



Universidad Católica Andrés Bello

Facultad de Ciencias Económicas y Sociales

Escuela de Economía

**FICHAJES Y SALARIOS COMO LOS FACTORES PREDICTIVOS DEL
ÉXITO DEPORTIVO DE LOS EQUIPOS DE FÚTBOL EUROPEOS
TEMPORADAS 2015-2016 HASTA 2018-2019**

RICARDO PANTALEO AÑEL

Tutor: Guillermo Guerrero

Caracas, 04 de Diciembre de 2020

AGRADECIMIENTOS

Me complace en poder agradecer a todos aquellos que formaron parte e hicieron posible el desarrollo de esta tesis, tanto directamente como indirectamente, ya que de cualquier forma fue gracias a ustedes que he podido lograr esta meta.

Agradezco a mis Padres, gracias por su apoyo incondicional.

A mi novia, Michelle Morales por siempre ayudarme en cada uno de los momentos en que la necesite.

A mi tutor, Guillermo Guerrero, por su ayuda y dedicación en el desarrollo de esta tesis, por su tiempo, colaboración, por creer en mí y brindarme su amistad.

A mis compañeros, Gerónimo Mansutti y Sharon Lozada, por su tiempo y apoyo.

ÍNDICE

INTRODUCCIÓN	8
CAPÍTULO I. FUNDAMENTOS DE LA INVESTIGACIÓN	10
I.1 Planteamiento y Justificación del Problema	10
I.2 Hipótesis	12
I.3 Objetivos	13
I.3.1 Objetivo General	13
I.3.2 Objetivos Específicos	13
I.3.3 Limitaciones	14
CAPÍTULO II. MARCO TEÓRICO	16
II.1 Antecedentes Teóricos	16
II.2 Definiciones	21
II.2.1 Éxito	21
II.2.2 Fichajes	21
II.2.3 Balance Competitivo	22
II.2.4 Diferencias entre los Sistemas de Ligas Americanas y Ligas de Fútbol Europeo.	24
II.2.4.1 La Estructura del Campeonato.	24
II.2.4.2 La forma en que operan los mercados laborales.	25
II.2.4.3 El Draft	26
CAPÍTULO III: MARCO METODOLÓGICO	27
III.1. Nivel de la investigación	27
III.2 Diseño de la investigación	27
III.3 Población	28
III.4 Muestra	28
III.5 Variables	29
III.5.1 Variables deportivas	30

III.5.1.2 Puntos Obtenidos	30
III.5.2 Variables Financieras	30
III.5.2.1 Gastos Salariales Promedios Por Equipo	30
III.5.2.2 Gasto En Compras de Jugadores	31
III.5.2.3 Ingresos En Ventas por Jugador	31
III.5.2.4 Valor de Mercado del Equipo	31
III.5.3 Variables No Deportivas	32
III.5.3.1 Edad Promedio del Equipo	32
III.5.3.2 Estatura Promedio del Equipo	32
III.6 Metodología	32
III.6.1 Matriz de Correlación	32
III.6.2 Modelo y justificación	34
III.6.2.1 Coeficiente de determinación (R^2)	38
III.6.2.2 Nivel de significancia	39
III.6.2.3 Intervalos de confianza de los coeficientes	42
III.6.2.4 Pruebas de Homocedasticidad (Breusch-Pagan)	44
III.6.2.5 Pruebas de Autocorrelación de los residuos (Breusch-Godfrey-Pagan)	45
III.6.2.6 Pruebas de Normalidad de los residuos (Kolmogorov-Smirnov)	48
III.6.2.7 Pruebas de Multicolinealidad de los residuos (Factor de Inflación de la Varianza)	50
CAPÍTULO IV. ANÁLISIS DE RESULTADOS	53
IV.1 Matrices de Correlación	53
IV.2 Modelos de Regresión Lineal Multivariantes	56
IV.3 Pruebas de Homocedasticidad (Breusch-Pagan)	66
IV.4 Pruebas de Autocorrelación de los residuos (Breusch-Godfrey-Pagan)	68
IV.5 Pruebas de Normalidad de los residuos (Kolmogorov-Smirnov)	70
IV.6 Pruebas de Multicolinealidad de los residuos (Factor de Inflación de la Varianza)	72
CAPÍTULO V: CONCLUSIONES	74

CAPÍTULO VI: REFERENCIAS BIBLIOGRAFICAS	77
CAPÍTULO VII: ANEXOS	80

ÍNDICE DE TABLAS

<i>Tabla 1. Tipos de variables</i>	29
<i>Tabla 2. Matriz de Correlación Liga Inglesa</i>	53
<i>Tabla 3. Matriz de Correlación Liga Española</i>	53
<i>Tabla 4. Matriz de Correlación Liga Francesa</i>	54
<i>Tabla 5. Matriz de Correlación Liga Alemana</i>	54
<i>Tabla 6. Matriz de Correlación Liga Italiana</i>	54
<i>Tabla 7. Modelo Preliminar Liga Inglesa</i>	56
<i>Tabla 8. Modelo Definitivo Liga Inglesa</i>	57
<i>Tabla 9. Modelo Preliminar Liga Española</i>	58
<i>Tabla 10. Modelo Definitivo Liga Española</i>	59
<i>Tabla 11. Modelo Preliminar Liga Francesa</i>	60
<i>Tabla 12. Modelo Definitivo Liga Francesa</i>	60
<i>Tabla 13. Modelo Preliminar Liga Alemana</i>	62
<i>Tabla 14. Modelo Definitivo Liga Alemana</i>	63
<i>Tabla 15. Modelo Preliminar Liga Italiana</i>	64
<i>Tabla 16. Modelo Definitivo Liga Italiana</i>	65
<i>Tabla 17. Pruebas de Homocedasticidad (Breusch-Pagan) Liga Inglesa</i>	66
<i>Tabla 18. Pruebas de Homocedasticidad (Breusch-Pagan) Liga Española</i>	66
<i>Tabla 19. Pruebas de Homocedasticidad (Breusch-Pagan) Liga Francesa</i>	67
<i>Tabla 20. Pruebas de Homocedasticidad (Breusch-Pagan) Liga Alemana</i>	67
<i>Tabla 21. Pruebas de Homocedasticidad (Breusch-Pagan) Liga Italiana</i>	67
<i>Tabla 22. Pruebas de Autocorrelación de los residuos (Breusch-Godfrey-Pagan) Liga Inglesa</i>	68
<i>Tabla 23. Pruebas de Autocorrelación de los residuos (Breusch-Godfrey-Pagan) Liga Española</i>	68
<i>Tabla 24. Pruebas de Autocorrelación de los residuos (Breusch-Godfrey-Pagan) Liga Francesa</i>	69
<i>Tabla 25. Pruebas de Autocorrelación de los residuos (Breusch-Godfrey-Pagan) Liga Alemana</i>	69
<i>Tabla 26. Pruebas de Autocorrelación de los residuos (Breusch-Godfrey-Pagan) Liga Italiana</i>	69
<i>Tabla 27. Pruebas de Normalidad de los residuos (Kolmogorov-Smirnov) Liga Inglesa</i>	70

<i>Tabla 28. Pruebas de Normalidad de los residuos (Kolmogorov-Smirnov) Liga Española</i>	70
<i>Tabla 29. Pruebas de Normalidad de los residuos (Kolmogorov-Smirnov) Liga Francesa</i>	71
<i>Tabla 30. Pruebas de Normalidad de los residuos (Kolmogorov-Smirnov) Liga Alemana</i>	71
<i>Tabla 31. Pruebas de Normalidad de los residuos (Kolmogorov-Smirnov) Liga Italiana</i>	71
<i>Tabla 32. Pruebas de Multicolinealidad de los residuos (F.I.V) Liga Inglesa</i>	72
<i>Tabla 33. Pruebas de Multicolinealidad de los residuos (F.I.V) Liga Española</i>	72
<i>Tabla 34. Pruebas de Multicolinealidad de los residuos (F.I.V) Liga Francesa</i>	72
<i>Tabla 35. Pruebas de Multicolinealidad de los residuos (F.I.V) Liga Alemana</i>	72
<i>Tabla 36. Pruebas de Multicolinealidad de los residuos (F.I.V) Liga Italiana</i>	73
<i>Tabla 37. Intervalo de Confianza de los coeficientes Liga Inglesa</i>	85
<i>Tabla 38. Intervalo de Confianza de los coeficientes Liga Española</i>	86
<i>Tabla 39. Intervalo de Confianza de los coeficientes Liga Francesa</i>	86
<i>Tabla 40. Intervalo de Confianza de los coeficientes Liga Alemana</i>	86
<i>Tabla 41. Intervalo de Confianza de los coeficientes Liga Italiana</i>	86

INDICES DE GRAFICOS

<i>Grafico 1. Matriz de Correlación Liga Inglesa</i>	80
<i>Grafico 2. Matriz de Correlación Liga Española</i>	81
<i>Grafico 3. Matriz de Correlación Liga Francesa</i>	81
<i>Grafico 4. Matriz de Correlación Liga Alemana</i>	82
<i>Grafico 5. Matriz de Correlación Liga Italiana</i>	82
<i>Grafico 6. Distribución Normal. Grafico Q-Q Plot Liga Inglesa</i>	83
<i>Grafico 7. Distribución Normal. Grafico Q-Q Plot Liga Española</i>	83
<i>Grafico 8. Distribución Normal. Grafico Q-Q Plot Liga Francesa</i>	84
<i>Grafico 9. Distribución Normal. Grafico Q-Q Plot Liga Alemana</i>	84
<i>Grafico 10. Distribución Normal. Grafico Q-Q Plot Liga Italiana</i>	85

INTRODUCCIÓN

Fútbol, el rey de los deportes. Se estima que a nivel mundial alrededor de 4 mil millones de personas, casi la mitad de los habitantes del planeta, son seguidoras de este deporte. Tal es la pasión que desata, que cada 4 años el mundo se paraliza ante la llegada de la Copa del Mundo.

Si bien desde sus inicios el deporte no ha sufrido cambios significativos en la forma en que se practica, por el lado comercial si ha cambiado radicalmente, generando toda una industria multimillonaria que continúa creciendo de forma acelerada. Si nos centramos en el evento deportivo que supone el Mundial, durante el penúltimo celebrado, Brasil 2014, la Fédération Internationale de Football Association (FIFA) generó 4.800 millones de dólares en ingresos brutos, obteniendo un beneficio neto de 2.600 millones de dólares. Para el último mundial, Rusia 2018, los ingresos fueron de aproximadamente 6.400 millones de dólares, con un beneficio neto estimado de 5.300 millones de dólares.

El epicentro de toda esta evolución tiene su origen en los principales campeonatos de fútbol europeo; entre los que sobresale un grupo selecto denominado comúnmente como *“Las Cinco Grandes Ligas”*, nombre con el que se hace referencia a: La Premier League de Inglaterra, La Liga de España, La Serie A en Italia, Bundesliga en Alemania y La Ligue 1 en Francia. Estos son considerados históricamente como los mejores campeonatos debido a que los equipos que las integran (especialmente el Real Madrid, Juventus, Bayern de Múnich, PSG, entre otros) son los más laureados y reconocidos a nivel mundial.

Para tener una idea del potencial de estas mencionadas ligas, la Liga de Campeones de la Union des Associations Européennes de Football (UEFA) considerado el torneo de clubes más importante y de mayor prestigio a nivel mundial, ha sido ganada por equipos de

las denominadas Cinco Grandes Ligas en 50 de sus 63 ediciones, lo cual representa casi un 80% de éxito del total de veces que se ha disputado el torneo.

Este éxito deportivo conllevó al crecimiento acelerado de los ingresos por contratos televisivos y patrocinios, lo que ha despertado el interés de un considerable número de inversionistas privados que buscan hacerse cargo de equipos pertenecientes a dichas ligas. En consecuencia, los clubes con los mayores ingresos empezaron un proceso sin igual para la adquisición de nuevos talentos y jugadores más consagrados. Para tener una idea, en 2009 la adquisición por aproximadamente 94 millones de Euros de Cristiano Ronaldo por parte del Real Madrid rompió el récord del fichaje más caro de la historia, que había establecido el mismo club en el 2001. El récord ha sido restablecido un total de 3 veces desde entonces.

Motivado a este gasto sin precedentes, surge la interrogante de si los equipos están optimizando sus resultados mediante la asignación de sus limitados recursos financieros, bajo el supuesto de que los clubes actúan como firmas con un objetivo dual de generar beneficios financieros y resultados deportivos.

El presente trabajo busca cuantificar el impacto deportivo de algunas de las principales estrategias de asignación de recursos como paso preliminar para una investigación que busque atender el problema general de la optimización del uso de recursos financieros.

CAPÍTULO I. FUNDAMENTOS DE LA INVESTIGACIÓN

I.1 Planteamiento y Justificación del Problema

No es secreto que el fútbol es un deporte que mueve masas y desata pasiones. Asimismo, tampoco es un secreto que en el fútbol existen disparidades económicas muy acentuadas. Al observar cada una de Las Cinco Grandes Ligas, catalogadas como la élite del fútbol mundial, encontramos una inmensa diferencia entre la capacidad financiera de los equipos más poderosos y los equipos más pequeños. Para ilustrar mejor nuestro punto, solo falta notar que en la temporada 2018-2019 el FC Barcelona poseía la nómina más costosa de todos los equipos europeos, con un estimado de 13.755.000 de euros de salario promedio anual por jugador; mientras que el Nimes de la Ligue 1 fue el equipo con la plantilla más económica, habiendo pagado tan solo 300.000 Euros en promedio a cada jugador.

A nivel financiero, las ya mencionadas Cinco Grandes Ligas han experimentado un gran auge financiero durante el último par de décadas. Según The European Club Footballing Landscape, reporte anual de la UEFA, para la temporada 1998-1999 el total de ingresos brutos generado por Las Cinco Grandes Ligas fue de alrededor de 3.320 millones de Euros y para la temporada 2018-2019 la cifra aproximada fue de 21.000 millones de Euros para todas las organizaciones inscritas en la UEFA, de los cuales cerca de 16.000 millones pertenece a Las Cinco Grandes Ligas, lo cual representa poco más del 75% del total recaudado por todas las asociaciones de fútbol europeo.

Este crecimiento también se ha visto reflejado en el costo de los traspasos de los jugadores de un club a otro. En el año 1990, la Juventus de Turín estableció un récord del fichaje más caro al adquirir a Roberto Baggio, quien es considerado por muchos el mejor jugador de su generación, por un estimado de 8.800.000 Euros (alrededor de unos 20.140.000

ajustados por inflación). En 2017, el Paris Saint-Germain sacudió el mercado de fichajes con la operación más grande registrada hasta los momentos, la compra de Neymar Jr. por una cifra cercana a los 222.000.000 Euros. En los 27 años entre los dos fichajes, el récord del fichaje más caro de la historia se multiplicó por más de 25 en términos nominales y por 10 veces en términos reales.

No es casualidad que las ligas que representan a los países de mayor tradición y éxito futbolístico sean las que obtienen mayor atención por parte del público en general, lo cual permite que generen una mayor cantidad de ingresos, que se traducen en presupuestos más grandes para fichajes de jugadores y mejores contratos salariales. El crecimiento comercial tanto de estas ligas, así como de los equipos que las integran ha desencadenado una batalla constante por ver quién obtiene a los mejores jugadores que se encuentren disponibles en el mercado.

Es precisamente esta denominada “guerra de fichajes” la que ha dado cabida a una gran cantidad de trabajos y artículos relacionados que sirvieron como fuente de inspiración para abordar una serie de inquietudes que surgen a la hora de realizar un análisis; ¿Es verdaderamente necesario para obtener éxito deportivo realizar fichajes cuyos costos exceden por mucho el costo promedio de los jugadores? ¿Estas inversiones generan un verdadero impacto en el desempeño del equipo? ¿Cuál es el costo asociado a mejorar posiciones en la tabla de la liga respecto de la temporada anterior? ¿Está justificado este importante incremento en el récord del fichaje más caro? ¿Neymar Jr. ha producido o producirá un incremento suficientemente importante en el desempeño deportivo del Paris Saint-Germain para justificar el importante gasto que realizó el equipo?

En aras de poder esclarecer estas y otras interrogantes sobre el tema, el presente trabajo pretende extender estudios previos generando un modelo que integra variables

deportivas, no deportivas y financieras de los equipos que tuvieron participación en Las Cinco Grandes Ligas europeas en las últimas 4 temporadas para comprender si en efecto los gastos deportivos mantienen una relación positiva con el desempeño de los equipos a nivel deportivo.

La gran mayoría de los trabajos relacionados con el tema estudian el Balance Competitivo dentro de ligas particulares. El Balance Competitivo es un concepto usado extensivamente en los estudios de economía deportiva centrada en las ligas norteamericanas y establece que a mayor equilibrio de gastos salariales entre los equipos, mayor incertidumbre y por ende un campeonato más reñido. Por el contrario, el presente trabajo pretende realizar un análisis completo que integre todas las Cinco Grandes Ligas y ayude a determinar cuáles son los factores financieros y deportivos clave en el desempeño de los equipos que las conforman.

