

**UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
FACULDADE DE CIÊNCIAS ECONÔMICAS
DEPARTAMENTO DE ECONOMIA E RELAÇÕES INTERNACIONAIS
CURSO DE CIÊNCIAS ECONÔMICAS**

HENRIQUE D'ÁVILA NUNES

**ESTIMATIVA DO VALOR DE MERCADO DE UM JOGADOR DE FUTEBOL
PROFISSIONAL DA LIGA INGLESA**

Porto Alegre

2020

HENRIQUE D'ÁVILA NUNES

**ESTIMATIVA DO VALOR DE MERCADO DE UM JOGADOR DE FUTEBOL
PROFISSIONAL DA LIGA INGLESA**

Trabalho de conclusão submetido ao Curso de Graduação em Economia da Faculdade de Ciências Econômicas da UFRGS, como requisito parcial para obtenção do título Bacharel em Ciências Econômicas.

Orientador: Prof. Dr. Sérgio Marley Modesto Monteiro.

Porto Alegre

2020

HENRIQUE D'ÁVILA NUNES

**ESTIMATIVA DO VALOR DE MERCADO DE UM JOGADOR DE FUTEBOL
PROFISSIONAL DA LIGA INGLESA**

Trabalho de conclusão submetido ao Curso de Graduação em Economia da Faculdade de Ciências Econômicas da UFRGS, como requisito parcial para obtenção do título Bacharel em Ciências Econômicas.

Aprovado em: Porto Alegre, ____ de _____ de 2020.

BANCA EXAMINADORA:

Prof. Dr. Sérgio Marley Modesto Monteiro - Orientador
UFRGS

Prof. Dr. Andres Ernesto Ferrari Haines
UFRGS

Prof. Dr. Sabino Porto Júnior
UFRGS

AGRADECIMENTOS

Em primeiro lugar, direciono meus agradecimentos a toda a minha família, em especial minha mãe, Inês Suzete Silveira D'Ávila, e meu pai, Valmir da Silva Nunes, que nunca mediram esforços para me proporcionar, desde a pré-escola, sempre as melhores possibilidades de estudo e aprendizado, muitas vezes abdicando de vontades pessoais para conseguir colocar eu e meu irmão - Guilherme D'Ávila Nunes, a quem também cito meus agradecimentos -, no lugar onde hoje estamos. Se hoje sou o que sou, devo isso ao apoio, incentivo, dedicação, esforço, suporte e sacrifício deles.

Agradeço também à minha namorada e aos meus amigos, da faculdade e de fora dela, que fizeram com que essa trajetória tenha sido seguida com muito mais leveza e prazer. Muitas vezes quando a vontade de ir à aula não era das maiores, era por causa da companhia e incentivo deles que eu acabava comparecendo.

Ao meu professor orientador, Prof. Dr. Sérgio Monteiro, que sempre se mostrou disponível e, além disso, com um método de ensino prático, respeitoso, divertido e efetivo a respeito do comportamento da economia na vida real, principalmente em suas aulas de Micro II, às terças e quintas, 20:30, do rigoroso inverno porto-alegrense, durante os horários de jogos do Inter na sonolenta série B, também proporcionava uma vontade natural de comparecimento à faculdade e de estudar.

Aos membros da banca, Prof. Dr. Andres Ernesto Ferrari Haines e Prof. Dr. Sabino Porto Júnior, por prontamente aceitarem o convite de auxiliar nesta etapa.

À UFRGS e à Faculdade de Ciências Econômicas pelo ensino de excelência que me foi ofertado e pelos excelentes profissionais que lá trabalham, em diferentes setores, e proporcionam o ambiente incrível que vivenciei em todos estes anos.

A todos os professores que passaram por minha vida, desde a pré-escola, por todos os ensinamentos compartilhados.

Obrigado!

“Futebol se joga no estádio? Futebol se joga na praia, futebol se joga na rua, futebol se joga na alma.” (Carlos Drummond de Andrade.)

RESUMO

Este trabalho tem como objetivo analisar a composição do valor de mercado dos atletas profissionais da Liga Inglesa de Futebol (Barclays Premier League), por meio do método de Regressão Linear Múltipla, que permite identificar quais são as variáveis que mais impactam a variável dependente em análise. Para isso, foram utilizados dados de 387 atletas, dos 20 clubes da Liga Inglesa, referentes à temporada 2018/2019, extraídos dos *websites Transfermarkt e Whoscored*. Concluiu-se que grande parte dos resultados encontrados eram os esperados e assemelham-se aos que foram abordados na revisão da literatura, como é o caso da relação negativa entre a variável *Idade* e o valor de mercado. O destaque fica para a relação positiva entre a variável *Instagram_followers* e a variável dependente, indicando que, dadas as demais variáveis constantes, quanto maior for o ganho com imagem do atleta, maior é o seu valor de mercado. Também foi possível observar que a hipótese de haver discriminação racial ou étnica na precificação dos atletas não foi confirmada no estudo.

Palavras-chave: Economia do Esporte. Estatística do Esporte. Economia do Futebol. Valor de Mercado. Liga Inglesa de Futebol.

Classificação JEL: Z20 – Sports Economics: General

ABSTRACT

This research aims to analyze the composition of the market value of professional athletes from the English Football League (Barclays Premier League), using the Multiple Linear Regression method, which allows identifying the variables that most impact the dependent variable under analysis. For this, data from 387 athletes from the 20 clubs of the English League for the 2018/2019 season were used, extracted from the *Transfermarkt* and *Whoscored* websites. It was concluded that most of the results found were those expected and are similar to those that were addressed in the literature review, as is the case of the negative relationship between the variable *age* and the market value. The highlight is the positive relationship between the *Instagram_followers* variable and the dependent variable, indicating that, given the other constant variables, the greater the gain with the athlete's image, the greater the market value. It was also possible to observe that the hypothesis of racial or ethnic discrimination in athletes' pricing was not confirmed in the study.

Keywords: Sport Economics. Sport Statistics. Football Economics. Market value. English Football League.

LISTA DE GRÁFICOS

Gráfico 1: Histograma dos resíduos.....	44
Gráfico 2: Gráfico Q-Q.....	44
Gráfico 3: Resíduos versus Valores Ajustados.	46
Gráfico 4: Análise de DFFITS, n = 387	47
Gráfico 5: Análise de DFFITS, n = 380	48

LISTA DE TABELAS

Tabela 1: Variáveis do modelo de Wicker, et. al (2013), segundo Majewski (2016, pg. 96).	21
Tabela 2: Variáveis do modelo e origem dos dados.....	26
Tabela 3: Diferença entre o valor estimado pelo portal Transfermarkt e o valor efetivamente pago pelos atletas transferidos.....	28
Tabela 4: Os vinte atletas mais valiosos do estudo.	29
Tabela 5: Os vinte atletas que mais se valorizaram.	30
Tabela 6: Os vinte melhores atletas.	31
Tabela 7: Distribuição dos atletas por continente, considerando a nacionalidade principal.....	32
Tabela 8: Distribuição dos atletas por continente de origem.	33
Tabela 9: Estatísticas descritivas da amostra.	34
Tabela 10: Número de observações das variáveis <i>Dummy</i>.	34
Tabela 11: Modelos e resultados.	37
Tabela 12: Teste RESET de Ramsey.	42
Tabela 13: Fator de Inflação da Variância	43
Tabela 14: Teste qui-quadrado de Pearson.....	45
Tabela 15: Teste de Goldfeld-Quandt.....	46
Tabela 16: Observações não usuais.....	47
Tabela 17: Comparação entre os modelos.....	49
Tabela 18: Modelos e resultados, n = 380.....	50
Tabela 19: Interpretação das variáveis independentes do Modelo.....	51

SUMÁRIO

1 INTRODUÇÃO	9
2 REVISÃO DA LITERATURA	11
3 MODELO E BANCO DE DADOS	22
3.1 APRESENTAÇÃO DO MÉTODO	22
3.1.1 A REGRESSÃO LINEAR E O MÉTODO STEPWISE	22
3.1.2 O CRITÉRIO AIC (AKAIKE INFORMATION CRITERION)	24
3.2 APRESENTAÇÃO DO BANCO DE DADOS	25
3.3 O SOFTWARE R-STUDIO	35
4 APRESENTAÇÃO DOS RESULTADOS.....	36
4.1 APRESENTAÇÃO DOS MODELOS	36
4.1.1 A APLICAÇÃO DO MÉTODO STEPWISE POR AIC E OS RESULTADOS.....	39
4.1.2 O MODELO COM A RETIRADA DAS VARIÁVEIS NÃO SIGNIFICATIVAS	40
4.1.3 ANÁLISE DO R ² AJUSTADO E DO TESTE F.....	41
4.1.4 TESTE RESET DE RAMSEY PARA MÁ ESPECIFICAÇÃO DA FORMA FUNCIONAL	42
4.1.5 TESTE PARA IDENTIFICAÇÃO DE <i>MULTICOLINEARIDADE</i>	42
4.1.6 TESTE DE NORMALIDADE DOS RESÍDUOS	43
4.1.7 TESTE PARA HOMOCEDASTICIDADE.....	45
4.1.8 OBSERVAÇÕES NÃO USUAIS.....	46
5 CONSIDERAÇÕES FINAIS.....	54
REFERÊNCIAS	57
APÊNDICE A – ROTINA UTILIZADA NO SOFTWARE R STUDIO.....	59

1 INTRODUÇÃO

A Economia do Esporte tornou-se uma ferramenta de grande importância para pesquisadores do mundo todo que buscam entender como os modelos da economia se comportam ao serem aplicados no mercado de trabalho das principais ligas desportivas do mundo. Rottenberg (1956) foi o pioneiro ao analisar o mercado de trabalho da liga profissional americana de Baseball. Seu trabalho abriu oportunidades de exploração e aplicação dos conceitos da economia ao esporte.

O Brasil é popularmente chamado de “país do futebol”. Embora a expressão seja quase um clichê, esse esporte está enraizado na cultura da sociedade brasileira desde os primórdios do século passado. Além disso, o comércio de jogadores vem se tornando uma das mais impactantes rubricas na economia dos clubes, principalmente no Brasil. Não é preciso ir muito longe para visualizar isso. Grêmio e Internacional, os dois clubes de maior receita do Rio Grande do Sul, tiveram no ano de 2019, respectivamente, 26,4% e 30,7% do total da receita bruta oriunda da negociação de atletas, em ambos os casos ultrapassando a casa dos R\$ 100 milhões, de acordo com dados extraídos das demonstrações financeiras dos clubes. Entretanto, a reduzida literatura nacional, a falta de dados confiáveis disponíveis gratuitamente e de transparência por parte dos clubes prejudicam análises mais completas sobre os determinantes de valor de mercado de jogadores do futebol brasileiro.

A Barclays Premier League (liga profissional inglesa de futebol) é uma das principais ligas de futebol do mundo, conta com uma ampla variedade de atletas com características variadas (nacionalidade, por exemplo) e, principalmente, uma maior disponibilidade de dados confiáveis. Por essas razões, será o objeto de análise do presente trabalho, utilizando-se para tanto um banco de dados contendo informações de atletas profissionais oriundos dos vinte clubes que participaram da temporada de 2018/2019, extraídos do portal alemão “*Transfermarkt*”, e do website “*Whoscored.com*”.

O presente trabalho, com base na liga inglesa, visa identificar quais são as variáveis que mais impactam o valor de mercado de um atleta profissional de futebol e quais são as relações existentes entre estas. Com o modelo obtido, será possível inferir o valor de mercado de um atleta com base nas características que se tem conhecimento deste. Além disso, a verificação do impacto ou não de questões raciais e características físicas nos valores das transferências e a necessidade de entender melhor o funcionamento da Barclays Premier League e da relação entre seus clubes e atletas são também objetivos do estudo.

