



ESCUELA SUPERIOR POLITÉCNICA DEL LITORAL

Facultad de Ciencias Naturales y Matemáticas

“Estudio del efecto del público en los partidos de fútbol ecuatoriano de la
Serie A”

PROYECTO INTEGRADOR

Previo la obtención del Título de:

INGENIERO EN ESTADÍSTICA INFORMÁTICA

Presentado por:

Briceño Carchi Jhuliana Pamela

Padilla Navarro Ashley Lissette

GUAYAQUIL - ECUADOR

Año: 2021

DEDICATORIA

Este proyecto se lo dedicamos a Dios principalmente, y a nuestros padres Manuel Padilla, Ketty Navarro, Homero Briceño y Rita Carchi que nos han apoyado en todos los sentidos durante toda nuestra vida a cumplir nuestras metas. A María Moreira Cuadros por ser una mujer fuerte, trabajadora y feliz, siempre brindando su sonrisa cuando se la necesitaba.

AGRADECIMIENTOS

A Dios, por siempre guiarnos y proteger cada paso que hemos dado en estos años.

A nuestros padres y familia que siempre nos han apoyado en todos los sentidos posibles, les agradecemos por brindarnos seguridad y confianza necesaria para culminar una etapa más en nuestra formación académica.

A nuestro tutor, Holger Cevallos, por su voluntad de ayudar y ser una guía indispensable en el desarrollo de este proyecto.

A Allisson Morales por brindarme su apoyo y su amistad durante todos los años de carrera.

DECLARACIÓN EXPRESA

“Los derechos de titularidad y explotación, nos corresponde conforme al reglamento de propiedad intelectual de la institución; nosotras, Jhuliana Pamela Briceño Carchi y Ashley Lissette Padilla Navarro damos nuestro consentimiento para que la ESPOL realice la comunicación pública de la obra por cualquier medio con el fin de promover la consulta, difusión y uso público de la producción intelectual”



Ashley Padilla Navarro



Jhuliana Briceño Carchi

EVALUADORES



SANDRA LORENA
GARCIA BUSTOS

Ph.D. Sandra García Bustos

PROFESOR DE LA MATERIA



HOLGER GEOVANNY
CEVALLOS
VALDIVIEZO

Ph.D. Holger Cevallos

PROFESOR TUTOR

RESUMEN

Por varios años se había deseado evaluar la situación de que en los partidos de fútbol no haya público presente y esto sólo había sido posible mediante simulaciones. En el 2020 se declaró al Covid-19 pandemia mundial, afectando a todos los sectores, el sector deportivo fue uno de ellos, lo que causó que normas de bioseguridad se implementaran, siendo una de estas los partidos a puerta cerrada, dando paso a que sea posible obtener datos de partidos sin público. Existen diferentes estudios que postulan que la ventaja del local se pierde cuando no hay público, así como también señalan que el público afecta en las decisiones arbitrales, lo cual se observó en las ligas menores de fútbol. Se observó que las variables utilizadas no seguían los supuestos de un modelo lineal, comportamiento que se intentó corregir con transformaciones de escalas, sin embargo, no cambió la naturaleza de los datos. Por este motivo, se aplicó el modelo lineal mixto, utilizando la función `lme()` en el lenguaje de programación R. Se construyeron varios modelos y se observó que los resultados obtenidos no fueron los esperados, puesto que, indican que cuando no hay público el equipo local tiende a anotar más goles en los partidos en comparación a cuando hay público. También se logró observar que cuando hay público presente el equipo local tiende en promedio a obtener más tarjetas, y el equipo visitante cuando no hay público presente en el estadio éste es sancionado en promedio con más tarjetas por el cuerpo arbitral.

Palabras claves: Ventaja en casa, Modelo Lineal Mixto, Sesgo arbitral.

ABSTRACT

For several years it had been of interest to evaluate the situation where there is no public present at soccer matches and this had only been possible through simulations. In 2020, the Covid-19 was declared as a global pandemic, affecting all sectors, the sports sector was one of them, which caused biosecurity standards to be implemented, one of these being the closed-door matches, making it is possible to obtain data from matches without an audience. Different studies postulate that the home advantage is lost when there is no public. The public affects arbitration decisions, which are executed in the minor soccer leagues. It was realized that the variables used did not follow the assumptions of a linear model, a behavior that was intended to correct with scale transformations, however, the nature of the data did not change. For this reason, the mixed linear model was applied, using the `lme ()` function in the programming language R. Several models were created, but the results were not the expected ones. The result shows that when there is no public the local team tends to score more goals in matches compared to when there is an audience. It was also observed that when there is an audience present the home team tends on average to obtain more cards, and the visiting team when there is no public present in the stadium is sanctioned on average with more cards by the referee.

Keywords: Home Advantage, Linear Mixed Model, Referee Bias

ÍNDICE GENERAL

1. CAPÍTULO 1	15
1.1 Introducción y Problema	15
1.2 Antecedentes	17
1.2.1 Árbitro.....	17
1.2.2 Clubes Deportivos Ecuatorianos que intervienen en el estudio.....	18
1.3 Objetivos	24
1.3.1 Objetivo General.....	24
1.3.2 Objetivos Específicos.....	24
1.4 Marco Teórico.....	25
1.4.1 Modelo Lineal.....	25
1.4.2 Modelo Lineal Mixto.....	27
1.4.3 Criterios de Información de Akaike (AIC).....	28
1.4.4 El criterio de información corregido de Akaike (AICC).....	28
1.4.5 El criterio de información bayesiano de Schwarz (BIC).....	29
1.4.6 Teorema del límite central.....	29
1.4.7 Estado del arte.....	29
2. CAPÍTULO 2	32
2.1 Metodología	32
2.1.1 Recolección de datos.....	32
2.1.2 Población objetivo.....	33
2.1.3 Muestra.....	33

2.1.4	Software Utilizado	35
3.	CAPÍTULO 3	36
3.1	Análisis Descriptivo	36
3.2	Ventaja en Casa.....	40
3.2.1	Modelo final.....	43
3.2.2	Evaluación de los Supuestos.....	46
3.3	Sesgo Arbitral.....	48
3.3.1	Modelo final.....	51
3.3.2	Evaluación de los supuestos.....	53
4.	CAPÍTULO 4	56
4.1	Conclusión.....	56
4.2	Recomendación.....	57
5.	BIBLIOGRAFÍA.....	58
6.	ANEXOS	62

ÍNDICE DE GRÁFICOS

Gráfico 1 Popularidad de los equipos de fútbol.....	23
Gráfico 2 Diagramas de caja.....	37
Gráfico 3 Histogramas de frecuencia.....	38
Gráfico 4 Gráficos QQ.....	39
Gráfico 5 Diagrama de barras.....	40
Gráfico 6 Matriz de correlación.....	43
Gráfico 7 Histograma de interceptos.....	46
Gráfico 8 Independencia de residuos.....	46
Gráfico 9 Histograma de residuos.....	47
Gráfico 10 Gráfico QQ de residuos.....	47
Gráfico 11 Interceptos del modelo.....	48
Gráfico 12 Matriz de correlación.....	50
Gráfico 13 Histograma de interceptos.....	53
Gráfico 14 Independencia de residuos.....	53
Gráfico 15 Histograma de residuos.....	54
Gráfico 16 Gráfico QQ de residuos.....	54
Gráfico 17 Intercepto del modelo.....	55

INDICE DE TABLAS

Tabla 1 Tipos de modelos	26
Tabla 2 Descripción de variables cualitativas.....	33
Tabla 3 Descripción de variables cuantitativas.....	34
Tabla 4 Medidas de tendencia central.....	36
Tabla 5 Modelos de Ventaja en Casa.....	41
Tabla 6 Estimaciones del modelo	44
Tabla 7 Resumen de resultados de los partidos	49
Tabla 8 Modelos de Sesgo Arbitral	49
Tabla 9 Estimaciones del modelo	51

ÍNDICE DE ANEXOS

Anexo A Resumen del modelo de Ventaja de casa	62
Anexo B Resumen del modelo de Sesgo Arbitral	62

INDICE DE FIGURAS

Figura 1	18
Figura 2	19
Figura 3	19
Figura 4	20
Figura 5	20
Figura 6	21
Figura 7	22
Figura 8	22
Figura 9	23

ABREVIATURAS

OMS	Organización Mundial de la Salud
COE	Comité de Operaciones de Emergencias
LigaPro	Liga Profesional de Fútbol Ecuatoriano
FEF	Federación Ecuatoriana de Fútbol
FIFA	La Federación Internacional de Fútbol Asociación
RAE	La Real Academia Española
UCLA	Universidad de California en Los Ángeles

1. CAPÍTULO 1

1.1 Introducción y Problema

El 11 de marzo del 2020, el Dr. Tedros Adhanom, director general de la Organización Mundial de la Salud (OMS), anunció que el virus COVID-19 se caracteriza como una pandemia, desencadenando que todas las naciones se declaren en estado de emergencia sanitaria y tomen medidas de bioseguridad para enfrentar al virus (Organización Panamericana de la Salud, 2020). A raíz de esto, todas las actividades del país se vieron afectadas, incluyendo el Campeonato Nacional de Fútbol, el cual tuvo que suspender sus actividades deportivas y programar los partidos cuando el COE Nacional anunció la reapertura de algunas actividades tomando medidas de bioseguridad correspondientes.

Después del duro golpe que enfrentó Ecuador debido al COVID-19 se implementaron varias medidas de bioseguridad las cuales llevaron a suspender varios eventos sociales. Se conoce que en nuestro país el fútbol es el deporte con mayor cantidad de aficionados ¿Cómo habrá afectado la cancelación de eventos que implicaban la concentración de personas? el campeonato de fútbol ecuatoriano se suspendió el 14 de marzo del 2020 (EL UNIVERSO, 2021), situación que duró algunos meses. Mediante un comunicado oficial el COE Nacional informó que la LigaPro continuaría el campeonato 2020 el 15 de agosto del mismo año (EL COMERCIO, 2020).

La pandemia no sólo afectó el sector de la salud, puesto que también provocó grandes pérdidas económicas a nivel mundial, y el fútbol no fue la excepción; por mencionar, la mayoría de los clubes del fútbol europeo tomaron medidas para mitigar las millonarias pérdidas como acudir al programa Plan de Apoyo covid-19(FIFA), abstenerse de grandes fichajes, reducir gastos salariales y tomar pocos derechos de retransmisión (Pérez, 2021). Mientras tanto, la LigaPro mediante protocolos y regímenes sancionadores, dio a conocer las

nuevas disposiciones reglamentarias en el fútbol profesional antes, durante y después de cada partido; entre los que se pueden mencionar son: todos los jugadores, árbitros, miembros del cuerpo técnico y demás personas que ingresen a los estadios en los partidos se deben realizar una prueba rápida de Covid-19 un día antes del partido, cada camerino deberá estar equipado adecuadamente para que los jugadores puedan desinfectarse y desechar el material de bioseguridad, se permitirán hasta 5 cambios por equipo durante el partido, evitar aglomeraciones y queda prohibido el ingreso a personas particulares (LigaPro, 2020).

García et al. (2012) mencionan que en el mundo del fútbol se da por hecho que el equipo local tiende a ganar con más frecuencia que el equipo visitante, dicho fenómeno ocurre dado el efecto que tienen los aficionados sobre el partido (Saavedra García, Gutiérrez Aguilar, Fernández Romero, & Sa Marques, 2012), sin embargo, en años anteriores se han realizado análisis en base a simulaciones, tomando datos de 17 ligas europeas que concluyeron que no en todas las ligas se hace presente la ventaja de jugar en casa (Benz & Lopez, 2021). Ahora bien, con las nuevas medidas implementadas en el Ecuador, los partidos de fútbol desde la sexta fecha de la temporada 2020 se jugaron sin público presente, lo que permite en la actualidad recolectar datos reales y evaluar el efecto de jugar sin público en el dominio del partido del equipo local y en las decisiones arbitrales.

Considerando esta problemática, en este trabajo se pretende estudiar los resultados del fútbol ecuatoriano de la Serie A desde el 2017 hasta la primera etapa del 2021 para observar el efecto del público en el rendimiento de los equipos y el sesgo arbitral. Se inicia con una búsqueda por internet de los datos correspondientes, luego de evaluar varios artículos y decidir el modelo a utilizar para el correcto análisis de los datos.

Comentado [HGCV1]: ¿Quién da por hecho?

Comentado [HGCV2]: ¿qué resultados se obtuvieron en estas simulaciones?