I.2 Hipótesis

Para los equipos pertenecientes a Las Cinco Grandes Ligas Europeas, el gasto absoluto en fichajes y salarios son las variables más significativas para la determinación de los resultados deportivos y la posición conseguida por cada equipo en su respectivo campeonato.

I.3 Objetivos

I.3.1 Objetivo General

Identificar y analizar cuáles son las variables financieras y no deportivas que tienen mayor influencia en el éxito deportivo de los equipos pertenecientes a las Cinco Grandes Ligas de fútbol europeo, caso de estudio temporadas 2015-2016 hasta 2018-2019.

I.3.2 Objetivos Específicos

- Identificar las variables financieras y no deportivas que impactan los resultados deportivos de los clubes.
- Cuantificar y definir el nivel de éxito de los equipos pertenecientes a las Cinco Grandes Ligas europeas en términos deportivos desde las temporadas 2015-2016 hasta 2018-2019
- Medir el impacto marginal de cada variable sobre el éxito deportivo.
- Comprobar si mayores gastos asociados a la adquisición y salarios de jugadores tienen impactos reales en los cambios de ubicación de los clubes en la tabla de posición respecto al año pasado.
- Calcular el impacto conjunto que tienen las variables seleccionadas sobre el desempeño deportivo del club.

I.3.3 Limitaciones

El estudio presenta algunas limitaciones, por lo que es prudente que la interpretación de los resultados se haga de manera cautelosa. El presente trabajo está centrado en los equipos pertenecientes a las denominadas Cinco Grandes Ligas Europeas, por lo cual el alcance y las conclusiones se restringen a los equipos que conforman las ligas de España, Italia, Alemania, Inglaterra y Francia; sin embargo, esta metodología podría ser replicada con otra muestra, a fin de expandir el alcance de la investigación.

Es prudente mencionar que, de los equipos que fueron considerados para la elaboración del presente trabajo, no todos los informes financieros son de uso público. Motivado a esto, la recopilación de la información se limitará a fuentes cuyos trabajos y publicaciones estuviesen disponibles de forma pública. Destacan trabajos hechos por consultoras, publicaciones elaboradas por organismos reguladores del deporte y artículos de prensa de diarios deportivos, así como de semanarios de información general. Es por esto que algunos de los valores y cifras utilizados en nuestro trabajo pueden ser aproximados o estimados de las cifras reales manejadas de forma interna por los equipos.

En cuanto a los salarios se refiere, debido a que no son cifras de uso público, se usarán los sueldos promedios estimados por equipos de periodistas y estudios especializados en el tema. Cualquier bonificación que esté incluida en los contratos no será tomada en consideración debido a que son incentivos en base al cumplimiento de objetivos.

Para los propósitos de esta investigación, no se considerarán los niveles de impuestos salariales de cada país y su influencia en sus respectivas ligas. Si bien las diferencias impositivas podrían resultar en ventajas de los equipos de unas ligas sobre otras, todos los equipos pertenecientes a una misma liga se enfrentan a las mismas tasas, de forma que esto

no tendrá efecto sobre nuestra investigación, esto debido a que la evaluación de los resultados será realizado por liga y no comparativamente.

CAPÍTULO II. MARCO TEÓRICO

El presente capítulo tiene como finalidad exponer todas aquellas definiciones, conceptos, teorías y estudios previos, que puedan aportar y apoyar al presente trabajo. En primer lugar, se reseñan investigaciones relacionadas con el objeto de estudio, se hará una breve reseña de la historia del deporte, posteriormente se abordarán los diferentes conceptos y definiciones necesarios para contextualizar la investigación, así como una revisión histórica de los acontecimientos pertinentes. Por último, se expondrán las teorías en las que se basa la investigación.

II.1 Antecedentes Teóricos

Si bien el deporte ha sido parte de las diferentes culturas de la humanidad a lo largo de varios siglos, los estudios económicos del deporte son relativamente recientes. Considerado generalmente como el padre de la Economía Deportiva, Rottenberg (1956) en su artículo *The Baseball Player's Labor Market*, detalla la composición y el funcionamiento del mercado laboral de los beisbolistas de la Major League Baseball (MLB), haciendo especial énfasis en la forma en cómo operan los diferentes mecanismos que pueden ser empleados por los equipos para contratar a los jugadores a lo largo de su carrera.

Los mercados de fútbol europeo y de las ligas deportivas norteamericanas están estructurados de forma completamente opuesta. Esta diferencia podría caracterizarse como de “libre mercado” vs. “mercado regulado”. Debido a esto, solo mencionamos el trabajo de Rottenberg (1956) a manera de reconocer que fue uno de los primeros esfuerzos por entender la economía detrás del deporte. Es prudente destacar que este trabajo ha sido fuente de inspiración de la gran mayoría de los trabajos enfocados en el estudio de la economía detrás del deporte.

Extendiendo la investigación de Rottenberg (1956), Neale (1964) expone que una liga deportiva es una “firma” de equipos y es por esto que su competencia son otros deportes, en un escenario donde la preferencia o demanda del fanático se vea comprometida y deba decidir entre uno de los dos deportes. Por ende, plantea que la liga es un monopolio en sí, dado que es la única que puede ofrecer ese producto en particular, lo que lleva a la sugerencia de que los equipos pertenecientes a la liga van a, en cierta forma, cooperar entre sí para poder ofrecer un mejor espectáculo en pro del beneficio de la liga. Es aquí donde el autor plantea la “Paradoja de Lois-Schmeling”, según la cual los monopolios deportivos no son tan beneficiosos como cualquier otro monopolio en empresas. Para argumentar esto, Neale (1964) concuerda con la hipótesis planteada por Rottenberg (1956) en donde plantea que una competición deportiva mucho más balanceada va a recibir un mayor grado de interés por parte del fanático. Neale (1956) explica que al mejor equipo de una liga deportiva le va a interesar ser el mejor, pero no completamente superior que el resto de los equipos. También expone que, mientras menos predecible sea el resultado del evento, mayor beneficio para el espectador.

El-Hodiri y Quirk (1971) fueron los primeros en estudiar la estructura económica de los deportes profesionales, desarrollando el que es considerado el primer modelo matemático de toma de decisiones dinámico aplicado a las ligas deportivas de los Estados Unidos.

En su trabajo establecen que dada la naturaleza única de la estructura económica de las ligas de deportes profesionales, los equipos pertenecientes a dichas ligas presentan problemas de toma de decisiones bajo incertidumbre, lo cual está directamente relacionado a los ingresos (i) que pueden generar los equipos, debido a que mencionados ingresos no dependen exclusivamente de las variables de decisión del equipo, sino que también dependerán de las variables de decisión de los otros equipos. El-Hodiri y Quirk (1971) en concordancia con lo planteado por Rottenberg (1956) y Neale (1964), establecen que los

beneficios de venta de entradas para un evento deportivo se ven afectados negativamente siempre que la probabilidad de que uno de los equipos sea el vencedor tienda hacia uno. Sin embargo, al igual que lo planteado por Neale (1964), los autores explican que un equipo va a tener fuertes motivos para ser superior al resto de los equipos, pero no demasiado superior. Es por esto que en su modelo el problema de toma de decisiones conjuntas de los equipos es descrito como un juego suma cero de “n” equipos.

Estos primeros trabajos son considerados como los pioneros de la economía deportiva y han fungido de inspiración para una gran cantidad de nuevas investigaciones en diferentes áreas no exploradas anteriormente. Una gran parte de las investigaciones se centra en la determinación de si una liga tiene mejores beneficios al tener un mayor balance deportivo entre los equipos que la conforma. Este tema de investigación por lo general es utilizado por autores en las ligas deportivas americanas, las cuales difieren en muchas características de las ligas europeas, como en la forma en como su mercado laboral opera, la estructura de la liga, la estructura del campeonato, entre otros. Por años los autores han debatido si es una obligación de las ligas y de los equipos que la conforman crear un mayor balance competitivo para generar mejores rendimientos para todos los integrantes. Este tema no será abordado a cabalidad, sin embargo, no será descartado por completo al ser uno de los temas que mayor contenido literario posee.

El que es hoy día considerado el máximo exponente en cuanto a investigación de economía deportiva se refiere, Stefan Szymanski junto con Ron P. Smith basaron su estudio en el trade-off inherente que realizan los equipos entre el desempeño en el campo de juego y el desempeño financiero. El enfoque de Szymanski & Smith (1997) va dirigido hacia la toma de decisión de los dueños de los equipos cuyas funciones objetivo dependen tanto de los ingresos como de la posición que ocupa cada equipo en la tabla. La maximización de la función objetivo, sujeta a la restricción beneficio económico-posición en la tabla,

determinará el nivel óptimo de sueldos, los beneficios del club y la posición que debería ocupar cada equipo ante los niveles individuales de cada equipo.

La función de producción de la industria corresponde a la cantidad de talento que los equipos adquieran en el mercado de jugadores. La cantidad de talento que los clubes adquieran va a determinar la posición de dicho equipo en la tabla. Por otra parte, la posición del año anterior o de años anteriores va a determinar, al menos parcialmente, los ingresos por venta de tickets al estadio, derechos de transmisión, etc; lo cual corresponde a la función de demanda de la industria (ambas funciones fueron determinadas por Szymanski et al).

Al combinar las funciones de producción y de demanda con las restricciones presupuestarias se obtiene el trade-off empírico que cada equipo enfrenta entre la maximización de las ganancias y la posición de la liga. En los resultados obtenidos presentan una pendiente negativa, lo cual denota que el incremento del gasto en salarios no es autofinanciable a través de mayor rendimiento ni tampoco de mayores ingresos.

Dietl et al (2008) determinaron que los equipos compiten a través de las inversiones que realizan y esa precisa interacción entre equipos es lo que genera la sobreinversión, lo cual lleva en muchas situaciones a diluir completamente todos los premios monetarios obtenidos por cada equipo en su participación en sus respectivas ligas. Canes (1974) plantea que cualquier mejora en la adquisición de talento por parte de un equipo inequívocamente va a generar externalidades negativas (desmejora) llevando a que los otros equipos participantes busquen adquirir o sobre-invertir en talentos. White (1993) cataloga como “competencia destructiva” a este tipo de práctica empleada en las ligas deportivas.

En el mismo orden de ideas planteadas por Dietl et al (2008) y Canes (1974), varios autores han adaptado en sus publicaciones la analogía empleada por Akerlof (1976) para el desarrollo de su modelo sobre el mercado laboral. Akerlof (1976) acuña la expresión

“Carrera de Ratas”, basándose en una serie de experimentos en donde un grupo control de ratas son puestas en laberintos con el fin de llegar primero a un trozo de queso, estableciendo que las ratas aumentan sus posibilidades de obtener la recompensa (trozo de queso) siempre que aumenten su velocidad. Akerlof (1976) establece que, por mucho que las ratas hagan un sobre esfuerzo por arribar primero al trozo de queso, no se va a producir ninguna dotación o unidad extra de queso. Esta analogía ha sido trasladada al modelo económico de las ligas europeas en donde, dada una cantidad de premios económicos predeterminados, cualquier gasto mayor en fichajes y salarios por parte de los equipos no va a generar ninguna unidad adicional de beneficio agregado. Tampoco influye si un equipo en particular tiene un desempeño sobresaliente, ya que al final el premio se encuentra preestablecido; el juego es suma cero. Por ende, la competencia entre los equipos de las ligas se termina convirtiendo en una “Carrera de Ratas” en donde, bajo estos supuestos, los equipos generan muy poca rentabilidad para la gran inversión que realizan, lo cual hace que la liga no sea autosostenible.

Rhode & Breuer (2016) extienden la gran mayoría de los trabajos previamente mencionados construyendo un modelo empírico enfocado en el desempeño financiero. Una objeción que encontramos pertinente mencionar, es el hecho de que los autores únicamente toman en consideración el uso de data de los 30 equipos que se encuentran reflejados en el listado anual publicada por Deloitte. (Deloitte Football Money League).

II.2 Definiciones

A continuación, se dará paso a la sección de definiciones, con la que se pretende dar un mayor entendimiento de los conceptos y de las teorías económicas que estarán en el centro de la discusión del trabajo.

II.2.1 Éxito

Schulz y Heckhausen (1996), define el éxito como “el cumplimiento de objetivos es un concepto subjetivo e individualizado que se refiere a la consecución de los resultados deseados y a la evitación de resultados no deseados”.

II.2.2 Fichajes

La Real Academia de la Lengua Española define el término fichaje como “acción y efecto de fichar a un jugador, atleta o técnico deportivo”, dicho de otra manera, se puede definir como la acción de contratar los servicios deportivos de un jugador o técnico. La forma en la que se ejecutan los fichajes suele ser entre 3 partes involucradas; el equipo comprador, el equipo vendedor y el jugador en cuestión; en donde el equipo vendedor recibe una compensación de dinero por la ficha o contrato (de aquí proviene el nombre). El equipo comprador debe primero establecer y acordar los lineamientos del contrato con el jugador.

Existen casos excepcionales o no tan usuales, un ejemplo son los fichajes a través de una cláusula de rescisión, donde un equipo y un jugador establecen una cifra de grandes proporciones en aras de mantener los servicios del jugador por un mayor tiempo, siendo el único inconveniente para el equipo dueño de la ficha la imposibilidad de rechazar la oferta

que iguale o active dicha cláusula. De forma resumida, el jugador tendrá un precio de salida fijo que puede ser muy elevado, pero en caso de que un equipo decida pagarlo el equipo dueño de la ficha no podrá contra ofertar.

Otro caso algo más común que fichajes por cláusula de rescisión es el fichaje de “costo cero”. Se le denomina así porque se realiza entre dos partes, el club comprador y el jugador. Este caso de fichaje sucede cuando un jugador que no posee un contrato vigente con alguna institución deportiva formaliza un contrato con un equipo interesado en sus servicios deportivos. Se le dice “costo cero” porque el club o equipo que contrata no hace un pago compensatorio a otro equipo, únicamente pagará el salario acordado con el jugador.

II.2.3 Balance Competitivo

Uno de los principales conceptos que es evaluado en la gran mayoría de los trabajos que sirvieron de base e inspiración para la elaboración del presente proyecto es el Balance Competitivo entre los equipos de las ligas. En economía deportiva, el Balance Competitivo es definido como una situación de mercado en la cual ningún negocio es tan grande que posee una ventaja injusta o desproporcionada. En deportes este concepto fue extendido por Michie & Oughton (2004) como “El balance entre las capacidades deportivas. Mientras más balanceados se encuentren los equipos, mayor será la incertidumbre del resultado de los partidos”. Los autores exponen que la forma más simple de determinar si una liga deportiva presenta un mayor Balance Competitivo es determinando la habilidad de que un equipo tenga una competencia o un competidor potencial en dicha liga. Tal como fue mencionado anteriormente en el presente trabajo, Neale (1964) expone que el mejor equipo va a querer ser superior, solo que no tan superior para mantener el interés de los espectadores. En caso

contrario, al existir un monopolio puro en la competición se pierde el factor de incertidumbre y con ello el interés del fanático.

Rottenberg (1956) y Neale (1964) a manera de poder explicar las consecuencias de la ausencia o de un Balance Competitivo, desarrollaron el concepto de la “La Hipótesis de Incertidumbre del Resultado”, con el que ellos describen la importancia de los factores que los espectadores encuentran fascinantes o interesantes para asistir a un evento deportivo u observarlo en transmisión televisiva. Exponen que si los fanáticos tuviesen certeza de cuál sería el resultado de un evento deportivo, los beneficios económicos serían muy bajos respecto a otro encuentro en el cual su resultado sea impredecible.

Por otro lado, Kuypers (1997) lo clasifica en tres partes:

- La atracción generada por la incertidumbre del resultado.
- Lo reñido del campeonato.
- La ausencia de un equipo dominante durante los años.

Rottenberg (1956) es el primero en emplear el concepto de Balance Competitivo. Desde entonces, este concepto ha sido empleado en una gran cantidad de trabajos que comparan las diferencias estructurales entre el sistema de Ligas Deportivas de Estados Unidos y las Ligas de Fútbol Europeo, siendo muy recurrente la comparación de cuál sistema es más eficiente en términos de competitividad.

Debido a la interrogante inherente que surge al momento de analizar cuál sistema de competición es menos disparejo, si el Sistema de Ligas Americano o Las Ligas de Fútbol Europeo. Motivado a que son sistemas completamente diferentes siempre van a existir sólidos argumentos de ambas partes. Aun cuando no nos centraremos en este aspecto en particular, es pertinente esclarecer las diferencias que rigen ambos sistemas. Principalmente porque los clubes más poderosos de las Cinco Grandes Ligas pagan cantidades elevadas de

dinero debido a la ausencia de un límite salarial, la cual es una de las grandes diferencias entre los ya mencionados sistemas.

II.2.4 Diferencias entre los Sistemas de Ligas Americanas y Ligas de Fútbol Europeo.

II.2.4.1 La Estructura del Campeonato.

Una de las diferencias más marcadas entre las Ligas Americanas (NBA, NFL, MLS, MLB y NHL) y la mayoría de los campeonatos de Fútbol Europeo, es el formato de la competición. Por un lado, las ligas americanas poseen un orden por conferencias o ligas, en donde los mejores equipos posicionados al final de la temporada regular clasifican a los “Playoffs” o Posttemporada, donde los equipos, dependiendo el lugar dentro de sus conferencias, serán emparejados para encuentros de eliminación directa, los cuales pueden ser partidos únicos o una serie de partidos.

