.00
186	2	'proporción visitante puntos en sitio'	0.01	0.01	1.46	0.15
187	2	'proporción visitante puntos en general'	-0.06	0.01	-7.59	0.00
188	2	'porcentaje local mas 1,5 en sitio'	0.02	0.03	0.80	0.42
189	2	'porcentaje local mas 1,5 en general'	-0.02	0.04	-0.38	0.70
190	2	'porcentaje visitante mas 1,5 en sitio'	0.04	0.03	1.16	0.24
191	2	'porcentaje visitante mas 1,5 en general'	0.04	0.04	0.92	0.36
192	2	'porcentaje local mas 2,5 en sitio'	0.05	0.03	1.69	0.09
193	2	'porcentaje local mas 2,5 en general'	0.00	0.03	0.08	0.94
194	2	'porcentaje visitante mas 2,5 en sitio'	0.06	0.02	2.25	0.02
195	2	'porcentaje visitante mas 2,5 en general'	-0.04	0.04	-1.01	0.31
196	2	'porcentaje local mas 3,5 en sitio'	0.04	0.03	1.48	0.14
197	2	'porcentaje local mas 3,5 en general'	-0.07	0.04	-1.55	0.12
198	2	'porcentaje visitante mas 3,5 en sitio'	0.02	0.03	0.58	0.56
199	2	'porcentaje visitante mas 3,5 en general'	-0.00	0.04	-0.01	0.99
200	2	'porcentaje local mas 4,5 en sitio'	0.09	0.03	2.82	0.00
201	2	'porcentaje local mas 4,5 en general'	-0.04	0.04	-0.99	0.32
202	2	'porcentaje visitante mas 4,5 en sitio'	0.00	0.03	0.09	0.93
203	2	'porcentaje visitante mas 4,5 en general'	0.09	0.04	2.21	0.03
204	2	'proporción local goles totales en sitio'	-1.17	0.69	-1.68	0.09
205	2	'proporción local goles totales en general'	-1.03	0.49	-2.12	0.03
206	2	'proporción local goles marcados en sitio'	-0.62	0.70	-0.89	0.37
207	2	'proporción local goles marcados en general'	-1.02	0.36	-2.86	0.00
208	2	'proporción local goles encajados en sitio'	-0.50	0.71	-0.70	0.48
209	2	'proporción local goles encajados en general'	-0.02	0.52	-0.04	0.97
210	2	'proporción visitante goles totales en sitio'	-1.45	0.67	-2.16	0.03
211	2	'proporción visitante goles totales en general'	-0.24	0.39	-0.60	0.55
212	2	'proporción visitante goles marcados en sitio'	0.21	0.71	0.30	0.76
213	2	'proporción visitante goles marcados en general'	0.22	0.43	0.52	0.60
214	2	'proporción visitante goles encajados en sitio'	-1.61	0.68	-2.38	0.02
215	2	'proporción visitante goles encajados en general'	-0.52	0.42	-1.24	0.22
216	2	'porcentaje local mas 0,5 marcados en sitio'	0.02	0.02	1.02	0.31
217	2	'porcentaje local mas 0,5 marcados en general'	0.03	0.03	0.93	0.35
218	2	'porcentaje local mas 1,5 marcados en sitio'	-0.04	0.03	-1.31	0.19
219	2	'porcentaje local mas 1,5 marcados en general'	0.06	0.04	1.59	0.11
