

Estimación de resultados de encuentros de la liga Santander haciendo uso de un modelo lineal generalizado mixto

Joan Sebastian Wilches Castro ^a
joanwilches@usantotomas.edu.co

Wilmer Pineda Rios^b
wilmerpineda@usantotomas.edu.co

Resumen

El presente artículo propone hacer uso de un modelo Poisson mixto con el fin de estimar los resultados de los partidos de fútbol de la liga española y entender los efectos que conlleva jugar contra los diferentes equipos que componen la liga. Como variables regresoras se hace uso de las estadísticas de los encuentros (posesión del balón de los equipos, cantidad de tiros de esquina, tarjetas, faltas, entre otros) desde el año 2016 a la actualidad, adicional a esto, mediante análisis de componente principales se construye indicadores de poder ofensivo y de poder defensivo, con las características de los jugadores que militan actualmente en los equipos, gracias a la base del juego FIFA 19 de Ea sports, estos indicadores funcionan como covariables en el modelo. En el trabajo se ilustra como la cantidad de goles que puede marcar uno equipo cambia según sus características y las del rival.

Palabras clave Modelo Poisson, Modelos mixtos, Fútbol, análisis de componentes principales.

Abstract

This article proposes to use a mixed Poisson model in order to estimate the results of Spain soccer league and understand the effects of playing against the different teams that make up the league. The covariates are the statistics of the matches (possession of the ball, number of corner shots, cards, fouls, among others) from 2016 to the present, in addition to this, construction of indicators through Principal component analysis, Indicators of offensive power and defensive power, with the characteristics of the players currently playing in the teams, thanks to the base of the FIFA 19 game of Ea sports, these indicators function as covariates in the model. The work illustrates how the number of goals can change according to their characteristics and those of the opponent.

Keywords Poisson Models, Mix Models, Soccer, Principal component analysis.

1. Introducción

El trabajo desarrollado a continuación está dividido en secciones, donde en primer lugar se encuentra la introducción y un abre bocas de los trabajos ya desarrollados con la misma finalidad, seguido por lo que es el fútbol, una tercera sección de las apuestas deportivas, la cuarta sección con una breve descripción de lo que es la liga Santander ya que es el torneo base del desarrollo del modelo y una última sección con la teoría del mismo, como lo es:

^aEstudiante pregrado Estadística, U. Santo Tomás, sede Bogotá

^bDocente Facultad de Estadística, U. Santo Tomás, sede Bogotá

- Modelos lineales Generalizados
- Distribución y regresión Poisson
- Modelo lineal Mixto
- Modelo Mixto Poisson

En el final del documentó se encuentra los resultados y discusión de los mismos.

El propósito de este trabajo es estimar los resultados de los partidos de fútbol de la liga Santander mediante un modelo Poisson mixto, teniendo en cuenta como efecto aleatorio el equipo contra el que se juega el partido (equipo visitante), resultados anteriores e indicadores de potencial de ataque, defensa, entre otros. Calculados mediante métodos multivariados. Esto para robustecer las diferentes estimaciones que se hallaron.

Para poder empezar con el desarrollo, se dejan en claro algunos conceptos y un contexto general de lo que es la actualidad del fútbol. La finalidad del deporte, es marcar en la puerta del equipo contrario y evitar que marquen en propia puerta, para así conseguir el triunfo. Esto es dado gracias a estrategias de juego, contratar jugadores con características superiores que la de los contrincantes, entre otros.

Hoy en día este deporte es uno de los negocios más grandes del mundo, por su popularidad, se asegura que llega a mover hasta 500.000 millones de dólares anuales mediante la venta de boletería, transferencias entre equipos por los jugadores y premios de los diferentes torneos a nivel mundial. Esto sin contar los ingresos de los jugadores por publicidad y diferentes patrocinios. Lo que ha llevado a que se habrán una gran cantidad de apuestas distintas por cada partido que hay en el mundo. Rojas Diusabá Victor,1/1/2018.El fútbol es el negocio del siglo. Semana

En la actualidad la predicción de eventos deportivos está en su auge ya que es algo que intenta determinar las personas del común hasta grandes equipos de investigación, todos con el mismo fin, pronosticar el ganador de un partido, campeón de un torneo, entre otras cosas.

Las personas del común ven esto como una oportunidad de ganar dinero de manera rápida y un pasatiempo, los grandes grupos de investigación intentan determinar esto, en la mayoría de casos para lograr establecer las cuotas de pago en las diferentes casas de apuestas o publicar resultados de un modelo el cual ayude a la comunidad a ganar dinero.

Hace años se han realizado diferentes intentos de estimación con múltiples técnicas y ajustes de las mismas, para tener estimaciones precisas como Dyte,D.y Clarke, S.(2000) quienes proponen un método para predecir la distribución de los marcadores en los partidos de fútbol internacionales, tratando los goles de cada equipo, como variables independientes Poisson que dependen de la calificación de la Federación Internacional de Fútbol (FIFA) para cada equipo y el lugar del partido.

Otro de los diferentes desarrollos fue Harvard, R., Oyvind, S.(1997) quienes sugirieron una dinámica bayesiana, de un modelo lineal generalizado para estimar las habilidades dependientes del tiempo de todos los equipos en una liga, y predecir los partidos de fútbol del siguiente fin de semana.

Adicional a este Goddard, J., Asimakopoulos, I.(2002) hicieron uso de un modelo de regresión pro-bit ordenado, utilizando datos de 15 años para modelar los Resultados de partidos de fútbol de la liga inglesa. mediante resultados de partidos pasados, la importancia del partido para los resultados de la liga al final de la temporada; la participación de los equipos en la competición de copa; la distancia geográfica entre las ciudades de origen de los dos equipos; y la asistencia promedio de los dos equipos.

Como ultima referencia Karlis y Ntzoufras (2003) estimarán la cantidad de goles marcados por un equipo mediante una distribución de Poisson bivariada aplicada a los resultados de partidos de fútbol.Para esto hicieron uso de variables tales como la localia, el rendimiento de defensa y ataque en situación de localia y visita.