Por su parte, el formato europeo es mucho más simple. Todos los equipos pertenecen a una liga única sin divisiones y tendrán dos enfrentamientos con cada uno de los integrantes de dichas ligas, una jugando como local y otra como visitante, siendo el que mayor cantidad de puntos obtenga el campeón. Mientras que el formato de ligas europeas premia la consistencia durante todo el año, el formato de ligas americanas premia el desempeño en la posttemporada, sujeto a la clasificación durante la temporada de regular.

Dentro de la estructura de las ligas existe otra gran diferencia, el sistema americano es lo que se conoce como “Ligas Cerradas” lo que quiere decir no tienen promoción ni descenso. La única forma de que equipos que no pertenecen a las ligas puedan entrar es a través de la adquisición de una franquicia/equipo perteneciente a la liga y renombrarla, o que

se realice una expansión de la liga; de otra forma, ningún nuevo equipo puede formar parte de la liga. Por otro lado, las ligas europeas son un conjunto de “Ligas Abiertas” que si cuentan con promociones y descensos, en donde los peores equipos de la temporada (varía el formato y la cantidad según las ligas) descienden a la división o categoría inferior a la que se encuentran y en contraparte los mejores equipos de la división inferior ascienden. Esto permite a equipos pequeños tener la oportunidad de participar en el mismo torneo que los equipos élite.

II.2.4.2 La forma en que operan los mercados laborales.

Si bien fuera de temporada el accionar de los clubes en ambos sistemas se basa en el objetivo común de obtener los mejores prospectos a disposición en el mercado, la forma en que los mercados operan es muy diferente. Para empezar, el Sistema Americano opera bajo el mecanismo de intercambios de jugadores, es decir, un equipo que denominaremos “A” busca hacerse con los servicios de un jugador del equipo “B”; el equipo “A” podrá ofrecer parte de los jugadores que pertenecen a su plantilla a cambio del jugador del equipo “B” sin la necesidad de movilizar cantidades reales de dinero, la única cifra que debe ser negociada es el monto del salario que se le pagará al jugador. Contrario al sistema americano, el mercado de las Ligas de Fútbol de Europa opera bajo el sistema de los traspasos de jugadores por una cantidad establecida por el club dueño de la ficha del jugador, lo cual no excluye que dos equipos estén dispuestos a intercambiar jugadores, solo que esta práctica no es tan común dada la discrepancia en la valoración del jugador por parte de los equipos.

Otra gran diferencia radica en la parte salarial. En las ligas europeas no existe un tope o límite salarial que restrinja los salarios por equipo, lo cual da libertad a estos de costear altos salarios de múltiples jugadores siempre que su presupuesto lo permita. En este sentido, dentro del formato europeo de ligas, la única limitante al gasto deportivo son los ingresos

del club. En contraposición, las ligas americanas incorporan un tope salarial, diseñado para potenciar el Balance Competitivo. Cada liga posee su propio sistema de top salarial.

II.2.4.3 El Draft

El “Draft” es un método de asignación de jugadores universitarios a los equipos pertenecientes a las Ligas Deportivas Americanas, en el cual los peores equipos de la temporada pasada obtienen los primeros puestos de selección. Muy diferente a las Ligas de Fútbol Europeo en donde los peores equipos descienden a categorías inferiores y en consecuencia teniendo que afrontar la posible pérdida de sus mejores jugadores ante equipos que lograron la permanencia.

En Europa, debido al sistema de Ligas Abiertas, los equipos no tienen ninguna ventaja al ser los peores de la liga debido a que no van a tener una compensación por ello.

CAPÍTULO III: MARCO METODOLÓGICO

III.1. Nivel de la investigación

Arias (2012) define el nivel de investigación como "el grado de profundidad con que se aborda un fenómeno u objeto".

Bajo este concepto, la presente investigación está enmarcada en un nivel descriptivo, el cual se distingue por la caracterización de un fenómeno, individuo o grupo. Arias (2012) subdivide este tipo de diseño, siendo la del tipo correlacional la empleada en el presente trabajo, en la cual se determinan las asociaciones causales existentes entre dos o más variables.

III.2 Diseño de la investigación

Los datos utilizados en la presente investigación son de procedencia exclusiva de fuentes secundarias, lo cual sitúa al presente estudio dentro del orden de investigación documental o diseño documental. Arias (2012) define a la investigación documental como "...un proceso basado en la búsqueda, recuperación, análisis, crítica e interpretación de datos secundarios, es decir, los obtenidos y registrados por otros investigadores en fuentes documentales: impresas, audiovisuales o electrónicas."

Arias (2012) establece que la investigación documental puede realizarse de tres formas o subniveles, siendo la correlación a partir de datos secundarios el que será implementado en el presente trabajo.

III.3 Población

Arias (2012) define la población como “conjunto finito o infinito de elementos con características comunes para los cuales serán extensivas las conclusiones de las investigaciones. Ésta queda delimitada por el problema y por los objetivos del estudio”.

El presente estudio toma como población, a todos los equipos pertenecientes a las Cinco Grandes Ligas de Fútbol Europeo durante las temporadas comprendidas desde la 2015-2016 hasta la 2018-2019. Debido que las cinco ligas mantienen un formato de Ligas Abiertas, existe una cantidad considerable de equipos que no participaron en todas las ediciones de los mencionados torneos por haber descendido de categoría. La población total para el período de estudio será de 128 equipos que participaron al menos una temporada en una de las Cinco Grandes Ligas, esto debido a que el estudio se centrará en un estudio de resultados por equipos y no desempeño individual.

III.4 Muestra

Para Arias (2012) “la muestra es un subconjunto representativo y finito que se extrae de la población accesible”. A su vez, una muestra representativa es aquella que, dado su tamaño y características similares a la población, permite realizar inferencias estadísticas, con un determinado margen de error.

El presente trabajo utilizará una muestra de más de 600 observaciones, segmentadas en valores deportivos y valores no deportivos entre los que se encuentran valores financieros de cada equipo.

III.5 Variables

Arias (2012) define variable como “una característica o cualidad, magnitud o cantidad que puede sufrir cambios y que es objeto de análisis, medición manipulación o control en una investigación”.

Así mismo, explica que según la naturaleza de las variables, estas pueden ser:

Tabla 1. Tipos de variables

Cuantitativas: toda aquella variable que está expresada en términos numéricos	Discretas: son aquellas que asumen valores enteros
	Continuas: son aquellas que asumen valores con decimales
Cualitativas: hacen referencia a características o atributos que no pueden ser expresados de forma numérica.	Dicotómicas: son aquellas en las que sólo existen dos categorías posibles
	Politómicas: son aquellas en las que existen múltiples categorías.

Fuente: elaboración propia a partir de Arias (2012)

Por otro lado, según la complejidad de las variables estas se clasifican en dos tipos: las simples y las complejas. El primer tipo de variables se refiere a aquellas que pueden ser medidas directamente o, a través de indicadores.

El segundo tipo de variables, las complejas, “son aquellas que se pueden descomponer en dos dimensiones como mínimo. Luego se determinan los indicadores para cada dimensión”.

III.5.1 Variables deportivas

III.5.1.2 Puntos Obtenidos

Queda definida como la variable dependiente del estudio, por ser considerado el criterio de medición del éxito deportivo más certero. Dado que la cantidad de partidos ganados o perdidos y la cantidad de goles (anotados o recibidos) no establecen un campeón, los puntos son el único criterio que definen que equipo será campeón al final de la temporada, así como todas las posiciones del campeonato. A pesar de que se obtienen mayor cantidad de puntos por partidos, no es garantía que el equipo con mayor partidos ganados sea el campeón al final de la temporada. El equipo campeón puede resultar de una combinación de partidos ganados, empatados y perdidos tal que aun no siendo el equipo con mayor cantidad de victorias, dicho equipo resultase campeón. Para los efectos de los modelos, la variable quedará expresada como “*Puntos*”.

III.5.2 Variables Financieras

III.5.2.1 Gastos Salariales Promedios Por Equipo

Es el gasto destinado por cada uno de los equipos al pago de su nómina. Únicamente se están considerando los salarios percibidos por los jugadores, los directivos y gastos salariales de staff médico, de mantenimiento y gerencia no serán incluidos. Los salarios fueron estudiados en promedio por equipo para evitar tener un número exagerado de observaciones. La variable queda expresada como en los modelos obtenidos como “*Salarios*”

III.5.2.2 Gasto En Compras de Jugadores

Son los gastos que realizan los equipos para la adquisición de nuevos jugadores (talentos) para reforzar la plantilla o nómina. Los gastos en compras representan los totales de los gastos realizados por compras de jugadores. Los gastos en compra quedan expresados para efectos de los modelos obtenidos como “*Compras*”.

III.5.2.3 Ingresos En Ventas por Jugador

Corresponde a todos los devengos percibidos por la venta de uno o varios jugadores por parte de un determinado equipo. Los ingresos quedaran expresados en sus valores totales bajo el título de “*Ventas*” para efectos de la presente investigación.

III.5.2.4 Valor de Mercado del Equipo

Es la suma del valor de mercado de cada a uno de los jugadores pertenecientes a un equipo. En otras palabras, representa la cantidad de dinero que en teoría obtendría un equipo por la venta de cada uno de sus jugadores en un precio de mercado establecido según el desempeño del jugador, edad del jugador, equipo al que pertenece el jugador, entre otros; los cuales permitirán asignar un valor de mercado aproximado. Es importante aclarar que, un valor de mercado no necesariamente va a ser el precio al que un jugador es adquirido o vendido, va a representar únicamente una estimación de cuál podría ser un precio potencial de dicho jugador. Representado en los modelos como “*Valor T*”.

III.5.3 Variables No Deportivas

III.5.3.1 Edad Promedio del Equipo

Representa el promedio de edades en años por equipo. Representado en los modelos como “*Edad*”.

III.5.3.2 Estatura Promedio del Equipo

Representa el promedio de estaturas en centímetros por equipo. Representado en los modelos como “*Estatura*”.

III.6 Metodología

A continuación, serán explicados todos los procesos metodológicos utilizados para la obtención de los modelos así como todas las pruebas estadísticas empleadas, no solo para confirmar que cada modelo cumple con los criterios básicos, sino también para medir la capacidad que tienen los modelos explicar la variabilidad de la variable observada.

En la presente sección únicamente se expondrán los procesos que fueron llevados a cabo; los resultados y sus análisis serán presentados en la siguiente sección y se encontrarán divididos por ligas.

III.6.1 Matriz de Correlación

Para la determinación del modelo que mejor se acople a las variables y los datos utilizados se procedió a la elaboración de una matriz de correlación la cual, tal como lo indica

su nombre, es una matriz simétrica en la cual se muestra la correlación existente entre dos de las variables seleccionadas. Para la elaboración de la matriz de correlación las variables son colocadas en un orden igual de columnas y filas, en donde los valores de la diagonal principal serán iguales a uno siendo estos valores las correlaciones entre la misma variable. Una gran correlación entre varios pares de variables sugiere que los estimadores no serán confiables, esto como consecuencia de que algunos estimadores pudiesen estar midiendo la misma característica. Por ende, se busca que las variables independientes de una matriz no estén altamente correlacionados entre ellas, caso contrario se busca que las variables independientes mantengan una alta correlación con la variable dependiente.

La estructura de una matriz de correlación queda representada de la siguiente forma:

$$\rho = \begin{pmatrix} 1 & r_{12} & \cdots & r_{1p} \\ r_{21} & 1 & \cdots & r_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ r_{p1} & r_{p2} & \cdots & 1 \end{pmatrix}$$

Los valores de correlaciones mostrados en la matriz pueden ser obtenidos a través del método de correlación Pearson o de Spearman, los cuales miden el grado de relación lineal entre cada par de variables. Para efectos del presente trabajo, fue empleado el método de Pearson, debido a que mide la asociación lineal entre variables cuantitativa aleatorias, diferente del método de Spearman, el cual mide el nivel de asociación entre variables ordinales. Los valores obtenidos se ubican entre -1 y +1, representando la magnitud o fuerza, y la orientación (positiva o negativa) en la que cada par de variables se encuentran correlacionadas. Los resultados que se buscan obtener de los pares de variables independientes observados es que no mantengan correlaciones positivas ni cercanas a +1, lo cual podría indicar un solapamiento en términos de que el par de variables evaluado podría estar midiendo la misma característica.

Según la bibliografía consultada, el valor de referencia utilizado para considerar la existencia de una alta correlación entre dos variables independientes es de ± 0.7 ; para efectos de la investigación este valor será tomado como el de referencia. Es prudente mencionar que, el valor puede variar según las consideraciones pertinentes de las áreas de los estudios que se estén realizando.

Una vez finalizado el análisis de la matriz de correlación se define el modelo preliminar, del cual se obtendrá un modelo definitivo después de ser sometido un proceso de selección de las variables más significativas y principio de criterios de información. Este proceso será expandido en el siguiente apartado.

III.6.2 Modelo y justificación

Con el fin de identificar cuáles de las variables explicativas presentan una mayor influencia y poder predictivo en el éxito deportivo de los equipos de las principales ligas europeas, se empleará un modelo de regresión lineal múltiple. A través de la regresión lineal múltiple se puede generar un modelo lineal con el cual es posible medir la influencia de las variables independientes sobre la variable dependiente o describir de manera simplificada complejas relaciones entre la variable dependiente y algunas de las variables independientes; es por estas razones y por ser un modelo que utiliza variables continuas, que fue seleccionado el modelo multivariable de regresión.

Otras formas de modelo fueron consideradas, tales como panel data. Según Hsiao et al. (1993), Panel data tiene una inferencia más precisa de los parámetros del modelo dado que generalmente contienen más grados de libertad y más variabilidad muestral que datos transversales (cross section) y que las series de tiempo, mejorando la eficiencia de las estimaciones econométricas del modelo; si bien esto permite tener mayor precisión presenta

una serie de aspectos negativos que fueron tomados en consideración y por los cuales fue descartado su uso.

Un conjunto panel data contiene n individuos, cada uno de los cuales incluye T observaciones las cuales son medidas en un período de tiempo. Tenemos entonces que, el número total de observaciones en un panel data es $n*T$. Según Baltagi (2005), existe un problema en las series temporales con dimensiones cortas, las cuales únicamente cubren lapso de tiempo reducido para cada individuo o componente que se está evaluando, en donde explica que "...los argumentos asintóticos se basan fundamentalmente en que el número de individuos necesariamente tiene que tender al infinito". En otras palabras, al querer trabajar con series temporales de muy pocos plazos de tiempo, bien estos plazos estén expresados de forma anual, mensual, bimensual, etc.; para que el modelo no pierda sus capacidades predictivas, el número de observaciones de los individuos debe ser considerablemente más grande.

Según Hill et al. (2020), otra deficiencia que presenta panel data es la alta tendencia que se cometa con mayor frecuencia errores de tipo II, lo cual puede provocar que existan coeficientes sesgados y errores estándar imprecisos. Según Cameron & Trivedi (2009) existe el panel corto, el cual tiene muchos individuos (n grande) pero muy pocos períodos de tiempo (T pequeña), y a su vez existe su contraparte, el panel largo con muchos períodos de tiempo (T grande) pero pocas entidades. Cameron & Trivedi (2009) explican que el tamaño de n no debe ser demasiado pequeño (error de tipo I) ni tampoco extremadamente grande (error de tipo II). Asimismo, Hill et al. (2020) también exponen que los p valores pueden resultar engañosos con lo cual añaden que se debe considerar la posible presencia de heterogeneidad no observada o variabilidad entre los efectos evaluados en otros modelos de regresión, siendo este posiblemente el mayor de los desafíos que presenta el uso de panel data.

Luego de consultar la bibliografía y dados los grandes problemas que se afrontaron para la recopilación de la información y que solo se pudieron recolectar o estaban a disposición datos de 4 temporadas (o años), y que a su vez el número de información de los equipos (individuos) no es abrumadoramente grande para compensar la falta de una serie temporal más extensa, en aras de poder elaborar un modelo que tenga la mayor capacidad predictiva se optó por utilizar un modelo de regresión lineal multivariable.

No se busca desestimar que este o futuros trabajos no sean realizados utilizando panel data, pero consideramos que es prudente revisar a fondo todas las implicaciones que pueden existir a pesar de probablemente ser un método más poderoso que la regresión lineal multivariable.

Una vez establecido el tipo de modelo que será empleado, tenemos que la ecuación de un modelo de regresión lineal múltiple mantiene la siguiente estructura:

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \dots + \beta_n X_{ni} + \varepsilon_i$$

En donde:

Y_i : variable dependiente.

β_0 : representa el valor intercepto o valor constante del modelo. Este valor es el que adopta la variable dependiente Y_i cuando las variables independientes son cero.

β_i : son los coeficientes parciales de regresión de cada variable explicativa X_i , los cuales denotan vectores que representan el efecto promedio generado por el incremento en una unidad de la variable explicativa X_i sobre la variable dependiente Y_i , manteniéndose el

resto de las variables constantes (*ceteris paribus*). El cálculo de los coeficientes del modelo se realiza a través del proceso de mínimos cuadrados ordinarios (MCO)¹.