220	2	'porcentaje local mas 2,5 marcados en sitio'	-0.02	0.03	-0.52	0.60
221	2	'porcentaje local mas 2,5 marcados en general'	0.05	0.04	1.23	0.22
222	2	'porcentaje local mas 0,5 encajados en sitio'	-0.03	0.02	-1.12	0.26
223	2	'porcentaje local mas 0,5 encajados en general'	0.03	0.03	0.90	0.37
224	2	'porcentaje local mas 1,5 encajados en sitio'	-0.03	0.03	-1.01	0.31
225	2	'porcentaje local mas 1,5 encajados en general'	0.04	0.04	1.03	0.30
226	2	'porcentaje local mas 2,5 encajados en sitio'	-0.03	0.03	-0.82	0.41
227	2	'porcentaje local mas 2,5 encajados en general'	0.05	0.05	0.94	0.35
228	2	'porcentaje visitante mas 0,5 marcados en sitio'	-0.01	0.03	-0.55	0.58
229	2	'porcentaje visitante mas 0,5 marcados en general'	-0.02	0.03	-0.75	0.45
230	2	'porcentaje visitante mas 1,5 marcados en sitio'	-0.01	0.03	-0.45	0.65
231	2	'porcentaje visitante mas 1,5 marcados en general'	-0.01	0.04	-0.21	0.84
232	2	'porcentaje visitante mas 2,5 marcados en sitio'	-0.05	0.03	-1.70	0.09
233	2	'porcentaje visitante mas 2,5 marcados en general'	-0.01	0.04	-0.27	0.79
234	2	'porcentaje visitante mas 0,5 encajados en sitio'	-0.02	0.02	-0.76	0.45
235	2	'porcentaje visitante mas 0,5 encajados en general'	0.03	0.03	1.00	0.32
236	2	'porcentaje visitante mas 1,5 encajados en sitio'	-0.00	0.03	-0.14	0.89
237	2	'porcentaje visitante mas 1,5 encajados en general'	-0.02	0.04	-0.47	0.64
238	2	'porcentaje visitante mas 2,5 encajados en sitio'	0.03	0.03	1.07	0.29
239	2	'porcentaje visitante mas 2,5 encajados en general'	-0.01	0.04	-0.25	0.80
240	2	'proporción local amarillas en sitio'	-0.03	0.43	-0.06	0.95
241	2	'proporción local amarillas en general'	-0.95	0.53	-1.80	0.07
242	2	'proporción visitante amarillas en sitio'	0.64	0.47	1.38	0.17
243	2	'proporción visitante amarillas en general'	-0.57	0.53	-1.07	0.28
244	2	'proporción local rojas en sitio'	0.24	0.82	0.29	0.77
245	2	'proporción local rojas en general'	0.71	0.50	1.43	0.15
246	2	'proporción visitante rojas en sitio'	0.51	0.85	0.60	0.55
247	2	'proporción visitante rojas en general'	0.69	0.53	1.32	0.19
248	2	'proporción local cambios en sitio'	0.46	0.64	0.73	0.47
249	2	'proporción local cambios en general'	-0.26	0.63	-0.42	0.67
250	2	'proporción visitante cambios en sitio'	-0.35	0.58	-0.60	0.55
251	2	'proporción visitante cambios en general'	0.40	0.68	0.59	0.56