A dificuldade na busca por referências no Brasil para o tema demonstra a importância de abordá-lo. Atletas profissionais de futebol movimentam cada vez mais dinheiro em transações que envolvem cifras astronômicas e crescem exponencialmente com o tempo. A fama alcançada pelos atletas, a valorização das marcas, a utilização de sua imagem para fins comerciais e o envolvimento dos torcedores impactam diretamente a demanda pelos produtos que os cercam e impactam diretamente a economia, através do efeito “Apelo de Bilheteria” descrito por Rosen (1981) e aprofundado por Lucifora e Simmons (2003).

Para alcançar os objetivos citados acima, o trabalho foi dividido em três capítulos. O primeiro busca retratar a literatura existente, apresentando conceitos, ideias e metodologias comumente utilizadas em estudos semelhantes e que serão adaptados ao presente trabalho. No segundo capítulo será apresentado o banco de dados, a metodologia e o *software* utilizado na análise de regressão. Por fim, no terceiro capítulo, serão apresentados e analisados os resultados obtidos.

2 REVISÃO DA LITERATURA

A economia do esporte teve seu marco inicial com o trabalho de Rottenberg (1956). Em *The Baseball Players' Labor Market*, o autor desenvolve sua teoria a respeito da, ainda embrionária, construção de um mercado de trabalho na Liga Nacional de Baseball norte-americana. Ao tratar de temas como o ajustamento e a padronização de contratos, a função determinante da presença de público, as condições de trabalho abusivas dos atletas profissionais, a comparação da estrutura de mercado da liga profissional com as ligas amadoras, as variáveis determinantes do salário dos jogadores e o caráter monopsonista das relações entre atletas e clubes, Rottenberg cria um ambiente imenso de oportunidades de exploração para posteriores autores desenvolverem suas ideias, aplicando conceitos das Ciências Econômicas ao esporte, caracterizando o que hoje chamamos de Economia do Esporte.

Para que pudesse aplicar os conceitos da teoria econômica, principalmente da microeconomia, Rottenberg descreve o ambiente de negócios do Baseball norte-americano como sendo uma indústria. O principal contraponto de seu trabalho consiste em analisar a liberdade entre a compra e venda de direitos econômicos de atletas, por parte dos clubes. Composto por uma grande estrutura de ligas amadoras (33 ao todo), contratos complexos que regem as transferências entre times e jogadores, e controlado por regras advindas de sete longos documentos, o mercado de trabalho do Baseball estadunidense da época serviu como um modelo inicial para os atuais formatos de mercado de transferências, de diferentes modalidades de esportes, no mundo todo.

Outra importante contribuição para o desenvolvimento das mais recentes análises da economia no esporte é a de Rosen (1981). No artigo, intitulado de *The Economics of Superstars*, Rosen descreve que há algo que vai além da simples relação entre os ganhos com a sua profissão (seja ela de escritor, comediante, esportistas, artistas, etc) e o talento para desempenhá-la. Rosen chama essa qualidade de “apelo de bilheteria”, que se traduz como sendo a habilidade para atrair audiência e gerar um grande volume de transações, sendo mais comumente observável em esportes profissionais e na área do entretenimento. É nessas áreas, em que o talento é supervalorizado pelos consumidores, que pequenas variações marginais no talento geram grandes variações nos salários, já que a demanda dos consumidores é uma função do quão talentosos são os produtos – leia-se aqui atletas, comediantes, etc – não podendo ser suprida por incremento em novas tecnologias, por exemplo. Nessas profissões, o valor de mercado de uma estrela é altamente influenciado pela disposição a pagar dos consumidores, sendo esta

determinada pelo preço extra pago para vê-la atuar e pelo número de consumidores dispostos a pagarem tal valor. Dessa forma, o aumento da demanda não é rígido apenas pelo mecanismo de preços, mas sim, pelo “apelo de bilheteria” descrito anteriormente.

Tal característica é ainda observável no atual contexto de mercado de transferências de atletas, no qual cada vez mais os clubes procuram profissionais cujas transações envolvem cifras milionárias, como forma de auferir ganhos esportivos, mas principalmente como forma de ampliar a visibilidade sobre a própria marca. Rosen afirma ainda ser difícil definir o que leva a gerar tal qualidade, classificando como uma combinação entre “talento e carisma em proporções incertas”. Bali, R. (2012) acrescenta, ainda, que apenas um pequeno grupo de clubes no mundo é capaz de oferecer altos salários e competir para contratar jogadores que detêm o “Fenômeno da Superestrela”, tornando este mercado caracteristicamente oligopolista, com grande apelo do público e com uma alta procura dos clubes pelos serviços destes atletas. Clubes menores, entretanto, procuram outros atletas que se encaixem em sua curva de restrição orçamentária para obter seus ganhos esportivos, gerando uma assimetria marcante entre as curvas de indiferença e funções de utilidade dos atletas deste grupo.

Ainda nesse sentido, Lucifora e Simmons (2003) aprofundam, particularmente ao futebol, o estudo do efeito das superestrelas no salário, usando como base os dados da Liga Italiana de Futebol de 1995 e 1996. Um ponto importante levantado por eles consiste no efeito da reputação na determinação dos ganhos obtidos pelos atletas, usando como exemplo o caso de Baggio na liga italiana de 1995. Sua maior reputação – aqui tratada como uma variável independente – fez com que seu salário, quando comparado a outros desportistas com desempenho semelhante, fosse muito maior, tornando os outros jogadores substitutos imperfeitos. Ainda, segundo os autores, há outros fatores que particularizam os efeitos das superestrelas no futebol quando comparados aos das demais profissões. Primeiramente, há uma lealdade dos torcedores com a sua equipe, criando um efeito irracional, que faz com que este torcedor seja influenciado na hora de decidir quem é o melhor jogador (difícilmente falará que o jogador do time rival é mais competente). Segundo, há uma dificuldade em medir qual a contribuição individual de um atleta no desempenho geral da equipe, o que problematiza uma avaliação eficiente por parte dos consumidores. Além disso, os torcedores podem ter gostos diferentes quanto ao estilo de jogo dos atletas e preferir, por exemplo, um atleta que seja de uma posição defensiva a um atleta de uma posição mais ofensiva. Por último, a preferência dos consumidores de assistirem as partidas ao vivo, sendo que cada uma é diferente da outra, o que pode levar a diferentes padrões de atuação e prejudicar a avaliação do público. Todas estas

particularidades do esporte podem fazer com que os efeitos gerados por uma superestrela sejam mais suaves do que quando aplicados em outras profissões, sendo necessário testá-los empiricamente.

Ainda que sua variedade de dados não tenha sido tão vasta como as que hoje estão disponíveis, através do método de Regressão por Mínimos Quadrados Ordinários, Lucifora e Simmons (2003) concluem que as medidas de desempenho de assistências e gols são as que mais impactam o “efeito superestrela”, sendo mais influentes nos jogadores de posições ofensivas, quais sejam, os atacantes. Sendo assim, através da análise destas e de outras variáveis, o que restou evidente na pesquisa é que jogadores que marcaram mais gols e deram mais assistências para seus companheiros, não necessariamente foram os que mais contribuíram para um melhor desempenho da equipe, mas invariavelmente foram os que mais atraíram reconhecimento dos torcedores e, por consequência, geraram mais receitas. Por conta disso, o efeito superestrela estava presente em jogadores como Roberto Baggio, Gabriel Batistuta e Giuseppe Signori, com média de gols superior a 0,5 por partida (havia apenas outros 6 neste patamar), que detinham os maiores salários, dado que contribuíam, em maior parte, para atrair audiência, seja ela nos estádios ou por meio de transmissões da mídia.

No estudo “*Valuations of Soccer Players from Statistical Performance Data*”, Radu S. Tunaru e Howard P. Viney (2010) utilizam o banco de dados, de desempenho de atletas e de times de futebol, OPTA Sports, para estimar o valor financeiro de jogadores profissionais de futebol para seus novos e potenciais clubes. O modelo reconhece que o valor de um atleta é resultado de vários fatores, tais como lesões, performance dentro de campo, comportamento disciplinar e fatores extracampo, como a imagem do atleta. Para isso, utiliza como exemplo o desempenho do jogador Thierry Henry e do seu clube (Arsenal), medido em OPTA points, na temporada de 2003-04. O estudo supõe ainda que há três objetivos primordiais a serem perseguidos pelos clubes, quais sejam, ganhar competições, operar com lucro ou manter seu *status* na hierarquia. Como são poucos os clubes que têm a oportunidade de ganhar títulos e operar com lucro, a maioria busca manter seu *status* na hierarquia do futebol, mesmo que isso signifique deixar de lado os outros dois objetivos. Dessa forma, Radu S. Tunaru e Howard P. Viney (2010) nos mostram que há um maior interesse pelo crescimento das receitas - e não da rentabilidade - e, por conta disto, um maior dispêndio de energia, pelos clubes, na busca por recursos e atletas (fora de campo) do que na disputa por troféus no âmbito, o que dimensiona o tamanho da importância dos negócios no mundo do futebol.

Carmichael e Thomas (1993) examinam o mercado de transferências da Liga Inglesa de futebol da temporada de 1990-1991, sob o ponto de vista da teoria da barganha entre agentes, com o mesmo objetivo de analisar os determinantes dos valores envolvidos nas trocas. Aqui, mais uma vez, os termos dos contratos são abordados como fatores limitantes para as transações. O grande marco para a liberdade no mercado de transferências inglês surgiu com o fim do sistema de “reter e transferir” nos anos 1963-64, vigente desde a invenção da liga de futebol no país, em 1888. Tal sistema consistia em “amarrar” os jogadores profissionais aos seus clubes e não permitir suas transações sem a permissão destes. Em 1977-78, leis foram modificadas em favor dos atletas, permitindo que pudessem negociar transferências a partir do último ano de contrato. Já as transações entre jogadores com contrato eram feitas a partir dos clubes, sendo proibido o contato direto entre atletas e clubes interessados, bem como agentes envolvidos na negociação.

Muitos são os motivos que levam clubes e atletas a efetivarem negociações. Os estímulos podem partir tanto dos clubes compradores, como dos clubes vendedores e, em alguns casos, dos próprios atletas. Clubes compradores buscam acrescentar ganhos esportivos e, conseqüentemente, ganhos financeiros com a aquisição de um novo atleta, mesmo que haja riscos inerentes nesse processo. Diversas são as variáveis que determinam o valor da transação e o valor pago deve estar relacionado às expectativas de ganhos futuros do atleta envolvido durante a vigência de seu contrato. Do ponto de vista do clube vendedor, entretanto, a venda pode acontecer de forma involuntária, como uma imposição, ou pode ocorrer voluntariamente, por diversos motivos, inclusive financeiros, para compra de outro atleta ou por pressão imposta pelo jogador em questão em decorrência do final de seu contrato.

O problema de barganha de Nash (1950, 1953), que consiste na ideia de que há fundamento para uma negociação quando pelo menos dois agentes têm a possibilidade de aumentar o seu estado de satisfação, caso cheguem a um acordo entre eles, é aplicado por Carmichael e Thomas (1993), em um teste de regressão contendo características específicas para os clubes compradores e vendedores, bem como atletas envolvidos em uma negociação. Os resultados sugerem que o problema da barganha está presente de forma significativa no mercado de transferências entre atletas e que pode não ser realista tratar os clubes compradores e os clubes vendedores simetricamente em análises desse sentido. O poder de barganha dos clubes vendedores foi identificado basicamente pelas características esportivas dos atletas e a capacidade de atrair público, o que fora chamado de “apelo de bilheteria” anteriormente por Rosen (1981). Já do lado do clube comprador, a relação positiva aparece quando analisada a

sua rentabilidade, sugerindo que a possibilidade de contratar um jogador alternativo aumenta o seu poder de negociação.