ε_i : representa el residuo o error idiosincrático estimado del modelo, mide la diferencia entre el valor estimado por el modelo y el valor observado.

Una vez definido el modelo, este será sometido a los pruebas de rigor para evaluar su nivel predictivo y de esta manera confirmar que cumple con todos los principios para ser aceptado como un instrumento. Es prudente mencionar que, a través de los Criterios de Información de Akaike (AIC)² basados en el principio de parsimonia, será seleccionado el modelo definitivo el cual debe tener el mejor nivel de estimación ante la menor cantidad de estimadores. Esto hace referencia a que el mejor modelo es aquel capaz de explicar con mayor precisión la variabilidad observada en la variable dependiente empleando el menor número de predictores o variables independientes.

Adicionalmente, el programa que será utilizado para generar los diferentes modelos predictivos será *R-Studio*. La razón principal es que *R-Studio* es un conjunto de programas para el manejo de datos en el que se pueden realizar rápidamente simulaciones, procesar modelos lineales y no lineales, y que no opera bajo un sistema de codificación sino con un sistema de paquetes o “*scripts*” pre-codificados que permiten automatizar los procesos y realizar las estimaciones de los modelos de manera más rápida pero con el mismo nivel de precisión. Otra ventaja de *R-Studio* es que se puede conseguir sin costo adicional al igual que sus paquetes o “*scripts*”.

¹ El cálculo de estimadores por el proceso de mínimos cuadrados utiliza como criterio la minimización de la suma de los cuadrados residuales (SCR). Los valores de los coeficientes deben satisfacer la siguiente expresión:

$Minimizar_{\beta_0, \beta_1} SCR = \sum_{i=1}^n \hat{u}_i^2$ para n = variables escogidas para el modelo.

² El método de Akaike proporciona una medida que penaliza el sobre ajuste del modelo permitiendo seleccionar entre dos modelos el que mejor ajuste tiene con la menor cantidad de parámetros

$AIC = -2 \log(\mathcal{L}(\hat{\theta})) + 2K$

III.6.2.1 Coeficiente de determinación (R^2)

El coeficiente de determinación (R^2), tal como lo define Gujarati & Porter (2008), “... es una medida comprendida que dice cuán bien se ajusta la línea de regresión muestral a los datos”, en otras palabras, el coeficiente de determinación mide que proporción o porcentaje de la variación de la variable dependiente Y es explicada por el modelo de regresión. También se puede expresar como el porcentaje de la variación de la variable dependiente Y que se puede explicar a través de las variables independientes del modelo. El coeficiente de determinación (R^2) es la medida de bondad del ajuste del modelo comúnmente empleada para determinar cuánto se ajustan los datos al modelo de regresión.

El coeficiente de determinación (R^2) queda expresado con la siguiente fórmula:

$$R^2 = 1 - \frac{SCR}{SCT}$$

En donde:

SCR : es la suma de los cuadrados de los residuos o la suma de los cuadrados no explicada por el modelo, entendiéndose como la variación no explicada de los valores de Y , la cual viene a ser representada de la siguiente manera:

$$SCR = \sum_{i=1}^n (Y_i - \hat{Y}_i)^2$$

SCT: es la suma total de los cuadrados respecto a la media, dicho de otra manera, es la variación de los valores reales de Y medido respecto a la media de $\bar{Y}$, y queda representado de la siguiente forma:

$$SCT = \sum_{i=1}^n (Y_i - \bar{Y}_i)^2$$

El resultado obtenido por el coeficiente de determinación será un valor ubicado entre 0 y 1, por lo que podemos definir los límites del resultado de la siguiente manera:

$$0 \leq R^2 \leq 1$$

En donde $R^2 = 1$ significaría un ajuste perfecto de los datos con el modelo, mientras que el caso contrario de $R^2 = 0$ implicaría la no existencia de ningún ajuste de los datos; por ende lo que se busca el modelo tenga un coeficiente de determinación cercano o igual a 1, o en su defecto que la suma cuadrada de los residuos sea cero o lo más cercana a cero posible para poder obtener un coeficiente más grande.

Para efectos del presente trabajo, los valores que estén cerca de 0.7 serán considerados como suficientemente explicativos por parte de las variables independientes X_i respecto a las variaciones que tenga la variable dependiente Y_i ; tal evidencian otras investigaciones consultadas en el área deportiva.

III.6.2.2 Nivel de significancia

La significancia estadística viene a ser definida como la probabilidad de que una o varias variables explicativas o independientes tengan influencia significativa sobre la

variable explicada o dependiente. Es decir, mide la probabilidad de que la relación de las variables independientes del modelo respecto a la variable dependiente no se debe al azar y queda demostrado matemáticamente que el estimador es confiable.

El nivel de significancia va a ser expresado con la letra α y su valor va a ser asignado de manera arbitraria según la discreción del investigador. Por lo general, el valor de α va a ser un valor igual a 0.01, 0.05 o 0.1. La razón para la asignación de estos valores está relacionada con la prevención de cometer los errores de Tipo I y Tipo II.³

Para realizar el cálculo de la significancia conjunta de los estimadores es necesario realizar un contraste de hipótesis en donde:

$$H_0 : \beta_i = \beta_{i+1} = \beta_{i+2} = \dots = \beta_k = 0$$

$$H_1 : \beta_i \neq \beta_{i+1} \neq \beta_{i+2} \neq \dots \neq \beta_k \neq 0$$

En este contraste de hipótesis se busca un resultado tal que podamos rechazar la hipótesis nula H_0 la cual establece que ninguno de los estimadores calculados tiene una influencia significativa sobre la variable Y . Para esto es necesario estimar el *p-valor* del estadístico de la prueba F . La razón por la que se utiliza la prueba F y no la prueba t se debe a que se realizará un contraste conjunto y no varios contrastes individuales, siendo la prueba F la que permite realizar dicho contraste. Explicado de otra forma, no es similar un contraste conjunto a dos contrastes individuales

³ El error de tipo I queda definido como la consecuencia de rechazar la hipótesis nula H_0 siendo esta verdadera. Por otro lado, el error de tipo II se comete cuando se acepta la hipótesis nula siendo esta falsa. Si se disminuye el nivel de significancia, disminuye al mismo tiempo la probabilidad de cometer el error de tipo I pero aumenta la probabilidad de cometer el error de tipo II. Pasa lo contrario al aumentar el nivel de significancia.

$$H_0 : \beta_i = \beta_{i+1} = \beta_{i+2} = \dots = \beta_k = 0$$

$$H_0 : \beta_i = 0 ; H_0 : \beta_{i+1} = 0$$

Tal como lo definen Gujarati & Porter (2008), “el *p-valor* es el valor de significancia más bajo al cual puede rechazarse la hipótesis nula”; en otras palabras, dado un valor de significancia α , siempre que el *p-valor* de la prueba sea un número igual o inferior se puede rechazar la hipótesis nula, y queda representado de la siguiente forma:

$$\alpha \geq p \text{ valor de } F$$

Si bien con el cálculo del *p-valor* se puede rechazar la hipótesis nula a niveles menores del 90% de confianza ($1 - \alpha$), en el presente trabajo asignaremos un valor de significancia $\alpha = 0.05$. Dicho esto, con un *p-valor* menor de 0.05, la hipótesis nula será rechazada con lo que se demuestra que alguna de los estimadores del modelo es significativo respecto a la variable dependiente.

El nivel de significancia no será un valor estándar y su valor será establecido según la prueba que se estén realizando. Debido a que las pruebas realizadas al modelo utilizan contrastes de hipótesis y que a través del *p-valor* podemos encontrar el nivel de significancia más bajo, este será el valor utilizado a lo largo del proyecto de investigación para confirmar si se rechaza o acepta la hipótesis nula en cada una de las pruebas.

III.6.2.3 Intervalos de confianza de los coeficientes

En los procesos de investigación en los cuales se utilizan modelos de regresión no basta con realizar la estimación de una cifra puntual de parámetros desconocidos. La existencia de fluctuaciones en la muestra pueden generar que un valor estimado sea diferente del valor verdadero, aun cuando en la práctica se espera que el promedio de los valores estimados sea igual al valor verdadero, es necesario en aras de tomar las decisiones de investigación más acertadas establecer intervalos de confianza. Siendo el error estándar de un estimador la forma de medir que tan confiable es dicho estimador, es posible construir un intervalo de confianza de un estimador puntual dentro de un rango de “x” errores estándar por lo que, el intervalo tendrá una probabilidad de incluir el verdadero valor del parámetro desconocido según la cantidad de errores estándar que se utilicen para su elaboración.

En la práctica no es usual que se conozca el verdadero valor de la varianza poblacional, por consiguiente se utiliza una distribución t con $n - 2$ grados de libertad para la construcción del intervalo en lugar de una distribución normal, siendo la muestra suficientemente grande para que se cumpla el principio básico del Teorema del Límite Central, por consiguiente tendremos que la aproximación a la distribución normal queda representada de la siguiente manera:

$$\bar{X} \rightarrow N\left(\mu, \frac{s}{\sqrt{n}}\right)$$

Como se mencionó anteriormente, la varianza no es conocida por lo que se representa la cuasidesviación típica de la muestra. Por consiguiente, centrando el estimador cambiando tanto el origen como la escala:

$$Z = \frac{\bar{x} - \mu}{s \sqrt{n}} \rightarrow N(0; 1)$$

Por lo que queda definido el intervalo de confianza como:

$$\left[\bar{x} - z_{\frac{\alpha}{2}} \frac{s}{\sqrt{n}} ; \bar{x} + z_{\frac{\alpha}{2}} \frac{s}{\sqrt{n}} \right]$$

En donde bajo la α es el valor de la significancia y tendrá un valor de 0.05, siendo el valor de $z_{\frac{\alpha}{2}}$ el valor que se ubica en la tabla de distribución de t .

Tal como lo exponen Gujarati & Porter (2011), dada una significancia de $\alpha = 0.05$, en 95 de 100 de las pruebas que se estimen, los intervalos contendrán el verdadero valor del parámetro, pero con esto no se puede afirmar que el intervalo específico de la ecuación con la que quedan definidos los intervalos de confianza contenga el verdadero valor del parámetro estimado con un 95% de probabilidad, esto se debe a que el intervalo deja de ser aleatorio para ser fijo, por ende, la probabilidad de que el verdadero valor del parámetro esté contenido en el intervalo es de 1 o 0. En otras palabras, una vez definido el intervalo para un parámetro estimado, el verdadero valor del parámetro se encontrará ubicado o no en el intervalo calculado.

III.6.2.4 Pruebas de Homocedasticidad (Breusch-Pagan)

Otro de los principios o criterios básicos que debe cumplir un modelo de regresión es el principio de homocedasticidad, el cual se basa en que los errores o perturbaciones u_i posean todos la misma varianza. Visto en forma de ecuación, el principio se define como:

$$E(u_i^2) = \sigma^2 \quad \text{en donde } i = 1, 2, \dots, n$$

En caso de no cumplirse este principio se podría tener el caso contrario denominado como heteroscedasticidad. Con ello se tendría un modelo menos fiable, debido a que los intervalos de confianza serían de un tamaño tan grande e innecesario que coeficientes que pueden ser significativos en apariencia no lo son debido a que los intervalos de confianza no son los correctos. Esto puede afectar la precisión de las pruebas de t y F . En otras palabras, de proseguir con las estimaciones con un modelo con ausencia de homocedasticidad los análisis podrían ser interpretaciones erróneas debido a resultados que no son tan acertados o lejanos de resultados reales.

Para lograr establecer si existe presencia o no de homocedasticidad, la prueba más aceptada es la prueba Breusch-Pagan-Godfrey en la que, a través de un contraste de hipótesis se busca aceptar la hipótesis nula en la que se plantea la presencia de homocedasticidad.

La prueba consiste en la construcción de un modelo alterno partiendo del modelo original que fue elaborado en el que a través del proceso de MCO se obtendrán los residuos $\hat{u}_1, \hat{u}_2, \dots, \hat{u}_n$:

$$\sigma^2 = \alpha_1 + \alpha_2 Z_{2i} + \dots + \alpha_n Z_{ni}$$

Siendo que $\alpha_2 = \alpha_3 = \dots = \alpha_n = 0$ y se cumple que $\sigma_i^2 = \alpha_1$ que representa la constante del modelo alternativo estimado, la hipótesis nula H_0 queda definida de la siguiente forma

$$H_0: \alpha_2 = \alpha_3 = \dots = \alpha_n = 0$$

Para poder determinar si se rechaza o acepta la hipótesis nula, el estadístico de la prueba viene a tener una distribución que va a tender a chi cuadrado χ_{n-1}^2 con $n - 1$ grados de libertad, denotado bajo la siguiente expresión:

$$\theta \sim \chi_{n-1}^2$$

Al igual que para la medición del nivel de significancia de los coeficientes, la aceptación o rechazo de la hipótesis nula dependerá del *p-valor* obtenido de la prueba χ_{n-1}^2 del estimador, siendo aceptada la hipótesis nula si el p-valor resulta mayor a 0.05, con lo que quedaría demostrada la presencia de homocedasticidad.

III.6.2.5 Pruebas de Autocorrelación de los residuos (Breusch-Godfrey-Pagan)

El término de correlación es definido por Gujarati & Porter (2008) como la “correlación entre miembros de series de observaciones ordenadas en el tiempo [como en datos de series de tiempo] o en el espacio [como en datos de corte transversal]”. En pocas palabras, es la correlación existente entre la perturbación o error de cualquier valor de la muestra respecto a cualquier otra perturbación o error de cualquier otro valor de la misma muestra. Puede ser del orden serial, para el caso de series de tiempo; o por su parte del orden

espacial, cuando se trabaja con muestras de datos de cortes transversal. Este representa otro de los supuestos básicos con los que debe contar el modelo para ser considerado de un alto poder predictivo debido a que no incurriría en el problema de sesgo.

El análisis de la autocorrelación es utilizado para demostrar que la perturbación de un valor de la muestra no será afectada o influenciada por la perturbación de otro valor de la muestra. Debido a que la autocorrelación puede ser del orden serial, la existencia de la misma puede no ser de valores que se siguen en términos ordinales, pero de valores de rezagos de la misma muestra, es decir, perturbaciones de valores rezagadas que afecten a otras perturbaciones de otros valores. La consecuencia de la presencia de autocorrelación, al igual que en el caso de la heteroscedasticidad, los estimadores dejan de ser eficientes debido a que se incumple el principio de mínima varianza.

La presencia de autocorrelación se puede deber a; la presencia de un componente cíclico o tendencial en la variable dependiente y este al no ser explicado por las variables independientes se encontrará contenido en las perturbaciones o residuos del modelo, otra razón se debe a los errores de especificación, bien sea por la omisión de variables relevantes o por la forma de la funcionalidad del modelo.

A estos efectos, el criterio de ausencia de autocorrelación queda definido como:

$$\text{cov}(u_i, u_j | x_i, x_j) = E(u_i, u_j) = 0 \quad \text{en donde } i \neq j$$

Caso contrario, la existencia de autocorrelación queda representada de la siguiente manera:

$$E(u_i, u_j) \neq 0 \quad \text{en donde } i \neq j$$

Para demostrar la ausencia o presencia de autocorrelación de los residuos existen múltiples pruebas, para efectos de este trabajo se realizará la prueba de Breusch-Godfrey también conocida como la prueba del multiplicador de LaGrange-Breusch-Godfrey, la cual consiste en una estimación aparte del modelo original en la que se calculan los errores $\hat{u}_t$ bajo un esquema de modelo de autorregresión de orden p, AR (p) del cual se obtendrá un nuevo R^2 .

El nuevo modelo queda definido de la siguiente manera:

$$u_t = \rho_1 u_t + \rho_2 u_{t-2} + \dots + \rho_p u_{t-p} + \varepsilon_t$$

En donde u_t son los errores estimados del modelo original, ε_t son los errores de ruido blanco (white noise) y los ρ_p son los niveles de covarianza existentes entre las nuevas variables determinadas. Dado este modelo alternativo, se establece entonces un contraste de hipótesis en el que se busca no rechazar la hipótesis nula

$$H_0: \rho_1 = \rho_2 = \dots = \rho_p = 0$$

En donde se tiene que, al ser todas las covarianzas de las nuevas variables estimadas igual a cero se demuestra que no se tiene evidencia de que los residuos de los valores estén influyendo entre ellos, por ello se aceptaría la hipótesis nula (H_0)

Con un tamaño de la muestra suficientemente grande se tiene que $(n-p)$ veces el valor del R^2 calculado de la regresión alterna se aproximará a una distribución chi cuadrada χ_p^2 de orden p . Queda expresado de la siguiente manera:

$$(n - p)R^2 \sim \chi_p^2$$

Si al calcular $(n - p) R^2$ el valor excede el valor crítico de la distribución chi cuadrado χ_p^2 dado un nivel de significancia, se tendría evidencia de la presencia de autocorrelación en los residuos, lo cual nos lleva a rechazar la hipótesis nula H_0 .