Comentado [HGCV3]: que concluyeron que

1.2 Antecedentes

El fútbol inició como un deporte oficial en el siglo XIX, el 26 de octubre de 1863 para ser exacto, se fundó en Inglaterra la 'Asociación de Fútbol' donde se crearon reglas con el principal objetivo de combatir la violencia. Sin embargo, se sitúan otros juegos desde la antigüedad que consistían en patear una pelota, por mencionar, en China durante la dinastía Han se practicaba el juego conocido como "Ts'uh Kúh" o "Cuju" (Profesor en línea, 2015).

En Ecuador este deporte es el más popular entre todos, por lo que la creación de la Federación Ecuatoriana de fútbol era inevitable, lo cual se dio en 1967. El fútbol aparece en el país debido a jóvenes que se formaron académicamente en Inglaterra, al regresar al país compartieron su conocimiento del deporte. En Guayas se realizó el primer torneo profesional en 1951, el campeonato fue organizado por algunos clubes y el equipo Río Guayas resultó como el primer campeón. Posteriormente, Emelec fue el primer equipo en ganar un campeonato nacional en 1957 y Barcelona quedó en segundo puesto (Montenegro Galárraga, 2020). Desde el 2019 la organización que dirige el fútbol nacional es la Liga Pro.

1.2.1 Árbitro

En los deportes es necesaria una persona que ayude a resolver conflictos que se presenten durante el juego, estas situaciones pueden ser difíciles y conflictivas, pero todo juego debe seguir las reglas establecidas, y el responsable de hacer que se respeten y cumplan las normas del juego se lo conoce como árbitro. La RAE (2015) define al árbitro como "Persona que, en algunas competiciones, normalmente deportivas, cuida de la aplicación del reglamento, sanciona las infracciones o fallos y valida los resultados".

El árbitro basa sus decisiones en el reglamento establecido de la competición o campeonato que se esté disputando, entre sus funciones se puede mencionar algunas:

- Observar qué jugador realiza acciones que violan el reglamento.

Comentado [HGCV4]: tilde

- Evaluar las acciones del jugador.
- Penalizar cuando se viola el reglamento de fútbol.

1.2.2 Clubes Deportivos Ecuatorianos que intervienen en el estudio.

Barcelona Sporting Club fue fundado el 1 de mayo de 1925 en la ciudad de Guayaquil. Desde 1957 ha formado parte de la Serie A ecuatoriana de fútbol, a lo largo de su carrera ha logrado obtener 16 títulos nacionales representados como estrellas en su escudo, 5 títulos provinciales y 1 amateur. Tiene su propio estadio, el “Estadio Monumental Isidro Romero Carbo” conocido popularmente en Ecuador como “El Monumental”, este fue inaugurado en 1987. Su mayor rival futbolístico es el Club Sport Emelec.

Figura 1

Barcelona Sporting Club



Tomada de (Barcelona Sporting Club, s.f.)

Delfín Sporting Club fue fundado en la ciudad de Manta en 1989, su primera participación la realizó en la Serie B ecuatoriana en el mismo año de su fundación, logró subir a la Serie A en el año 2016, y en el 2019 ganó el título de Campeón Nacional. Cuando juega de local, los partidos son en un estadio municipal de Manta llamado ‘Jocay’.

Figura 2

Delfín Sporting Club



Tomada de (PES Logos, s.f.)

Club Deportivo Cuenca organizado el 4 de marzo de 1971 en la ciudad de Cuenca, desde sus inicios formó parte de la Serie A. Este club de local juega en el estadio “Alejandro Serrano Aguilar”, su nombre es en honor a quien fue alcalde de Cuenca y presidente del club en el año 1971, el cual pertenece a la Federación Deportiva de Azuay. En el 2018, el equipo consiguió clasificar a la competición internacional Copa Sudamérica y llegar hasta a octavos de final.

Figura 3

Club Deportivo Cuenca



Tomada de (Club Deportivo Cuenca, s.f.)

Club Sport Emelec conocido comúnmente como Los Eléctricos, El Bombillo o Los Millonarios, fue fundado en Guayaquil el 28 de abril de 1929. Forma parte de la Serie A desde

1957. A lo largo de su carrera ha logrado adjudicarse 14 copas nacionales, así como también títulos provinciales y amateurs. Posee su propio estadio el 'Estadio George Capwell', popularmente llamado como "El Capwell". Su mayor rival deportivo desde su fundación es el Barcelona Sporting Club.

Figura 4

Club Sport Emelec



Tomada de (Club Sport Emelec, s.f.)

Guayaquil City se fundó el 7 de septiembre de 2007 bajo el nombre de Club Deportivo River Plate Ecuador, sin embargo, el 11 de julio del 2017 se realizó el cambio de nombre. El Club debutó en la segunda categoría del Guayas, en 2009 logró su ascenso a la Serie B, y en el 2014 ascendió a la Serie A del campeonato nacional. En su carrera deportiva el club no ha ganado títulos nacionales hasta la fecha. En la actualidad disputa sus partidos en el Estadio "Christian Benítez Betancourt" que se encuentra dentro del parque Samanes de Guayaquil, y es propiedad estatal.

Figura 5

Guayaquil City



Tomada de (Guayaquil City, s.f.)

El Club de Alto Rendimiento Especializado Independiente del Valle, más conocido como Independiente del Valle fue fundado el 1 de marzo de 1958 en la ciudad de Sangolquí. Posee un logro sobresaliente, que es haber ganado un campeonato nacional en el 2013, también llegó a ser subcampeón de la Copa Libertadores en 2016.

Figura 6

Independiente del Valle



Tomada de (Independiente del Valle , s.f.)

Liga Deportiva Universitaria de Quito frecuentemente identificado como LDU o Liga de Quito, inició como un equipo llamado Club Universitario en la Universidad Central del Ecuador” en 1918, sin embargo, fue organizado formalmente el 11 d enero de 1930. Comenzó a jugar desde 1960 en la serie A. Ha ganado campeonatos tanto nacionales como internacionales, tiene 11 torneos nacionales, 1 torneo de copa, 2 de supercopa, 6 torneos regionales y 3 amateurs. Cuando juega de local, disputa sus partidos en el estadio “Rodríguez Paz Delgado” más conocido como la Casa Blanca. Tiene una rivalidad deportiva con Sociedad Deportiva Aucas.

Figura 7

Liga Deportiva Universitaria de Quito



Tomada de (Liga Rey de Copas, s.f.)

Club Deportivo Macará usualmente conocido como Macará, organizado en la ciudad de Ambato el 25 de agosto de 1939. Forma parte de la Serie A desde 1960. Juega sus partidos de local en el Estadio de Bellavista propiedad de la Federación Deportiva de Tungurahua.

Figura 8

Club Deportivo Macará



Tomada de (Club Deportivo Macará, s.f.)

Club Deportivo de la Universidad Católica, originario en la ciudad de Quito en 1963, forma parte de la Serie A desde 1965 y hasta la fecha no posee ningún campeonato ganado. Los partidos que disputa como local toman lugar en el Estadio “Olímpico Atahualpa”.

Figura 9

Club Deportivo de la Universidad Católica



Tomada de (U. CATÓLICA OFICIAL, s.f.)

En el Gráfico 1 se observa una variable interesante para la construcción del modelo, se puede decir que la audiencia en un partido de fútbol depende bastante de la popularidad del equipo. Como se muestra en la Gráfica 1, equipos como Barcelona, Emelec y LDU Quito se pueden considerar los más populares en el Ecuador, lo que se refleja en las redes sociales cuando algún dirigente o futbolista de uno de los equipos anteriormente mencionados realiza alguna publicación referente al equipo.

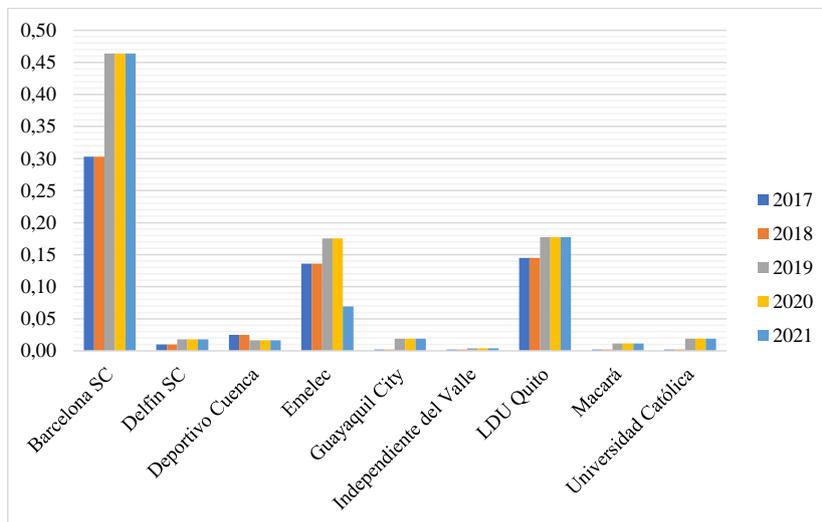


Gráfico 1 Popularidad de los equipos de fútbol

Fuente: Diario El Comercio. Equipos más populares del Ecuador

Comentado [HGCv5]: Mencionar la fuente de estos datos.

Comentado [HGCv6]: Me parece que la fuente debe ir por debajo del título del gráfico. Por favor revisar el formato y estilo que se requiere en la FCNM.

1.3 Objetivos

1.3.1 Objetivo General.

Investigar el efecto del público en los resultados del equipo local y en las decisiones del árbitro.

1.3.2 Objetivos Específicos.

- Seleccionar las variables necesarias para el análisis correspondiente de los partidos del fútbol ecuatoriano en el periodo establecido.
- Identificar el modelo que mejor se ajuste a los datos con el fin de responder la pregunta de investigación.
- Construir un modelo que considere las características de los datos observados y ayude a responder las preguntas de investigación.

Comentado [HGCV7]: datos con el fin de responder a nuestra pregunta de investigación.

1.4 Marco Teórico

En esta sección se presentan los diferentes conceptos que se tomaron en cuenta al momento de construir el modelo que mejor se ajusta los datos recolectados. Para este estudio se utilizó el modelo lineal mixto, el cual es robusto, permite trabajar con variables discretas y no considera necesario el cumplimiento del supuesto de independencia.

1.4.1 Modelo Lineal.

Los modelos lineales aparecieron en el siglo XX, se fundamentan en la correlación parcial y en tres métodos: análisis de covarianza (ANCOVA), análisis de varianza (ANOVA) y análisis de varianza multivariado (MANOVA). Este modelo trabaja con variables predictoras continuas y con variables explicativas que pueden ser de clasificación o continuas. El objetivo general de este modelo es analizar los diferentes grupos que se presentan y observar su comportamiento, o también, en base a las variables explicativas obtener una predicción para la variable de respuesta. La ecuación del modelo lineal se representa de la siguiente forma:

$$E(\mathbf{y}) = \mathbf{X}\boldsymbol{\beta}, \quad \mathbf{y} \sim N(E(\mathbf{y}), \sigma^2 \mathbf{I})$$

Si se tienen datos no lineales, lo que implica que uno de los supuestos del modelo no se cumple, una de las posibles soluciones es optar por la transformación de la escala de los datos, aclarando que lo que se transforma es la escala y que los datos reales no se alteran. Una de las transformaciones más usadas es la de logaritmo, aunque se tienen diferentes formas de transformación, de las cuales se puede indicar las siguientes:

- $y = x^2$
- $y = \sqrt{n}$
- $y = \ln x$

- $y = \frac{1}{x}$

Sin embargo, aún con las transformaciones disponibles no siempre se logra el propósito de mejorar el ajuste de la variable, lo que representa un problema para los modelos lineales clásicos.