Tabla 12: Resumen de coeficientes del modelo Multinomial completo

Clase	Variable	Estimate	Std..Error	z.value	Pr...z..
252	'proporción local posesion en sitio'	0.03	0.07	0.36	0.72
253	'proporción local posesion en general'	-0.10	0.08	-1.24	0.21
254	'proporción visitante posesion en sitio'	0.08	0.07	1.11	0.27
255	'proporción visitante posesion en general'	-0.04	0.08	-0.55	0.58
256	'proporción local total tiros en sitio'	-0.07	0.15	-0.44	0.66
257	'proporción local total tiros en general'	0.05	0.21	0.25	0.80
258	'proporción visitante total tiros en sitio'	-0.12	0.17	-0.69	0.49
259	'proporción visitante total tiros en general'	0.18	0.19	0.97	0.33
260	'proporción local corners a favor en sitio'	0.15	0.24	0.63	0.53
261	'proporción local corners a favor en general'	-0.22	0.32	-0.67	0.50
262	'proporción visitante corners a favor en sitio'	-0.14	0.24	-0.58	0.56
263	'proporción visitante corners a favor en general'	0.15	0.32	0.48	0.63
264	'proporción local corners en contra en sitio'	-0.09	0.23	-0.39	0.69
265	'proporción local corners en contra en general'	0.03	0.27	0.11	0.91
266	'proporción visitante corners en contra en sitio'	-0.10	0.24	-0.40	0.69
267	'proporción visitante corners en contra en general'	-0.10	0.33	-0.30	0.76
268	'proporción local cambios lesionados sitio'	-0.47	0.71	-0.67	0.51
269	'proporción local cambios lesionados en general'	0.01	0.82	0.01	0.99
270	'proporción visitante cambios lesionados en sitio'	-0.59	0.72	-0.83	0.41
271	'proporción visitante cambios lesionados en general'	-0.32	0.78	-0.40	0.69
272	'proporción local cambios amarillas sitio'	1.15	0.64	1.79	0.07
273	'proporción local cambios amarillas en general'	-0.15	0.80	-0.19	0.85
274	'proporción visitante cambios amarillas en sitio'	-1.07	0.66	-1.62	0.10
275	'proporción visitante cambios amarillas en general'	-1.12	0.75	-1.49	0.14
276	'proporción local cambios goleadores sitio'	0.31	0.71	0.44	0.66
277	'proporción local cambios goleadores en general'	0.34	0.78	0.44	0.66
278	'proporción visitante cambios goleadores en sitio'	0.12	0.68	0.17	0.86
279	'proporción visitante cambios goleadores en general'	-0.23	0.69	-0.33	0.74
280	'proporción local cambios asistentes sitio'	0.65	0.68	0.96	0.34
281	'proporción local cambios asistentes en general'	-0.65	0.59	-1.11	0.27
282	'proporción visitante cambios asistentes en sitio'	-0.45	0.77	-0.59	0.56
283	'proporción visitante cambios asistentes en general'	-0.61	0.69	-0.88	0.38
284	'media local cambios minutos sitio'	-0.03	0.10	-0.32	0.75
285	'media local cambios minutos en general'	-0.02	0.11	-0.18	0.85
286	'media visitante cambios minutos sitio'	0.14	0.10	1.36	0.17
287	'media visitante cambios minutos en general'	-0.10	0.12	-0.90	0.37
288	'proporción local cambios delanteros a centrocampistas sitio'	-0.05	0.66	-0.08	0.94
289	'proporción local cambios delanteros a centrocampistas en general'	0.12	0.74	0.17	0.87
290	'proporción visitante cambios delanteros a centrocampistas en sitio'	0.25	0.64	0.38	0.70
291	'proporción visitante cambios delanteros a centrocampistas en general'	0.31	0.65	0.47	0.64
292	'proporción local cambios delanteros a defensas sitio'	0.14	0.57	0.25	0.80
293	'proporción local cambios delanteros a defensas en general'	-0.41	0.56	-0.73	0.47
294	'proporción visitante cambios delanteros a defensas en sitio'	-0.56	0.66	-0.85	0.39
295	'proporción visitante cambios delanteros a defensas en general'	0.04	0.68	0.05	0.96
296	'proporción local cambios centrocampistas a delanteros sitio'	0.87	0.74	1.19	0.24
297	'proporción local cambios centrocampistas a delanteros en general'	-0.43	0.81	-0.53	0.60
298	'proporción visitante cambios centrocampistas a delanteros en sitio'	-0.54	0.59	-0.92	0.36
299	'proporción visitante cambios centrocampistas a delanteros en general'	-0.59	0.71	-0.83	0.41
300	'proporción local cambios centrocampistas a defensas sitio'	-0.45	0.73	-0.61	0.54
301	'proporción local cambios centrocampistas a defensas en general'	0.55	0.56	0.98	0.33
302	'proporción visitante cambios centrocampistas a defensas en sitio'	0.67	0.72	0.93	0.35
303	'proporción visitante cambios centrocampistas a defensas en general'	0.00	0.50	0.00	1.00
304	'proporción local cambios defensas a delanteros sitio'	0.26	0.74	0.35	0.73
305	'proporción local cambios defensas a delanteros en general'	-0.57	0.75	-0.76	0.45
306	'proporción visitante cambios defensas a delanteros en sitio'	-0.96	0.55	-1.74	0.08
307	'proporción visitante cambios defensas a delanteros en general'	-0.20	0.53	-0.38	0.71
308	'proporción local cambios defensas a centrocampistas sitio'	0.03	0.75	0.04	0.97
309	'proporción local cambios defensas a centrocampistas en general'	-1.23	0.55	-2.25	0.02
310	'proporción visitante cambios defensas a centrocampistas en sitio'	0.16	0.68	0.24	0.81
311	'proporción visitante cambios defensas a centrocampistas en general'	0.63	0.54	1.15	0.25
312	'proporción local cambios antes descanso sitio'	0.24	0.70	0.34	0.73
313	'proporción local cambios antes descanso en general'	0.23	0.59	0.40	0.69
314	'proporción visitante cambios antes descanso en sitio'	0.97	0.82	1.18	0.24