Al igual que como se ha trabajado en las pruebas anteriores, para efectos del presente trabajo, el proceso de evaluación del contraste de hipótesis será realizado contra el *p-valor* de la prueba de chi cuadrado χ_p^2 . De modo que, si el *p-valor* es mayor al término de significancia $\alpha = 0.05$ se acepta la hipótesis nula, quedando demostrado que no existe presencia de autocorrelación de los residuos.

III.6.2.6 Pruebas de Normalidad de los residuos (Kolmogorov-Smirnov)

En la teoría clásica de los modelos de regresión lineal, tal como lo explican Gujarati & Porter (2008), los errores o perturbaciones u_i estarán distribuidos de forma normal de manera que:

Media: $E(u_i) = 0$

Varianza: $E(u_i^2) = \sigma^2$

Covarianza: $cov(u_i, u_j) : E(u_i, u_j) = 0 \quad i \neq j$

Por lo que se puede decir que:

$$u_i \sim N(0, \sigma^2)$$

Es importante resaltar que la distribución normal de los errores justifica la utilización de las pruebas t y F , las cuales pueden ser aplicadas con validez en los contrastes de hipótesis. Asimismo, con una distribución normal de los errores, los estimadores serán insesgados, eficientes y presentarán consistencia.⁴

Para determinar que los errores se encuentran distribuidos de forma normal es común el uso del análisis de los gráficos *qq-plot* y el cálculo de la prueba de bondad de ajuste Kolmogorov-Smirnov, esta última es una prueba que consiste en realizar un contraste o comparación de las frecuencias relativas acumuladas $F(x_i)$ de la distribución teórica $F_t(x_i)$ y de la distribución observada o de la muestra $F_o(x_i)$, en este caso el del modelo, cuantificando la distancia vertical máxima entre la diferencia de las frecuencias acumuladas.

⁴ Se define como consistencia a la característica que presentan los estimadores de converger hacia sus verdaderos valores poblacionales a medida que el tamaño de la muestra aumenta indefinidamente.

De forma que:

$$D = \max |F_t(x_i) - F_o(x_i)|$$

El cual será un estadístico que calculará el valor absoluto máximo de la diferencia de las frecuencias relativas acumuladas de las distribuciones teóricas y observadas para el mismo valor o intervalo de la variable.

Al igual que las pruebas anteriores, el contraste de hipótesis se realizará con el *p-valor* que arroje la prueba de Kolmogorov-Smirnov en donde un *p-valor* mayor a el nivel de significancia $\alpha = 0.05$ nos permite aceptar la hipótesis nula, con lo que se demuestra que los errores se encuentran distribuidos de forma normal.

III.6.2.7 Pruebas de Multicolinealidad de los residuos (Factor de Inflación de la Varianza)

Se define como multicolinealidad a la relación de dependencia lineal existente entre dos o más variables independientes de un modelo de regresión, la cual en caso de ser exacta o perfecta incumple el principio del teorema Gauss-Márkov.

Gujarati & Porter (2008) establecen que cuando la multicolinealidad es perfecta para un modelo con n número de variables si cumple con la siguiente condición:

$$\lambda_1 X_1 + \lambda_2 X_2 + \dots + \lambda_n X_n = 0$$

Siendo $\lambda_1, \lambda_2, \dots, \lambda_n$ constantes que no son iguales a cero de forma simultánea. En este sentido los coeficientes parciales de regresión de las variables independientes serán imposibles de determinar y sus errores estándar tenderán a un valor infinito.

Asimismo, también existe la multicolinealidad no perfecta, en donde una cantidad de las variables se encuentran correlacionadas pero no de forma exacta. La forma de expresarlo a través de una demostración es la siguiente:

$$\lambda_1 X_1 + \lambda_2 X_2 + \dots + \lambda_n X_n + v_i = 0$$

En donde v_i representa el error estocástico⁵. En este caso, los coeficientes parciales de regresión se pueden determinar pero mantienen errores estándar de gran tamaño, por lo que la precisión con la que son determinados los coeficientes no es buena.

Existen dos formas de medición que son las más comunes para medir el nivel de multicolinealidad de las variables independientes, los cuales son la Tolerancia (*TOL*) y el Factor de Inflación de la Varianza (*FIV*), en donde un nivel de Tolerancia (*TOL*) cercano de cero se representará mayor multicolinealidad y un número cercano a uno representará ausencia de multicolinealidad entre los regresores. Por su parte, un número cercano a cero en el Factor de Inflación de la varianza (*FIV*) representará ausencia de multicolinealidad, un valor comprendido entre cero y cinco será equivalente a una multicolinealidad considerable,

⁵ La palabra estocástico hace referencia a la aleatoriedad, por ende, el error estocástico queda definido como el error aleatorio.

por encima de cinco o cercano a diez demuestra alta multicolinealidad y un número igual o mayor a diez representará una multicolinealidad perfecta.

Para efectos del presente estudio, únicamente serán evaluados los Factores de Inflación de las Varianzas (*FIV*) de cada una de las ligas.

CAPÍTULO IV. ANÁLISIS DE RESULTADOS

Partiendo de la estructura base de modelo de regresión lineal, se procedió a estimar un modelo para cada una de las ligas que fueron consideradas en el presente estudio, con la finalidad de obtener un acercamiento sobre la relación que tienen los salarios con el desempeño de los equipos de fútbol. A continuación, se realizará un análisis de los resultados estadísticos obtenidos para cada uno de los modelos correspondientes a las diferentes ligas que se están evaluando.

IV.1 Matrices de Correlación

A continuación se presentan las matrices de correlación correspondientes a cada una de las ligas seleccionadas para la elaboración del presente estudio.

Tabla 2. Matriz de Correlación Liga Inglesa

	<i>Pts</i>	<i>Salarios</i>	<i>Compras</i>	<i>Ventas</i>	<i>Valor T.</i>	<i>Edad</i>	<i>Estatura</i>
Pts	1						
Salarios	0.7042	1					
Compras	0.5441	0.6210	1				
Ventas	0.3767	0.3963	0.6010	1			
Valor T.	0.7557	0.8257	0.5622	0.3887	1		
Edad	-0.2815	-0.2559	-0.2418	-0.3337	-0.2867	1	
Estatura	-0.2728	-0.1642	-0.2416	-0.1618	-0.1600	0.1980	1

Fuente: Elaboración propia con el uso de R-Studio

Tabla 3. Matriz de Correlación Liga Española

	<i>Pts</i>	<i>Salarios</i>	<i>Compras</i>	<i>Ventas</i>	<i>Valor T</i>	<i>Edad</i>	<i>Estatura</i>
Pts	1						
Salarios	0.7602	1					
Compras	0.6320	0.6975	1				
Ventas	0.5467	0.6469	0.8105	1			
Valor T	0.7603	0.9819	0.7087	0.6551	1		
Edad	0.0683	0.0911	0.0141	0.0366	0.0576	1	
Estatura	0.0333	-0.0751	-0.0763	-0.1007	-0.0610	-0.2457	1

Fuente: Elaboración propia con el uso de R-Studio

Tabla 4. Matriz de Correlación Liga Francesa

	<i>Puntos</i>	<i>Salarios</i>	<i>Compras</i>	<i>Ventas</i>	<i>Valor T.</i>	<i>Edad</i>	<i>Estatura</i>
Pts	1						
Salarios	0.7162	1					
Compras	0.6352	0.8567	1				
Ventas	0.3098	0.3915	0.6724	1			
Valor T.	0.6682	0.8863	0.9028	0.5244	1		
Edad	-0.2466	0.0700	-0.1781	-0.2394	-0.1586	1	
Estatura	-0.0832	-0.0524	0.0399	0.1063	-0.0224	0.0367	1

Fuente: Elaboración propia con el uso de R-Studio

Tabla 5. Matriz de Correlación Liga Alemana

	<i>Pts</i>	<i>Salarios</i>	<i>Compras</i>	<i>Ventas</i>	<i>Valor T</i>	<i>Edad</i>	<i>Estatura</i>
Pts	1						
Salarios	0.7027	1					
Compras	0.5049	0.5453	1				
Ventas	0.2456	0.2990	0.4713	1			
Valor T	0.7274	0.8996	0.6062	0.3811	1		
Edad	-0.0038	0.1456	-0.2526	-0.0935	0.1067	1	
Estatura	-0.2719	-0.1374	-0.2512	-0.1172	-0.2446	0.0672	1

Fuente: Elaboración propia con el uso de R-Studio

Tabla 6. Matriz de Correlación Liga Italiana

	<i>Pts</i>	<i>Salarios</i>	<i>Compras</i>	<i>Ventas</i>	<i>Valor T</i>	<i>Edad</i>	<i>Estatura</i>
Pts	1						
Salarios	0.8004	1					
Compras	0.7036	0.8954	1				
Ventas	0.6451	0.6456	0.6706	1			
Valor T	0.7818	0.9345	0.8745	0.6543	1		
Edad	0.0956	0.1146	0.0246	0.0301	0.0724	1	
Estatura	0.0760	0.0124	0.0899	0.1045	-0.0133	-0.0732	1

Fuente: Elaboración propia con el uso de R-Studio

Tal como se puede apreciar las tablas presentadas 2-6 representan las matrices de correlación correspondientes a cada una de las ligas estudiadas. En ellas se busca encontrar niveles de correlación bajo por cada par de variables independientes analizado. Al poseer un orden simétrico, cada matriz poseerá una diagonal principal la cual será igual a uno, mientras

que los resultados serán idénticos en ambos lados de la diagonal principal, razón por la cual se dejarán los campos del lado derecho sin valores.

Para el caso de Inglaterra, no existen niveles de alta correlación entre las variables independientes. Para los casos de las demás ligas, existen casos puntuales de niveles de correlación fuerte entre pares de variables independientes; en España se puede observar con el nivel correlación entre las variables *Salarios/Valor de Temporada* el cual es de $r = 0.9819$, en Francia con el par de variables *Compras/Valor de Temporada* con un $r = 0.9028$, por su parte en Alemania se tiene que *Salarios/Valor de Temporada* tiene un $r = 0.8996$, por último en Italia se tiene que los pares de variables *Salarios/Compras* con $r = 0.8954$ y *Salarios/Valor de Temporada* con $r = 0.9345$ representan los pares de variables de cada liga que poseen un coeficiente de correlación muy grande. Se asume la posibilidad que los pares mencionados podrían estar midiendo de manera independiente las mismas características. La variable que sea menos significativa será descartada del modelo, este se podrá apreciar con los modelos preliminares y los modelos definitivos sugeridos.

Vale la pena destacar que las matrices de correlación permiten dar una primera lectura de la relación de las variables independientes con la variable dependiente. Como se puede apreciar en cada una de las matrices, la variable salarios si bien presenta niveles de correlación considerablemente elevados, no en todas las ligas representa el mayor nivel de correlación con la variable *Puntos*. Para los casos de Alemania e Inglaterra, la variable *Valor de Temporada* presenta en ambas ligas un mayor nivel de correlación con la variable *Puntos*, lo cual puede ser un indicativo de que la variable *Salarios* no es la variable que mejor pueda explicar el desempeño de los equipos en esas ligas. La variable *Compras* no presenta niveles importantes de significancia respecto a la variable dependiente.

IV.2 Modelos de Regresión Lineal Multivariantes

En el presente apartado serán presentados los modelos de estimación preliminares como los modelos definitivos de cada una de la ligas bajo el mismo orden en que fueron presentadas las matrices de correlación. Únicamente serán presentados por cada liga el primer modelo sugerido, el cual cuenta con todas las variables independientes que fueron consideradas, y el modelo definitivo, siendo este último el que será presentado como el modelo sugerido para los análisis de los objetivos del presente estudio siguiendo los principios de criterios de información y parsimonia.

Tabla 7. Modelo Preliminar Liga Inglesa

Call:					
lm(formula = PtsInglaterra ~ SalarioInglaterra + ComprasInglaterra + VentasInglaterra + ValorTemporadaInglaterra + EdadInglaterra + EstaturaInglaterra)					
Residuals:					
Min	1Q	Median	3Q	Max	
-26.681	-6.788	-0.829	6.675	39.645	
Coefficients:					
	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	3.356e+02	1.674e+02	2.005	0.048689	*
SalarioInglaterra	1.527e-06	1.095e-06	1.394	0.167654	
ComprasInglaterra	2.812e-02	3.390e-02	0.829	0.409557	
VentasInglaterra	7.367e-03	4.040e-02	0.182	0.855804	
ValorTemporadaInglaterra	4.363e-02	1.106e-02	3.945	0.000182	***
EdadInglaterra	-6.481e-01	1.525e+00	-0.425	0.672178	
EstaturaInglaterra	-1.577e+00	9.202e-01	-1.713	0.090918	.

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1					
Residual standard error: 11.95 on 73 degrees of freedom					
Multiple R-squared: 0.62, Adjusted R-squared: 0.5888					
F-statistic: 19.85 on 6 and 73 DF, p-value: 1.319e-13					

Fuente: Elaboración propia con el uso de R-Studio

Tabla 8. Modelo Definitivo Liga Inglesa

```
Call:
lm(formula = PtsInglaterra ~ SalarioInglaterra + ValorTemporadaInglaterra +
    EstaturaInglaterra)

Residuals:
    Min       1Q   Median       3Q      Max
-26.169  -7.019  -1.362   4.698  38.695

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)    3.618e+02  1.621e+02   2.232   0.0285 *
SalarioInglaterra  1.913e-06  1.022e-06   1.871   0.0651 .
ValorTemporadaInglaterra  4.564e-02  1.079e-02   4.231  6.44e-05 ***
EstaturaInglaterra  -1.812e+00  8.867e-01  -2.043   0.0445 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 11.82 on 76 degrees of freedom
Multiple R-squared:  0.6126,    Adjusted R-squared:  0.5973
F-statistic: 40.05 on 3 and 76 DF,  p-value: 1.246e-15
```

Fuente: Elaboración propia con el uso de R-Studio

Como se puede apreciar en la tabla-7, se presenta el modelo para el estudio global de la Liga Inglesa con todas las variables independientes. Tal como se mencionó en el análisis de la matriz de correlación, la variable *Valor de Temporada* es la variable más significativa del modelo preliminar al poseer el menor *p-valor* (***) Siguiendo los principios de los criterios de información, se procede a través de los comandos de *R-Studio* a generar un modelo que presente la menor cantidad de criterios informativos, obteniendo como resultado el modelo expresado en la tabla-8, en el que, la variable *Valor de Temporada* sigue siendo la de mayor nivel de significancia y en donde la variable *Salarios* es la que cuenta con el menor nivel de significancia con *p-valor* mayor al nivel de significancia para un contraste de 95% de confianza ($p\text{-valor} = 0.0651 > \alpha = 0.05$), lo cual indica que los salarios no van a tener una fuerte influencia sobre los puntos obtenidos por los equipos de la liga inglesa. El coeficiente de determinación ajustado es de $R^2_{ajustado} = 0.5973$ por lo que el modelo sugerido explica cerca de un 60% las variaciones en los puntos obtenidos por los equipos de

la liga inglesa. Asimismo, el *p*-valor del modelo es inferior al nivel de significancia *p*-valor < 0.05, quedando demostrado que el modelo en su conjunto es significativo.

Tabla 9. Modelo Preliminar Liga Española

Call:				
lm(formula = PtsEspana ~ SalarioEsp + ComprasEsp + VentasEsp + ValorTemporadaEsp + EdadEsp + EstaturaEsp)				
Residuals:				
Min	1Q	Median	3Q	Max
-25.8455	-6.0098	0.8629	8.4894	25.5260
Coefficients:				
	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	-2.810e+02	2.381e+02	-1.180	0.242
SalarioEsp	2.176e-06	2.261e-06	0.962	0.339
ComprasEsp	6.996e-02	4.246e-02	1.647	0.104
VentasEsp	-1.710e-02	5.127e-02	-0.334	0.740
ValorTemporadaEsp	1.883e-02	2.855e-02	0.660	0.512
EdadEsp	7.595e-01	1.385e+00	0.548	0.585
EstaturaEsp	1.666e+00	1.251e+00	1.331	0.187
Residual standard error: 11.38 on 73 degrees of freedom				
Multiple R-squared: 0.6111, Adjusted R-squared: 0.5791				
F-statistic: 19.12 on 6 and 73 DF, p-value: 2.99e-13				

Fuente: Elaboración propia con el uso de R-Studio

Tabla 10. Modelo Definitivo Liga Española

```
Call:
lm(formula = PtsEspana ~ SalarioEsp + ComprasEsp)

Residuals:
    Min       1Q   Median       3Q      Max
-25.9443  -7.1651  -0.1028   9.1020  25.2465

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  4.079e+01  1.656e+00  24.638  < 2e-16 ***
SalarioEsp   3.583e-06  5.808e-07   6.169  2.96e-08 ***
ComprasEsp   6.175e-02  3.143e-02   1.965   0.0531 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 11.27 on 77 degrees of freedom
Multiple R-squared:  0.598,    Adjusted R-squared:  0.5876
F-statistic: 57.28 on 2 and 77 DF,  p-value: 5.77e-16
```

Fuente: Elaboración propia con el uso de R-Studio

Diferente a lo que se apreció en la Liga Inglesa, en el modelo preliminar de la Liga Española no se logra apreciar alguna variable significativa (tabla-9), aun cuando el modelo en su conjunto tiene un *p-valor* < 0.05 que permite demostrar que es significativo. Una vez obtenido el modelo definitivo (tabla-10) se puede observar que tanto la variable *Salarios* como el *Intercepto* (***) del modelo son significativos para un nivel de confianza de 95%, siendo la variable *Compras* no significativa para dicho nivel de confianza. Con un coeficiente de regresión para la variable Salarios de $3.58 \cdot 10^{-6}$, nos indica que por cada millón de euros que aumente el salario promedio de un determinado equipo de la Liga Española tendrá un impacto positivo de 3.58 puntos en su desempeño deportivo de ese equipo. El modelo arroja un R^2 ajustado = 0.5876 y un *p-valor* < 0.05, con lo que se puede afirmar que el modelo es significativo y en su conjunto puede explicar en un aproximado de 60% las variaciones de la variable dependiente.