Kahn (2000), assim como Rottenberg (1956), transforma o mundo dos esportes em um laboratório para o mercado de trabalho, onde é possível observar qualquer proposição econômica e, posteriormente, testar seu real impacto no mercado de trabalho como um todo. Para ele, não há nenhum outro cenário de investigação como o esporte, em que nome, rosto e histórico de vida de cada trabalhador de produção tenha tão fácil acesso. Além disso, as ligas esportivas profissionais experimentam grandes mudanças nas regras e nas estruturas do mercado de trabalho, criando experiências naturais interessantes que oferecem oportunidades ímpares para análise.

No artigo, são quatro os seus objetos de análise. Primeiramente, Kahn compara diferentes ligas esportivas norte-americanas e discorre sobre a redução ou não de salários dos atletas com a concessão da liberdade dos direitos econômicos – o que, analogamente, no mercado de trabalho tem relação com leis trabalhistas mais flexíveis. Segundamente, compara os salários dos atletas e sua produtividade para testar se há discriminação racial dentro das ligas. Após, apresenta um teste para o Teorema de Coase, nas principais ligas esportivas, que diz que, numa transação econômica com externalidades, se os direitos de propriedade forem bem definidos e se os custos de transação forem suficientemente baixos, então a solução privada é socialmente ótima, não havendo qualquer necessidade de intervenção do governo na correção da externalidade, que é uma falha de mercado. Por fim, estima o impacto de incentivos financeiros aos atletas em sua produtividade.

Em geral, as conclusões obtidas relacionadas ao esporte em seu estudo servem como base para o restante do mercado de trabalho. Importante destacar, aqui, a presença detectada de discriminação racial no esporte, o que será também um objeto de estudo do presente trabalho. Kahn sugere que times que deixaram a discriminação de lado e contrataram trabalhadores pela sua produtividade, tiveram resultados mais positivos, como é o caso da Liga Nacional de Baseball no final dos anos 1940 e 1950, na qual equipes que estavam mais dispostas a usarem jogadores negros realmente tiveram mais sucesso do que aquelas que não o fizeram. Para testar a hipótese, Kahn sugere que o indicador mais eficiente para qualificar a discriminação no esporte é o de discriminação salarial, onde salário é a variável dependente e os indicadores de

desempenho são as variáveis independentes, com uma variável *dummy*¹ para a raça branca. O esporte no qual foi detectada uma maior evidência de discriminação salarial foi o basquete profissional, onde os salários de atletas negros tinham, em média, reduções entre 11-25% quando comparados aos atletas brancos, na metade dos anos 1980.

Nesse sentido, em “*Racial discrimination in English professional football: evidence from an empirical analysis of players' career progression*”, John Goddard e John O. S. Wilson (2009) utilizam um banco de dados extraído da Associação de Futebol do Reino Unido e de fotografias lá publicadas para avaliarem a presença ou não de discriminação racial no mercado de transferências do futebol do país, no período entre 1986 e 2001. Tal período, foi marcado por mudanças institucionais relevantes no mercado de trabalho do futebol, como por exemplo o Caso Bosman², em 1995, que abriu pela primeira vez a possibilidade de jogadores se transferirem entre clubes de forma livre. Tal marco, elevou significativamente o fluxo de jogadores nascidos fora do Reino Unido para os clubes locais. O percentual de jogadores estrangeiros, que era de de 2,8% em 1986, se elevou para 17,7% em 2001, conforme apurou o estudo. Além disso, a participação de jogadores negros entre os anos de 1986 e 1996 subiu de 7,6% para 13,2%. O principal ponto levantado pelo estudo sugere ainda que jogadores negros que estiveram nas ligas de diferentes divisões inglesas, no período em estudo, tendiam a se transferir para equipes de divisões superiores, ou seja, teriam talento acima da média. Analisando tal resultado, John Goddard e John O. S. Wilson (2009, pg. 314), comentam que é possível supor uma maior dificuldade para jogadores negros chegarem ao futebol profissional do que jogadores brancos com as mesmas ou semelhantes características: “While the most

¹ As variáveis empregadas na análise de regressão são muitas vezes variáveis quantitativas, ou seja, as variáveis têm uma escala de medição bem definida. Variáveis como temperatura, distância, pressão e renda são variáveis quantitativas. Em algumas situações é necessário utilizar variáveis qualitativas ou categóricas como variáveis preditoras na regressão. Exemplos de variáveis qualitativas ou categóricas são operadores, situação de emprego (empregado ou desempregado), turnos (dia, noite ou noite) e sexo (masculino ou feminino). Em geral, uma variável qualitativa não tem escala natural de medição. Devemos atribuir um conjunto de níveis a uma variável qualitativa para explicar o efeito que a variável pode ter na resposta. Isso é feito através do uso de variáveis Dummy. (Montgomery, D. C.; Peck, E. A. e Vining, G. G. (2001), pp. 278)

² O caso do atleta Jean-Marc Bosman teve grandes consequências no futebol europeu e mundial. Bosman levou o seu descontentamento – e de grande parte dos atletas - com a regulamentação vigente na Europa até o Tribunal de Justiça da União Europeia. Em um nível europeu, a decisão revogou a regra que limitava o número de atletas estrangeiros que podiam disputar os campeonatos nacionais e organizados pela UEFA. No âmbito do futebol mundial, levou à aprovação, por parte da FIFA, de uma nova regulamentação a respeito do *status* e das transferências de jogadores. O caso marcou o início de uma abertura progressiva do mercado de transferências europeu e de uma maior independência dos atletas na negociação de seus contratos. (“*O Direito e a Copa Do Mundo de Futebol*”. Marques, J.; Faria da Silva, M. (2013))

talented tend to be able to overcome whatever barriers exist, less talented black footballers find it harder to do so than their (equally talented) white counterparts.”.

Por outro lado, Frick (2007) cita que não existem evidências suficientes para apontar a presença de discriminação sobre salários ou valores pagos nas transferências dos atletas. Analisando atletas nascidos na Alemanha e atletas estrangeiros que atuam em clubes da Bundesliga (Liga Alemã), apontou que houve uma queda significativa no número de jogadores nascidos na Alemanha atuando na liga, enquanto atletas vindos da Europa Oriental, Ocidental e América do Sul ganham, respectivamente, 15%, 30% e 50% mais do que os alemães. Tais incrementos, entretanto, não advêm de uma possível discriminação com os jogadores natais, mas deve-se ao fato de os estrangeiros atraírem mais público, venderem mais camisetas e terem maior procura de outros mercados (mecanismo de concorrência).

Frick também traz ao debate alguns dos principais motivos que causavam falta de interesse, até então, por parte dos economistas, em analisar o mercado de trabalho do futebol, quando comparado aos quatro principais esportes americanos (futebol americano, basquete, baseball e hockey). Primeiramente, reporta a dificuldade de acesso aos dados salariais dos atletas, em toda Europa (exceto na Inglaterra), o que não ocorria nos Estados Unidos. Em segundo lugar, até meados de 1990, havia excessivas restrições ao mercado de transferências de atletas no continente europeu, o que afastava qualquer semelhança com o mercado de trabalho geral, e desmotivava estudos científicos nesse sentido.

Entretanto, com o desaparecimento destas restrições, o número de economistas de toda Europa dispostos a estudar o esporte cresceu relevantemente. Além disso, há de se destacar a diferença, apontada por Frick, entre a maximização da utilidade por parte dos clubes na Europa, que é satisfeita através de ganhos esportivos, e o que é encontrado nos Estados Unidos, onde a maximização da utilidade dos clubes advém dos lucros obtidos. Por conta disso, não há muitos incentivos no velho continente para se diminuir os salários dos atletas e, por consequência, os altos valores envolvidos em transações esportivas, tornando-os cada vez maiores. Em seu trabalho, Frick também utiliza ferramentas econométricas para aprofundar os estudos sobre o futebol europeu e entender os determinantes das remunerações dos atletas, das taxas de transferência, da duração da carreira dos atletas, dos prazos contratuais e da existência ou não de discriminação.

Importante ressaltar que as variações observadas nas taxas de transferências puderam ser explicadas pelas mesmas variáveis que afetam o salário dos atletas, merecendo destaque a

variável que mensura o tempo de contrato restante do atleta, já que desde meados dos anos 1990, com as novas leis e a liberalização do mercado de transferências, esta tende a ser a variável mais importante na determinação dos valores das transferências. Isso se dá, pois quanto menor for o tempo de contrato restante, menor deverá ser o valor pago pela multa rescisória do atleta e vice-versa.

Na linha de Frick (2007), Majewski (2016) teve como objeto de estudo a identificação dos fatores determinantes do valor de mercado de jogadores de futebol. Sua pesquisa contou com o método da regressão linear múltipla e se restringiu a atletas da posição de ataque, tendo em vista a não padronização de características observadas nas diferentes posições de campo. A partir de dados das 150 maiores transferências para jogadores destas posições, extraídos do site alemão “Transfermarkt.de”, Majewski constrói um quadro de pesquisa, com variáveis divididas entre as características do atleta e de seu rendimento (idade, altura, reputação, tempo de contrato, experiência, gols, assistências, número de roubadas de bola, etc...), bem como características dos clubes que vendem os atletas (valor de seu elenco, posição na liga que joga, etc...) e características do país de nascimento.

As características do atleta foram subdivididas entre fatores pertencentes ao Capital Humano (idade, performance ou número de partidas jogadas na temporada, número de vezes entre os titulares e número de vezes substituído), Produtividade (gols feitos, assistências e número de cartões vermelhos e amarelos) e Capital Organizacional (tempo total em campo, valor do time, posição do time na liga e rank nacional da sua seleção natal). Além disso, foram utilizadas variáveis *dummy*¹ e testes para correção do modelo. Dentre as dezesseis variáveis testadas, as que mais se mostraram representativas no modelo, seguindo o critério de exclusão por nível de significância, foram, primeiramente, a soma entre os gols feitos e assistências, dada por X_{15} , bem como o valor ponderado do clube do atleta, que é ajustado pela pontuação da equipe no ranking da FIFA, dada por X_{16} . A variável *dummy*, dada por X_{17} , que separa os cinco atletas mais valiosos (Lionel Messi, Cristiano Ronaldo, Neymar, Gareth Bale e Edinson Cavani) dos demais, também se mostrou relevante no modelo, indicando uma espécie de “fator da marca” ou “efeito estrela”, como alguns autores também relataram na literatura. Dessa forma, a equação encontrada por Majewski em seu artigo é dada por

$$y = 19,953 + 0,814.X_{15} + 0,105.X_{16} + 38,832.X_{17}$$

onde, β_0 expressa que o valor médio da variável dependente (valor de mercado), dadas as demais variáveis nulas, é 19,9 milhões de euros. β_1 , denota que, caso haja uma variação de uma unidade

em X_{15} (de um gol ou assistência), o valor de mercado do atacante será impactado em 0,8 milhão de euros, dadas as demais variáveis constantes. Quanto à variável X_{16} , a interpretação é de que, havendo um aumento ou redução de uma unidade no “valor ponderado do clube”, o atleta será afetado, em igual sinal, em aproximadamente 0,11 milhão de euros, dado que as demais variáveis permaneçam constantes. Para Majewski, entretanto, a variável *dummy* (X_{17}) é a mais interessante dentre as testadas. Esta indica que, por fazer parte do seletivo “grupo das estrelas”, o valor de mercado do atleta será acrescentado em 38,8 milhões de euros, tendo as demais variáveis constantes.