Cuando se trabaja con variables discretas no es adecuado asumir distribución normal. En diferentes situaciones en los estudios se presentan factores fijos y factores aleatorios, los cuales han colaborado para la existencia de otros modelos, como los que se puede observar en la Tabla 1:

Tabla 1 Tipos de modelos

Tipo de Modelo	Similitudes	Diferencias
Lineales Generales	<ul style="list-style-type: none"> • Efectos Fijos • Distribución normal. • Variable de respuesta continua. • Variables predictoras: continua o discreta. 	<ul style="list-style-type: none"> • Estimación por mínimos cuadrados.
Lineales Mixtos		<ul style="list-style-type: none"> • Efectos aleatorios • Estimación por máxima verosimilitud
Lineales Generalizados Mixtos		<ul style="list-style-type: none"> • Distribución de la familia exponencial • Estimación de los efectos por mínimos cuadrados fijos y aleatorios • Función de ligamento no lineal • Variable de respuesta discreta

Fuente: (Bandera Fernández & Pérez Pelea, 2017)

1.4.2 Modelo Lineal Mixto.

El modelo lineal mixto es una forma extendida del modelo lineal general. Hay varias razones para tomar en cuenta el modelo mixto, entre las que se puede mencionar: son flexibles cuando existen varianzas y correlaciones no homogéneas, también son flexibles con modelos multinivel, modelos lineales jerárquicos y modelos con coeficientes aleatorios, (SPSS Statistics, s.f.). Estos modelos generalmente son aplicados en estudios que conllevan factores con niveles que se encuentran fuera de control del examinador ‘efectos aleatorios’, y factores con niveles que pueden ser controlados ‘efectos fijos’, es el mismo fenómeno que ocurre en la regresión lineal cuando se asume que los datos son variables aleatorias, los parámetros son efectos fijos. Sin embargo, en el modelo mixto los datos son variables aleatorias y los parámetros son variables aleatorias en un nivel, pero fijos en el nivel más alto (UCLA, s.f.). Los efectos jerárquicos evalúan las variables de respuesta en más de un nivel. La formulación general del modelo lineal mixto esta dado por la siguiente ecuación:

$$y_i = X_i\beta + Z_i b_i + e_i$$

Donde:

- i es la unidad de muestreo.
- y_i es el vector dimensional n_i con componentes, $j = 1, 2, \dots, n_i \quad i = 1, 2, \dots, N$.
- X_i la matriz de efectos fijos con dimensión $n_i \times p$.
- Z_i es la matriz de efectos aleatorios.
- β un vector de dimensión $p \times 1$, con parámetros desconocidos relacionados a los efectos fijos.
- b_i representa el vector de efectos aleatorios.

Entre las consideraciones para la aplicación de este modelo se tiene que la variable de respuesta debe ser cuantitativa, los factores pueden ser categóricos, valores en cadena o

numéricos. Uno de sus supuestos, es que la variable de respuesta es independiente con respecto a los factores aleatorios, fijos y covariables (Escobar, 2017). Este modelo permite trabajar con datos que, por naturaleza, no se ajustan a una distribución normal y que son correlacionados. Cuando se viola el supuesto de independencia, se pueden aplicar modelos lineales mixtos, puesto que estos se basan en la partición del error no explicado en una componente común (Evelyn Bandera-Fernández, 2018).

1.4.3 Criterios de Información de Akaike (AIC).

Este criterio es una herramienta para la selección del mejor modelo, no es una prueba de hipótesis del modelo ajustado es, más bien, un criterio comparativo entre los modelos, ya que se puede encontrar varios modelos que se ajusten a los datos. Este criterio se define de la siguiente manera:

$$AIC = -2 \log Lik + 2k$$

Siendo k el número de parámetros estimados. Para tomar la decisión del mejor modelo, se ranquea todos los AIC de los diferentes modelos postulados y se escoge el que tenga menor valor. Este criterio ayuda a identificar el mejor modelo en comparación a los demás, sin embargo, no identifica al modelo que mejor explica lo datos. (Juan Carlos Correa Morales, 2016)

1.4.4 El criterio de información corregido de Akaike (AICC).

Así mismo como el AIC , este criterio no es una prueba de hipótesis, es una herramienta comparativa para seleccionar modelos. Este criterio es una interpretación corregida del AIC explicado previamente. Tal como el AIC , para la elección del modelo, se ranquea todos los $AICC$ obtenidos. El criterio se formula de la siguiente manera:

$$AICC = AIC + \frac{2k(k+1)}{n-k-1}$$

1.4.5 El criterio de información bayesiano de Schwarz (BIC).

BIC es un criterio que, al igual que *AIC*, sirve para la selección de modelo, puesto que también evalúa los ajustes de los modelos que se están comparando, sin embargo, no quiere decir que sea el modelo exacto en cuanto a la descripción del comportamiento de los datos. Se selecciona el modelo que tenga el *BIC* más bajo, este criterio se define como:

$$BIC = -2 \log Lik + k \log n$$

1.4.6 Teorema del límite central.

Este teorema indica que si la muestra es lo suficientemente grande (por lo general, $n > 30$), indiferentemente de cuál sea la distribución de la media muestral, esta se aproximará a una distribución normal. Lo que indica que si se toman muestras n grandes y a estas muestras se les calculan los promedios muestrales, los promedios se comportarán de acuerdo con una distribución normal. En algunos modelos, el supuesto de la normalidad está presente y no siempre se cumple, por lo que el Teorema del Límite Central es muy útil al momento de aplicar diferentes procedimientos con poblaciones que no son normales (Minitab 18, 2019).

1.4.7 Estado del arte.

García et al. (2012) presentan un estudio sobre la ventaja de jugar en casa en el fútbol Español. Trabajan con una muestra de 22.015 partidos, los cuales se disputaron en 80 temporadas de la primera división desde 1928 hasta 2011, concluyendo que la ventaja de jugar en casa existe y es significativa, como también que la ventaja de local es decreciente, la cual desciende de un 70% hasta un 50% durante el periodo estudiado.

Benz y Lopez (2021) realizaron un estudio enfocado en estimar el cambio de la ventaja de jugar en casa durante la pandemia. Dado que los resultados de los partidos no son

lineales decidieron utilizar la distribución Poisson y demostraron mediante simulaciones que la regresión Poisson bivariada disminuye el sesgo al momento de estimar la ventaja local en un 85% en comparación a la regresión lineal. Con datos de 17 ligas europeas concluyeron que no en todas las ligas se hace presente la ventaja de jugar en casa.

Garicano et al. (2001) estudiaron el efecto de la presión social en el juego. Presentaron evidencia empírica que prueba que los árbitros de fútbol muestran favoritismo hacia los equipos locales para satisfacer al público presente en el estadio. Se conoce que los árbitros tienen la autoridad de aumentar tiempo en el partido, en este estudio se evidencia que los árbitros tienden a aumentar el tiempo de juego cuando equipo local va perdiendo o, por el contrario, terminar los partidos sin dar tiempo extra cuando el equipo local va ganando, estimaron que esta parcialidad ocurre aproximadamente en 2.5% de los juegos observados.

Jiménez et al. (2021) analizaron la diferencia de puntos y goles en función de la presencia o ausencia de público, también tomaron en consideración las decisiones arbitrales y el estilo de juego, con una muestra conformada por 8 ligas europeas (Alemania, España, Italia, Inglaterra y Austria) en la temporada 2019-2020. Concluyeron que no existe diferencia significativa entre jugar con o sin público, con excepción de la primera categoría alemana y española. Observaron que existe una disminución en la ofensiva y un aumento en la defensa cuando se juega sin espectadores, a raíz de ello las decisiones arbitrales se vieron afectadas, asignan menos faltas y tarjetas para los equipos visitantes excluyendo a Alemania y Austria.

Karlis y Ntzoufras (2000) investigan sobre la modelización de datos de fútbol. Entre sus hallazgos obtuvieron que la regresión Poisson puede ser considerada para aplicarse en datos de fútbol, debido a que es sencilla de aplicar. Determinaron también que el desempeño del equipo se puede medir según el número de goles, para demostrar esta idea realizaron

correlaciones entre el número de goles y la posición final de cada equipo, obteniendo correlaciones de hasta 0.85.

2. CAPÍTULO 2

2.1 Metodología

2.1.1 Recolección de datos.

Para la obtención de los datos se utilizaron tres fuentes, la primera fuente fue la página oficial de La “Liga Profesional de Fútbol Ecuatoriano” (LigaPro Ecuador, 2018), la misma que está conformada por 26 clubes de la serie A y serie B. Esta liga fue fundada el 20 de abril del 2018 y entre sus funciones se encuentran las de administrar, organizar torneos, controlar los derechos comerciales, estadios, tribunales de apelación y tribunal de seguridad. La segunda fuente, la página web “livefutbol.com” (worldfootball.net, 2000) la cual, es una revista electrónica que proporciona información detallada tanto de los partidos de fútbol como de las personas involucradas en el desarrollo de éstos a nivel mundial, y la tercera fuente, la aplicación de deportes “BeSoccer” (Heredia, 2006) que cuenta con más de 20 millones de usuarios, inició como una página web conocida como resultados-futbol.com y actualmente, posee información de fútbol de más de 200 países, convirtiéndose en el proyecto que posee la mayor base de datos del deporte más popular del mundo.

El periodo de estudio considerado fue desde la temporada 2017 hasta la primera etapa del 2021, obteniendo inicialmente 1028 registros. Sin embargo, luego de una revisión se identificaron 5 partidos que fueron cancelados debido a sanciones a uno de los equipos disputantes, dando a lugar a implementar los artículos correspondientes de los reglamentos de la FEF y LigaPro que indican que el resultado del partido será 0-3 (o viceversa) a favor del equipo no sancionado, por lo cual se decidió eliminar dichos partidos de la base de datos, ya que tales situaciones no representan lo que se desea estudiar. Además, para el año 2019 no se consideraron los partidos jugados en playoffs, es decir, cuartos y octavos de final. Y, por último, se eliminaron los partidos de los equipos que no permanecieron en la serie A durante

el periodo de estudio, por lo cual sólo se tomó el registro de 9 equipos, de los cuales se da una breve descripción en la sección 1.2.2 del capítulo anterior.

2.1.2 Población objetivo.

Se tomó como población objetivo a los partidos de la serie A de la Liga Profesional de Fútbol Ecuatoriano.

2.1.3 Muestra.

Entre los datos recolectados, se tiene 26 variables de las cuales 7 son cualitativas y 19 cuantitativas, y 950 observaciones.

En la Tabla 2 se enuncia y describe cada una de las variables categóricas

Tabla 2 Descripción de variables cualitativas

Variable	Descripción
Etapa	Representa las tres etapas de la competición: 1= Primera etapa/ida, 2= Segunda etapa/vuelta y 3=Final. En la etapa 3 juegan los equipos que quedaron en primer lugar tanto en la primera como en la segunda etapa.
Estación	Indica si los partidos se jugaron en invierno o verano.
Equipo	Los 9 equipos nacionales que participaron en las temporadas.
Status	Indica si el equipo está jugando como Local o Visitante.
Ciudad	La ciudad en la que llevaron a cabo los encuentros.
Región	Indica si los partidos han sido jugados en la sierra o costa.
Público	Indica si el partido se disputó con o sin asistentes.

Fuente: Elaboración propia

Comentado [HGCv8]: Sugiero que se expliquen únicamente las variables que se utilizó en el análisis.

Comentado [ALPN9R8]: Quitamos algunas, pero mencionamos también variables, con las que hicimos los modelos de prueba, estos modelos están en la tabla 8

Comentado [ALPN10R8]:

De igual manera, en la Tabla 3, se describen las variables cuantitativas.

Tabla 3 Descripción de variables cuantitativas

Variable	Descripción
Fecha	La jornada que juega cada equipo por etapa.
P_Gana	Probabilidad de ganar de cada equipo.
POS	Posesión del balón.
SOT	Remates a portería.
COR	Córner o saques de esquina.
ELO	Ranking de clasificación de fuerza del equipo.
EXP	Goles esperados.
Altura	Está dada por cada ciudad en la que se disputaron los partidos.
G	Goles en el partido.
A	Autogoles.
TA	Tarjetas amarillas.
DA	Tarjetas dobles amarillas.
TR	Tarjetas rojas.
GP	Goles por penales
PF	Penales fallidos
TE	Tiempo extra
Posición	Indica la posición en la que se encuentra el equipo en cada fecha.
Score	Es un puntaje de acuerdo con la posición del equipo.
Popularidad	Representa cuan conocido y apoyado es el equipo.

Fuente: Elaboración propia

2.1.4 Software Utilizado.

R es un lenguaje de programación que cuenta con una gran variedad de librerías donde se aplican diversas técnicas estadísticas, ideal para los diferentes análisis estadísticos. Interactúa de manera eficaz con otros lenguajes de programación y bases de datos, lo que permite una buena depuración y análisis preliminar de los datos. Para el desarrollo del estudio se utilizó RStudio, el cual es un entorno de desarrollo integrado (IDE) para R (Allaire, 2011).

Para la construcción del modelo, se utilizó la función `lme()` del paquete `nlme()`, la función “se ajusta a un modelo lineal de efectos mixtos, pero permitiendo efectos aleatorios anidados. Se permite que los errores dentro del grupo estén correlacionados y/o tengan varianzas desiguales” (Pinheiro, Bates, & R-core, 2021).

3. CAPÍTULO 3

Este capítulo está organizado de la siguiente manera: en la primera sección se muestra un análisis descriptivo de los datos recolectados, en la segunda sección se presenta el análisis realizado para la ventaja en casa, y al finalizar se encuentra el análisis del sesgo arbitral.