Con lo dicho anteriormente, se puede concluir que hay un sin número de trabajos enfocados a estimar el marcador de un partido, pero se encuentra un común denominador en cada uno de ellos, no se tiene en cuenta el equipo contra el que se juega, por esto, la metodología del presente documento mantiene la línea de una estimación mediante la distribución Poisson añadiéndole la componente mixta, que permite hacer estas mismas estimaciones, teniendo en cuenta el equipo contra el que se enfrenta.

2. Fútbol

El fútbol data del año 1863 con la separación del rugby-football en Inglaterra y se dio lugar a la primera asociación propia del deporte, llamada "Football Association", es un deporte de competencia conocido a nivel mundial y show principal del evento con mayor asistencia mundial.

Es jugado por dos equipos, compuestos por 11 jugadores iniciales en un césped de medidas variables de aproximadamente cien metros de largo por 75 metros de ancho, al momento de iniciar el encuentro, los jugadores se ubican estratégicamente en sectores del campo, lo cual se determina como posiciones, que permite aplicar diferentes metodologías de juego, los jugadores NO son ubicados aleatoriamente en las posiciones ya que deben tener características especiales en cada una de ellas, las características serán descritas a continuación.

Portero: Hoy en día si hay algo inexcusable en un portero, es que debe manejar el juego aéreo y dominar el juego con el pie, a partir de ahí, pueden entrarse a valorar si es más importante la agilidad o la flexibilidad, atendiendo a las características físicas, el dominio técnico o táctico de la situación de portería, etc.

Laterales defensivos: Hay laterales que forman parte de una línea de 4, que neutralizan el juego rival de su banda, ayudan a los centrales en las coberturas y limitan su participación a la faceta defensiva, son laterales que suelen ir bien también en el juego aéreo.

Laterales ofensivos: Existen laterales profundos, con recorrido, cuya función es doblar en ataque y desdoblarse en la fase defensiva, con gran actividad y que suelen ser pareja de baile de otro jugador de banda, que bien intercala, dentro o fuera, intentando no solapar los pasillos verticales para ganar más líneas de amplitud. Hay laterales que si bien son laterales, se incrustan en medio campo y los puedes ver evolucionar desde dentro o fuera, según intereses.

Centrales: Antes el alto era el central, ahora el listo es el central, ya que es muy importante el trabajo que realizan en la organización defensiva posicional y en la reorganización defensiva en transiciones, también es importante que, o bien manejen el espacio tácticamente hablando, o sean rápidos para subsanar lagunas tácticas, ya que la mayoría de futbolistas a neutralizar, son rápidos y listos.

Medios Centro: Aquí si que puedes encontrar de todo en el fútbol actual y que cada uno desde su faceta, sobreviven en el fútbol profesional. Así, tienes el recuperador, el organizador, el que dota de equilibrio táctico al equipo, el que maneja el ritmo del juego de tu equipo, el llegador, el replegador, que siempre llega a la ayuda, infinidad de aspectos, que hace que este puesto, tenga las exigencias de un futbolista completo. No obstante, hay medios centro que solo viven de su técnica exquisita, hay otros de su visión táctica, hay otros de sus condiciones físicas y muchos de su hegemonía para liderar el juego colectivo del equipo.

Interiores: Este puesto, al igual que los otros, también ha evolucionado y hoy en día, es muy frecuente verlos jugar a pierna cambiada, abandonar su posición de banda y venirse al centro, perder amplitud, para que la ganen otros o ganarla ellos, para que la pierdan otros. Utilizan movimientos verticales para dar profundidad y movimientos diagonales para buscar finalización en dirección al área, también pueden

formar parte de una línea de 3 atacante, que permuta posiciones. También puede ser un futbolista posicional, que busque el juego de 1 contra 1, 2 contra 1, provocando superioridad en los duelos en igualdad. Muchas veces su efectividad vendrá marcada por la sincronización con el futbolista que juegue en su banda y establezca el juego de pares, provocando que la táctica grupal, favorezca a la colectiva.

Delanteros: existen dos tipos, generales y otros secundarios, generales: 1) El futbolista que le gusta moverse al espacio de la espalda rival, buscando desmarques de ruptura y 2) El futbolista que le gusta moverse para fijar, atraer rivales y balones al pie, este segundo le gusta vivir cerca del área y el otro sale de esos espacios para llegar a ellos, el segundo está, mientras el primero aparece. Los secundarios son aquellos que están relacionados con el sistema, que acostumbran a ser segundo delantero, a jugar entre líneas, a caer a banda, a jugar como falso 9, innumerables aspectos que condicionan que el delantero sea multifactorial, pero siempre buscando el fin último, que es el gol.(Fútbol Ofensivo, 2016).

3. Apuestas Deportivas

Desde que se aprobó la Ley del juego en 2012 en España el negocio de las apuestas deportivas no ha parado de crecer. En el año 2018 las apuestas deportivas virtuales contaban con casi un millón y medio de jugadores activos y acaparaban más de la mitad de los ingresos del juego online, con 365 millones de euros descontados premios y reapuestas, lo que en la jerga se conoce como margen de juego o GGR (gross gaming revenue). Según la Dirección General de Ordenación del Juego (DGOJ), el primer trimestre de 2019 tampoco defraudó: el segmento avanzó un 26% con respecto al mismo periodo de 2018, con un margen de 100 millones.(Diario de Pontevedra,25 de Abril del 2018).

3.1. Tipos de apuestas de Fútbol

Apuesta simple: el apostador realiza una sola apuesta a un evento de su elección, la apuesta sencilla más común es el 1X2, es decir, apostar a quien va a ser el ganador del partido o si este terminará en empate, aunque en la actualidad se pueden realizar apuestas de todo tipo como, por ejemplo: qué equipo marcará primero, cuál será el resultado al final del primer tiempo o qué jugador marcará el primer gol.

Apuesta combinada: estas realizan varias predicciones sencillas y las cuotas de cada una de ellas se multiplican entre sí, es una de las opciones preferidas por los apostadores, lo más normal es que se elija 3 o más eventos que tengan un favorito claro, de este modo se minimiza el riesgo de que pierdas lo apostado.