Para medir a qualidade do ajuste obtido com o modelo, isto é, quanto da variável dependente consegue ser explicada pelo modelo ajustado, utiliza-se o valor de R^2 ajustado, ou coeficiente de determinação ajustado. O valor encontrado por Majewski (2016) do R^2 ajustado foi de 56,11%, indicando que há espaço para outras variáveis explicativas que não foram apresentadas por este modelo, já que 44% não conseguiram ser explicadas pelo modelo.

Geurts (2016) também utiliza o método de regressão linear múltipla e encontra evidências de que o desempenho do atleta tem relação positiva com o valor pago em sua transferência. Além disso, os termos do contrato do atleta também são variáveis altamente significativas aqui, principalmente no que se refere ao tempo restante do contrato em vigor, que tem forte relação positiva. Na pesquisa, Geurts (2016) também utiliza como base de dados o *site* alemão “Transfermarkt.de”, bem como o *site* que monitora performances dos atletas “whoscored.com” para modelar as transações das cinco principais ligas europeias (França, Espanha, Itália e Alemanha) da janela de transferências do verão do ano de 2015 e o desempenho destes atletas.

Ademais, Geurts (2016) utiliza como referência, para categorizar e avaliar o desempenho por posições destes jogadores em campo, uma estrutura projetada pelo Observatório do Futebol (CIES) que elenca as características necessárias dos atletas para cada uma das funções exercidas dentro das quatro linhas. Suas conclusões são de que existem evidências para afirmar que a performance dos atletas é o principal determinante de seus valores de mercado.

Com o descarte de algumas variáveis que se mostraram insignificantes, a reta encontrada por Geurts, seguindo um critério de 5% de nível de significância, é dada por

$$\text{Log(Fee)} = 14,5117 + 0,0406 \cdot \text{Timerem} + 0,0632 \cdot \text{Goals} + 0,0930 \cdot \text{Assists} + 1,3280 \cdot \text{Perf}$$

onde a aplicação de Log à variável dependente *Fee* (valor da transferência do atleta, em libras) permite interpretar os coeficientes das variáveis independentes percentualmente. Sendo assim, considerando todas as demais variáveis nulas, o β_0 indica que o valor médio das transferências observadas foi de $e^{14,5117}$, isto é, aproximadamente 2,01 milhões de libras. Por sua vez, uma variação de um mês no tempo de contrato restante do atleta, dado pela variável *Timerem*, impacta em aproximadamente 4,1% no valor de sua venda, dadas as demais variáveis constantes, sendo esta a variável mais significativa do modelo. Os gols marcados (variável *Goals*), as assistências (variável *Assists*) e a variável criada por Geurts para categorizar a performance, por posição, de cada atleta, dada por *Perf*, também apresentam forte relação positiva com o valor de mercado do atleta. Respectivamente, acrescentam 6,3%, 9,3% e 132,8% ao preço dos jogadores, dadas as demais variáveis constantes, todas com nível de significância abaixo de 1%. O valor de R^2 ajustado encontrado por Geurts (2016) foi relativamente baixo, comparado ao de Majewski (2016), indicando que apenas 45,97% das variações da variável dependente são explicadas pelo modelo encontrado.

Após avaliar os resultados encontrados, Geurts (2016) sugere que seria necessário que pesquisadores conseguissem combinar uma metodologia capaz de mesclar a teoria da barganha com a avaliação de performance dos atletas, com a finalidade de obter melhores resultados para explicar os determinantes do valor de mercado destes. Acrescenta ainda a necessidade de considerar o mercado do futebol como um negócio, no qual os jogadores são os ativos financeiros dos clubes e não mais trabalhadores inseridos em um mercado de trabalho, como são tratados na literatura.

Wicker, et. al (2013) também utiliza uma base de dados extraída do portal *Transfermarkt*, em conjunto com dados de desempenho extraídos do website da liga alemã, relativos à temporada 2011/2012, para testar qual o impacto do esforço físico individual dos atletas na variação do valor de mercado. Entretanto, Wicker, et. al (2013) excluiu de seu banco de dados atletas que foram transferidos durante a temporada, goleiros, por terem medidas de esforço incomparáveis com os demais atletas, bem como jogadores que jogaram menos do que noventa minutos (o tempo de uma partida completa). Ao final, o modelo apresentado por Wicker, et. al (2013) contou com dados de 446 atletas e foi apresentado conforme descrito abaixo:

$$\begin{aligned} MV_LOG = & \alpha_0 + \alpha_1 AGE + \alpha_2 AGE^2 + \alpha_3 GERMAN + \alpha_4 HIGH + \alpha_5 APPEAR + \alpha_6 TRANSFER + \alpha_7 \\ & TIME_CLUB + \alpha_8 SP + \alpha_9 EFFIC + \alpha_{10} TACK + \alpha_{11} NAT + \alpha_{12} NAT_TOP + \alpha_{13} RUN + \alpha_{14} RUN2 + \alpha_{15} \\ & RUN*SP + \alpha_{16} RUN*EEFIC + \alpha_{17} RUN*TACK + \sum_{i=1}^8 \alpha_i POSITION + \sum_{i=1}^{21} \alpha_i TEAM + \varepsilon \end{aligned}$$

onde, conforme cita Majewski (2016, pg. 96):

Tabela 1: Variáveis do modelo de Wicker, et. al (2013), segundo Majewski (2016, pg. 96).

<i>MV_LOG</i>	Logarítimo do valor de mercado do jogador;
<i>AGE</i>	Idade do jogador;
<i>GERMAN</i>	Variável Dummy (se alemão = 1);
<i>HIGH</i>	Altura do jogador (em cm);
<i>APPEAR</i>	Número de aparições em campo;
<i>TRANSFER</i>	Variável Dummy (Se foi transferido na metade da temporada = 1);
<i>TIME CLUB</i>	Tempo jogando no clube (em anos);
<i>SP</i>	Performance de gols e assistências somados, por minuto;
<i>EFFIC TACK</i>	Taxa de eficiência nos contatos de bola ([escapadas + passes certos + chutes]/contatos físicos);
<i>NAT</i>	Taxa de carrinhos certos;
<i>NAT TOP</i>	Variável Dummy (se o jogador joga pela seleção nacional = 1);
<i>RUN</i>	Variável Dummy (se o jogador joga por uma seleção nacional que está presente no TOP da FIFA = 1);
<i>POSITION</i>	Número médio de corridas intensas por jogo, e;
<i>TEAM</i>	Variável Dummy representando a posição do atleta (defesa central, lateral esquerdo, lateral direito, volante, meia-esquerdo, meia-direito, meia-atacante, atacante);
	Variável Dummy que representa os clubes.

No estudo, Wicker, et. al (2013), ao tentar medir o impacto do esforço dos atletas no valor de mercado, encontrou relativa insignificância nas variáveis testadas e, ainda, efeitos negativos significativos nesta relação. O autor encontrou correlação positiva apenas nas variáveis *RUN* e *TACK*. Identifica, portanto, um efeito interessante:

These results seem interesting in the view of the classical economic literature, which assumes that higher effort leads to higher performance and subsequent higher payment (Lazear & Shaw, 2007). The insignificant and negative effects of effort on player values seem surprising, although it has already been found in previous research (Treble, 2001). Moreover, several explanations for these findings can be advanced. (Pg. 16)

Primeiramente, os autores descrevem que um esforço maior, por parte do atleta, não necessariamente gera uma melhora no desempenho que seja percebida pelos gerentes. Segundamente, o esforço não compensa a falta de talento. Terceiro, alguns jogadores conseguem ler a partida melhor e, portanto, se posicionar melhor, obtendo resultados melhores com menos esforço. Quarto, alguns jogadores se sacrificam pelo time para que outros atletas, com características mais ofensivas e maior talento, consigam marcar gols. Sendo assim, o esforço destes atletas que contribuem para que o jogador talentoso marque os gols e ajude a equipe não é reconhecido em igual escala. Por fim, os autores sugerem que pode haver uma subvalorização do esforço pelo mercado de trabalho do futebol.

3 MODELO E BANCO DE DADOS

3.1 APRESENTAÇÃO DO MÉTODO

O método proposto consiste em testar, através de um modelo econométrico de regressão por Mínimos Quadrados Ordinários (MQO), as variáveis que melhor explicam o valor de mercado de um atleta de futebol da Barclays Premier League (Liga Profissional Inglesa) da temporada de 2018/19. Ao final, será estimado quanto do valor de mercado individual dos atletas pode ser explicado pelo modelo.

O critério para exclusão e inclusão das variáveis independentes no modelo será o método Stepwise, cuja aplicação se dará através de uma rotina no software R-Studio. Ademais, o banco de dados também será apresentado no presente capítulo.

3.1.1 A REGRESSÃO LINEAR E O MÉTODO STEPWISE

Em uma regressão linear múltipla, isto é, com mais de uma variável explicativa, a equação para estimar o valor da variável dependente é dada por

$$y = \beta_0 + \beta_1 \cdot X_1 + \beta_2 \cdot X_2 + \dots + \beta_n \cdot X_n + \varepsilon,$$

onde y é a variável dependente; β_0 é o parâmetro que indica o valor médio quando todas as variáveis explicativas são nulas; $\beta_1, \beta_2, \dots, \beta_n$ são coeficientes da regressão que indicam a mudança na variável dependente a cada variação unitária em X_1, X_2, \dots, X_n , respectivamente, mantidas as demais variáveis constantes, sendo n o número de variáveis explicativas; e ε é uma variável aleatória que representa o resíduo, também chamada de erro padrão.

Dado que a base de dados coletada para a realização da análise de regressão do presente trabalho conta com um expressivo número de variáveis, não se mostra viável selecioná-las para o modelo considerando o simples critério de eliminação por nível de significância. Isso se dá pela provável correlação entre as variáveis explicativas, podendo resultar em problemas de multicolinearidade³ ou outros problemas oriundos de uma má

³ A multicolinearidade é um problema comumente notado em análises de regressão, no qual as variáveis independentes possuem relações lineares exatas ou aproximadamente exatas entre si. As consequências da multicolinearidade em uma regressão são: erros-padrão elevados, no caso de multicolinearidade moderada ou severa e, até mesmo, a impossibilidade de qualquer estimação se a multicolinearidade for perfeita. O nível de multicolinearidade pode ser medido pelo Fator de Inflação da Variância (FIV).

especificação do modelo. Estes, por sua vez, serão testados e expostos no capítulo de apresentação dos resultados.

O critério escolhido para apontar o melhor ajustamento e selecionar as variáveis que devem integrar a regressão é o método *Stepwise*, juntamente com o critério de seleção de variáveis pelo menor valor de AIC (*Akaike Information Criterion*) sempre que uma variável for adicionada ou retirada do modelo. Segundo Montgomery, D. C.; Peck, E. A. e Vining, G. G. (2001), o método *Stepwise*, assim como outros, foi desenvolvido com o objetivo de avaliar computacionalmente modelos com um número elevado de possibilidades de regressões, de forma que não seja sobrecarregada a capacidade de processamento das máquinas. Estes modelos, conforme os autores, estão separados em três tipos, denominados como, *Forward Selection*, *Backward Selection* e *Stepwise Regression*.

O método *Forward* de seleção de variáveis consiste em adicionar regressores do início, isto é, a partir do momento em que não há nenhuma variável no modelo, apenas a constante. A partir daí, adiciona-se, segundo o critério escolhido, o regressor que melhor representar este critério. No exemplo de Montgomery, D. C.; Peck, E. A. e Vining, G. G. (2001), utiliza-se a seleção pelo maior valor de F , isto é, a variável que detiver a maior correlação com a variável dependente e estiver dentro do intervalo de significância escolhido. O segundo regressor a ser adicionado ao modelo é aquele que tem a maior correlação com a variável dependente, ajustada pelo efeito do primeiro regressor, e assim sucessivamente, até o momento em que a próxima variável testada não esteja mais respeitando o nível de significância escolhido.