Tabla 13: Resumen de coeficientes del modelo Multinomial completo

	Clase	Variable	Estimate	Std..Error	z.value	Pr...z..
315	2	'proporcion visitante cambios antes descanso en general'	0.62	0.63	0.98	0.33
316	2	'proporcion local cambios 45 a 60 sitio'	-0.52	0.53	-0.98	0.33
317	2	'proporcion local cambios 45 a 60 en general'	-0.59	0.69	-0.86	0.39
318	2	'proporcion visitante cambios 45 a 60 en sitio'	0.20	0.54	0.37	0.71
319	2	'proporcion visitante cambios 45 a 60 en general'	0.26	0.71	0.37	0.71
320	2	'proporcion local cambios 61 a 75 sitio'	-0.54	0.44	-1.22	0.22
321	2	'proporcion local cambios 61 a 75 en general'	-0.14	0.51	-0.27	0.78
322	2	'proporcion visitante cambios 61 a 75 en sitio'	-0.11	0.44	-0.26	0.80
323	2	'proporcion visitante cambios 61 a 75 en general'	0.47	0.57	0.82	0.41
324	2	'proporcion local cambios 76 a final sitio'	0.30	0.56	0.53	0.60
325	2	'proporcion local cambios 76 a final en general'	-0.46	0.55	-0.82	0.41
326	2	'proporcion visitante cambios 76 a final en sitio'	-0.44	0.56	-0.79	0.43
327	2	'proporcion visitante cambios 76 a final en general'	0.24	0.62	0.39	0.70
328	2	'proporcion local cambios alineacion defensa sitio'	0.56	0.66	0.84	0.40
329	2	'proporcion local cambios alineacion defensa en general'	-0.67	0.79	-0.85	0.40
330	2	'proporcion visitante cambios alineacion defensa en sitio'	0.57	0.64	0.89	0.37
331	2	'proporcion visitante cambios alineacion defensa en general'	0.98	0.72	1.38	0.17
332	2	'proporcion local cambios alineacion centrocampista sitio'	0.18	0.72	0.25	0.80
333	2	'proporcion local cambios alineacion centrocampista en general'	-0.31	0.77	-0.40	0.69
334	2	'proporcion visitante cambios alineacion centrocampista en sitio'	-0.12	0.58	-0.21	0.83
335	2	'proporcion visitante cambios alineacion centrocampista en general'	-0.29	0.75	-0.38	0.70
336	2	'proporcion local cambios alineacion delantero sitio'	0.23	0.73	0.32	0.75
337	2	'proporcion local cambios alineacion delantero en general'	-0.03	0.78	-0.03	0.97
338	2	'proporcion visitante cambios alineacion delantero en sitio'	0.07	0.67	0.11	0.92
339	2	'proporcion visitante cambios alineacion delantero en general'	-0.30	0.82	-0.37	0.71
340	2	cinco_cambios	0.52	0.89	0.59	0.56

Tabla 14: Resumen de coeficientes del modelo Multinomial completo

**.7. Modelo multinomial reducido**

	Clase	Variable	Estimate	Std..Error	z.value	Pr...z...
1	X	(Intercept)	-5.48	2.57	-2.13	0.03
2	X	'porcentaje local ganados en general'	-0.02	0.03	-0.82	0.41
3	X	'porcentaje visitante perdidos en sitio'	0.01	0.01	0.61	0.54
4	X	'proporción local puntos en general'	0.76	1.29	0.59	0.55
5	X	'proporción visitante puntos en sitio'	0.10	0.64	0.16	0.87
6	X	'proporción visitante puntos en general'	0.73	0.81	0.90	0.37
7	X	'porcentaje local mas 2,5 en sitio'	0.02	0.01	2.29	0.02
8	X	'porcentaje visitante mas 2,5 en sitio'	-0.00	0.01	-0.31	0.76
9	X	'porcentaje local mas 4,5 en sitio'	0.01	0.01	1.14	0.25
10	X	'porcentaje visitante mas 4,5 en general'	-0.02	0.02	-1.15	0.25
11	X	'proporción local goles totales en general'	-0.63	0.71	-0.89	0.38
12	X	'proporción local goles marcados en general'	-0.46	1.03	-0.45	0.65
13	X	'proporción visitante goles totales en sitio'	0.02	0.63	0.04	0.97
14	X	'proporción visitante goles totales en general'	0.15	0.74	0.20	0.84
15	X	'proporción visitante goles marcados en general'	0.18	1.32	0.13	0.89
16	X	'proporción visitante goles encajados en sitio'	0.16	0.98	0.16	0.87
17	X	'porcentaje local mas 2,5 marcados en sitio'	-0.02	0.01	-1.57	0.12
18	X	'porcentaje local mas 1,5 encajados en sitio'	-0.02	0.01	-1.81	0.07
19	X	'porcentaje local mas 1,5 encajados en general'	0.04	0.02	1.93	0.05
20	X	'porcentaje visitante mas 0,5 encajados en sitio'	-0.00	0.01	-0.47	0.64
21	X	'proporción visitante amarillas en sitio'	0.27	0.14	1.85	0.06
22	X	'proporción local rojas en general'	-1.29	1.05	-1.23	0.22
23	X	'proporción local cambios asistentes sitio'	0.82	0.50	1.64	0.10
24	X	'media visitante cambios minutos sitio'	0.05	0.02	2.08	0.04
25	X	'proporción local cambios centrocampistas a delanteros sitio'	-0.33	0.35	-0.94	0.35
26	X	'proporción local cambios centrocampistas a defensas sitio'	-0.98	0.93	-1.05	0.29
27	X	'proporción local cambios centrocampistas a defensas en general'	0.13	1.18	0.11	0.92
28	X	'proporción local cambios defensas a centrocampistas en general'	0.40	0.81	0.50	0.62
29	X	'proporción visitante cambios 61 a 75 en general'	0.21	0.25	0.83	0.40
30	X	'proporción visitante cambios alineación centrocampista en sitio'	-0.23	0.25	-0.93	0.35