Apuesta Live: son apuestas sencillas que se realizan mientras se está jugando un partido de fútbol, si un equipo se pone por delante es muy común que se realicen apuestas del tipo: el partido acabará en empate, no se marcarán más goles o el equipo que ha marcado primero se llevará la victoria, también se conoce las apuestas Live como combinadas con la salvedad de que no tienes por qué acertar todas las predicciones para ganar.

Apuesta de sistema(2/3): en ella el apostante realiza dos predicciones distintas sobre 3 eventos diferentes, la cantidad apostada se divide entre las tres apuestas y con solo acertar dos de los tres resultados se obtienen ganancias.

Apuestas a largo plazo: son aquellas que se realizan mucho antes de conocerse el resultado. Las apuestas más clásicas son predecir qué equipo se alzará con el título de liga o de Champions League ese año, aunque también se puede apostar a otro tipo de sucesos como qué equipo quedara segundo o cuáles serán los equipos que descenderán de categoría.(Diario de Pontevedra,25 de Abril del 2018)

4. La Liga Santander

La Primera División de España conocida como La Liga Santander y oficialmente como Campeonato Nacional de Liga de Primera División, es la máxima categoría masculina del sistema de ligas de España y la principal competición a nivel de clubes del país. La organiza desde 1984 la Liga de Fútbol Profesional (LFP) tras suceder a la Real Federación Española de Fútbol (RFEF). Empezó a disputarse en 1929 y su último campeón es el Fútbol Club Barcelona, fue interrumpida únicamente entre 1936 y 1939, debido a la Guerra Civil Española.

La liga es un torneo organizado y regulado (conjuntamente con la Segunda División) por la Liga Nacional de Fútbol Profesional (LFP), cuyos miembros son los propios clubes participantes. La competición se disputa anualmente, empezando a finales del mes de agosto o principios de septiembre, y terminando en el mes de mayo o junio del siguiente año.

El torneo consta de un grupo único integrado por veinte equipos, pertenecientes a clubes de fútbol o sociedades anónimas deportivas (SAD). Siguiendo un sistema de liga, los veinte equipos se enfrentan todos contra todos en dos ocasiones -una en campo propio y otra en campo contrario- sumando un total de 38 jornadas. El orden de los encuentros se decide por sorteo antes de empezar la competición. (Planeta fútbol, 2010)

5. Métodos

5.1. Modelo Lineal Generalizado (MLG)

Un modelo lineal generalizado (MLG), en su esencia son una extensión de los modelos lineales, solo que estos permiten el uso de distribuciones diferentes a la normal en las variables respuestas, como Binomiales, Poisson, Gamma y otras distribuciones pertenecientes a la familia exponencial. El MLG generaliza la regresión lineal al permitir que el modelo lineal esté relacionado con la variable de respuesta a través de una función de enlace y al permitir que la magnitud de la varianza de cada medición sea una función de su valor predicho. (Modelo lineal generalizado, s.f)

John Nelder y Robert Wedderburn (1972) formularon modelos lineales generalizados como una forma de unificar otros modelos estadísticos, como la regresión lineal, la regresión logística y la regresión de Poisson. Propusieron un método de mínimos cuadrados iterativamente ponderados para la estimación de máxima verosimilitud de los parámetros del modelo

Un MLG se compone por tres partes:

- **Componente Aleatorio:** Esta parte del modelo es la que identifica la variable respuesta y la distribución de probabilidad a la que esta asociada, la cual sera de la familia exponencial tales como normal, lognormal, Poisson, Gamma, etc
- **Componente sistemática:** Este compone las variables explicativas (independientes o predictoras) utilizadas en la función predictora lineal, también es conocido como el predictor lineal y generalmente se denota con η y corresponde al vector de n componentes.
- **Función de enlace:** La función de enlace proporciona la relación entre el predictor lineal y la

media de la función de distribución, siempre la función de enlace debe ser monótona y doblemente diferenciable.

Teóricamente un modelo lineal generalizado se puede expresar por medio de:

$$\mathbb{E}(\mathbf{Y}) = \mu = g^{-1}(\mathbf{X}\beta) \quad (1)$$

Donde :

$\mathbb{E}(\mathbf{Y})$ es el valor esperado de \mathbf{Y}

$(\mathbf{X}\beta)$ es el predictor lineal, o una combinación lineal de parámetros desconocidos β

g , es la función de enlace.

Existen varios tipos de modelos generalizados que dependen como ya se menciono anteriormente de la distribución de probabilidad de la variable respuesta, para efectos de el desarrollo de este trabajo el foco central esta en el modelo Poisson.

5.2. Distribución Poisson

En 1837, Poisson publicó la distribución que lleva su nombre, obteniéndola como una distribución límite de la distribución binomial. Si la variable aleatoria Y tiene una distribución de Poisson, $P(\cdot)$, con el parámetro $\mu > 0$, entonces su función de probabilidad se expresa como:

$$f(y; \mu) = \frac{e^{-\mu} \mu^y}{y!} \quad (2)$$

Su valor esperado y varianza dado x esta representado por:

$$\lambda = \mathbf{E}(y|x) = e^{x'\beta} \quad (3)$$

El modelo de Poisson juega un papel importante en el análisis de datos en forma de recuentos. Sus características principales son:

- Proporciona, en general, una descripción satisfactoria de datos experimentales cuya varianza es proporcional a la media.
- Puede deducirse teóricamente de principios elementales con un número mínimo de restricciones.
- Si los eventos ocurren de manera independiente y aleatoria a lo largo del tiempo, con una tasa promedio constante de ocurrencia, el modelo determina el número de eventos durante un intervalo de tiempo especificado.

El modelo de regresión de Poisson juega el mismo papel en el análisis de datos categóricos que en el modelo normal de análisis continuo. La diferencia fundamental es que la estructura multiplicativa para el modelo de Poisson es más apropiado que la estructura aditiva de los medios del modelo normal. Se ha encontrado en el análisis de datos categóricos que la media μ se obtiene generalmente como producto de otros medios marginales que se convierten en los parámetros lineales del modelo. (Cordeiro, G, 2008)

5.3. Modelo Poisson

Este modelo se utiliza cuando la variable respuesta es una cantidad discreta que se puede modelizar, suele representar el recuento de sucesos o hechos por ejemplo, el número de goles marcados en un partido que es la variable de respuesta de este trabajo.