O critério de seleção de variáveis *Backward*, por sua vez, tem uma ideia semelhante, entretanto, utiliza a direção contrária. Este consiste em retirar os regressores, conforme o critério escolhido, partindo de uma equação que já contenha todas as variáveis do modelo. Os autores utilizam o mesmo critério, relacionado ao valor de F , para demonstrar o método. O primeiro passo consiste em retirar a variável que tiver o menor valor de F , isto é, a variável que tiver menor correlação, desde que esta esteja fora do intervalo de significância escolhido. O segundo regressor a ser retirado será o que, novamente, tiver o menor valor de F , após a retirada da primeira variável, e não estiver respeitando o nível de significância escolhido. O processo só se encerra quando todas

variáveis que restarem, após ajustadas pela retirada do regressor do passo anterior, estiverem respeitando o nível de significância escolhido.

O *Stepwise Regression Algorithm of Efroyimson* (1960), entretanto, é uma junção entre os dois critérios. Montgomery, D. C.; Peck, E. A. e Vining, G. G. (pg. 366) apontam que ambos os métodos descritos acima permitem uma ampla possibilidade de combinações para o modelo. Isso ocorre porque, ao adicionar uma variável, outro regressor que já estava no modelo, no passo anterior, pode perder significância, como é o caso do método *Forward*. Já no critério *Backward*, ao retirar uma variável, outra que havia sido retirada anteriormente pode recuperar um certo grau de significância que a permita voltar ao modelo. Tal comportamento, explicam os autores, se deve ao grau de correlação entre os regressores do modelo.

No exemplo trazido pelos autores, a primeira etapa do método *Stepwise* inicia-se utilizando o método *Forward*, e, portanto, sem nenhum regressor no modelo, apenas a constante. Após isso, é acrescentada a variável que tiver a maior correlação com a variável dependente. O próximo passo é, a partir do ajuste do modelo, verificar se há alguma variável que não seja compatível com o nível de significância escolhido. Caso haja, esta variável é retirada. Do contrário, adiciona-se o próximo regressor. Sendo assim, a metodologia baseia-se em aplicar antes de cada etapa do método *Forward*, uma etapa do método *Backward*. O processo se encerra quando não há mais variáveis que necessitem ser retiradas ou adicionadas, respeitando o critério escolhido.

3.1.2 O CRITÉRIO AIC (AKAIKE INFORMATION CRITERION)

A escolha do melhor modelo é uma importante etapa dentro do processo de análise dos dados. Este, por sua vez, deverá ser selecionado dentro de um conjunto amplo de possibilidades. Deve-se buscar um modelo que tenha um número não tão elevado de parâmetros e, ao mesmo tempo, explique a variável dependente significativamente, respeitando o princípio científico fundamentado na parcimônia, ou simplicidade do modelo. São muitos os critérios que podem ser adotados para eleger o melhor modelo, sendo este um tópico com amplo debate na literatura. Dado o elevado número de variáveis a serem investigadas e o objetivo do modelo proposto, de estimar ou prever valores para atletas que se assemelhem ao banco de dados, o método que se mostrou mais apropriado para elaboração do trabalho é o de Regressão via *Stepwise*, utilizando o critério de Akaike (Akaike, 1974) para verificar a exclusão e inclusão de variáveis no modelo. Isso porque

o critério de seleção por AIC é uma técnica ajustada para encontrar o modelo com a máxima verossimilhança e, com isso, apropriado para estimar valores futuros a partir de uma amostra.

Segundo Bozdogan, H. (2000), o critério de Akaike admite que não há, dentre todos os conjuntos de modelos possíveis, algum que descreva perfeitamente a relação entre a variável dependente e as variáveis explicativas, tido como o “modelo verdadeiro”. Dessa forma, o critério busca encontrar o modelo com a menor divergência de Kullback-Leibler (K-L), ou máxima entropia, isto é, o que mais se aproxima do “modelo verdadeiro”, ou o que menos se afasta da realidade. O valor do AIC é dado por

$$AIC = 2k - 2\ln(L),$$

onde k é o número de parâmetros desconhecidos e estimados e L é a verossimilhança maximizada.

Conforme Emiliano, P.C. (2009), o AIC é um método que avalia a qualidade do ajuste do modelo paramétrico através da máxima verossimilhança, elegendo aquele com menor valor de AIC como o melhor. Sendo assim, o AIC não testa hipóteses e, com isso, não toma o nível de significância, ou o p-valor dos regressores como critério. O valor do AIC, por si só, não traz conclusões para a análise. Ele é utilizado apenas para eleger aquele modelo que detiver o menor valor. É possível também considerar os três primeiros valores como elegíveis. Ainda segundo o autor:

Na regressão múltipla, nem todas as variáveis explicativas necessariamente influenciarão significativamente a variável resposta. Um modelo estimado com um grande número de variáveis explicativas desnecessárias pode ser inviável. Selecionando o modelo com o menor AIC para todas as diferentes possíveis combinações da variável explicativa, espera-se obter um modelo razoável, que equilibre a qualidade do ajuste e a complexidade. Pg. 62

3.2 APRESENTAÇÃO DO BANCO DE DADOS

Majewski (2016) define o valor de mercado dos atletas como sendo uma ferramenta utilizada pelos clubes para contabilizar ativos intangíveis em seus balanços. Dentre muitas formas de contrato existentes no futebol, a maior parte delas têm cláusulas de rescisão para que um atleta possa sair de uma equipe e atuar em outra. Segundo ele, o clube que quiser contar com os ganhos esportivos de um jogador, deverá pagar esta

quantia ou negociar com o clube que detém os direitos deste atleta. Outro ponto levantado por Majewski (2016) é de que os jogadores representam uma questão comercial e seu valor de mercado, em geral, acaba sendo superior ao valor exposto nos balanços dos clubes, por diversos fatores.

O presente trabalho utilizará dados extraídos majoritariamente de duas fontes, sendo (1) o portal *Transfermarkt*, especializado em descrever o valor de mercado dos atletas e amplamente utilizado na literatura, conforme levantado nas referências do capítulo anterior, bem como (2) dados de desempenho individual dos atletas na temporada 2018/2019 da Barclays Premier League, estes retirados do portal *Whoscored*. Além disso, dados relativos ao número de seguidores dos atletas também foram extraídos do Instagram. Todas as variáveis utilizadas no modelo estão demonstradas e descritas na tabela a seguir:

Tabela 2: Variáveis do modelo e origem dos dados.

<i>Transfermarkt</i>	
Variável	Descrição
<i>Val_mercado</i>	Variável dependente. Valor de mercado do atleta estimado pelo portal <i>Transfermarkt</i> , em milhões de euros.
<i>Idade</i>	Idade do atleta, em anos.
<i>Altura</i>	Altura do atleta, em centímetros.
<i>Jogos_Selecao</i>	Número de partidas oficiais e amistosas disputadas pela seleção que defende.
<i>Gols_Jogos_Selecao</i>	Número de gols marcados em partidas oficiais e amistosas pela seleção que defende, em relação ao número de partidas disputadas.
<i>Titulos_idade</i>	Número de títulos conquistados pelo atleta, considerando todas as ligas, copas e torneios de base, em relação à idade.
<i>Copas (Dummy)</i>	Se o atleta conquistou uma Copa do Mundo = 1
<i>Champions (Dummy)</i>	Se o atleta conquistou a UEFA Champions League = 1.
<i>Premios_idade</i>	Número de prêmios individuais conquistados pelo atleta (ex: artilheiro de um torneio, melhor do torneio, melhor goleiro do torneio, bola de ouro, luva de ouro, etc.), em relação à idade.
<i>Clube (Dummy)</i>	Clube do atleta. Arsenal é a referência.
<i>Posicao (Dummy)</i>	Posição preferencial de atuação do atleta. Defensor é a referência. (Defensor, Lateral, Meia-defensivo, Meia-ofensivo, Extremo, Centroavante)
<i>Nacionalidade (Dummy)</i>	Nacionalidade de origem do atleta. Europeu é a referência. (Europa, África, América do Sul, América do Norte, América Central, Oceania, Ásia)
<i>Perna (Dummy)</i>	Perna preferencial do atleta. Ser destro é a referência. (Destro, Canhoto e Ambidestro)
<i>Temp_clube_years</i>	Tempo que o atleta está no clube, em anos.

<i>Whoscored</i>	
<i>Titular</i>	Número de vezes que o atleta começou um jogo entre os 11 iniciais, com relação a todas as partidas que jogou (titulares e reserva) na BPL 2018/19. (Partidas titular/[partidas titular + partidas reserva])
<i>Minutos_jogados</i>	Número de minutos jogados, na BPL 2018/19.
<i>Gols_assist_90minutos</i>	Soma entre gols e assistências por cada 90 minutos jogados, na BPL 2018/2019. $([gols + assistências]/[minutos jogados/90])$
<i>Melhor</i>	Número de vezes eleito o melhor em campo, com relação ao número de aparições (titular + reserva), na BPL 2018/19.
<i>Nota_media_jogo</i>	Medida de desempenho. Nota média apurada pelo portal Whoscored.com para o atleta, na BPL 2018/19.
<i>Instagram</i>	
<i>Instagram_followers</i>	Número de seguidores no Instagram, em milhões de seguidores.

Fonte: Elaboração própria.

Os atletas que não apresentaram todas as informações, bem como os que estiveram por menos de 270 minutos (3 partidas completas) em campo, foram extraídos do banco de dados. Goleiros, assim como em outros modelos apresentados, também foram descartados do modelo, isso porque não é possível comparar igualmente o desempenho destes com atletas de outras posições de campo. Ao final, de um total de 508 observações, o banco de dados restou com 387.

Desta feita, o modelo a ser estimado é:

$$\begin{aligned} \text{Log}(\text{Val_mercado}) = & \beta_0 + \beta_1 \text{Idade} + \beta_2 \text{Altura} + \beta_3 \text{Jogos_Selecao} + \beta_4 \\ & \text{Gols_Jogos_Selecao} + \beta_5 \text{Titulos_idade} + \beta_6 \text{Premios_idade} + \beta_7 \text{Temp_clube_years} \\ & + \beta_8 \text{Titular} + \beta_9 \text{Minutos_jogados} + \beta_{10} \text{Gols_assist_90minutos} + \beta_{11} \text{Melhor} + \beta_{12} \\ & \text{Nota_media_jogo} + \beta_{13} \text{Instagram_followers} + \sum \beta_i \text{copas} + \sum \beta_i \text{champions} + \sum \beta_i \\ & \text{Clube} + \sum \beta_i \text{Posição} + \sum \beta_i \text{Nacionalidade} + \sum \beta_i \text{Perna} + \varepsilon \end{aligned}$$

Os valores apresentados no banco de dados se referem ao final da temporada, isto é, em maio de 2019. O site conta com um portal, onde fãs e especialistas do mundo todo identificam e opinam a respeito do preço dos atletas, de modo que os administradores da página consigam elaborar metodologias específicas para precificar e atualizar o valor dos jogadores, conforme descrito no estudo de Wicker, et. al (2013):

The website www.transfermarkt.de is a German community-based information site that provides performance statistics as well as market values of soccer players. Market values are published at the end of each half season. They are assessed and discussed by community members and experts. Only proven and confident members are allowed to enter the discussion forum about market values. At the end of the discussion the head of the internet forum sets the final market value. Pg. 9

A tabela abaixo demonstra a diferença entre o valor efetivamente pago na última transferência e o valor estimado pelo *Transfermarkt* na data desta transferência. Para o cálculo, foram aproveitados os dados de 414 atletas (dentre os 508 que pertenciam inicialmente ao banco de dados) que tiveram transferências anteriores ao período de coleta registradas pela página. Dessa forma, é possível estimar com qual precisão o *website* calcula o valor dos atletas. Os dados foram divididos em faixas de valores de mercado:

Tabela 3: Diferença entre o valor estimado pelo portal Transfermarkt e o valor efetivamente pago pelos atletas transferidos.