3.1 Análisis Descriptivo

Antes de iniciar con el análisis del modelamiento, se realiza una breve descripción de los datos para conocer su comportamiento, identificar datos faltantes y posibles valores atípicos. En la Tabla 4 se observa un resumen de estadísticos descriptivos de las variables numéricas utilizadas en esta investigación.

Tabla 4 Medidas de tendencia central

	Min	1st Cuantil	Mediana	Media	3rd Cuantil	Max
Goles	0,0000	0,0000	1,0000	1,2390	2,0000	6,0000
Posición	1,0000	3,0000	5,0000	5,8380	9,0000	16,0000
Popularidad	0,0020	0,0020	0,0188	0,0820	0,1450	0,5140
Tiempo Extra	1,000	4,000	5,000	4,998	6,000	19,000
Tiros de Esquina	0,000	3,000	4,000	4,482	6,000	15,000
Penales	0,000	0,000	0,000	0,1611	0,000	3,000
Tarjetas	0,000	2,000	3,000	2,691	4,000	9,000

Fuente: Elaboración propia

En el Gráfico 2 se observa de manera general que ninguna variable es simétrica, lo que concuerda con los resultados de la Media y la Mediana presentados en la Tabla 4. Como se aprecia, solo en el gráfico que representa a la variable posición no hay posibles valores atípicos, mientras que en el diagrama de cajas de las variables: Goles, Popularidad, Tiempo Extra, Tiros de esquina, Penales y Tarjetas, se muestran algunos posibles valores atípicos.

Comentado [HGCV11]: Pueden mencionar que este análisis lo hacen para explorar y describir sus datos, antes de realizar el análisis del modelamiento. Que es útil para describir la distribución de las variables, para identificar datos faltantes y posibles valores atípicos.

Comentado [HGCV12]: se muestran

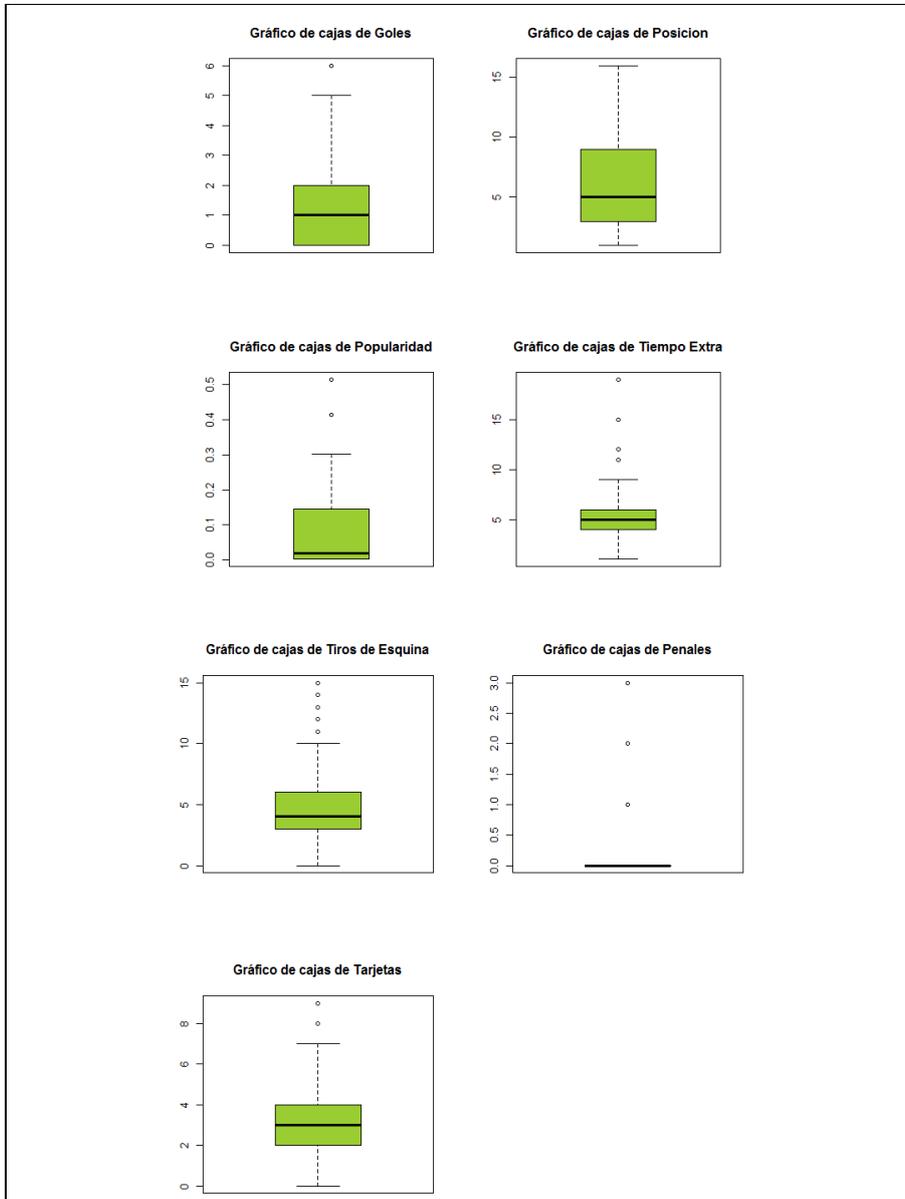


Gráfico 2 Diagramas de caja

Fuente: Elaboración propia

Como se aprecia en el Gráfico 3 los histogramas presentan una evidente desviación de la normal, una asimetría positiva y una baja densidad en la parte derecha.

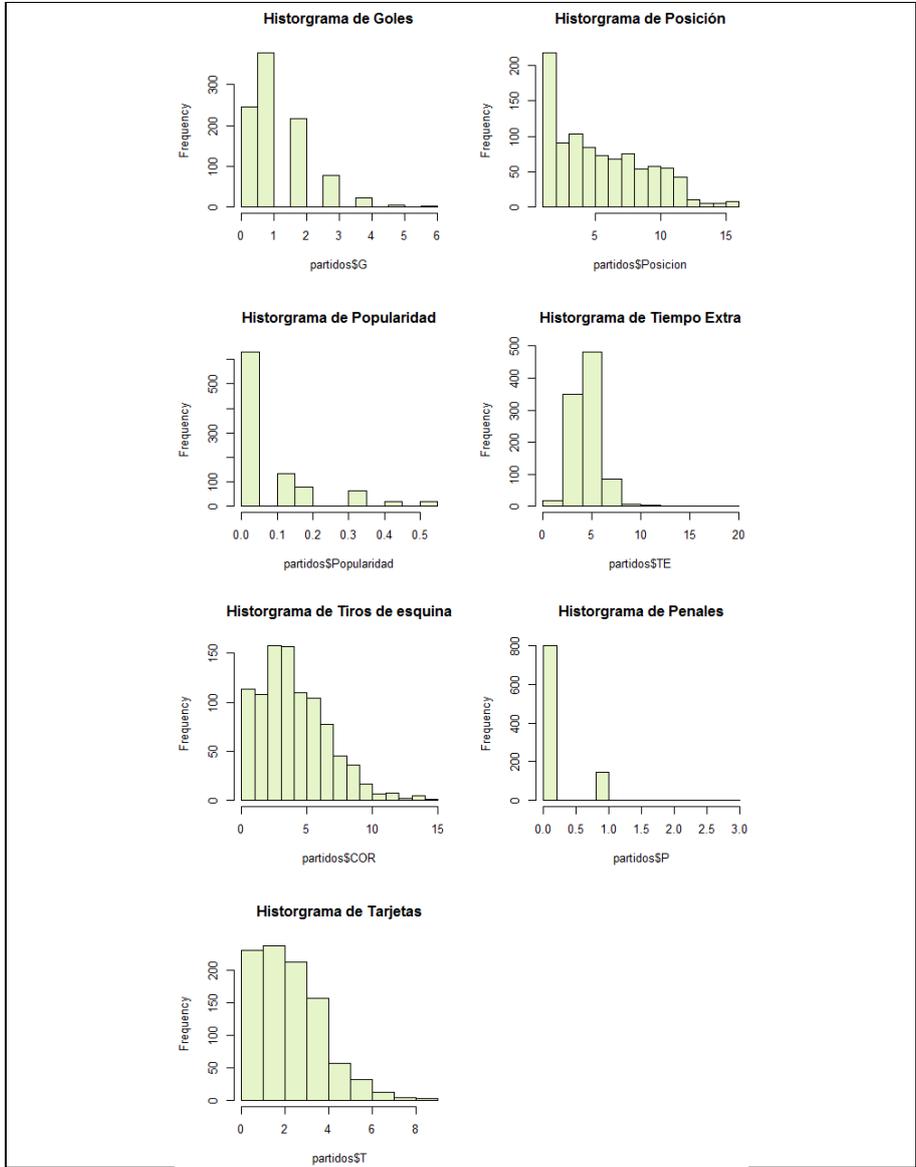


Gráfico 3 Histogramas de frecuencia

Fuente: Elaboración propia

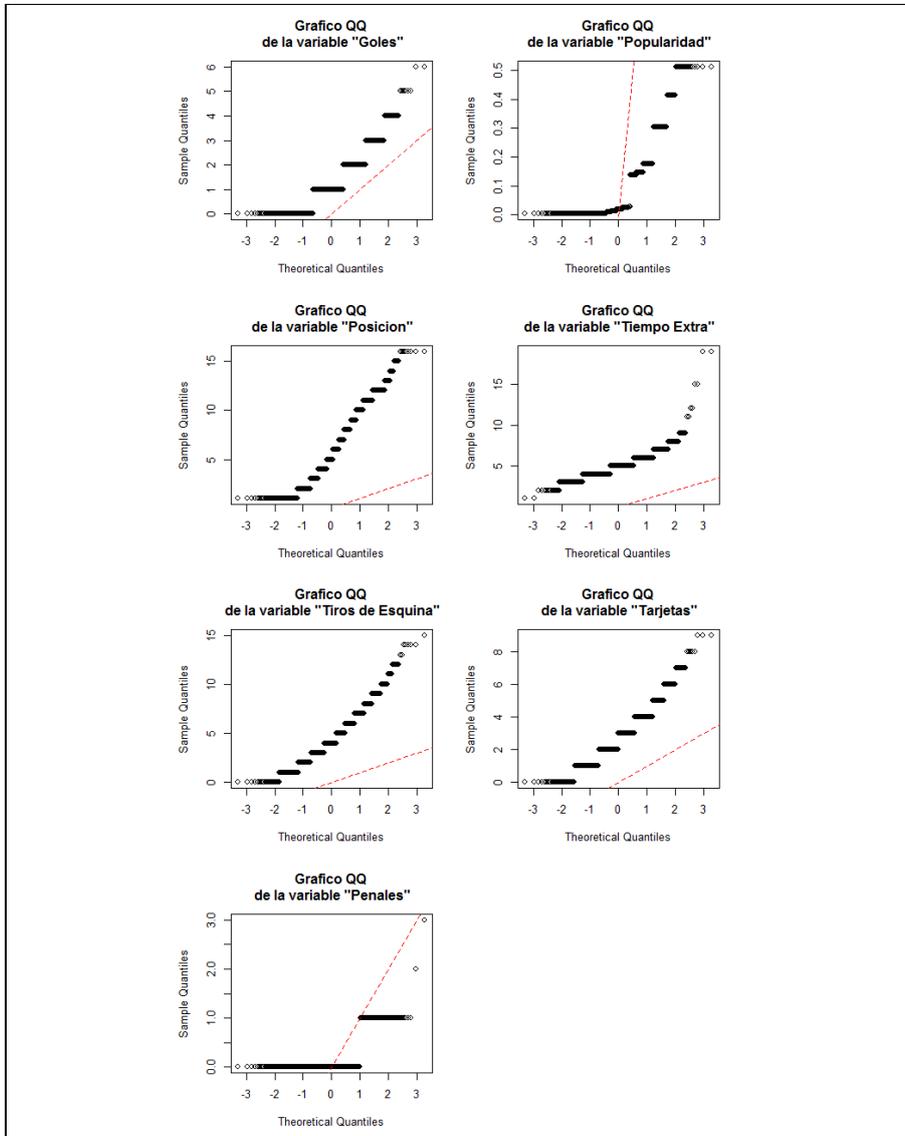


Gráfico 4 Gráficos QQ

Fuente: Elaboración propia

El Gráfico 5 representa las frecuencias de la variable público, donde se observa que en los datos recolectados los partidos sin público son menores en comparación de los partidos

Comentado [HGCV13]: tilde

con público, puesto que estos fueron tomados desde la sexta jornada de la primera etapa de la temporada del 2020 hasta la primera etapa de la temporada 2021.