Tabla 15: Resumen de coeficientes del modelo Multinomial reducido

	Clase	Variable	Estimate	Std..Error	z.value	Pr...z..
31	2	(Intercept)	1.28	2.66	0.48	0.63
32	2	'porcentaje local ganados en general'	-0.04	0.03	-1.38	0.17
33	2	'porcentaje visitante perdidos en sitio'	0.00	0.01	0.28	0.78
34	2	'proporcion local puntos en general'	1.25	1.39	0.90	0.37
35	2	'proporcion visitante puntos en sitio'	0.14	0.68	0.21	0.83
36	2	'proporcion visitante puntos en general'	0.67	0.86	0.78	0.44
37	2	'porcentaje local mas 2,5 en sitio'	0.01	0.01	1.64	0.10
38	2	'porcentaje visitante mas 2,5 en sitio'	0.02	0.01	1.71	0.09
39	2	'porcentaje local mas 4,5 en sitio'	0.03	0.01	1.90	0.06
40	2	'porcentaje visitante mas 4,5 en general'	0.03	0.02	1.40	0.16
41	2	'proporcion local goles totales en general'	-0.75	0.77	-0.97	0.33
42	2	'proporcion local goles marcados en general'	-0.44	1.13	-0.39	0.70
43	2	'proporcion visitante goles totales en sitio'	-1.01	0.66	-1.53	0.13
44	2	'proporcion visitante goles totales en general'	-0.81	0.81	-1.01	0.31
45	2	'proporcion visitante goles marcados en general'	1.34	1.40	0.96	0.34
46	2	'proporcion visitante goles encajados en sitio'	0.72	1.04	0.70	0.49
47	2	'porcentaje local mas 2,5 marcados en sitio'	-0.02	0.01	-1.43	0.15
48	2	'porcentaje local mas 1,5 encajados en sitio'	-0.03	0.01	-2.46	0.01
49	2	'porcentaje local mas 1,5 encajados en general'	0.04	0.02	2.12	0.03
50	2	'porcentaje visitante mas 0,5 encajados en sitio'	-0.00	0.01	-0.52	0.60
51	2	'proporcion visitante amarillas en sitio'	-0.08	0.16	-0.48	0.63
52	2	'proporcion local rojas en general'	-0.86	1.09	-0.78	0.43
53	2	'proporcion local cambios asistentes sitio'	0.78	0.54	1.43	0.15
54	2	'media visitante cambios minutos sitio'	-0.02	0.02	-0.65	0.52
55	2	'proporcion local cambios centrocampistas a delanteros sitio'	-0.04	0.37	-0.10	0.92
56	2	'proporcion local cambios centrocampistas a defensas sitio'	-2.13	1.00	-2.14	0.03
57	2	'proporcion local cambios centrocampistas a defensas en general'	2.09	1.23	1.70	0.09
58	2	'proporcion local cambios defensas a centrocampistas en general'	-0.36	0.87	-0.42	0.68
59	2	'proporcion visitante cambios 61 a 75 en general'	0.13	0.27	0.49	0.63
60	2	'proporcion visitante cambios alineacion centrocampista en sitio'	-0.10	0.26	-0.37	0.71

Tabla 16: Resumen de coeficientes del modelo Multinomial reducido