Faixa de valores estimados pelo portal Transfermarkt, em milhões de euros	Número de observações	Erro de estimação do Transfermarkt (%)
66 +	2	-0,7%
61 até 65	3	9,6%
56 até 60	1	4,2%
51 até 55	0	0,0%
46 até 50	4	21,3%
41 até 45	2	34,4%
36 até 40	7	33,0%
31 até 35	9	16,3%
26 até 30	14	33,9%
21 até 25	18	49,0%
16 até 20	37	22,1%
10,5 até 15	42	45,0%
5,5 até 10	90	69,3%
2,5 até 5	90	53,6%
0 até 2,4	95	149,4%

Fonte: Transfermarkt. Elaboração própria.

Ainda que o *website* não aponte de forma altamente precisa o valor de mercado dos atletas, como demonstrado acima, esta é a base de dados disponível e utilizada comumente na literatura. É possível identificar que os desvios e o número de transferências, em geral, diminuem conforme o preço do jogador aumenta, apontando uma maior assertividade do portal quando os jogadores são mais valiosos e uma maior raridade nas transferências que envolvem altas cifras. Outra característica observada aponta que as expectativas dos agentes quanto ao valor dos atletas tendem a ser inferiores ao valor que os clubes efetivamente desembolsam nas contratações, isto é, há uma subnotificação do valor do passe dos atletas pelo portal.

Ainda, a tabela abaixo demonstra o *ranking* dos vinte atletas mais valiosos dentre todos os 387 que farão parte do estudo, isto é, excluídos os atletas que jogaram menos de 270 minutos, os goleiros e aqueles que não tinham informações suficientes para serem estudados. A soma entre eles alcança – e ultrapassa – os dois bilhões de euros.

Tabela 4: Os vinte atletas mais valiosos do estudo.

Ranking	Nome	Clube	Valor de mercado estimado (em milhões de euros)
1	Harry Kane	Tottenham	150,00
2	Eden Hazard	Chelsea	150,00
3	Mohamed Salah	Liverpool	150,00
4	Raheem Sterling	Manchester City	140,00
5	Kevin De Bruyne	Manchester City	130,00
6	Sadio Mané	Liverpool	120,00
7	Leroy Sané	Manchester City	100,00
8	Christian Eriksen	Tottenham	100,00
9	Bernardo Silva	Manchester City	100,00
10	Paul Pogba	Manchester United	100,00
11	N'Golo Kanté	Chelsea	100,00
12	Dele Alli	Tottenham	90,00
13	Virgil van Dijk	Liverpool	90,00
14	Son Heung-Min	Tottenham	80,00
15	Marcus Rashford	Manchester United	80,00
16	Trent Alexander-Arnold	Liverpool	80,00
17	Roberto Firmino	Liverpool	80,00
18	Romelu Lukaku	Manchester United	75,00
19	Aymeric Laporte	Manchester City	75,00
20	Gabriel Jesus	Manchester City	70,00

Fonte: Transfermarkt. Elaboração própria.

Algumas características interessantes são perceptíveis na tabela acima. Apenas cinco clubes figuram na lista dos atletas mais valiosos da liga. São eles: Manchester City (6 atletas), Liverpool (5 atletas), Tottenham (4 atletas), Manchester United (3 atletas) e Chelsea (2 atletas). Com exceção do Arsenal, que será o clube escolhido como base para a *dummy Clube* nos testes do modelo, todos os times considerados favoritos, anualmente, em competições da Inglaterra, têm atletas entre os vinte mais valiosos da liga em seu elenco. Outra característica interessante é que há apenas um goleiro, com valor estimado de 80 milhões de euros à época, que figuraria na tabela acima: Alisson, do Liverpool. A

posição que mais apresenta jogadores é a de atacantes (12 atletas), seguida de meio-campistas (5 atletas) e de defensores e laterais (3 atletas), completando o grupo.

Outra relação interessante é a dos atletas que mais se valorizaram durante a temporada em análise. Para isso, foram extraídos dados do valor de mercado dos atletas no início da temporada (maio de 2018) e os mesmos dados ao final da temporada (maio de 2019):

Tabela 5: Os vinte atletas que mais se valorizaram.

Ranking	Nome	Clube	Valor de mercado em mai/2018 (em milhões de euros)	Valor de mercado em mai/2019 (em milhões de euros)	Variação (%)
1	Harvey Barnes	Leicester	0,25	12,00	4700,0%
2	Sean Longstaff	Newcastle United	0,15	7,00	4566,7%
3	Aaron Wan-Bissaka	Crystal Palace	1,00	35,00	3400,0%
4	Karlan Grant	Huddersfield	0,20	5,00	2400,0%
5	Matteo Guendouzi	Arsenal	2,50	35,00	1300,0%
6	Rúben Vinagre	Wolverhampton Wanderers	0,50	7,00	1300,0%
7	Morgan Gibbs-White	Wolverhampton Wanderers	1,00	8,00	700,0%
8	Matt Doherty	Wolverhampton Wanderers	2,50	15,00	500,0%
9	Jack Simpson	Bournemouth	0,25	1,50	500,0%
10	Ryan Fraser	Bournemouth	6,00	30,00	400,0%
11	Domingos Quina	Watford	1,00	5,00	400,0%
12	Hamza Choudhury	Leicester	1,00	5,00	400,0%
13	Declan Rice	West Ham	10,00	45,00	350,0%
14	Ben Chilwell	Leicester	8,00	35,00	337,5%
15	Chris Mepham	Bournemouth	3,50	13,00	271,4%
16	Diego Rico	Bournemouth	2,50	9,00	260,0%
17	Phil Foden	Manchester City	7,00	25,00	257,1%
18	Ken Sema	Watford	1,00	3,50	250,0%
19	David Brooks	Bournemouth	9,00	30,00	233,3%
20	Trent Alexander-Arnold	Liverpool	25,00	80,00	220,0%

Fonte: Transfermarkt. Elaboração própria.

A tabela acima indica que há maior espaço para valorização de atletas nos pequenos clubes, já que há apenas dois jogadores (Phil Foden, do Manchester City e Trent Alexander-Arnold, do Liverpool) presentes na lista acima que pertencem a algum dos

cinco clubes que figuravam na Tabela 3. Esta é, talvez, uma explicação para o grande número de empréstimos de jogadores jovens, cujos direitos de transferência pertencem a grandes clubes, para clubes menores.

Os atletas que obtiveram o melhor desempenho na temporada, seguindo o critério de maior nota média⁵ pelo portal *Whoscored*, cujo valor é calculado com base nas estatísticas do atleta na partida, estão mostrados na tabela abaixo:

Tabela 6: Os vinte melhores atletas.

Ranking	Nome	Nota média	Posição	Nacionalidade
1	Eden Hazard	7.81	Extremo	Europa
2	Raheem Sterling	7.61	Extremo	Europa
3	Mohamed Salah	7.57	Extremo	África
4	Sergio Agüero	7.53	Centroavante	América do Sul
5	Virgil van Dijk	7.44	Defensor	Europa
6	Sadio Mané	7.43	Extremo	África
7	Harry Kane	7.38	Centroavante	Europa
8	Trent Alexander-Arnold	7.36	Lateral	Europa
9	Lucas Digne	7.36	Lateral	Europa
10	Ricardo Pereira	7.34	Lateral	Europa
11	Paul Pogba	7.31	Médio Ofensivo	Europa
12	Fernandinho	7.31	Médio Defensivo	América do Sul
13	Fabian Schär	7.30	Defensor	Europa
14	Benjamin Mendy	7.29	Lateral	Europa
15	Aaron Wan-Bissaka	7.27	Lateral	Europa
16	Bernardo Silva	7.26	Extremo	Europa
17	Leroy Sané	7.26	Extremo	Europa
18	David Silva	7.26	Médio Ofensivo	Europa
19	Felipe Anderson	7.23	Extremo	América do Sul
20	Wilfried Zaha	7.21	Extremo	África

Fonte: Transfermarkt. Whoscored.com. Elaboração própria.

Dentre os vinte melhores atletas da temporada, conforme apontado pelo *Whoscored*, a nacionalidade mais frequente é a de europeus, com quatorze atletas. Três são sul-americanos e outros três são africanos. No que se refere ao posicionamento dos

⁵ Ratings are based on a unique, comprehensive statistical algorithm, calculated live during the game. There are over 200 raw statistics included in the calculation of a player's/team's rating, weighted according to their influence within the game. Every event of importance is taken into account, with a positive or negative effect on ratings weighted in relation to its area on the pitch and its outcome", (extraído de www.whoscored.com/Explanations).

atletas, a maioria dos jogadores (12) de melhor desempenho é de uma posição ofensiva (Extremo, Centroavante e Médio Ofensivo). A lista é completada por cinco laterais, dois defensores e um volante (Médio Defensivo).

Quando considerada toda a amostra, entretanto, a distribuição das nacionalidades principais⁶ dos atletas se dá conforme a tabela abaixo:

Tabela 7: Distribuição dos atletas por continente, considerando a nacionalidade principal.

Continente	Número de observações	%
Europa	292	75,5%
África	38	9,8%
América do Sul	35	9,0%
Ásia	10	2,6%
América Central	5	1,3%
América do Norte	5	1,3%
Oceania	2	0,5%
TOTAL	387	100,0%

Fonte: Transfermarkt. Elaboração própria.

A informação referente à nacionalidade é uma importante ferramenta para testar a existência de discriminação, seja ela racial ou relacionada à origem do atleta. Entretanto, não consideraremos, para fins de testagem na amostra, a nacionalidade principal apontada no portal *Transfermarkt*. Por vezes, o site também aponta a nacionalidade de origem do atleta. Portanto, será feito um ajuste na variável, com a finalidade de considerar a nacionalidade de origem do atleta e não a nacionalidade da seleção que o atleta escolheu defender.

Para exemplificar, utilizaremos os dados do atleta Aaron Wan-Bissaka, do Crystal Palace, à época. O *website* identifica o atleta como inglês, entretanto, aponta que há uma segunda nacionalidade, da República Democrática do Congo. Neste caso, para verificar existência de discriminação, o mais apropriado seria considerar a origem do atleta e não a nacionalidade atual, ainda que o atleta tenha frequentado seleções de base na Inglaterra

⁶ É dada pelo país que o atleta escolheu defender a seleção, caso haja mais de uma nacionalidade apontada no banco de dados.

(ver Figura 1, no Anexo). Para a variável *dummy Nacionalidade*, portanto, será considerado o atleta como oriundo do continente Africano.

O ajuste é necessário já que há muitos jogadores, principalmente oriundos da África – mas também de outros continentes –, que se naturalizam europeus ainda jovens. Em geral isso se dá para que possam jogar pelas seleções de base europeias ou por conta dos pais que, em decorrência de fluxos migratórios se estabeleceram na Europa.

Após estas alterações, a distribuição de nacionalidades, por origem, é apresentada abaixo:

Tabela 8: Distribuição dos atletas por continente de origem.