La regresión de Poisson asume:

- Y tiene una distribución Poisson
- $\log(\mathbf{E}(y|x)) = x'\beta$ el modelo de regresión de Poisson se refiere a veces como el modelo log-lineal
- La probabilidad viene dada por:

$$Pr(y|x) = \frac{e^{yx'\beta} e^{-e^{x'\beta}}}{y!} \quad (4)$$

- La log-verosimilitud es:

$$L(b) = \sum_{i=1}^N y_i x_i' b - e^{x_i' b} - \log(y_i!) \quad (5)$$

por lo tanto,

$$\mathbf{E}(y|x) = e^{x'\beta} = e^{\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k} \quad (6)$$

Un cambio de una unidad en el regresor j conduce a un cambio en la media condicional de β_j x $\mathbf{E}(y_i|x_i)$, Otra forma de decir esto es que un cambio de una unidad en el regresor j conlleva un cambio proporcional de $\mathbf{E}(y_i|x_i)$ de β_j y si el regresor está en logaritmos entonces β_j es una elasticidad. (Mora. 2018. Poisson Regression)

5.4. Modelo Lineal Mixto

Torres-Saavedra (2018) afirma que un modelo mixto permite analizar una variable aleatoria Y modelizando simultáneamente el valor esperado del fenómeno estudiado y su variabilidad, es decir se modela un efecto fijo a todos los sujetos de estudio y otro efecto aleatorio asociado a cada uno de los sujetos el cual esta representado por la ecuación 3:

$$Y_{ij} = (\beta_0 + \beta_1 t_{ij}) + (b_{i0} + b_{1i} t_{ij}) + \varepsilon_{ij}. \quad (7)$$

y matricialmente esta representado por:

$$\mathbf{Y}_i = \mathbf{X}_i \beta + \mathbf{Z}_i \mathbf{b}_i + \varepsilon_i \quad (8)$$

Los supuestos que se establecen para este modelo son:

- $\mathbf{b}_i \sim N_2(\mathbf{0}, D)$
- $\varepsilon_{ij} \sim N_2(\mathbf{0}, \sigma^2 \mathbf{V}_i)$
- La matriz D tiene dimensión 2×2 mientras que V_i es de dimensión 5×5

5.4.1. Estimación en un MLM

En un MLM se puede obtener la distribución conjunta de (Y_i, b_i) en términos del producto de distribuciones conocidas. Esto es,

$$f_y(\mathbf{y}_i) = \int f(\mathbf{y}_i, \mathbf{b}_i) d\mathbf{b}_i = \int f_{y|b}(\mathbf{y}_i|\mathbf{b}_i) f_b(\mathbf{b}_i) d\mathbf{b}_i \quad (9)$$

Por lo tanto, la densidad marginal para los datos usando un MLM se obtiene integrando los efectos aleatorios

$$f(\mathbf{y}_i, \mathbf{b}_i) = f_{y|b}(\mathbf{y}_i|\mathbf{b}_i) f_b(\mathbf{b}_i) = \text{dist. condicional} \times \text{dist. efectos aleatorios.} \quad (10)$$

Esta distribución esta dada por:

$$\mathbf{Y}_i \sim N_5(\mathbf{X}_i\beta, \Sigma_i) \quad (11)$$

donde $\Sigma_i = \mathbf{Z}_i D \mathbf{Z}_i^t + \sigma^2 \mathbf{V}_i$. La función de verosimilitud para los datos bajo este modelo se puede construir tal como se discutió en los modelos marginales. Por lo tanto, los métodos ML y REML se pueden usar para estimar los parámetros en un LML con sus respectivas propiedades.

5.4.2. Predicción de los Efectos Aleatorios

En un MLM las componentes de \mathbf{b}_i son los efectos aleatorios. Estas cantidades se pueden predecir de acuerdo a la esperanza condicional del \mathbf{b}_i dado \mathbf{y}_i , la cual toma la siguiente forma

$$\hat{\mathbf{b}}_i = \hat{\mathbf{D}} \mathbf{Z}_i^t \hat{\Sigma}_i^{-1} (\mathbf{y}_i - \mathbf{X}_i \hat{\beta}) \quad (12)$$

Estas cantidades se conocen como las estimaciones empíricas de Bayes.

5.4.3. Residuales

En un MLM se pueden obtener dos tipos de residuales: condicionales y marginales. Los residuales condicionales estan dados por

$$\hat{\varepsilon}_{(c)i} = \mathbf{Y}_i - \mathbf{X}_i \hat{\beta} - \mathbf{Z}_i \hat{\mathbf{b}}_i \quad (13)$$

mientras que los marginales se definen como

$$\hat{\varepsilon}_{(m)i} = \mathbf{Y}_i - \mathbf{X}_i \hat{\beta}. \quad (14)$$

Estos residuales crudos se pueden usar para detectar varianza marginal o condicional no constante. Sin embargo, normalidad y detección de datos atípicos usando estos residuales o versiones escaladas no son apropiados debido a la correlación entre los residuales.

5.5. Modelo lineal mixto tipo Poisson

(Correa y Salazar.2006) sustentan que los modelos lineales mixtos generalizados se basan en especificar el modelo en términos de la distribución de los datos, de la función de la media y expresar esta media en función de alguna transformación de $\mathbf{X}\beta + \mathbf{Z}\mathbf{b}$. La distribución se toma como un miembro de la

familia exponencial y la relación entre la media $\mu = \mathbf{E}(Y)$ y el predictor lineal $\mathbf{X}\beta + \mathbf{Zb}$ debe ser una transformación monótona. Es común denotar este predictor lineal como:

$$\eta = \mathbf{X}\beta + \mathbf{Zb} \tag{15}$$

y la transformación de la media como $g(\mu)$. Esta función $g(\mu)$ es la Función Link porque vincula el modelo mixto con la media de Y . Puesto que la función de vínculo es monótona, la relación entre μ y η se puede expresar en términos de la inversa de la función de vínculo:

$$g(\mu) = \mathbf{X}\beta + \mathbf{Zb} = \eta \tag{16}$$

$$\mu = g^{-1}(\mathbf{X}\beta + \mathbf{Zb}) \tag{17}$$

Para las variables respuestas que contengan conteos, es preferible usar como distribución la Poisson, cuya media condicional está dada por λ , la función de vínculo $\eta = \log(\lambda) = \mathbf{X}\beta + \mathbf{Zb}$ y la inversa de la función de vínculo $g^{-1}(\eta) = e^\eta$.