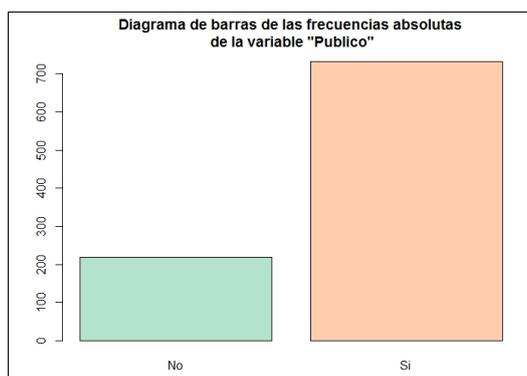


Gráfico 5 Diagrama de barras

Fuente: Elaboración propia

3.2 Ventaja en Casa

Para construir el modelo, se consideró la variable 'Goles' como la variable que mide la ventaja en casa y se decidió investigar el efecto de interacción entre dos variables importantes las cuales son Status que indica si el equipo está jugando de local o visitante, y una segunda variable público la cual indica si en el partido hubo o no público presente. Se tiene interés en investigar esta interacción puesto que uno de los objetivos es determinar si ocurren cambios significativos en cuanto a la ventaja en casa en los partidos de fútbol cuando no hay público, fenómeno que se presentó desde que se declaró el toque de queda a raíz de la pandemia por COVID-19.

Luego de contemplar la naturaleza de los datos, realizar las pruebas para evaluar los supuestos de normalidad e independencia y tener evidencia estadística para concluir que los datos no cumplen con ninguno de los supuestos de un modelo lineal clásico, se decidió trabajar con Modelos Lineales Mixtos.

Comentado [HGCV14]: Se debe especificar que se utilizó la variable "goles" para medir la ventaja de jugar en casa.

Comentado [HGCV15]: Investigar el efecto del efecto de interacción

Comentado [HGCV16]: indica

Comentado [HGCV17]: Tenemos interés en investigar esta interacción...

Comentado [HGCV18]: cambios significativos en cuanto a la ventaja en casa...

En el Software RStudio, se utiliza la función “lme()” con la cual se ajusta al modelo lineal de efectos mixtos, permite que los errores dentro del grupo estén correlacionados y / o tengan variaciones desiguales.

Comentado [HGCV19]: con la cual ...

```
>>> lme(fixed, data, random, correlation, weights, subset, method,
        na.action, control, contrasts = NULL, keep.data = TRUE)
```

Se realizaron 20 diferentes modelos con intercepto aleatorio para los equipos, los mismos que se pueden apreciar en la Tabla 5.

Tabla 5 Modelos de Ventaja en Casa

Modelos lineales mixtos con intercepto aleatorio	AIC	BIC
G ~ Público + Status + (Status * Público) + Posición	2723,252	2757,21
G ~ Público + Status + (Status * Público) + Posición + Popularidad	2725,392	2764,193
G ~ Público + Status + (Status * Público) + Posición + Región	2726,741	2765,542
G ~ Público + Status + (Status * Público) + Posición + Estación	2728,558	2767,359
G ~ Público + Status + (Status * Público) + Posición + Fecha	2733,198	2771,999
G ~ Público + Status + (Status * Público) + Posición + Región + Popularidad	2728,705	2772,347
G ~ Público + Status + (Status * Público) + Posición + P_Gana	2734,016	2772,817
G ~ Público + Status + (Status * Público) + Posición + Fecha + Popularidad	2735,393	2779,035
G ~ Público + Status + (Status * Público) + Posición + Fecha + Estación	2735,875	2779,516
G ~ Público + Status + (Status * Público) + Posición + Región + Fecha	2736,827	2780,469
G ~ Público + Status + (Status * Público) + Posición + Fecha + Popularidad + Estación	2738,15	2786,631

Comentado [HGCV20]: tilde

G ~ Público + Status + (Status * Público) + Posición + Región + Fecha + Estación	2739,532	2788,012
G ~ Público + Equipo + (Equipo * Público)	2759,24	2788,354
G ~ Público + Status + (Status * Público) + Posición + Fecha + Altura	2752,457	2796,099
G ~ Público+ Status + (Status*Público) +Estación+ Región+ Fecha + Posición + Score	2744,398	2797,714
G ~ Público + Status + (Status * Público) + Score + ELO	2762,105	2800,906
G ~ Público + Status + (Status * Público) + Score + P_Gana	2762,217	2801,018
G ~ Público + Status + (Status * Público) + P_Gana	2768,728	2802,687
G ~ Público + Status + (Status * Público) + ELO	2769,06	2803,018
G ~ Público + Status + (Status * Público) + P_Gana + Popularidad	2768,129	2806,.93

Fuente: Elaboración propia

En el Gráfico 6 se observa el valor de cada correlación entre todas las variables numéricas que se utilizaron en los modelos de la Tabla 5, de los cuales se puede destacar que hay una correlación estadísticamente significativa y negativa entre posición y score del equipo, y una correlación estadísticamente significativa y positiva entre popularidad y ELO, es decir, -0.70 y 0.54, respectivamente. Es necesario aclarar que la variable score se relaciona de manera inversa con la variable posición, puesto que el equipo con mayor valor de score (valor 10) se encuentra en la primera posición en la tabla (posición 1).

Comentado [HGCV21]: ¿Cuál es la ilustración 2?

Comentado [HGCV22]: estadísticamente significativa

Comentado [HGCV23]: estadísticamente significativa

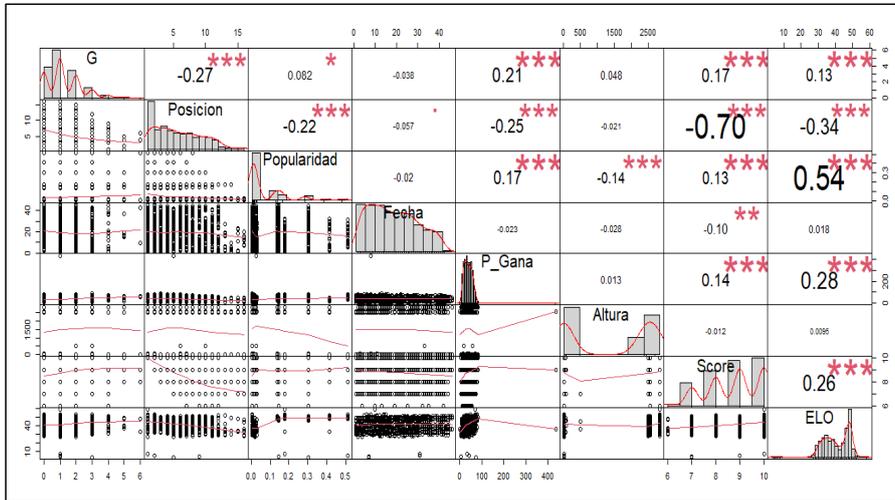


Gráfico 6 Matriz de correlación

Fuente: Elaboración propia

En base a los criterios de decisión AIC y BIC se seleccionó el modelo que describe mejor la relación que se investiga. A continuación, se presenta a detalle el modelo seleccionado para describir la ventaja en casa:

Comentado [HGCV24]: seleccionó

3.2.1 Modelo final.

$$\text{modelo1} < - \text{lme}(G \sim \text{Publico} + \text{Status} + (\text{Status} * \text{Publico}) + \text{Posicion},$$

$$\text{random} = \sim 1 | \text{Equipo}, \text{data} = \text{partidos})$$

3.2.1.1 Supuestos del modelo

Para un modelo lineal mixto, los términos de los errores deben cumplir con los siguientes supuestos:

Comentado [HGCV25]: términos de los errores

- Homogeneidad de la varianza,
- Estar aproximadamente distribuidos de manera normal y;
- Ser independientes.

En el Anexo A se puede observar un resumen detallado de los resultados del modelo, el cual indica que las estimaciones se realizan considerando como nivel de referencia para la variable Público cuando el partido se juega sin público, mientras que para la variable Status se consideró como nivel de referencia cuando el equipo juega de local. Además, el parámetro de interés, es decir, el término de interacción (Status*Público) está representado por β_4 en el modelo que se presenta a continuación:

$$\text{modelo: } \text{Goles} = \beta_0 + \beta_1 \text{Publico} + \beta_2 \text{Status} + \beta_3 \text{Posicion} + \beta_4 (\text{Status} * \text{Publico}) + e, \sim (1 | \text{Equipo})$$

Tabla 6 Estimaciones del modelo

	Estimaciones	Valor p
\hat{b}_0	2.0916559	0.0000
\hat{b}_1	-0.2799257	0.0103
\hat{b}_2	-0.6360224	0.0000
\hat{b}_3	-0.0737404	0.0000
\hat{b}_4	0.2892051	0.0614

Fuente: Elaboración propia

Considerando un nivel de significancia de $\alpha = 10\%$, en la Tabla 6 se observa diferentes estimaciones, donde \hat{b}_1 estima que el equipo local marca en promedio 0.28 goles menos cuando juega con público en comparación cuando juega sin público. El \hat{b}_2 estima que cuando no hay público, el equipo visitante convierte 0.64 goles menos que el equipo local. La interacción, es decir, el \hat{b}_4 estima que cuando hay público versus cuando no hay público presente en el estadio la diferencia de goles entre el equipo visitante y el equipo local aumenta 0.28 en promedio. Además, se puede observar que todas las variables consideradas para construir el modelo son estadísticamente significativas y aportan información para el

Comentado [HGCV26]: como nivel de referencia para la variable X cuando....

Comentado [HGCV27]: Mientras que para la variable X se consideró como nivel de referencia cuando el....

Comentado [HGCV28]: sin el sombrero

Comentado [HGCV29]: betas sin sombreros!!

Comentado [HGCV30]: Modelo 1: Goles = Además, el modelo no está correctamente escrito. Por favor incluir el intercepto aleatorio del equipo, el término de los errores epsilon y los supuestos del modelo.. Incluir también cuales son los niveles de referencia de las variables indicadoras. Es también importante que especifiquen que el parámetro de interés es el parámetro del término de interacción.

efecto de interés. En el caso de la interacción de Status y Público se consideran los mismos niveles de referencia anteriormente mencionados, es decir, para la variable Público cuando el partido se juega sin público y para la variable Status cuando el equipo juega de local.

Se muestra que no se cumple con la creencia común que plantea que al jugar en casa el equipo local tiene más ventaja, puesto que cuenta con “las porras de los aficionados”. Sin embargo, en base a los datos recolectados se aprecia que, en la Serie A del fútbol ecuatoriano, cuando no hay público el Equipo Local tiende a anotar una cantidad mayor de goles que el visitante, estos resultados contrarrestan la ideología popular de que el local convierte menos goles sin la ayuda del “jugador 12”, pero como se menciona en el estudio similar sobre la ventaja de jugar en casa realizado por (Benz & Lopez, 2021) y (Jiménez Sánchez, Lavín, & Endara, 2021), los cuales indican que en las ligas menores ocurre el mismo comportamiento de que ha visto con los datos de la Serie A ecuatoriana, es decir, el público no influye lo suficiente en los partidos para determinarlo como un factor significativo en el resultado del encuentro.

Variabes como SOT, COR y POS también fueron consideradas en el estudio como variables de respuesta de los modelos, sin embargo no hubo significancia de las variables predictoras en los modelos y a su vez, no se obtuvo un efecto significativo de la interacción entre Público y Status que son las variables de interés principal.

Comentado [HGCV32]: ¿cuáles son esos estudios?

Comentado [HGCV33]: ¿cuáles variables?

Comentado [HGCV34]: Quiso decir: ¿No se obtuvo un efecto significativo?