Continente de origem	Número de observações	%
Europa	217	56,1%
África	86	22,2%
América do Sul	43	11,1%
América Central	24	6,2%
América do Norte	10	2,6%
Ásia	5	1,3%
Oceania	2	0,5%
TOTAL	387	100,0%

Fonte: Transfermarkt. Elaboração própria.

Quanto aos dados extraídos do *Instagram*, estes têm por finalidade testar o impacto do valor da marca dos atletas em seu valor de mercado. Assim como Rosen (1981), que descreve esta qualidade como “apelo de bilheteria” e a qualifica como uma habilidade para atrair audiência e gerar um grande volume de transações, o presente trabalho também testará os impactos deste efeito, utilizando como base o número de seguidores dos atletas na rede social, já que esta é, hoje, uma das ferramentas mais utilizadas para medir o potencial de alcance e de engajamento das marcas. Rosen (1981) afirma, ainda, que há uma combinação difícil de ser classificada que resulta no efeito descrito acima, acreditando, entretanto, que tal relação provenha do talento e do carisma dos atletas. Neste sentido, outros autores também buscaram entender tal relação, como é o caso de Lucifora e Simmons (2003), que concluíram que medidas de desempenho são as que mais impactam no “efeito superestrela”, sendo mais influentes nos jogadores de posições ofensivas.

Ademais, com o intuito de melhor apresentar a amostra, outras características, incluindo as estatísticas descritivas, estão expostas nas duas tabelas subsequentes:

Tabela 9: Estatísticas descritivas da amostra.

Variável	Valor mín.	Valor máx.	Desvio Padrão	Variância	Média
<i>Val_mercado</i>	0,75	150,00	24,3964	595,1825	21,02
<i>Idade</i>	18,52	36,76	3,5412	12,5403	27,16
<i>Altura</i>	163,00	199,00	6,6611	44,3706	181,57
<i>Jogos_Selecao</i>	0,00	131,00	29,2271	854,2221	24,97
<i>Gols_Jogos_Selecao</i>	0,00	1,00	0,1518	0,0231	0,09
<i>Titulos_idade</i>	0,00	0,79	0,1539	0,0237	0,12
<i>Premios_idade</i>	0,00	0,36	0,0335	0,0011	0,01
<i>Temp_clube_years</i>	0,18	14,37	2,3953	5,7374	2,93
<i>Titular</i>	0,04	1,00	0,2245	0,0504	0,77
<i>Minutos_jogados</i>	271,00	3420,00	893,0953	797619,2698	1725,37
<i>Gols_assist_90minutos</i>	0,00	1,05	0,2211	0,0489	0,23
<i>Melhor</i>	0,00	0,38	0,0493	0,0024	0,03
<i>Nota_media_jogo</i>	6,03	7,81	0,2997	0,0898	6,72
<i>Instagram_followers</i>	0,00	37,72	3,6762	13,5147	1,35

Fonte: Transfermarkt. Whoscored. Instagram. Elaboração própria.

Tabela 10: Número de observações das variáveis *Dummy*.

Variáveis <i>Dummy</i>	
Pé preferencial	Observações
Destro	276
Canhoto	93
Ambidestro	18
Clube	Observações
Arsenal	19
Bournemouth	21
Brighton	18
Burnley	16
Cardiff	19
Chelsea	18
Crystal_Palace	17
Everton	20
Fulham	20
Huddersfield	21
Leicester	20
Liverpool	18
Manchester_City	20
Manchester_United	22

Newcastle_United	22
Southampton	20
Tottenham	21
Watford	18
West_Ham	21
Wolverhampton_Wanderers	16
Copa do mundo FIFA	Observações
Ganhou	12
Não ganhou	375
Champions League UEFA	Observações
Ganhou	31
Não ganhou	356
Posição	Observações
Defensor	72
Lateral	75
Meia_Defensivo	32
Meia_Ofensivo	93
Extremo	60
Centroavante	55

Fonte: Transfermarkt. Whoscored. Elaboração própria.

3.3 O SOFTWARE R-STUDIO

Para realizar a testagem e a adequação do melhor modelo, será utilizado o *software* R-Studio. O *software*, por ser de linguagem aberta, conta com a possibilidade de desenvolvimento e compartilhamento de funções entre os usuários, através da linguagem R de programação:

RStudio is an integrated development environment (IDE) for R. It includes a console, syntax-highlighting editor that supports direct code execution, as well as tools for plotting, history, debugging and workspace management. Click here to see more RStudio features. RStudio is available in open source and commercial editions and runs on the desktop (Windows, Mac, and Linux) or in a browser connected to RStudio Server or RStudio Server Pro (Debian/Ubuntu, Red Hat/CentOS, and SUSE Linux). Extraído de: <https://rstudio.com/products/rstudio/>.

A rotina utilizada no estudo pode ser encontrada no apêndice. Os resultados obtidos, bem como testes para alguns subgrupos da amostra, serão apresentados no capítulo a seguir.

4 APRESENTAÇÃO DOS RESULTADOS

O presente capítulo tem como objetivo apresentar os resultados obtidos e analisá-los com base na literatura. Será aplicado o método descrito no capítulo anterior ao banco de dados contendo todas as 387 observações.

4.1 APRESENTAÇÃO DOS MODELOS

Seguindo a especificação *log-linear*⁷ apresentada no capítulo anterior, que contém a variável dependente *Val_mercado* e todos os regressores, os resultados primários para a Regressão Linear Múltipla encontrados pelo *software R-Studio* são os apresentados como modelo (1), na Tabela 11 abaixo.

Entretanto, é necessário ajustar o modelo com o objetivo de encontrar aquele que melhor explique a relação entre a variável dependente e as variáveis explicativas. Os modelos (2) e (3), também expostos na tabela, serão detalhados subsequentemente. Enquanto o modelo (2) apresenta o resultado da aplicação do método *Stepwise*, o (3) será utilizado com o objetivo de inferência, considerando-se apenas as variáveis obtidas pelo método com nível de significância menor ou igual a 5%.

Para todos os modelos encontrados, os valores de R^2 ajustado e do teste F se mostraram adequados. Os demais testes e gráficos, quais sejam, o teste RESET de Ramsey para má especificação da forma funcional, teste para identificação de multicolinearidade, testes para normalidade dos resíduos, para heterocedasticidade e de observações não usuais (*outliers*) serão realizados sobre o modelo (3), já que é o mais adequado para inferência.

⁷ A regressão pelo modelo log-linear se mostrou a mais adequada para o presente trabalho. Também foram avaliadas as especificações log-log e linear-linear.

Tabela 11: Modelos e resultados.

Variável	Modelo 1 (Modelo inicial)		Modelo 2 (Estimativa)		Modelo 3 (Inferência)	
	Coefficiente	P-valor	Coefficiente	P-valor	Coefficiente	P-valor
<i>Intercept</i>	2,8587	0,039569 *	4,3693	< 2e-16 ***	4,3724	< 2e-16 ***
<i>Idade</i>	-0,1067	< 2e-16 ***	-0,1014	< 2e-16 ***	-0,1015	< 2e-16 ***
<i>Altura</i>	-0,0011	0,803508	-	-	-	-
<i>Titular</i>	0,7667	0,000166 ***	0,8729	8,59e-07 ***	0,8767	8,72e-07 ***
<i>Gols_assist_90minutos</i>	-0,1892	0,330553	-	-	-	-
<i>Minutos_jogados</i>	0,0003	3,95e-13 ***	0,0003	< 2e-16 ***	0,0003	< 2e-16 ***
<i>Melhor</i>	-0,4209	0,573492	-	-	-	-
<i>Nota_Media_jogo</i>	0,2889	0,119238	-	-	-	-
<i>Jogos_Selecao</i>	0,0017	0,168247	-	-	-	-
<i>Gols_Jogos_Selecao</i>	0,6819	0,000813 ***	0,7551	9,93e-05 ***	0,8267	1,33e-05 ***
<i>Titulos_idade</i>	-0,2588	0,289680	-	-	-	-
<i>copas</i>	-0,2227	0,129641	-0,2250	0,11330	-	-
<i>champions</i>	0,0741	0,614499	-	-	-	-
<i>Premios_idade</i>	1,1929	0,167408	1,3254	0,09017	-	-
<i>Instagram_followers</i>	0,0219	0,013039 *	0,0227	0,00415 **	0,0222	0,00385 **
<i>Temp_clube_years</i>	-0,0285	0,018707 *	-0,0242	0,03237 *	-0,0262	0,01951 *
<i>Bournemouth</i>	-0,3580	3,04e-06 ***	-0,4154	8,85e-11 ***	-0,4144	1,09e-10 ***
<i>Brighton</i>	-0,3330	1,41e-09 ***	-0,3719	9,92e-16 ***	-0,3706	1,30e-15 ***
<i>Burnley</i>	-0,2325	3,13e-08 ***	-0,2519	1,50e-12 ***	-0,2510	1,97e-12 ***
<i>Cardiff</i>	-0,3208	< 2e-16 ***	-0,3418	< 2e-16 ***	-0,3436	< 2e-16 ***
<i>Chelsea</i>	0,0665	0,011441 *	0,0510	0,01782 *	0,0464	0,03101 *
<i>Crystal_Palace</i>	-0,0910	8,55e-05 ***	-0,1050	8,64e-08 ***	-0,1065	5,20e-08 ***
<i>Everton</i>	-0,0349	0,060190 ,	-0,0463	0,00324 **	-0,0452	0,00412 **
<i>Fulham</i>	-0,1015	6,83e-09 ***	-0,1163	3,63e-15 ***	-0,1174	2,52e-15 ***

<i>Huddersfield</i>	-0,1430	< 2e-16 ***	-0,1571	< 2e-16 ***	-0,1588	< 2e-16 ***
<i>Leicester</i>	-0,0478	0,000693 ***	-0,0522	7,46e-06 ***	-0,0528	5,98e-06 ***
<i>Liverpool</i>	0,0318	0,080788 ,	0,0322	0,00283 **	0,0344	0,00140 **
<i>Manchester_City</i>	0,0476	0,000142 ***	0,0424	8,60e-06 ***	0,0416	1,38e-05 ***
<i>Manchester_United</i>	0,0128	0,217479	-	-	-	-
<i>Newcastle_United</i>	-0,0576	2,66e-08 ***	-0,0606	1,31e-12 ***	-0,0607	1,18e-12 ***
<i>Southampton</i>	-0,0395	5,09e-05 ***	-0,0427	1,28e-07 ***	-0,0432	9,27e-08 ***
<i>Tottenham</i>	0,0190	0,028454 *	0,0160	0,02552 *	0,0165	0,02217 *
<i>Watford</i>	-0,0486	6,05e-08 ***	-0,0552	4,32e-13 ***	-0,0551	5,26e-13 ***
<i>West_Ham</i>	-0,0265	0,001092 **	-0,0309	3,69e-06 ***	-0,0304	5,47e-06 ***
<i>Wolverhampton_Wanderers</i>	-0,0323	7,80e-05 ***	-0,0374	1,43e-07 ***	-0,0380	8,93e-08 ***
<i>Lateral</i>	-0,1829	0,036249 *	-0,1650	0,02710 *	-0,1682	0,02497 *
<i>Meia_Defensivo</i>	0,1949	0,055782 ,	0,1865	0,05579 ,	0,1944	0,04720 *
<i>Meia_Ofensivo</i>	0,3197	0,000660 ***	0,3157	4,03e-05 ***	0,3153	4,24e-05 ***
<i>Extremo</i>	0,4889	3,85e-05 ***	0,4449	9,71e-07 ***	0,4495	8,54e-07 ***
<i>Centroavante</i>	0,6044	2,69e-06 ***	0,5297	2,38e-07 ***	0,5388	1,69e-07 ***
<i>Canhoto</i>	0,0595	0,305905	-	-	-	-
<i>Ambidestro</i>	0,1083	0,377635	-	-	-	-
<i>Nac_africa</i>	-0,0785	0,225527	-	-	-	-
<i>Nac_america_sul</i>	-0,0447	0,607159	-	-	-	-
<i>Nac_america_norte</i>	-0,0958	0,657276	-	-	-	-
<i>Nac_america_central</i>	0,0422	0,683405	-	-	-	-
<i>Nac_asia</i>	0,0349	0,836413	-	-	-	-
<i>Nac_oceania</i>	0,2938	0,379363	-	-	-	-
Residual standard error	0,4486 on 339 DF		0,4459 on 355 DF		0,4483 on 357 DF	
Multiple R-squared	0,8262		0,8202		0,8173	
Adjusted R-squared	0,8022		0,8045		0,8024	

Nota: Códigos para a significância: p-valor < 0,001 '***'; 0,001 < p-valor < 0,01 '**'; 0,01 < p-valor < 0,05 '*'; 0,05 < p-valor < 0,1 '.'.