6. Resultados y Discusión

6.1. Variables

Para el desarrollo, se hizo uso de los datos más cercanos a la realidad, los cuales provienen del juego FIFA de EA sport games, donde se encuentran estimaciones ponderadas puntuales de todos los jugadores y variables importantes de los clubes, adicional a esto, mediante un algoritmo de web scraping se obtuvo el resultado histórico de los encuentros junto con sus estadísticas (desde inicia de la temporada 2016 a la actualidad).

<i>Variables de jugadores</i>	
Característica	Significado
Ball control	Control del balón
Crossing	La capacidad para tirar buenos centros desde la banda
Curve	Capacidad para darle rosca o efecto al balón a la hora de patear
Dribbling	Capacidad del jugador para cortar la cancha con regates
Finishing	La habilidad del jugador para poder definir frente al arco
Free Kicks	Precisión que tiene el jugador para los Tiros Libres
Heading	Capacidad para cabecear
Long Passing	Habilidad del jugador para los pases en profundidad
Long Shots	Precisión de los tiros lejanos
Acceleration	Velocidad con la que el jugador arranca con la pelota
Agility	Habilidad para desmarcarse
Balance	Habilidad que tiene el jugador para no caerse ante el primer golpe
Stamina	Es la resistencia del jugador para jugar 90 minutos
Interceptions	Capacidad del jugador para interceptar pelotas sueltas
Positioning	Es la habilidad del jugador para estar siempre donde se le necesita
Strength	Es la fuerza del jugador

En la tabla 1 se muestran algunas de las característica obtenidas por cada uno de los jugadores perteneciente a los equipos de la liga española.

En la tabla 2 se encuentran algunas de las variables que se obtuvieron para el estudio, a nivel de equipo, cabe resaltar que el individuo esta marcado por el equipo local y los promedios de dichas variables son

<i>Variables de equipos por encuentros</i>	
variable	Significado
Team_one	Equipo local
Team_two	Equipo visitante
Goles_local	Promedio de goles anotados por el equipo local en encuentros anteriores
Goles_visitante	Promedio de goles anotados por el equipo visitante en encuentros anteriores
Faltas_local	Promedio de faltas hechas por el equipo local en encuentros anteriores
Faltas_visitante	Promedio de faltas hechas por el equipo visitante en encuentros anteriores
Local_corner	Promedio de tiros de esquina a favor del equipo local en encuentros anteriores
Visitante_corner	Promedio de tiros de esquina a favor del equipo visitante en encuentros anteriores
Posesion_Local	Promedio de posesión del equipo local en encuentros anteriores
Posesion_visitante	Promedio de posesión del equipo visitante en encuentros anteriores

tabla 2.

calculados de los encuentros disputados desde la temporada del 2016 a la actualidad en las mismas condiciones(local y visitante)

6.2. Ajuste de variables

Teniendo en cuenta la cantidad de variables mostradas anteriormente y la premisa fundamental del estudio "los individuos son los equipos" se busco una estrategia para convertir las variables de los jugador a variables del equipo.

6.3. Análisis de Componentes Principales

El método de componentes principales tiene por objeto transformar un conjunto de variables, a las que se denomina originales, en un nuevo conjunto de variables denominadas componentes principales. Estas últimas se caracterizan por estar incorrelacionadas entre sí y, además, pueden ordenarse de acuerdo con la información que llevan incorporada(Fernandez, 2011)

A continuación se presentan las variables que se deben tener en cuenta para el análisis:

6.3.1. ACP Jugadores

Como resultado se obtiene que:

Componente 1: están ubicados los jugadores que tienen puntaje alto en características tales como Dribbling, Positioning, LongShots, BallControl, Vision, Curve, Penalties, propias de un jugador ofensivo, por tal motivo esta componente será llamada poder ofensivo. Los jugadores que resaltan en este componente son Messi, Suarez, Griezman y Modiric de los cuales se sabe que son jugadores elites de la liga en posiciones ofensivas

Componente 2: están ubicados los jugadores que tienen puntaje alto en características tales como Interceptions, StandingTackle, SlidingTackle, Marking, Strength, las cuales componen un jugador defensivo, por tal motivo esta componente será llamada poder defensivo. Los jugadores que resaltan en este componente son Casemero, Godin, Ramos y Busquets de los cuales se sabe que son jugadores elites de la liga en posiciones defensivas