3.2.2 Evaluación de los Supuestos

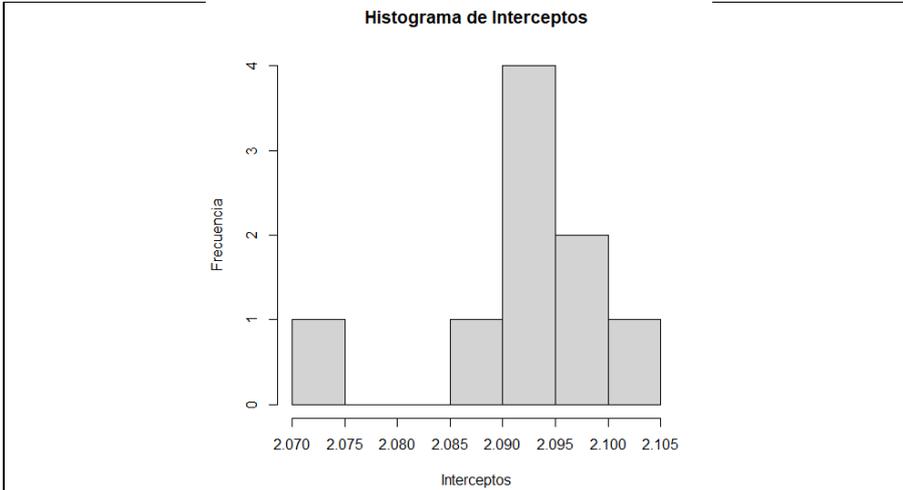


Gráfico 7 Histograma de interceptos

Fuente: Elaboración propia

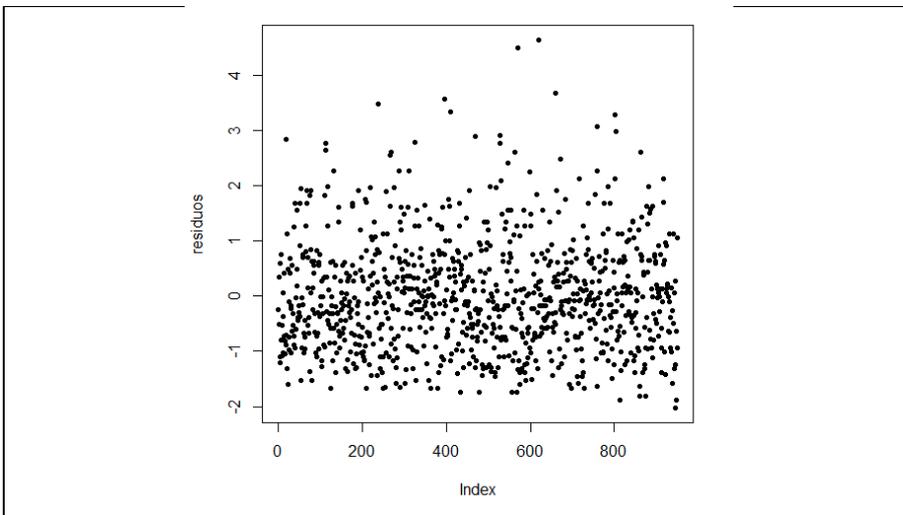


Gráfico 8 Independencia de residuos

Fuente: Elaboración propia

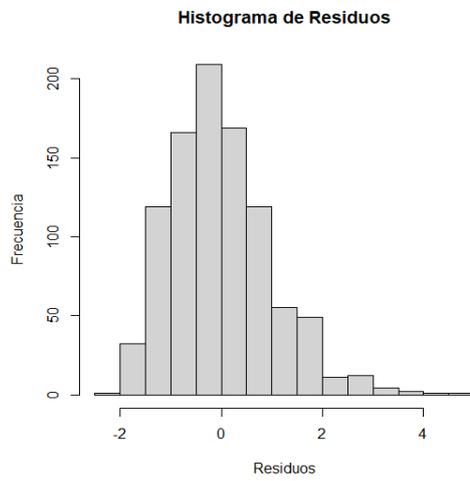


Gráfico 9 Histograma de residuos

Fuente: Elaboración propia

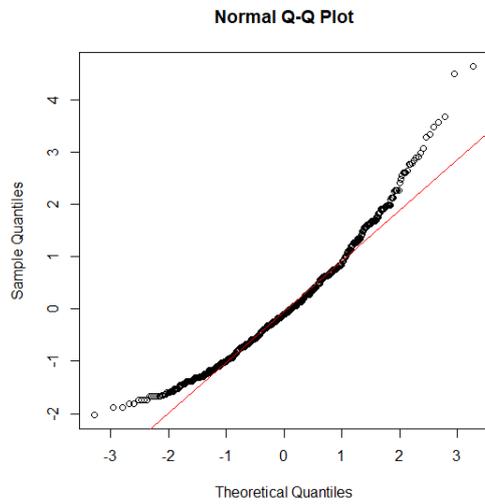


Gráfico 10 Gráfico QQ de residuos

Fuente: Elaboración propia

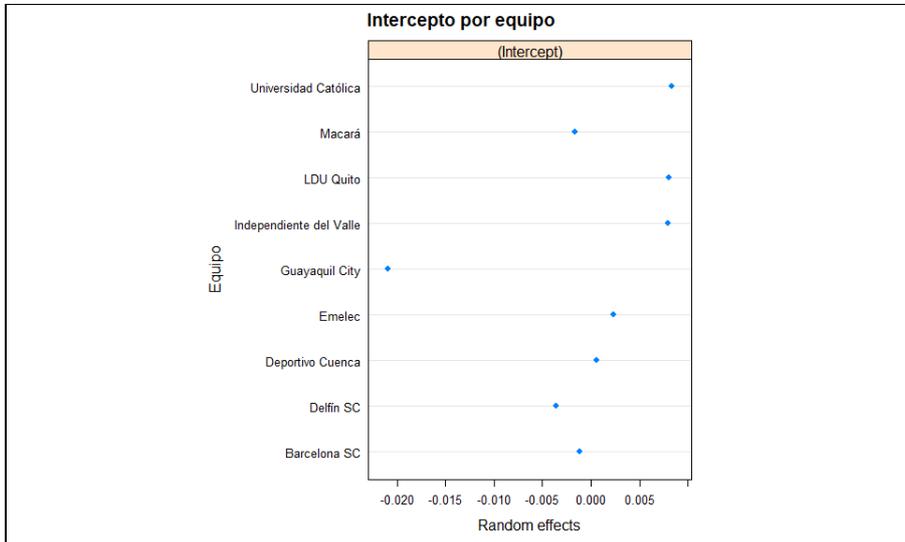


Gráfico 11 Interceptos del modelo

Fuente: Elaboración propia

En el Gráfico 7 se aprecia que los interceptos del modelo no presentan simetría, por lo que se observa que no hay normalidad, sin embargo se puede utilizar el Teorema del Límite Central en estos datos, debido a que el tamaño de muestra es lo suficientemente grande. De igual manera, en la Gráfica 9 se observa desviaciones de la normalidad en los residuos del modelo, se presenta una tendencia positiva. En la Gráfica 8 no se aprecia ningún patrón claro para los residuos de los modelos, lo que indicaría que no hay dependencia.

Como se observa en el Gráfico 10, en los extremos se desvian los residuos, es decir, todos los valores no se ajustan a la recta, sin embargo, este modelo es robusto por lo que aunque existan pequeñas desviaciones de los supuestos el modelo produce inferencias aproximadamente válidas.

3.3 Sesgo Arbitral

Para construir el modelo se consideró la variable Tarjetas totales 'T' como variable que mide el sesgo arbitral en favor del equipo local, también se consideró importante

Comentado [HGCV35]: Los gráficos no hacen inferencia

Comentado [HGCV36]: Sin embargo, podemos utilizar el teorema del límite central en estos datos.

Comentado [HGCV37]: Desviaciones de la normalidad

Comentado [HGCV38]: Gráfica 8

Comentado [HGCV39]: produce inferencias aproximadamente válidas.

investigar el efecto de interacción entre las variables Público y Status debido a que uno de los objetivos del estudio es conocer si el público influye de manera significativa en las decisiones del árbitro, las variables predictoras: Fecha, Penales ‘P’, Popularidad, Tiempo extra ‘TE’, Posesión del balón ‘POS’, Córner ‘COR’, Remates a portería ‘SOT’, Público y Status. Al igual que en el caso de Ventaja en Casa, se utilizó el modelo lineal mixto para el estudio del efecto de interés, se usó la función ‘lme()’ para ajustar al modelo

Comentado [HGCV40]: colocar esto al inicio de la oración, esto es lo más importante que deben de dar a conocer al lector!

Comentado [HGCV41]: del efecto de interés (indicar cuál es el parámetro que se investiga)

En la Tabla 7 se logra visualizar las variables de interés para el Sesgo Arbitral, de la que se puede destacar que en el 2019 no se presentaron partidos sin público.

Tabla 7 Resumen de resultados de los partidos

Año	Público	Goles		Tarjetas Amarillas		Tarjetas Rojas		Goles Penales		Penales Fallidos		Tiempo Extra
		Local	Visitante	Local	Visitante	Local	Visitante	Local	Visitante	Local	Visitante	
2017	No	7	2	10	12	0	0	2	1	0	0	18
	Si	195	138	366	379	24	25	14	9	0	0	641
2018	No	8	5	12	8	1	2	2	0	0	0	24
	Si	207	152	299	343	19	24	21	23	7	3	687
2019	Si	96	74	187	212	4	4	14	10	0	0	384
2020	No	109	60	145	155	10	8	10	11	1	1	360
	Si	14	17	23	34	1	1	3	2	0	1	50
2021	No	58	35	85	78	6	1	6	5	5	2	210

Fuente: Elaboración propia

Se realizaron 8 modelos con intercepto aleatorio para los equipos, los que se pueden apreciar en la Tabla 8.

Tabla 8 Modelos de Sesgo Arbitral

Modelos lineales mixtos con intercepto aleatorio	AIC	BIC
T ~ Público + Status + (Status * Público) + Posición	3613,74	3647,699
T ~ Público + Status + (Status * Público) + Posición + Fecha	3624,184	3662,985
T ~ Público + Status + (Status * Público) + Posición + Popularidad	3609,33	3648,131

Comentado [HGCV42]: tilde

T ~ Público + Status + (Status * Público) + Posición + POS + SOT + P +Popularidad	3583,909	3632,389
T ~ Público + Status + (Status * Público) + Posición + P + TE + COR + Popularidad	3586,401	3634,881
T ~ Público + Status + (Status * Público) + P + TE + Popularidad	3582,571	3626,212
T ~ Público + Status + (Status * Público) + P + TE + COR	3590,769	3634,411
T ~ Público + Status + (Status * Público) + P + TE	3586,745	3625,546

Fuente: Elaboración propia

En la Gráfica 12 se tiene las correlaciones de las variables cuantitativas que se consideraron para el modelo del sesgo arbitral, de las que destacan la correlación positiva de 0.38 entre SOT y COR, la correlación positiva de 0.36 entre POS y COR, y la correlación entre POS y SOT de 0.31. Se observa que las demás variables no tienen una correlación mayor entre ellas.

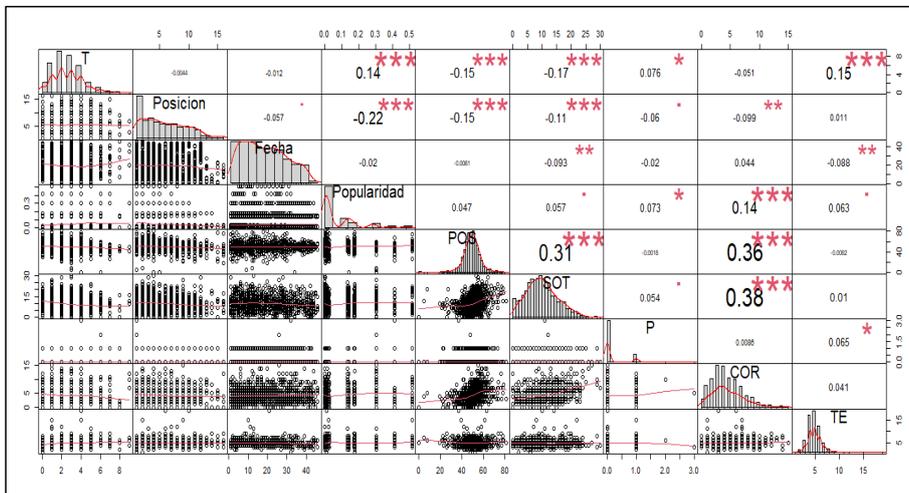


Gráfico 12 Matriz de correlación

Fuente: Elaboración propia

Se seleccionó el modelo con los menores valores de AIC y BIC, el cual presenta una mejor descripción del sesgo arbitral. A continuación, se presenta el modelo:

3.3.1 Modelo final.

$$\text{modelo2} = \text{lme}(T \sim \text{Público} + \text{Status} + (\text{Público} * \text{Status}) + P + TE + \text{Popularidad},$$

$$\text{random} = \sim 1 | \text{Equipo}, \text{ data} = \text{partidos})$$

3.3.1.1 Supuestos del modelo

Para un modelo lineal mixto, los términos de los errores deben cumplir con los siguientes supuestos:

- Homogeneidad de la varianza,
- Estar aproximadamente distribuidos de manera normal y;
- Ser independientes.