Tabla 11. Modelo Preliminar Liga Francesa

```
Call:
lm(formula = PtsFrancia ~ SalarioFrancia + ComprasFrancia + VentasFrancia +
  ValorTemporadaFrancia + EdadFrancia + EstaturaFrancia)

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)   2.008e+02  1.976e+02   1.016  0.31301
SalarioFrancia  5.231e-06  1.621e-06   3.227  0.00188 **
ComprasFrancia -8.036e-03  8.767e-02  -0.092  0.92721
VentasFrancia  -6.722e-03  3.826e-02  -0.176  0.86104
ValorTemporadaFrancia 1.372e-02  2.750e-02   0.499  0.61920
EdadFrancia    -2.516e+00  1.060e+00  -2.373  0.02030 *
EstaturaFrancia -5.096e-01  1.090e+00  -0.468  0.64143
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 11.32 on 73 degrees of freedom
Multiple R-squared:  0.555,    Adjusted R-squared:  0.5184
F-statistic: 15.18 on 6 and 73 DF,  p-value: 3.388e-11
```

Fuente: Elaboración propia con el uso de R-Studio

Tabla 12. Modelo Definitivo Liga Francesa

```
Call:
lm(formula = PtsFrancia ~ SalarioFrancia + EdadFrancia)

Residuals:
    Min       1Q   Median       3Q      Max
-26.993  -6.273   0.718   5.667  34.800

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)   1.099e+02  2.590e+01   4.243 6.09e-05 ***
SalarioFrancia  5.790e-06  6.305e-07   9.183 5.32e-14 ***
EdadFrancia    -2.572e+00  9.966e-01  -2.580  0.0118 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 11.06 on 77 degrees of freedom
Multiple R-squared:  0.5517,    Adjusted R-squared:  0.5401
F-statistic: 47.39 on 2 and 77 DF,  p-value: 3.835e-14
```

Fuente: Elaboración propia con el uso de R-Studio

Para el modelo preliminar correspondiente a la Liga de Francia se observa en la tabla-11 que las variables *Salarios* y *Edad* son significativas para un nivel de confianza del 95% respectivamente, siendo estas las únicas variables significativas. En la tabla-12 se aprecia que la variable *Salarios*, la variable *Edad* y el *Intercepto* del modelo son significativas con 95% de nivel de confianza. El modelo definitivo que se obtiene presenta un p-valor del estadístico *F* menor al nivel de significancia ($p\text{-valor} < 0.05$) lo que demuestra que el modelo en su conjunto es significativo con un R^2 ajustado de 0.5401. La variable *Salarios* posee un coeficiente de $5.79 \cdot 10^{-6}$, lo que indica que un aumento promedio de un millón de euros destinada a los salarios realizado por cualquier equipo perteneciente a la Liga de Francia tendría un impacto positivo de 5.79 puntos. Vale la pena mencionar que la variable *Edad* tiene un coeficiente negativo de - 2.572, por lo que un aumento de un año en la edad promedio por parte de cualquier equipo de la Liga de Francia tendría un efecto negativo en el rendimiento deportivo de 2.6 puntos aproximadamente.

Tabla 13. Modelo Preliminar Liga Alemana

```
Call:
lm(formula = PtsAlemania ~ SalarioAlemania + ComprasAlemania +
  VentasAlemania + ValorTemporadaAlemania + EdadAlemania +
  EstaturaAlemania)

Residuals:
    Min       1Q   Median       3Q      Max
-22.184  -5.619   1.245   6.798  21.142

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)    4.042e+02  2.620e+02   1.543   0.1278
SalarioAlemania  2.504e-06  1.564e-06   1.601   0.1143
ComprasAlemania  3.499e-02  6.237e-02   0.561   0.5767
VentasAlemania  -1.737e-02  3.001e-02  -0.579   0.5646
ValorTemporadaAlemania  4.051e-02  2.041e-02   1.985   0.0514 .
EdadAlemania    -1.081e+00  1.335e+00  -0.809   0.4212
EstaturaAlemania -1.863e+00  1.418e+00  -1.314   0.1935
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 10.24 on 65 degrees of freedom
Multiple R-squared:  0.5667,    Adjusted R-squared:  0.5267
F-statistic: 14.17 on 6 and 65 DF,  p-value: 3.048e-10
```

Fuente: Elaboración propia con el uso de R-Studio

Tabla 14. Modelo Definitivo Liga Alemana

```
Call:
lm(formula = PtsAlemania ~ SalarioAlemania + ValorTemporadaAlemania)

Residuals:
    Min       1Q   Median       3Q      Max
-20.954  -6.290   1.067   7.269  26.261

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)   3.473e+01  1.856e+00  18.714  < 2e-16 ***
SalarioAlemania  2.052e-06  1.510e-06   1.359  0.17864
ValorTemporadaAlemania 4.941e-02  1.847e-02   2.675  0.00933 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 10.22 on 69 degrees of freedom
Multiple R-squared:  0.5414,    Adjusted R-squared:  0.5281
F-statistic: 40.72 on 2 and 69 DF,  p-value: 2.094e-12
```

Fuente: Elaboración propia con el uso de R-Studio

En el modelo preliminar de la Liga Alemana (tabla-13) se observa que no hay variables significativas bajo un nivel de confianza de 95%, sin embargo la variable *Valor de Temporada* es significativa para un nivel de confianza del 90% ($p\text{-valor} = 0.1$). A pesar de no tener variables significativas para los estándares de significancia de $\alpha = 0.05$, el modelo en su conjunto es significativo con un $p\text{-valor}$ menor al 0.05. Tal como se había anticipado en el análisis de las matrices de correlación (tabla-14) y al igual que en el caso de la Liga Inglesa, el análisis preliminar que mostraba que la correlación entre la variable dependiente *Puntos* y la variable independiente *Valor de Temporada* sugería que esta última sería la variable con el mayor nivel de significancia o la variable que podría explicar mejor los cambios en la variables dependiente; si bien el *Intercepto* del modelo presenta mejores niveles comparativos, la variable *Valor de Temporada* es la variable independiente con mejor nivel significativo con un $p\text{-valor}$ menor al 0.05. La variable Salarios, a pesar de ser parte del modelo sugerido, no es significativa ni siquiera para un nivel de confianza del 90%.

La variable Valor de Temporada posee un coeficiente de regresión parcial de $4.92 \cdot 10^{-2}$ lo que indica que ante una variación de un millón del valor promedio un equipo dado de la Liga Alemana, ese aumentó representaría un aumento de un estimado de 4.92 puntos. El modelo posee un p-valor menor al nivel de significancia $\alpha = 0.05$, confirmando que el modelo es significativo con un R^2 ajustado = 0.5281.

Tabla 15. Modelo Preliminar Liga Italiana

Call:					
lm(formula = PtsItalia ~ SalarioItalia + ComprasItalia + VentasItalia + ValorTemporadaItalia + EdadItalia + EstaturaItalia)					
Residuals:					
Min	1Q	Median	3Q	Max	
-24.4733	-4.8340	-0.7622	4.8470	30.3367	
Coefficients:					
	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	-2.265e+02	2.462e+02	-0.920	0.36063	
SalarioItalia	6.398e-06	2.155e-06	2.969	0.00404	**
ComprasItalia	-8.588e-02	5.583e-02	-1.538	0.12831	
VentasItalia	1.020e-01	4.102e-02	2.486	0.01522	*
ValorTemporadaItalia	3.376e-02	2.445e-02	1.381	0.17152	
EdadItalia	1.426e-01	1.045e+00	0.136	0.89182	
EstaturaItalia	1.397e+00	1.325e+00	1.054	0.29547	

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1					
Residual standard error: 10.99 on 73 degrees of freedom					
Multiple R-squared: 0.6864, Adjusted R-squared: 0.6607					
F-statistic: 26.63 on 6 and 73 DF, p-value: 2.2e-16					

Fuente: Elaboración propia con el uso de R-Studio

Tabla 16. Modelo Definitivo Liga Italiana

```
Call:
lm(formula = PtsItalia ~ SalarioItalia + VentasItalia)

Residuals:
    Min       1Q   Median       3Q      Max
-24.2814  -6.3917  -0.9476   4.4897  31.2145

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  3.412e+01  1.923e+00  17.737 < 2e-16 ***
SalarioItalia 6.745e-06  8.799e-07   7.666 4.47e-11 ***
VentasItalia  9.985e-02  3.896e-02   2.563  0.0123 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 11 on 77 degrees of freedom
Multiple R-squared:  0.6688,    Adjusted R-squared:  0.6602
F-statistic: 77.76 on 2 and 77 DF,  p-value:  2.2e-16
```

Fuente: Elaboración propia con el uso de R-Studio

Para el modelo preliminar de la Liga Italiana (tabla-15), las variables Salarios y Ventas son las únicas variables significativas para un nivel de confianza del 95%. En el modelo definitivo sugerido (tabla-16), el R^2 ajustado = 0.6602 es el mayor registrado entre todas las ligas evaluadas en el presente estudio. Al igual que en el resto de las ligas, el *p*-valor del estadístico *F* del modelo es menor que el nivel de significancia de $\alpha = 0.05$, demostrando que el modelo en conjunto es significativo. Un aumento del salario promedio para un equipo determinado de la Liga Italiana tendrá un impacto de 6.75 puntos por cada millón que incremente el salario promedio de ese equipo, asimismo, un aumento de las ventas de jugadores en un equipo dado tendrá un impacto positivo de 10 puntos aproximadamente, por cada millón que se incremente en promedio.

IV.3 Pruebas de Homocedasticidad (Breusch-Pagan)

En el siguiente apartado se procede a evaluar los resultados obtenidos en la prueba Breusch-Pagan para confirmar la presencia de homocedasticidad en cada uno de los modelos. Vale la pena recordar que la evaluación del resultado consiste en un contraste de hipótesis en donde aceptar la hipótesis nula (H_0) significa la comprobación de la presencia de homocedasticidad en el modelo. Caso contrario, rechazar la hipótesis nula (H_0) significaría la ausencia de homocedasticidad, lo que equivale a decir que el modelo presenta heteroscedasticidad. Para aceptar la hipótesis nula (H_0) el *p-valor* de la prueba debe ser mayor al nivel de significancia $\alpha = 0.05$.

Tabla 17. Pruebas de Homocedasticidad (Breusch-Pagan) Liga Inglesa

```
studentized Breusch-Pagan test
data:  modelo.lm.Inglaterra1
BP = 0.56984, df = 3, p-value = 0.9033
```

Fuente: Elaboración propia con el uso de R-Studio

Tabla 18. Pruebas de Homocedasticidad (Breusch-Pagan) Liga Española

```
studentized Breusch-Pagan test
data:  modelo.lm.espana1
BP = 3.4155, df = 2, p-value = 0.1813
```

Fuente: Elaboración propia con el uso de R-Studio

Tabla 19. Pruebas de Homocedasticidad (Breusch-Pagan) Liga Francesa

```
studentized Breusch-Pagan test

data:  modelo.lm.Francia1
BP = 3.5702, df = 2, p-value = 0.1678
```

Fuente: Elaboración propia con el uso de R-Studio

Tabla 20. Pruebas de Homocedasticidad (Breusch-Pagan) Liga Alemana

```
studentized Breusch-Pagan test

data:  modelo.lm.Alemania1
BP = 0.26087, df = 2, p-value = 0.8777
```

Fuente: Elaboración propia con el uso de R-Studio

Tabla 21. Pruebas de Homocedasticidad (Breusch-Pagan) Liga Italiana

```
studentized Breusch-Pagan test

data:  modelo.lm.Italia1
BP = 2.5125, df = 2, p-value = 0.2847
```

Fuente: Elaboración propia con el uso de R-Studio

Dados los resultados de las tablas 4.1 – 4.5, se acepta la hipótesis nula (H_0) en todas las ligas evaluadas ($p\text{-valor} > 0.05$) por lo que existe evidencia suficiente para comprobar que los modelos son homocedásticos.

IV.4 Pruebas de Autocorrelación de los residuos (Breusch-Godfrey-Pagan)

En este apartado se examinarán los resultados que se obtuvieron de las pruebas Breusch-Godfrey-Pagan en cada uno de los respectivos modelos. Al igual que con la prueba de Homocedasticidad, se busca a través de un contraste de hipótesis confirmar la presencia o ausencia de autocorrelación de los residuos, para todo $p\text{-valor} > \alpha = 0.05$ se acepta la hipótesis nula (H_0) confirmando la ausencia de autocorrelación de los residuos. Caso contrario, para todo $p\text{-valor} < \alpha = 0.05$ se rechaza la hipótesis nula (H_0) con lo que quedaría determinada la autocorrelación de los residuos en el modelo evaluado.

Tabla 22. Pruebas de Autocorrelación de los residuos (Breusch-Godfrey-Pagan)
Liga Inglesa

Breusch-Godfrey test for serial correlation of order up to 10 data: Residuals LM test = 8.822, df = 10, p-value = 0.5491
--

Fuente: Elaboración propia con el uso de R-Studio

Tabla 23. Pruebas de Autocorrelación de los residuos (Breusch-Godfrey-Pagan)
Liga Española

Breusch-Godfrey test for serial correlation of order up to 10 data: Residuals LM test = 31.867, df = 10, p-value = 0.0004214
--

Fuente: Elaboración propia con el uso de R-Studio

Tabla 24. Pruebas de Autocorrelación de los residuos (Breusch-Godfrey-Pagan)
Liga Francesa

Breusch-Godfrey test for serial correlation of order up to 10
data: Residuals
LM test = 32.971, df = 10, p-value = 0.0002755

Fuente: Elaboración propia con el uso de R-Studio

Tabla 25. Pruebas de Autocorrelación de los residuos (Breusch-Godfrey-Pagan)
Liga Alemana

Breusch-Godfrey test for serial correlation of order up to 10
data: Residuals
LM test = 24.29, df = 10, p-value = 0.006867

Fuente: Elaboración propia con el uso de R-Studio

Tabla 26. Pruebas de Autocorrelación de los residuos (Breusch-Godfrey-Pagan)
Liga Italiana

Breusch-Godfrey test for serial correlation of order up to 10
data: Residuals
LM test = 19.603, df = 10, p-value = 0.03324

Fuente: Elaboración propia con el uso de R-Studio

De los resultados obtenidos, únicamente el modelo de Inglaterra (tabla-22) presenta un *p-valor* superior al nivel de significancia $\alpha = 0.05$ por lo que se acepta la hipótesis nula (H_0) quedando demostrado que hay ausencia de autocorrelación de los residuos. Por su parte, en el caso de Italia (tabla-26), se observa que a un nivel de confianza de 99% ($\alpha = 0.01$) el *p-valor* del estadístico del modelo permite rechazar la hipótesis nula (H_0). Para el resto de los modelos, los *p-valor* son menores para cualquier nivel de significancia que le sea otorgado a α lo cual nos lleva a rechazar la hipótesis nula (H_0) confirmando la presencia de autocorrelación entre los residuos.

IV.5 Pruebas de Normalidad de los residuos (Kolmogorov-Smirnov)

A continuación serán analizados los resultados de las pruebas de normalidad de los residuos de los errores, conocida como prueba Kolmogorov-Smirnov. Al igual que en las pruebas anteriores, se someterá a un contraste de hipótesis en donde al aceptarse la hipótesis nula (H_0) se comprobará que los residuos se encuentran distribuidos de forma normal. La aceptación de la hipótesis nula (H_0) dependerá del p-valor del estadístico de la prueba, el cual al ser mayor que el nivel de significancia $\alpha = 0.05$ permitirá aceptar la hipótesis nula (H_0), caso contrario si el *p-valor* llegase a ser inferior se rechazaría la hipótesis nula H_0 , prueba suficiente de que los errores del modelo no se distribuyen de manera normal.