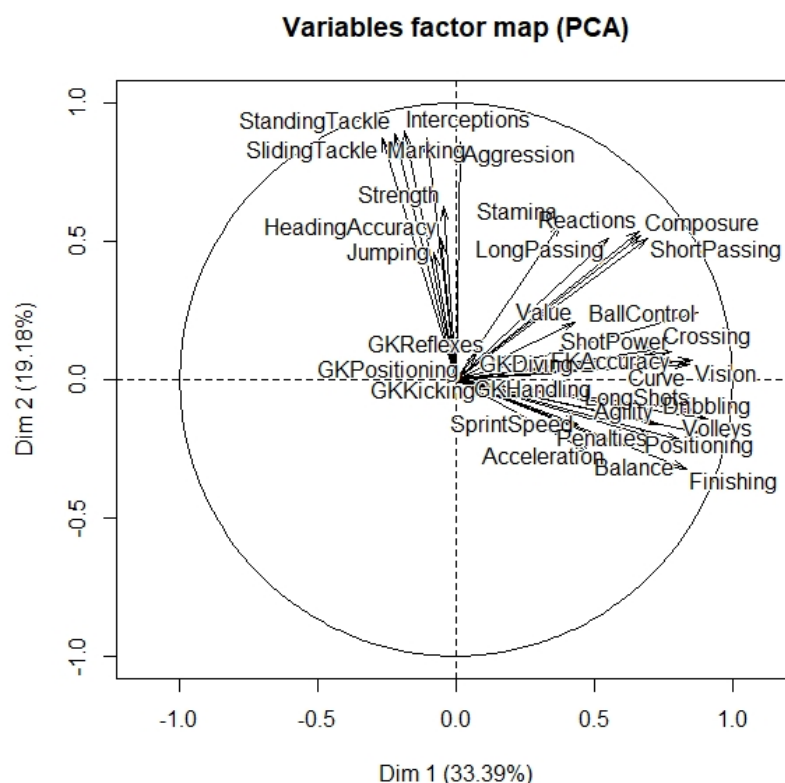


Figura 1: ACP de jugadores

6.3.2. ACP Arqueros

Realizando el proceso anterior de la misma manera para los arqueros, se encontró que en el componente número 1, están ubicados los arqueros con puntajes más altos en GKPositioning, GKKicking, GKHandling, GKReflexes, las cuales son únicas en los porteros.

6.4. Indicadores

Para el desarrollo de los indicadores se realizaron los siguientes pasos:

1. En las coordenadas obtenidas de los componentes (poder ofensivo y poder defensivo) se aplicó la función anti-logit con el fin de dejar todos los valores entre 0 y 1.
2. Seleccionar los mejores 8 jugadores por equipo en cada una de las componentes y calcular el promedio de los mismos, para así tener un único valor por equipo en las componentes.

Con estos pasos se logró obtener el poder promedio ofensivo y el poder promedio defensivo de cada uno de los equipos, de igual manera se realizó con los arqueros de cada equipo seleccionando únicamente a los dos mejores del componente obteniendo así el poder medio de la portería.

6.5. Comparación entre modelos

6.5.1. Modelo 1

$$goleslocal \sim 1 + (ofensivo + defensivo | team_{one}) + (Faltas_{local} + Localyellow + possession_{local} + Localcorner | team_{two}) \quad (18)$$

6.5.2. Modelo 2

$$goleslocal \sim 1 + (ofensivo + defensivo | team_{one}) + (Faltas_{local} + Localyellow + Localcorner + Localshot | team_{two}) \quad (19)$$

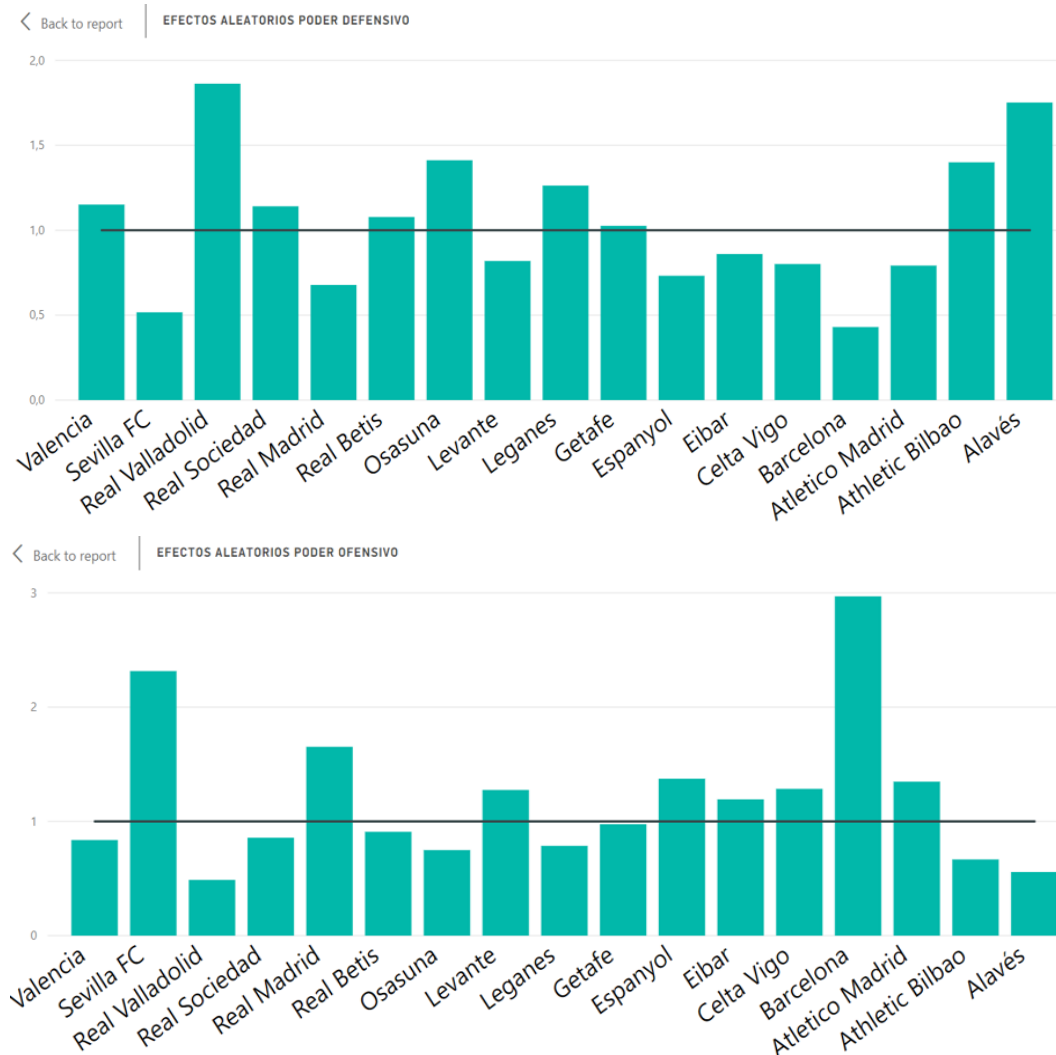
6.5.3. Modelo 3

$$goleslocal \sim 1 + (ofensivo | team_{one}) + (Localcorner + Localshot | team_{two}) \quad (20)$$