En el Anexo B se muestra con detalle el resultado del modelo, el mismo que para las estimaciones se considera como nivel de referencia para la variable Público cuando los partidos se llevan a cabo sin público y para la variable Status se considera como nivel de referencia cuando el equipo juega de local. El parámetro de interés es la interacción de Status y Público, el cual está representado por β_6 y se consideró como nivel de referencia para Status cuando el equipo es local y para Público cuando el partido se juega sin espectadores. El modelo está presentado de la siguiente manera:

$$\text{modelo: Tarjetas} = \beta_0 + \beta_1 \text{Público} + \beta_2 \text{Status} + \beta_3 P + \beta_4 TE + \beta_5 \text{Popularidad} + \beta_6 (\text{Publico} * \text{Status}) + e, \sim (1 | \text{Equipo})$$

Tabla 9 Estimaciones del modelo

	Estimaciones	Valor p
\hat{b}_0	1.3136711	0.0000

Comentado [HGCV43]: Se debe explicar qué variable se utilizó para medir el sesgo arbitral.

Comentado [HGCV44R43]: No se ha considerado este comentario.

Comentado [JB45R43]: Se encuentra en la introducción del sesgo arbitral.

Comentado [HGCV46]: Reescribir esto como lo hicieron en la sección de ventaja en casa.

Comentado [JB47R46]: El modelo está descrito después.

Comentado [HGCV49]: Corregir esto como en la sección de ventaja en casa.

Comentado [HGCV48]: Se debe escribir el modelo seleccionado, para que el lector pueda seguir la interpretación que hacen sobre los parámetros del modelo.

\hat{b}_1	0.2982302	0.0867
\hat{b}_2	0.0085974	0.9678
\hat{b}_3	0.2877530	0.0340
\hat{b}_4	0.1697232	0.0000
\hat{b}_5	1.6741602	0.0002
\hat{b}_6	0.2874902	0.2361

Fuente: Elaboración propia

Trabajando con un nivel de significancia de $\alpha=10\%$, en la Tabla 9 se observan algunas estimaciones, de las cuales \hat{b}_1 estima que el equipo local obtiene en promedio 0.30 tarjetas más cuando juega con público en comparación cuando juega sin público. El \hat{b}_2 estima que cuando se juega sin público, el equipo visitante es sancionado en promedio con 0.009 tarjetas más que el equipo local. La estimación de \hat{b}_3 indica que un incremento de un punto en la popularidad del equipo se asocia con un incremento de 1.67 en el número de tarjetas recibidas en promedio por partido. La interacción, representada por el \hat{b}_6 , estima que cuando hay público a diferencia de cuando no hay público presente en el estadio la diferencia del número de tarjetas entre equipo visitante y local aumenta en 0.29 tarjetas en promedio. A su vez, se puede observar que algunas de las variables consideradas para construir el modelo son estadísticamente significativas. Sin embargo, se tiene variables como Status y la interacción entre Status y Público que no son estadísticamente significativas.

3.3.2 Evaluación de los supuestos

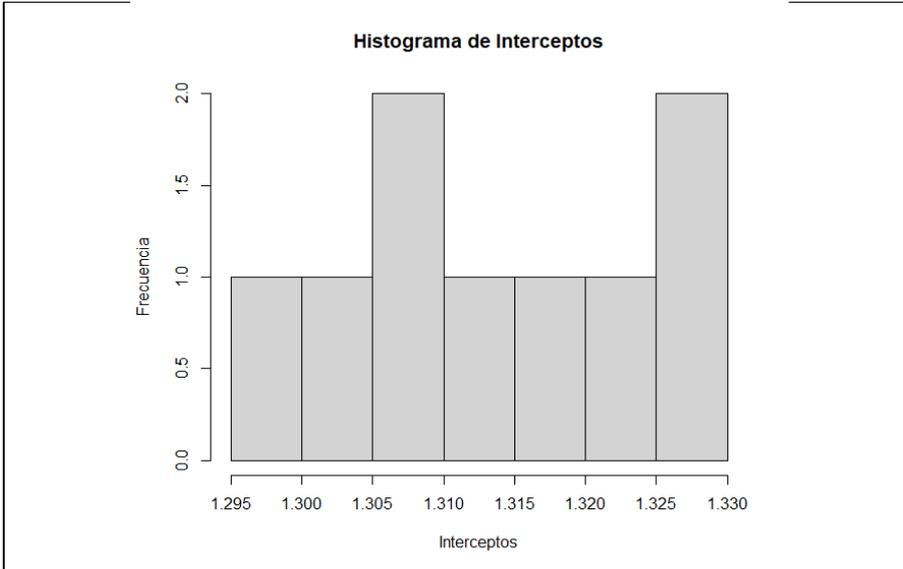


Gráfico 13 Histograma de interceptos

Fuente: Elaboración propia

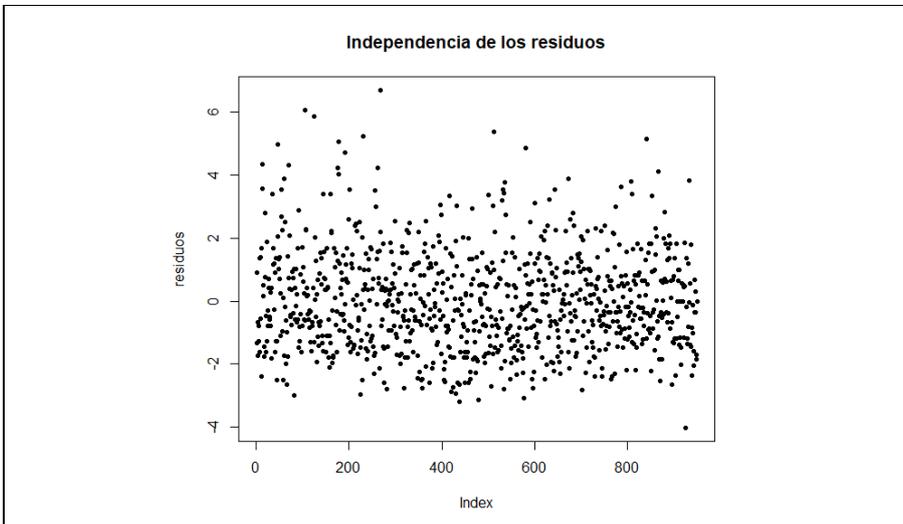


Gráfico 14 Independencia de residuos

Fuente: Elaboración propia

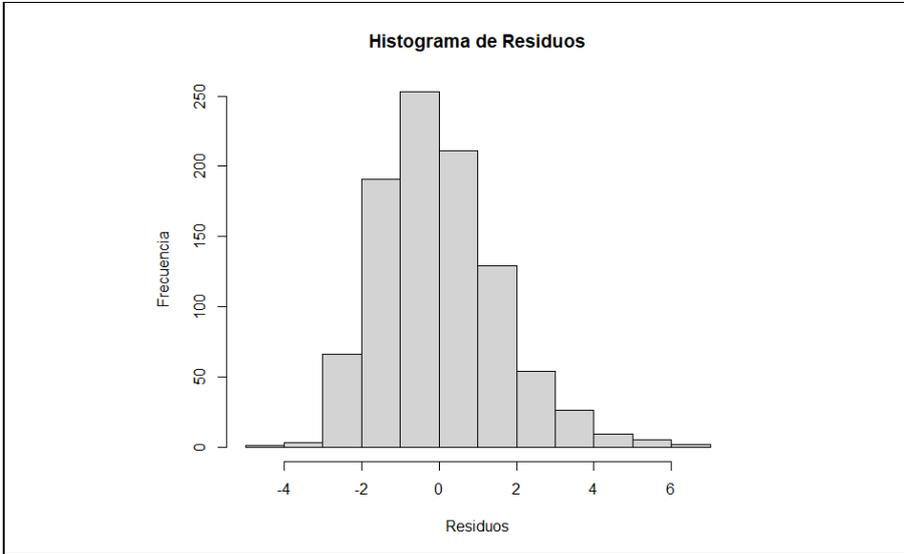


Gráfico 15 Histograma de residuos

Fuente: Elaboración propia

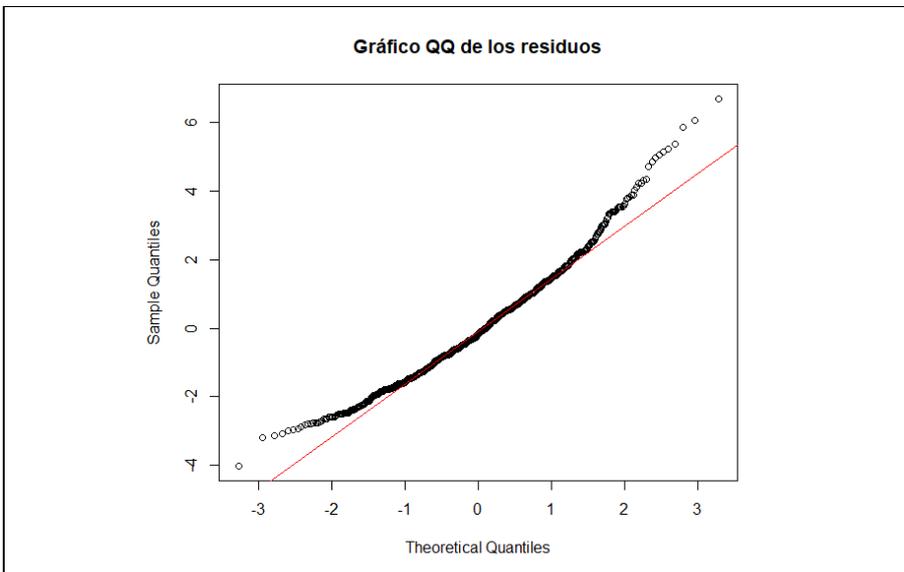


Gráfico 16oie Gráfico QQ de residuos

Fuente: Elaboración propia

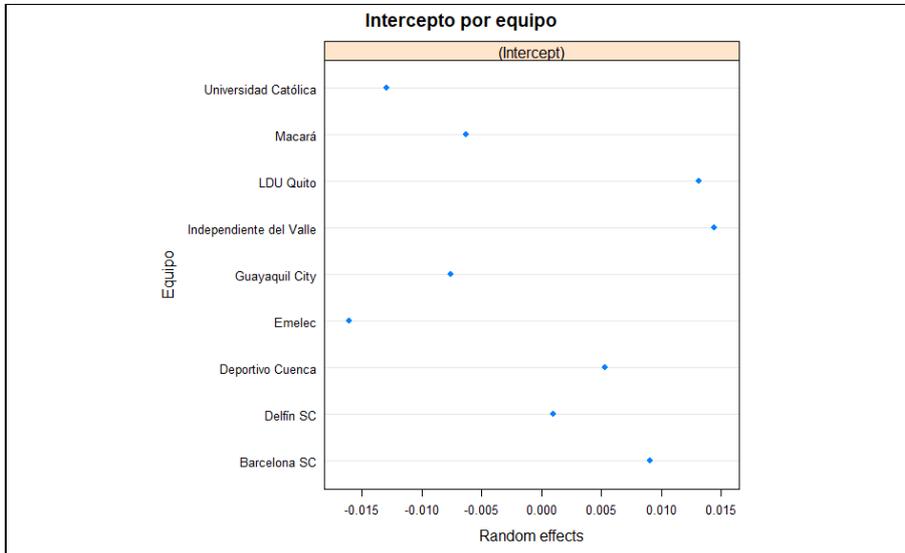


Gráfico 17 Intercepto del modelo

Fuente: Elaboración propia

En el Gráfico 13 se observa que los interceptos del modelo no presentan simetría, lo que indicaría que no hay normalidad, sin embargo, en base a lo enunciado en el Capítulo 1 se utiliza el Teorema del Limite Central. Además, en la Gráfica 15 se presentan desviaciones de normalidad en los residuos. En cuanto a la independencia de los residuos, en la Gráfica 14 se aprecia que no hay un patrón en los residuos, por lo que se puede inferir que no hay dependencia.

En el Gráfico 16 se contempla que es en los extremos donde se desvían los residuos, es decir, todos los valores no se ajustan a la recta, sin embargo, pese a que existan pequeñas desviaciones de los supuestos, el modelo se caracteriza por producir inferencias aproximadamente válidas.

Comentado [HGCV50]: se caracteriza por producir inferencias aproximadamente válidas

Comentado [HGCV51]: aquí también tomar en cuenta los comentarios de la sección 3.2.4

4. CAPÍTULO 4

La metodología para construir los modelos planteados para observar la ventaja en casa y el sesgo del árbitro, se basó en evaluar datos del 2017 hasta la primera temporada del 2021. Se trabajó con la interacción de las variables Público y Status lo que permitió determinar el efecto que tienen estas variables sobre los resultados del juego.