Tabla 27. Pruebas de Normalidad de los residuos (Kolmogorov-Smirnov) Liga Inglesa

```
One-sample Kolmogorov-Smirnov test  
data: residuals(modelo.lm.Inglaterra1)  
D = 0.13074, p-value = 0.1185  
alternative hypothesis: two-sided
```

Fuente: Elaboración propia con el uso de R-Studio

Tabla 28. Pruebas de Normalidad de los residuos (Kolmogorov-Smirnov) Liga Española

```
One-sample Kolmogorov-Smirnov test  
data: residuals(modelo.lm.espana1)  
D = 0.088853, p-value = 0.5238  
alternative hypothesis: two-sided
```

Fuente: Elaboración propia con el uso de R-Studio

Tabla 29. Pruebas de Normalidad de los residuos (Kolmogorov-Smirnov) Liga Francesa

One-sample Kolmogorov-Smirnov test

data: residuals(modelo.lm.Francia1)

D = 0.083087, p-value = 0.6088

alternative hypothesis: two-sided

Fuente: Elaboración propia con el uso de R-Studio

Tabla 30. Pruebas de Normalidad de los residuos (Kolmogorov-Smirnov) Liga Alemana

One-sample Kolmogorov-Smirnov test

data: residuals(modelo.lm.Alemania1)

D = 0.07757, p-value = 0.7497

alternative hypothesis: two-sided

Fuente: Elaboración propia con el uso de R-Studio

Tabla 31. Pruebas de Normalidad de los residuos (Kolmogorov-Smirnov) Liga Italiana

One-sample Kolmogorov-Smirnov test

data: residuals(modelo.lm.Italia1)

D = 0.11647, p-value = 0.2108

alternative hypothesis: two-sided

Fuente: Elaboración propia con el uso de R-Studio

Como se puede apreciar en los resultados (tabla-27 a la tabla-31), todos los *p-valor* del estadístico de la prueba son mayores a el nivel de significancia $\alpha = 0.05$, por lo que todos los residuos de los modelos se distribuyen normalmente con un nivel de confianza de 95%.

IV.6 Pruebas de Multicolinealidad de los residuos (Factor de Inflación de la Varianza)

Por último, tenemos el análisis de la prueba de multicolinealidad la cual se mide a través del Factor de Inflación de la Varianza (FIV). En esta prueba no se realiza un contraste de hipótesis. El resultado obtenido será un número ubicado entre 0 y 10, en donde, cualquier número cercano o igual a cero representa la ausencia de multicolinealidad y cualquier número cercano o igual a diez representa una presencia fuerte de multicolinealidad.

Tabla 32. Pruebas de Multicolinealidad de los residuos (F.I.V) Liga Inglesa

SalarioInglaterra	valorTemporadaInglaterra	EstaturaInglaterra
3.153511	3.149089	1.029691

Fuente: Elaboración propia con el uso de R-Studio

Tabla 33. Pruebas de Multicolinealidad de los residuos (F.I.V) Liga Española

SalarioEsp	ComprasEsp
1.9475	1.9475

Fuente: Elaboración propia con el uso de R-Studio

Tabla 34. Pruebas de Multicolinealidad de los residuos (F.I.V) Liga Francesa

SalarioFrancia	EdadFrancia
1.00493	1.00493

Fuente: Elaboración propia con el uso de R-Studio

Tabla 35. Pruebas de Multicolinealidad de los residuos (F.I.V) Liga Alemana

SalarioAlemania	valorTemporadaAlemania
5.241137	5.241137

Fuente: Elaboración propia con el uso de R-Studio

Tabla 36. Pruebas de Multicolinealidad de los residuos (F.I.V) Liga Italiana

SalarioItalia	VentasItalia
1.71466	1.71466

Fuente: Elaboración propia con el uso de R-Studio

De los resultados obtenidos que se muestran desde la tabla-31 hasta la tabla-36 se puede concluir que en todos los modelos se presencia una multicolinealidad leve, siendo el caso de Alemania (tabla 7.4) el que mayor Factor de Inflación de la Varianza presenta pero manteniéndose dentro de los parámetros aceptados de niveles de multicolinealidad.

CAPÍTULO V: CONCLUSIONES

El principal objetivo de esta investigación consistió en identificar y analizar cuáles son las variables financieras y no deportivas que tienen mayor influencia en el éxito deportivo de los equipos pertenecientes a las Cinco Grandes Ligas de fútbol europeo, caso de estudio temporadas 2015-2016 hasta 2018-2019.

La totalidad de los objetivos del estudio fueron cumplidos de la siguiente manera:

Se identificaron las variables financieras y no deportivas mediante la investigación de publicaciones y estudios previos. A través de ellos se establecieron una cantidad de posibles variables y posteriormente fueron sometidas a una serie de pruebas básicas para establecer cuáles de las variables eran las más apropiadas para el estudio.

El proceso de cuantificación del éxito se estableció como la cantidad de puntos que obtuvieron los equipos a lo largo de las temporadas utilizadas para el presente trabajo. La razón de la selección de los puntos se debió a que es el criterio máximo para determinar el orden de la clasificación de los equipos al final de cada temporada. Los demás criterios, si bien presentaban relación con el éxito de cada equipo, no iban a ser determinantes de las posiciones de los equipos.

Para realizar la medición de los impactos marginales de las variables, se elaboraron modelos de regresión lineal para cada una de las Cinco Grandes Ligas del fútbol europeo. A partir de estimadores de mínimos cuadrados ordinarios se obtuvieron los coeficientes de regresión parcial de cada una de las variables, con lo cual se puede medir el impacto marginal de cada una de ellas respecto al éxito deportivo de los equipos pertenecientes a cada liga. Una vez determinados los modelos, se procedió a seleccionar, bajo los criterios de

información de Akaike y principio de parsimonia, los modelos para cada liga que con menor cantidad de estimadores tuviesen el mejor poder de predicción o estimación.

De los modelos finales seleccionados para cada una de las ligas podemos concluir que, únicamente en lo modelo de la Liga Española comprueba lo planteado en la hipótesis del estudio, siendo las variables más significativas para la estimación del éxito deportivo medido a través de los puntos obtenidos por los equipos los Gastos salariales promedio por equipo y Gastos destinados a compras de jugadores.

Con respecto a los modelos definitivos obtenidos en el resto de las ligas; en la Liga Inglesa se evidencia que el *Valor de Temporada* va a ser la variable con mayor nivel significativo y los *Salarios*, si bien forma parte del modelo definitivo, representa la variable independiente que menor influencia presenta sobre la variable dependiente. En la Liga de Francia, la variable *Salarios* es la variable independiente que mayor influencia tiene sobre la obtención de puntos, teniendo a su vez a la variable Edad con impacto negativo por cada aumento que tenga. En la Liga Alemana, curiosamente la variable *Salarios* pertenece al modelo final sin ser una variable significativa, bajo los criterios de Akaike podría ser eliminada del modelo pero con la variable *Salarios* en el modelo se obtuvieron mejores resultados en los R^2 , por lo que fue mantenida; la variable más significativa es Valor de Temporada. Por último, en la Liga Italiana, la variable *Salarios* si bien mantiene una nivel significativo importante, el *Intercepto* tiene un nivel más significativo en el modelo definitivo. Vale la pena mencionar que en la Liga Italiana sucede un caso curioso en el que la variable *Ventas*, que podría ser vista como la contraparte de la variable *Compras*, presenta un nivel significativo considerable y mantiene una influencia positiva sobre los puntos que generan los equipos conforme aumente dicha variable.

Un punto muy importante que se debe mencionar es la presencia de autocorrelación en varios de los modelos, esto podría deberse a la omisión de alguna variable relevante en los modelos preliminares lo que podría tener como consecuencia que algunos de los β_i de los modelos sean sesgados. En caso de que se quiera replicar el estudio o extenderlo se sugiere considerar otras variables en el proceso para evitar el problema de autocorrelación.

CAPÍTULO VI: REFERENCIAS BIBLIOGRAFICAS

- Akerlof, G. (Noviembre de 1976). The Economics of Caste and of the Rat Race and Other Woeful Tales. *The Quarterly Journal of Economics*, 90, 599-617. Obtenido de <http://www.jstor.org/stable/1885324>
- Aparicio, J., Asuncion, M., & Morales, J. (s.f.). Obtenido de <https://umh3067.edu.umh.es/wp-content/uploads/sites/240/2013/02/Modelos-Lineales-Aplicados-en-R.pdf>
- Arcidiacono, P., & Kimbrough,, K. (s.f.). Obtenido de <https://sites.duke.edu/djepapers/files/2017/06/grantshorin-dje.pdf>
- Arias, F. (2012). *El Proyecto de Investigación. Introducción a la metodología científica* (6ta ed.). Caracas: Episteme.
- Baltagi, B. (2005). *Econometric Analysis of Panel Data (3era Edición)*. Londres: John Wiley & Sons Ltd.
- Cameron, C. &. (2009). *Microeconometrics Using Stata (2da edición)*. Texas: Stata Press.
- El-Hodiri, M., & Quirk, J. (Noviembre de 1971). An Economic Model of a Professional Sports League. *The Journal of Political Economy*, 79(6), 1302-1319.
- Gujarati, D., & Porter, D. (2008). *Basic Econometrics*. McGraw Hill.
- Heckhausen, J., & Schulz, R. (1996). *A life span model of successful aging*. *American Psychologist*.
- Hill, T., Davis, A., & Micah Ross, J. (Junio de 2020). *Limitations of Fixed-Effects Models for Panel Data. Sociological Perspectives*. Obtenido de ResearchGate: https://www.researchgate.net/publication/334000163_Limitations_of_Fixed-Effects_Models_for_Panel_Data
- Hsiao, C., T.W. Appelbe, & Dineen, C. (1993). A General Framework for Panel Data Analysis—With an Application to Canadian Customer Dialed Long Distance Service. *Journal of Econometrics*, 59, 63-86.
- Intelligence, S. (2015-2019). <https://www.globalsportssalaries.com/>. Obtenido de <https://www.globalsportssalaries.com/>

- Kuypers, T. (1997). *The beautiful game? an econometric study of audiences, gambling and efficiency in English football*. Londres: University of London.
- López Casuso, R. (2009). *Cálculo de Probabilidades e Inferencia Estadística con tópicos de Econometría*. Caracas: Publicaciones UCAB.
- Martinez, D., Albin, J., & Cabaleiro, J. (Septiembre de 2009). Obtenido de https://www.researchgate.net/profile/Tomas_Pena/publication/236279245_El_criterio_de_informacion_de_Akaike_en_la_obtencion_de_modelos_estadisticos_de_Rendimiento/links/58904fa3aca272bc14be3600/El-criterio-de-informacion-de-Akaike-en-la-obtencion-de-modelo
- Michie, J., & Oughton, C. (2004). *Competitive Balance in Football: Trends and Effects*. Londres: Sports Nexus.
- Montero Granado, R. (2016). Obtenido de https://www.ugr.es/~montero/matematicas/regresion_lineal.pdf
- Montgomery, E., & Vining, D. (2006). *Introducción Al Análisis de Regresión Lineal*. 3ed ed. México: Cecsa. Obtenido de https://fhernanb.github.io/libro_regresion/r_lm.html
- Neale, W. C. (Febrero de 1964). The Peculiar Economics of Professional Sports: A Contribution to the Theory of the Firm in Sporting Competition and in Market Competition. *The Quarterly Journal of Economics*, 78(1), 1-14.
- Novalés, A. (20 de Septiembre de 2010). Obtenido de <https://www.ucm.es/data/cont/docs/518-2013-11-13-Analisis%20de%20Regresion.pdf>
- Rohde, M., & Breuer, C. (Junio de 2016). Europe's Elite Football: Financial Growth Sporting Success, Transfer Investment, and Private Majority Investors. *International Journal of Financial Studies*.
- Rosen, S., & Sanderson, A. (s.f.). *LABOR MARKETS IN PROFESSIONAL SPORTS*. Obtenido de https://www.nber.org/system/files/working_papers/w7573/w7573.pdf
- Scmicht, S., & Torgler, S. (2005). *Crema Research*. Obtenido de <http://www.crema-research.ch/papers/2005-25.pdf>
- Stephen M, S. (1986). *The History of Statistics: The Measurement of Uncertainty Before 1900*. Harvard University Press.

Szymanski, S., & Smith, R. (1997). The English Football Industry: profit, performance and industrial structure. *International Review of Applied Economics*, 11(1), 135-153.

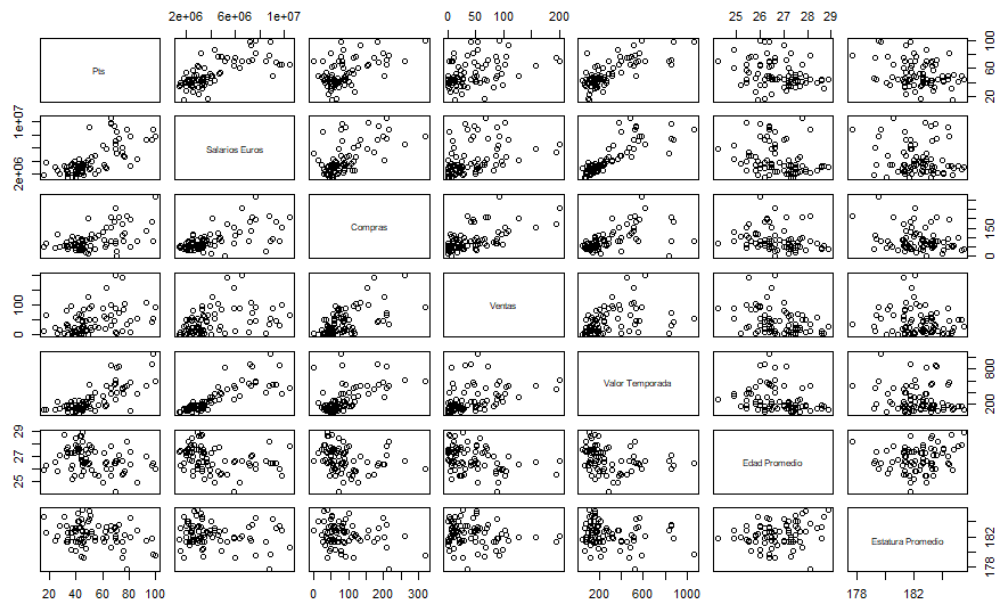
Transfermarkt.com. (2020). Obtenido de <https://www.transfermarkt.com/>

WorldFootball.net. (2020). Obtenido de <https://www.worldfootball.net>

CAPÍTULO VII: ANEXOS

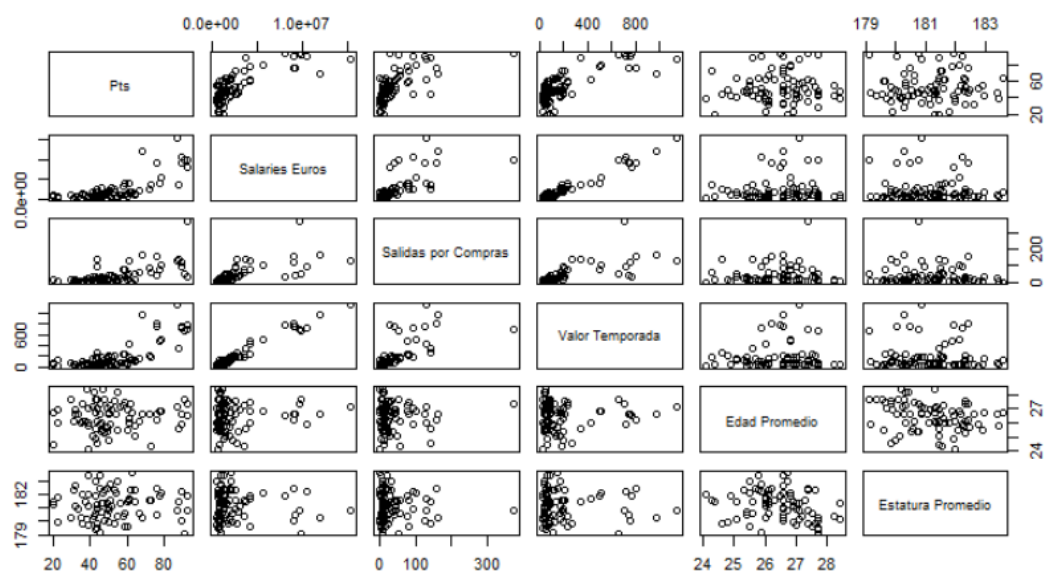
Matrices de Correlación

Grafico 1. Matriz de Correlación Liga Inglesa



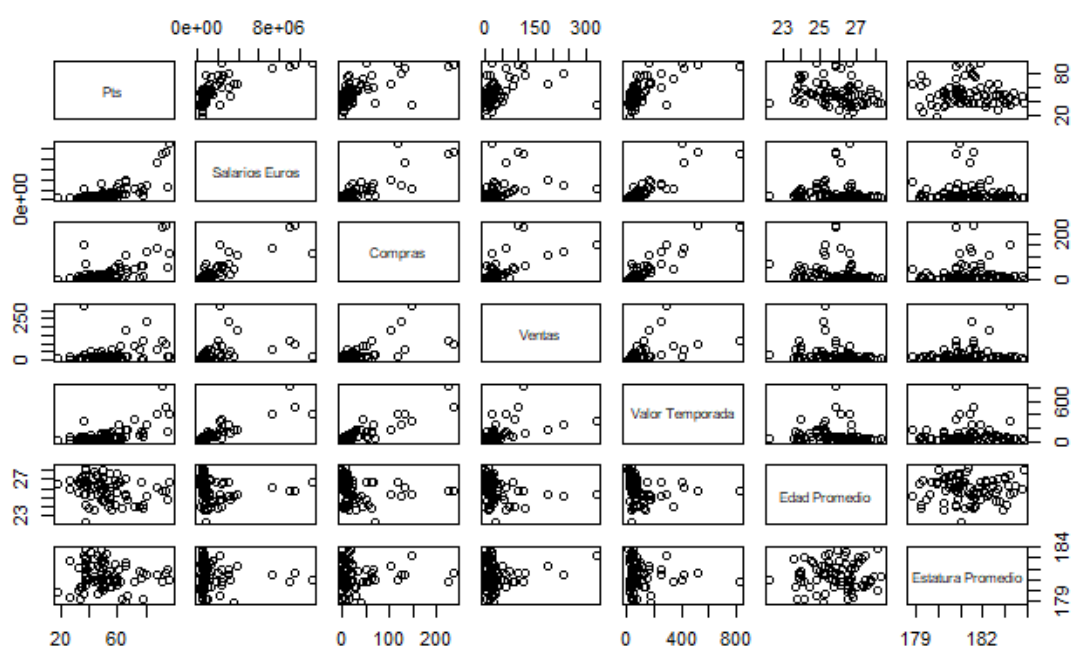
Fuente: Elaboración propia con el uso de R-Studio