6.5.4. Decisión

Al probar diferentes estimadores de ajuste para los diferentes modelos, se encuentra que los mejores modelos son los nombrados anteriormente sin una diferencia significativa entre ellos, sin embargo al realizar la prueba de sobre dispersión el único modelo que no sufre de esto es el número uno, por tal motivo se procede a realizar el estudio de los efectos del mismo

6.6. Efectos Aleatorios propios



Como resultado de los efectos aleatorios propios del equipo, se llega a conclusiones tales como:

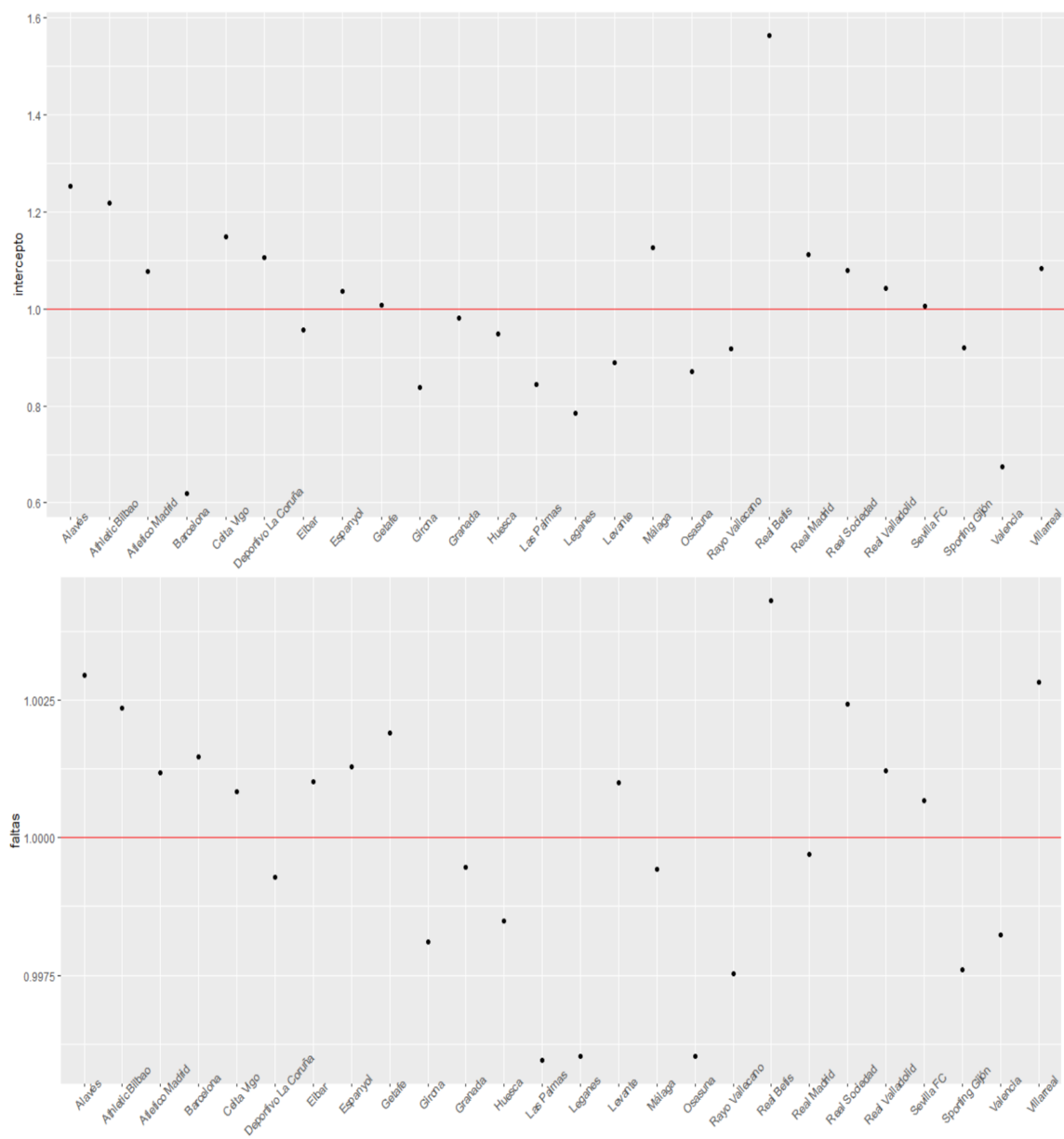
1. Por cada punto que aumente en el indicador de poder Defensivo el Athletic de Bilbao anota casi 1.1 veces más su valor esperado de gol en Ceteris paribus.
2. Por cada punto que aumente en el indicador de poder Defensivo el Real Valladolid anota 1.8 veces más su valor esperado de gol, esta cifra puede parecer bastante alta, pero es dado a las características del equipo y su baja oportunidad de marcar un gol en Ceteris paribus.