4.1 Conclusión

- El modelo lineal mixto, permite analizar los datos de fútbol recolectados, a pesar de que se encontró ligeras desviaciones de los supuestos de este modelo, se lo aplicó considerando la estructura de los datos y tratando de corregir con variables de confusión.
- El modelo construido para la ventaja en casa señala un comportamiento contrario a no esperado. En los resultados se aprecia que en la serie A de fútbol ecuatoriano cuando el equipo local no tiene público tiende de forma significativa a convertir más goles que cuando tiene público, a un nivel de significancia del 0.1.
- La posición del equipo está presente en los modelos finales presentados en la ventaja en casa, y es significativa, lo que indica que esta variable aporta información importante al modelo.
- A pesar del efecto de la interacción no ser significativo en el sesgo arbitral con un nivel de significancia de 0.1, la estimación del efecto sugiere que el árbitro es influenciado por la presencia de público en el partido, es decir, cuando hay público presente en el estadio el árbitro tiende a mostrar más tarjetas a los jugadores en cambio, cuando no hay público las sanciones por tarjetas a jugadores disminuye.

Comentado [HGCV52]: No veo que en el documento haya evaluado los supuestos del modelo con efectos mixtos.

Comentado [HGCV53]: A pesar que se encontró ligeras desviaciones de los supuestos de este modelo

Comentado [JB54R53]: sale un mensaje de gramática para corregir por 'a pesar de que.'

Comentado [HGCV55]: ¿en cuál modelo?

Comentado [HGCV56]: la estimación del efecto sugiere que el

Comentado [HGCV57]: ¿Este efecto fue significativo o no? ¿Qué nivel de significancia se utilizó?

4.2 Recomendación

- Para mejores resultados, conseguir más variables que traten de representar el problema, dado que, actualmente, la Liga Pro no publica los registros de los partidos de la Serie A, se dificultó la obtención de más variables que pudieron haber contribuido de manera positiva en el modelo planteado para los dos casos a estudiar.
- Los resultados obtenidos pueden ser objeto de sesgo por la falta de datos de calidad disponibles, para analizar a fondo el problema planteado.
- Consultar y comparar resultados con estudios, similares hechos ya sea de años recientes o anteriores, ya que dan una pauta de qué camino tomar para la construcción del modelo y de las variables a considerar.

Comentado [HGCV58]: Pueden comentar sobre el hecho de que los resultados que obtuvieron pueden ser objeto de sesgo por la falta de datos de calidad disponibles para analizar el problema que se investigó.

5. BIBLIOGRAFÍA

- Barcelona Sporting Club. (s.f.). *Barcelona Sporting Club*. Obtenido de <https://barcelonasc.com.ec/>
- Allaire, J. (28 de Febrero de 2011). *RStudio*. Obtenido de RStudio: <https://www.rstudio.com/>
- Bandera Fernández, E., & Pérez Pelea, L. (14 de Diciembre de 2017). *Instituto Nacional de Ciencias Agrícolas*. Obtenido de Instituto Nacional de Ciencias Agrícolas: <https://www.redalyc.org/journal/1932/193260614019/>
- Benz, L., & Lopez, M. (2021, Mayo 28). *Cornell University*. Retrieved from Cornell University: <https://arxiv.org/abs/2012.14949>
- Club Deportivo Cuenca. (s.f.). *Club Deportivo Cuenca*. Obtenido de <https://clubdeportivocuenca.com/>
- Club Deportivo Macará. (s.f.). *Club Deportivo Macará*. Obtenido de Club Deportivo Macará: <https://clubmacara.ec/>
- Club Sport Emelec. (s.f.). *Club Sport Emelec*. Obtenido de Club Sport Emelec: <https://emelec.com.ec/>
- De la Cruz, F. (2008). Modelos multinivel. *Revista Peruana de Epidemiología*.
- EL COMERCIO. (3 de 8 de 2020). La LigaPro publicó el calendario para la reanudación del campeonato nacional de fútbol el viernes 14 de agosto. *EL COMERCIO*.
- EL UNIVERSO. (14 de 03 de 2021). El fútbol de Ecuador: un año sin público en estadios.
- Escobar, G. A. (2017). *Un modelo lineal mixto con covariable funcional aplicado a datos de concentraci3n de clorofila*. Santiago de Cali: Universidad del Valle.

Evelyn Bandera-Fernández, L. P.-P. (2018). *Los modelos lineales generalizados mixtos. Su aplicación en el mejoramiento de plantas*. La Habana: Universidad de La Habana.

Garicano, L., Palacios, I., & Prendergast, C. (2001). FAVORITISM UNDER SOCIAL PRESSURE. *NATIONAL BUREAU OF ECONOMIC RESEARCH*.

Guayaquil City. (s.f.). *Guayaquil City*. Obtenido de Guayaquil City:

<https://www.guayaquilcityfc.com/>

Heredia, M. (Noviembre de 2006). *BeSoccer*. Obtenido de BeSoccer: <https://es.besoccer.com/>

Independiente del Valle . (s.f.). *Independiente del Valle* . Obtenido de Independiente del

Valle : <http://www.independientedelvalle.com/>

Jiménez Sánchez, Á., Lavín, J., & Endara, F. (24 de 01 de 2021). Repercusiones de jugar sin público en la ventaja local, las decisiones arbitrales y en los componentes del juego.

Cuadernos de Psicología del Deporte, 198-212.

Juan Carlos Correa Morales, J. C. (2016). *Introducción a los modelos mixtos*. Medellín:

Universidad Nacional de Colombia Sede Medellín.

Karlis, D., & Ntzoufras, I. (2000). On Modelling Soccer Data. *Department of Statistics,*

Athens University of Economics and Business.

Liga Rey de Copas. (s.f.). *LIGA DEPORTIVA UNIVERSITARIA*. Obtenido de

<https://ldu.com.ec/web/>

LigaPro. (20 de Mayo de 2020). *Protocolo de Medidas de Bioseguridad para las*

Competiciones Organizadas por la LigaPro. Obtenido de LigaPro:

<https://ligapro.ec/legals>

LigaPro Ecuador. (2018). *Liga Profesional de Fútbol del Ecuador*. Obtenido de Liga Profesional de Fútbol del Ecuador: <https://www.ligapro.ec/history/season/2019>

Minitab 18. (2019). *Soporte de Minitab 18*. Obtenido de <https://support.minitab.com/es-mx/minitab/18/help-and-how-to/statistics/basic-statistics/supporting-topics/data-concepts/about-the-central-limit-theorem/>

Montenegro Galárraga, P. (09 de Febrero de 2020). *Amanecer Deportivo*. Obtenido de Amanecer Deportivo: <https://amanecerdeportivoec.com/uncategorized/historia-del-futbol-ecuatoriano/>

Organizacion Panamericana de la Salud. (11 de 03 de 2020). *Organizacion Panamericana de la Salud*. Obtenido de <https://www.paho.org/es/noticias/11-3-2020-oms-caracteriza-covid-19-como-pandemia>

Paredes Bonifaz, A. S. (18 de Febrero de 2020). *Universidad Nacional de Chimborazo*. Obtenido de <http://dspace.unach.edu.ec/bitstream/51000/6695/1/UNACH-EC-FCEHT-DS%C3%91-GRF-2020-000005.pdf>

Pérez, G. R. (2021 de Marzo de 2021). *EL PAÍS*. Obtenido de La pandemia pincha la burbuja económica del fútbol: <https://elpais.com/economia/2021-03-06/la-pandemia-pincha-la-burbuja-economica-del-futbol.html>

PES Logos. (s.f.). Obtenido de <https://peslogos.blogspot.com/2017/07/delfin-sporting-club.html>

Pinheiro, J., Bates, D., & R-core. (04 de Febrero de 2021). *R-project.org*. Obtenido de <https://cran.r-project.org/web/packages/nlme/>

Profesor en línea. (2015). *Profesor en línea*. Obtenido de Profesor en línea: <https://www.profesorenlinea.cl/Deportes/FutbolHistoria.htm>

Saavedra García, M., Gutiérrez Aguilar, O., Fernández Romero, J., & Sa Marques, P. (2012).

VENTAJA DE JUGAR EN CASA EN EL FÚTBOL ESPAÑOL (1928-2011).

Revista Internacional de Medicina y Ciencias de la Actividad Física y del Deporte,
15.

SPSS Statistics. (s.f.). IBM. Obtenido de IBM: [https://www.ibm.com/docs/es/spss-](https://www.ibm.com/docs/es/spss-statistics/25.0.0?topic=SSLVMB_25.0.0/spss/advanced/idh_mix1.html#idh_mix1)

[statistics/25.0.0?topic=SSLVMB_25.0.0/spss/advanced/idh_mix1.html#idh_mix1](https://www.ibm.com/docs/es/spss-statistics/25.0.0?topic=SSLVMB_25.0.0/spss/advanced/idh_mix1.html#idh_mix1)

U. CATÓLICA OFICIAL. (s.f.). U. CATÓLICA OFICIAL. Obtenido de U. CATÓLICA

OFICIAL: <https://twitter.com/ucaticolicaec?lang=es>

UCLA. (s.f.). *Institute for Digital Research/ Education Statistical Consulting*. Obtenido de

Institute for Digital Research/ Education Statistical Consulting:

<https://stats.idre.ucla.edu/other/mult-pkg/introduction-to-linear-mixed-models/>

Vacchiano, M., & Merino Noé, J. (2018). *INCASI Working Paper Series*. Obtenido de

INCASI Working Paper Series:

https://ddd.uab.cat/pub/worpaper/2018/188612/INCASI_WPSa2018n2.pdf

worldfootball.net. (2000). *worldfootball.net*. Obtenido de worldfootball.net:

<https://ecu.worldfootball.net/competicion/ecu-serie-a/>

6. ANEXOS

Anexo A Resumen del modelo de Ventaja de casa

```

Linear mixed-effects model fit by REML
Data: partidos
      AIC      BIC    logLik
2723.252 2757.21 -1354.626

Random effects:
Formula: ~1 | Equipo
(Intercept) Residual
StdDev: 0.03068717 0.9979442

Fixed effects: G ~ Publico + Status + (Status * Publico) + Posicion
              Value Std.Error DF t-value p-value
(Intercept)  2.0916559 0.10927339 937 19.141494 0.0000
PublicoSi    -0.2799357 0.10891587 937 -2.570201 0.0103
StatusVisitante -0.6360224 0.13573669 937 -4.685707 0.0000
Posicion     -0.0737404 0.00908045 937 -8.120783 0.0000
PublicoSi:StatusVisitante 0.2892051 0.15439548 937 1.873145 0.0614
Correlation:
              (Intr) Publics SttsVs Posicn
PublicoSi    -0.773
StatusVisitante -0.574 0.617
Posicion     -0.475 0.010 -0.089
PublicoSi:StatusVisitante 0.509 -0.705 -0.878 0.069

Standardized Within-Group Residuals:
      Min       Q1       Med       Q3       Max
-2.02441126 -0.70798823 -0.09873784 0.60171634 4.64146843

Number of Observations: 950
Number of Groups: 9

```

Fuente: Elaboración propia

Anexo B Resumen del modelo de Sesgo Arbitral

```

Linear mixed-effects model fit by REML
Data: partidos
      AIC      BIC    logLik
3582.571 3626.212 -1782.285

Random effects:
Formula: ~1 | Equipo
(Intercept) Residual
StdDev: 0.04367041 1.570641

Fixed effects: T ~ Publico + Status + (Status * Publico) + P + TE + Popularidad
              Value Std.Error DF t-value p-value
(Intercept)  1.3136711 0.2449770 935 5.362426 0.0000
PublicoSi    0.2982302 0.1739279 935 1.714677 0.0867
StatusVisitante 0.0085974 0.2129704 935 0.040369 0.9678
P            0.2877530 0.1355447 935 2.122937 0.0340
TE           0.1697232 0.0336493 935 5.043888 0.0000
Popularidad  1.6741602 0.4402096 935 3.803098 0.0002
PublicoSi:StatusVisitante 0.2874902 0.2424857 935 1.185597 0.2361
Correlation:
              (Intr) Publics SttsVs P      TE      Pp1rdd
PublicoSi    -0.661
StatusVisitante -0.440 0.614
P            -0.084 0.050 0.034
TE           -0.756 0.149 -0.003 -0.047
Popularidad  -0.136 0.046 0.022 -0.064 -0.048
PublicoSi:StatusVisitante 0.385 -0.698 -0.878 -0.013 0.002 -0.020

Standardized Within-Group Residuals:
      Min       Q1       Med       Q3       Max
-2.5574766 -0.7250452 -0.1157203 0.5964625 4.2603567

Number of Observations: 950
Number of Groups: 9

```

Fuente: Elaboración propia