Grafico 2. Matriz de Correlación Liga Española



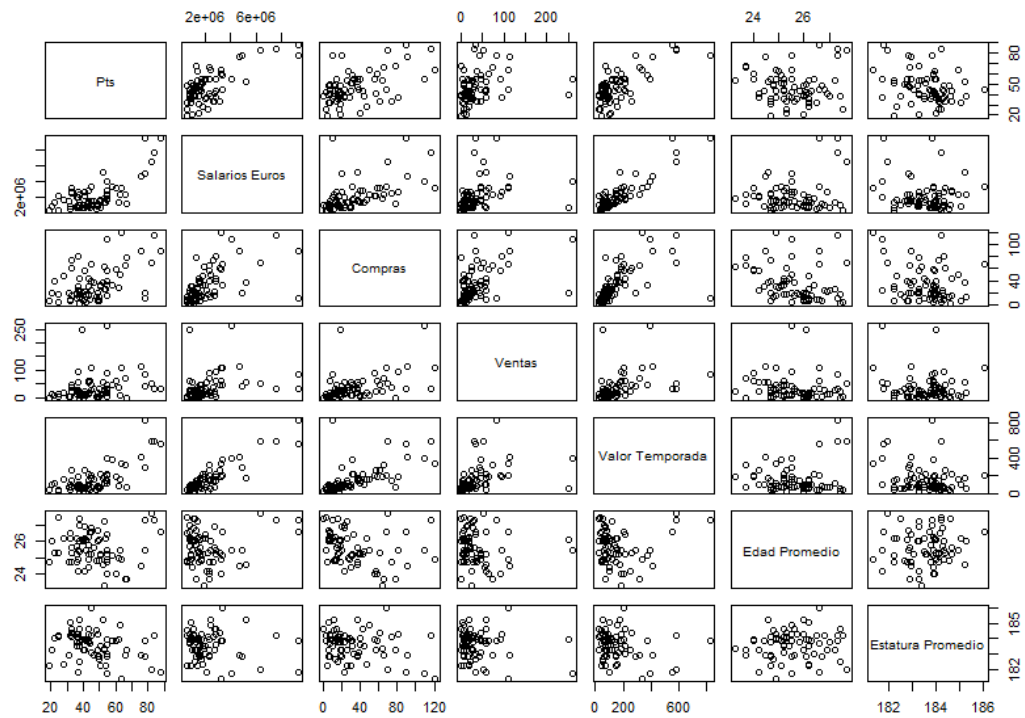
Fuente: Elaboración propia con el uso de R-Studio

Grafico 3. Matriz de Correlación Liga Francesa



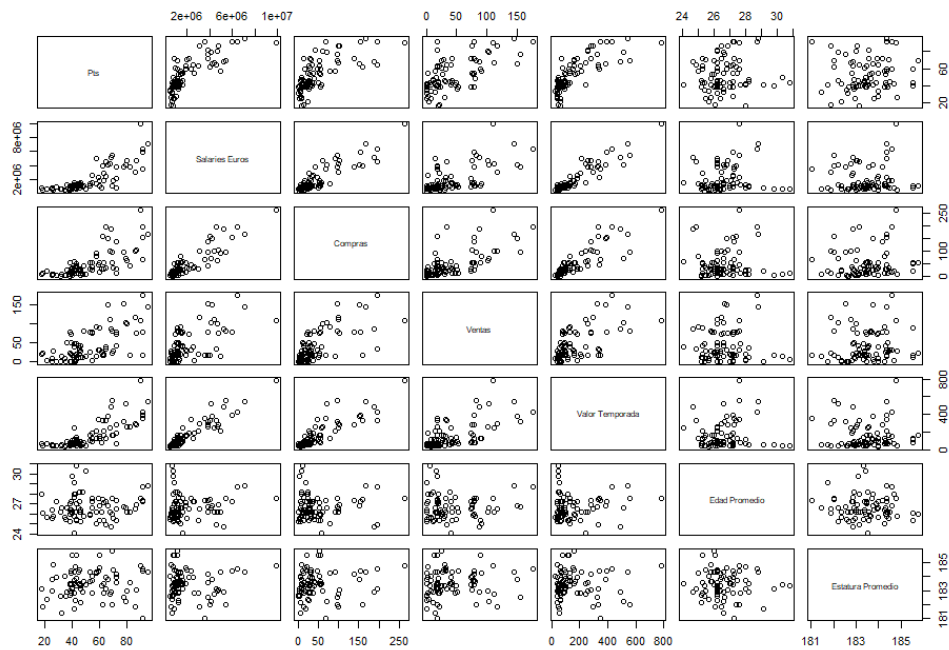
Fuente: Elaboración propia con el uso de R-Studio

Grafico 4. Matriz de Correlación Liga Alemana



Fuente: Elaboración propia con el uso de R-Studio

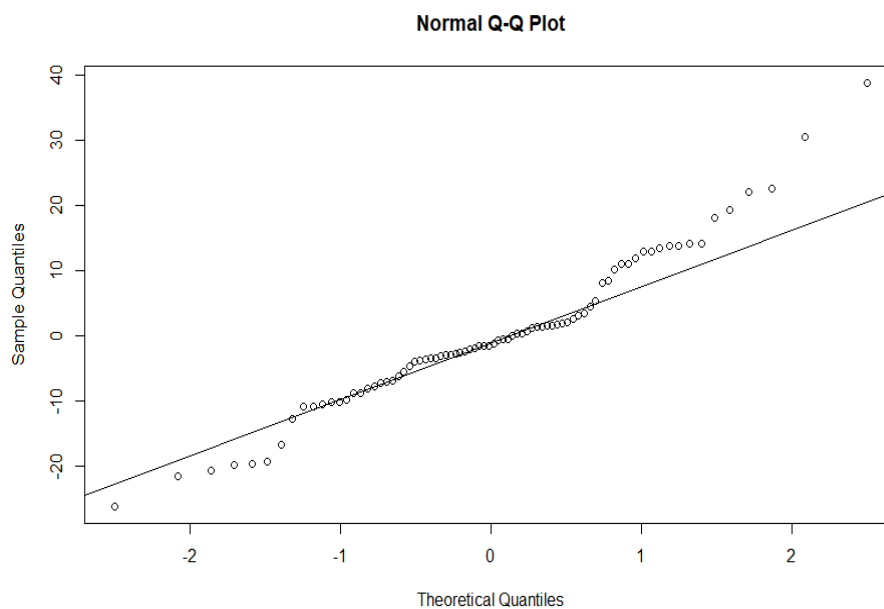
Grafico 5. Matriz de Correlación Liga Italiana



Fuente: Elaboración propia con el uso de R-Studio

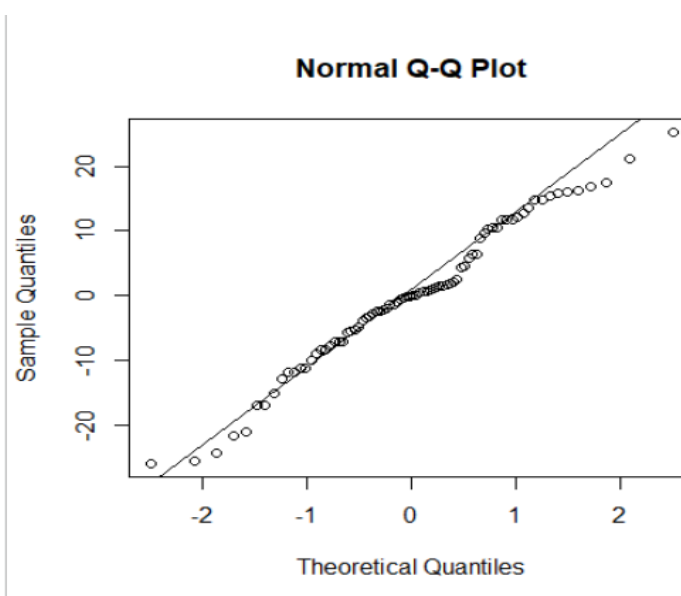
Distribución Normal. Grafico Q-Q Plot

Grafico 6. Distribución Normal. Grafico Q-Q Plot Liga Inglesa



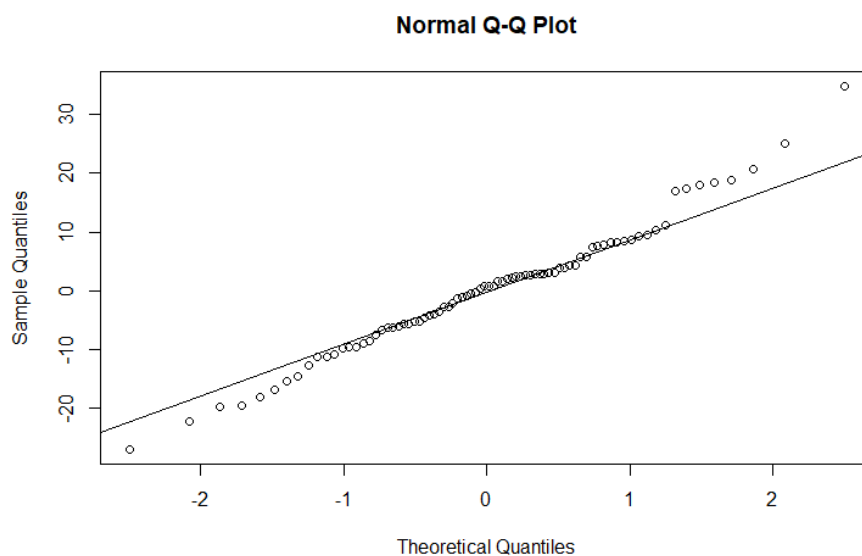
Fuente: Elaboración propia con el uso de R-Studio

Grafico 7. Distribución Normal. Grafico Q-Q Plot Liga Española



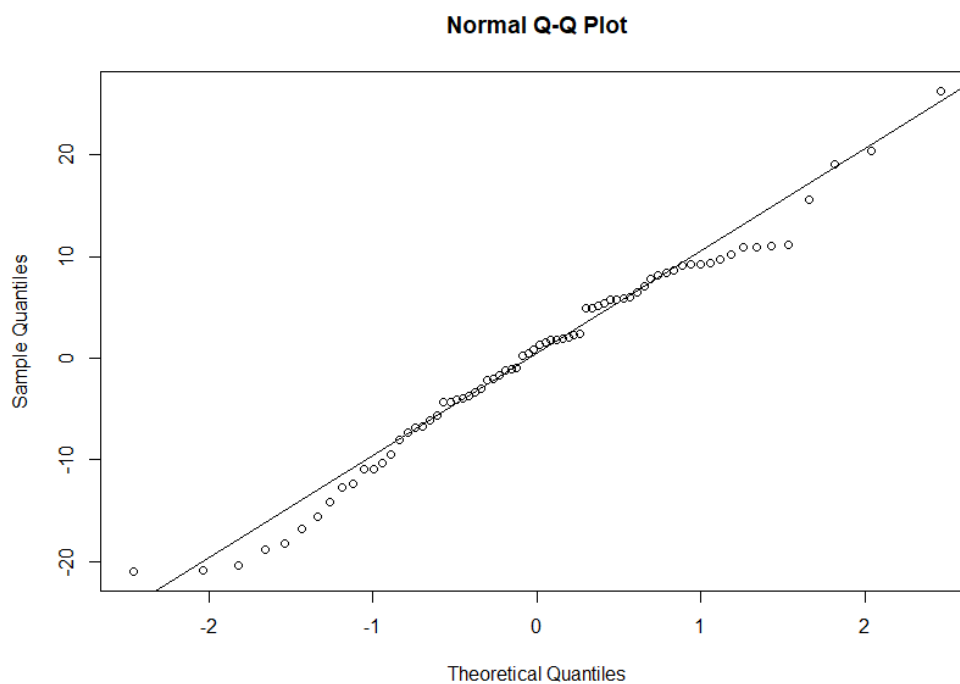
Fuente: Elaboración propia con el uso de R-Studio

Grafico 8. Distribución Normal. Grafico Q-Q Plot Liga Francesa



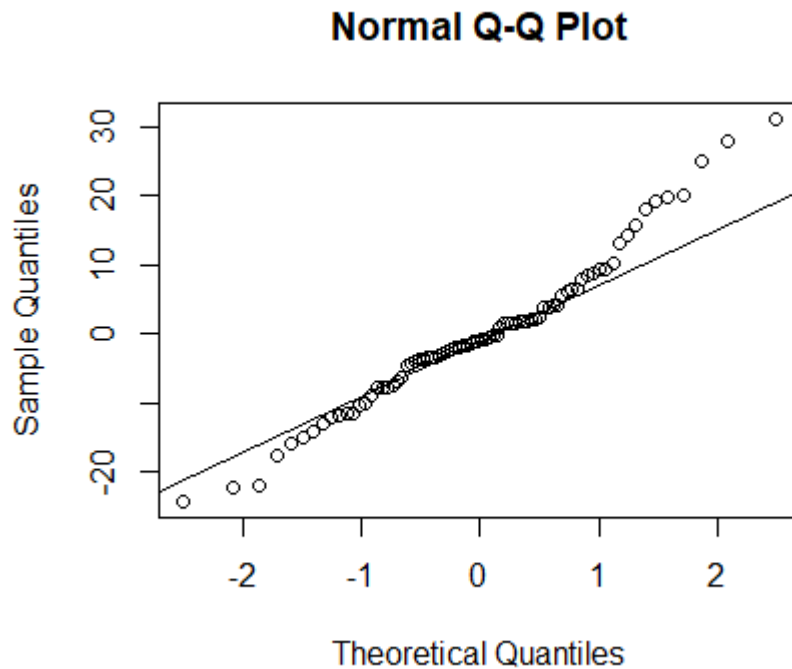
Fuente: Elaboración propia con el uso de R-Studio

Grafico 9. Distribución Normal. Grafico Q-Q Plot Liga Alemana



Fuente: Elaboración propia con el uso de R-Studio

Grafico 10. Distribución Normal. Grafico Q-Q Plot Liga Italiana



Fuente: Elaboración propia con el uso de R-Studio

Intervalo de Confianza de los Coeficientes

Tabla 37. Intervalo de Confianza de los coeficientes Liga Inglesa

	2.5 %	97.5 %
(Intercept)	3.899804e+01	6.845169e+02
SalarioInglaterra	-1.229083e-07	3.949251e-06
valorTemporadaInglaterra	2.415328e-02	6.712312e-02
EstaturaInglaterra	-3.577718e+00	-4.583963e-02

Fuente: Elaboración propia con el uso de R-Studio

Tabla 38. Intervalo de Confianza de los coeficientes Liga Española

	2.5 %	97.5 %
(Intercept)	3.749310e+01	4.408639e+01
SalarioEsp	2.426674e-06	4.739913e-06
ComprasEsp	-8.384481e-04	1.243317e-01

Fuente: Elaboración propia con el uso de R-Studio

Tabla 39. Intervalo de Confianza de los coeficientes Liga Francesa

	2.5 %	97.5 %
(Intercept)	5.832187e+01	1.614746e+02
SalarioFrancia	4.534562e-06	7.045529e-06
EdadFrancia	-4.556288e+00	-5.872469e-01

Fuente: Elaboración propia con el uso de R-Studio

Tabla 40. Intervalo de Confianza de los coeficientes Liga Alemana

	2.5 %	97.5 %
(Intercept)	3.102858e+01	3.843322e+01
SalarioAlemania	-9.605550e-07	5.063975e-06
valorTemporadaAlemania	1.255640e-02	8.625741e-02

Fuente: Elaboración propia con el uso de R-Studio

Tabla 41. Intervalo de Confianza de los coeficientes Liga Italiana

	2.5 %	97.5 %
(Intercept)	3.028612e+01	3.794618e+01
SalarioItalia	4.992754e-06	8.496937e-06
VentasItalia	2.227085e-02	1.774327e-01

Fuente: Elaboración propia con el uso de R-Studio