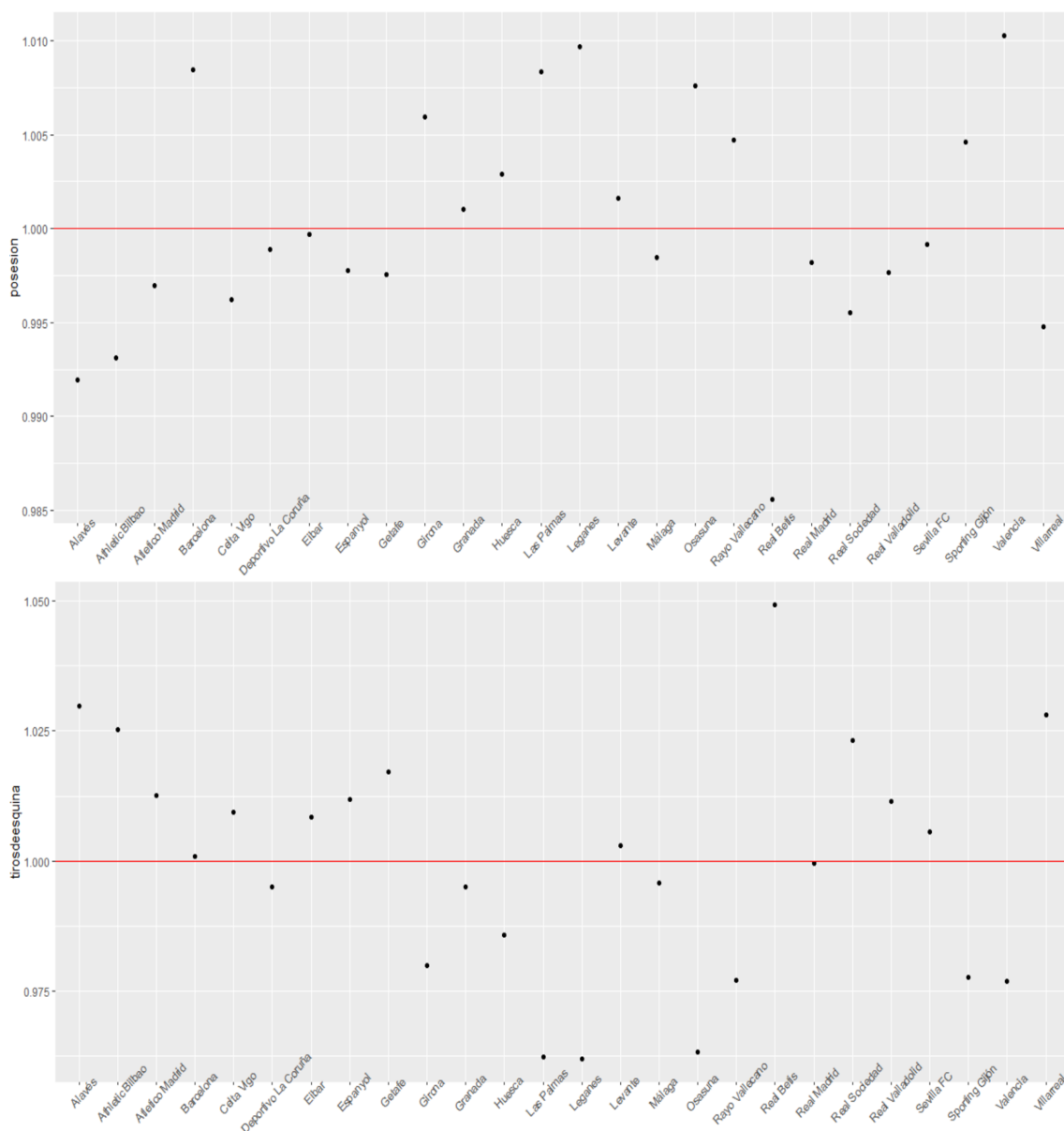
Por otra parte también se encuentra que el Barcelona por cada punto que aumente en el indicador de poder ofensivo marca casi 3 veces su valor esperado de gol en Ceteris paribus, esto es dado a que el equipo tiene el indicador más alto en poder ofensivo, por tal motivo es muy difícil que llegue a subir un punto más en dicho indicador.

6.7. Efectos Aleatorios contra el equipo jugado

Al hablar de la cantidad de goles marcados por un equipo se debe tener en cuenta contra quien se estaba jugando, ya que no es lo mismo la cantidad de goles que espera marcar el Barcelona a un equipo como

el levante a la cantidad de goles que espera marcarle al Real Madrid.





teniendo en cuenta lo anterior se encontró casos a resaltar tales como:

1. por cada punto porcentual que aumente cualquier equipo en la posesión del balón al jugar contra al valencia hace un uno por ciento más de goles en Ceteris paribus.
2. por cada tiro de esquina de más que tenga a favor cualquier equipo al jugar contra el real betis, logra marcar 5 por ciento más de goles.

Estos análisis son algunos de los que se pueden realizar con el desarrollo de un modelo poisson mixto además de estimar el resultado de un encuentro.

Referencias

- [1] Cordeiro, G. and Demétrio, C. (2008) *Modelos Lineares Generalizados e Extensões*.
- [2] Correa, J, Salazar J. (2016), *Introducción a los modelos mixtos*, Medellin Colombia, Centro Editorial.
- [3] Diario de Pontevedra. (24 de junio de 2018), *Los diferentes tipos de apuestas deportivas*, Recuperado de <https://www.diariodepontevedra.es/articulo/deporte-general/diferentes-tipos-apuestas-deportivas/20180425192316977407.html>.
- [4] Dyte, D. y Clarke, S. J Oper Res Soc (2000) 51: 993. <https://doi.org/10.1057/palgrave.jors.2600997>
- [5] Fútbol ofensivo. (18 de Septiembre de 2016), *Características y funciones de los puestos en un equipo de fútbol*, Recuperado de <http://www.futbolofensivo.com/2016/09/caracteristicas-y-funciones-de-los-puestos-en-un-equipo-de-futbol/>.
- [6] Goddard, J., Asimakopoulos, I. (2002). *Modelling football match results and the efficiency of fixed-odds betting* . Swansea: University of Wales Swansea.
- [7] Karlis, D. Ntzoufras, I. (2011). *Robust fitting of football prediction models* . IMA Journal Management Mathematics.
- [8] Nelder, J.A. and Wedderburn, R.W.M. (1972) *Generalized Linear Models*. *Journal of the Royal Statistical Society Series A*, 135, 370-384.
- [9] Planeta fútbol. (2010), *Liga Española*, Recuperado de <https://sites.google.com/site/planetafutbol51/futbol-europeo/liga-espanola>
- [10] Rojas Diosabá Victor,(1 de Enero de 2018). *El fútbol es el negocio del siglo.Semana*. Recuperado de <https://www.semana.com/contenidos-editoriales/ahora-si-rusia/articulo/el-futbol-es-el-negocio-del-siglo/551550>
- [11] Torres-Saavedra, P. A. (2018). *Modelos Estadísticos Avanzados*. <http://pegasus.uprm.edu/pedro.torres/book/references.html>.