Visando encontrar o modelo mais apropriado para estimar resultados, o primeiro passo consiste em aplicar o método *Stepwise* utilizando o critério de Akaike (AIC), conforme apresentado no capítulo anterior.

4.1.1 A APLICAÇÃO DO MÉTODO STEPWISE POR AIC E OS RESULTADOS

O método consiste em retirar e aplicar regressores até o momento em que seja encontrado o modelo com menor valor de AIC. O modelo resultante terá, portanto, a máxima entropia, sendo este o que mais se aproxima do “modelo verdadeiro”, que não é viável de ser encontrado, segundo o critério de Akaike (1974).

Nessa linha, os resultados obtidos através do *software R-Studio* que mostram quais são os regressores que melhor explicam a variável dependente *Val_mercado*, bem como o modelo encontrado, estão apresentados abaixo:

$$\begin{aligned} \text{Log}(\text{Val_mercado}) = & 4,3693 - 0,1014.\text{Idade} + 0,8729.\text{Titular} + 0,0003.\text{Minutos_jogados} + \\ & 0,7551.\text{Gols_Jogos_Selecao} - 0,2250.\text{copas} + 1,3254.\text{Premios_idade} + 0,0227.\text{Instagram_followers} - \\ & 0,0242.\text{Temp_clube_years} - 0,4154.\text{Bournemouth} - 0,3719.\text{Brighton} - 0,2519.\text{Burnley} - 0,3418.\text{Cardiff} + \\ & 0,0510.\text{Chelsea} - 0,1050.\text{Crystal_Palace} - 0,0463.\text{Everton} - 0,1163.\text{Fulham} - 0,1571.\text{Huddersfield} - \\ & 0,0522.\text{Leicester} + 0,0322.\text{Liverpool} + 0,0424.\text{Manchester_City} - 0,0606.\text{Newcastle_United} - \\ & 0,0427.\text{Southampton} + 0,0160.\text{Tottenham} - 0,0552.\text{Watford} - 0,0309.\text{West_Ham} - \\ & 0,0374.\text{Wolverhampton_Wanderers} - 0,1650.\text{Lateral} + 0,1865.\text{Meia_Defensivo} + 0,3157.\text{Meia_Ofensivo} + \\ & 0,4449.\text{Extremo} + 0,5297.\text{Centroavante} + \varepsilon \end{aligned}$$

De acordo com o método *Stepwise*, com base na amostra utilizada, este é o modelo mais apropriado para estimar o valor de mercado dos atletas que tenham as mesmas características requeridas pela seleção da amostra.

O modelo *log-linear* permite interpretar o coeficiente de cada um dos regressores como variações percentuais sobre a variável dependente, dadas as demais variáveis constantes. Em geral, os resultados encontrados mostram-se de acordo com o esperado. A exceção se dá pela relação negativa entre jogadores que conquistaram a copa do mundo e o seu valor de mercado. A variável *Dummy copas* indica que atletas que conquistaram o torneio tem o valor seu valor de mercado reduzido em 22,50%. Outro ponto a destacar é o da exclusão da variável *Nota_Media_jogo* do modelo, sendo esta a principal variável de medida de desempenho dos atletas na liga.

Também foram descartadas, a partir do método *Stepwise*, as seguintes variáveis:

- ***Jogos_Selecao***, número de partidas disputadas pela seleção nacional;
- ***Titulos_idade***, número de títulos com relação à idade do atleta;

- ***Melhor***, número de vezes eleito o melhor em campo, com relação ao número de aparições na liga;
- ***Gols_Assist_90minutos***, soma entre gols e assistências para cada 90 minutos em campo;
- ***Altura***, altura do atleta;
- ***Dummys de Nacionalidade***, dadas por Europa, América do Sul, América do Norte, América Central, Ásia, África e Oceania. Europa é o padrão;
- ***Dummys de Perna preferencial***, dadas por Canhoto, Destro e Ambidestro. Ser Destro é o padrão;
- ***Dummy champions***, atletas que conquistaram a UEFA Champions League. Não conquistar o título é o padrão;
- ***Dummy para o clube Manchester_United***, atletas do clube Manchester United. Ser do Arsenal é o padrão.

Dentre os modelos apresentados, este também é o que apresenta o maior valor para o R^2 ajustado, sugerindo que 80,45% das variações na variável dependente podem ser explicadas pelos regressores do modelo.

No entanto, é necessário ajustar o modelo acima para que seja possível inferir, a partir da amostra, o comportamento da população. Para isso, devem ser retiradas também as variáveis que não se mostraram significativas para o modelo.

4.1.2 O MODELO COM A RETIRADA DAS VARIÁVEIS NÃO SIGNIFICATIVAS

Seguindo o critério padrão de 5% de nível de significância para determinar a permanência ou não das variáveis no modelo, e conforme apresentado na Tabela 10, as variáveis *copas*, *Premios_idade* e *Meia_Defensivo* não permitem a rejeição da hipótese nula de que não há relação com a variável dependente. Sendo assim, a exclusão se dará uma a uma, por ordem de maior *p-valor*, a partir do método *Backward*. Após a retirada das variáveis *copas* e *Premios_idade*, a variável *Meia_Defensivo* passou a rejeitar a hipótese nula e, portanto, permanecerá no modelo.

Sendo assim, o melhor modelo encontrado para inferência é dado por:

$$\begin{aligned} \text{Log}(\text{Val_mercado}) = & 4,3724 - 0,1015.\text{Idade} + 0,8767.\text{Titular} + 0,0003.\text{Minutos_jogados} + \\ & 0,8267.\text{Gols_Jogos_Selecao} + 0,0222.\text{Instagram_followers} - 0,0262.\text{Temp_clube_years} - 0,4144.\text{Bournemouth} \\ & - 0,3706.\text{Brighton} - 0,2510.\text{Burnley} - 0,3436.\text{Cardiff} + 0,0464.\text{Chelsea} - 0,1065.\text{Crystal_Palace} - \\ & 0,0452.\text{Everton} - 0,1174.\text{Fulham} - 0,1588.\text{Huddersfield} - 0,0528.\text{Leicester} + 0,0344.\text{Liverpool} + \\ & 0,0416.\text{Manchester_City} - 0,0607.\text{Newcastle_United} - 0,0432.\text{Southampton} + 0,0165.\text{Tottenham} - \\ & 0,0551.\text{Watford} - 0,0304.\text{West_Ham} - 0,0380.\text{Wolverhampton_Wanderers} - 0,1682.\text{Lateral} + \\ & 0,1944.\text{Meia_Defensivo} + 0,3153.\text{Meia_Ofensivo} + 0,4495.\text{Extremo} + 0,5388.\text{Centroavante} + \varepsilon \end{aligned}$$

Dessa forma, após realizados os testes para medir a qualidade deste modelo - que serão apresentados no próximo item do capítulo -, será possível inferir sobre determinados comportamentos populacionais a partir dos coeficientes apresentados acima.

4.1.3 ANÁLISE DO R² AJUSTADO E DO TESTE F

Uma das formas de avaliar a qualidade do ajuste do modelo é através do coeficiente de determinação (R^2). Em geral referimo-nos ao R^2 como a variação da variável dependente que pode ser explicada pelos regressores do modelo. O R^2 ajustado, por sua vez, penaliza a inclusão de variáveis independentes no modelo e é o mais apropriado para analisar uma Regressão Linear Múltipla. Como mostrado na Tabela 10, o R^2 ajustado é dado por 0,8024. Isso implica que 80,24% da variabilidade do valor de mercado dos atletas em análise pode ser explicada pelas variáveis independentes do modelo.

Comparativamente aos demais trabalhos que foram abordados na literatura, o resultado é significativamente alto, já que Majewski (2016) e Geurts (2016) encontraram, respectivamente, valores para o R^2 ajustado de 56,11% e 45,97% para seus modelos. Wicker et. al. (2013) ao tentar medir os impactos do esforço físico do atleta no valor de mercado, encontrou um R^2 de aproximadamente 20%. Considerando-se o número de variáveis explicativas, o coeficiente de determinação mais alto apresentado pelo modelo pode estar relacionado ao tamanho da amostra utilizada nesse estudo.

Quanto ao teste F , este serve para verificar se o modelo é ou não globalmente significativo. Para isso, ele indica se a hipótese nula, dada por,

$$H_0: \beta_1 = \beta_2 = \dots = \beta_n = 0,$$

deve ou não ser rejeitada. Se H_0 for rejeitada, temos que pelo menos uma das variáveis do modelo é razoável para explicar a variável dependente. Para este modelo, temos que $F = 55,06$ e $p\text{-valor} (F) < 2.2e-16$. Sendo assim, como o $p\text{-valor} (F)$ é menor do que o nível de significância de 5%, concluímos que o modelo é globalmente significativo.

4.1.4 TESTE RESET DE RAMSEY PARA MÁ ESPECIFICAÇÃO DA FORMA FUNCIONAL

O Teste RESET de Ramsey é o mais conhecido para verificar erros de especificações em modelos. Tais erros podem ocorrer por omissão de variável relevante, inclusão de variável irrelevante, adoção de forma funcional inadequada ou erros de medida. O teste RESET adiciona polinômios na equação para detectar má especificação de formas funcionais. Em geral, são verificados possíveis erros de especificação para termos quadráticos e cúbicos. De acordo com Christiaan Heij, et. al. (2004), entretanto é possível selecionar termos com ordem mais elevada. Para o modelo em análise, foram testados possíveis erros de especificação para termos quadráticos e cúbicos. O resultado, obtido através do *software R-studio*, está apresentado abaixo:

Tabela 12: Teste RESET de Ramsey.

TESTE RESET				
Ordem	F	df1	df2	p-valor
Termos quadráticos	1,46	1	356	0,2270
Termos quadráticos e cúbicos	0,85	2	355	0,4287

Elaboração própria.

O Teste RESET apresentou o *p-valor* superior ao nível de significância de 5%, portanto não permite rejeitar a hipótese nula de que o modelo está corretamente especificado. Dessa forma, para o modelo em análise, é possível aceitar a hipótese nula.

4.1.5 TESTE PARA IDENTIFICAÇÃO DE MULTICOLINEARIDADE

A *multicolinearidade* é um problema comumente notado em regressões múltiplas, nas quais as variáveis independentes possuem correlações lineares entre si. Um bom indício para caracterização deste problema ocorre quando, ao mesmo tempo em que o modelo apresenta um R^2 elevado, os coeficientes da regressão não são estatisticamente significativos (segundo o *teste individual t*). As consequências deste variam desde erros-padrão elevados, no caso de *multicolinearidade* moderada ou severa e, até mesmo, à impossibilidade de qualquer estimação se a *multicolinearidade* for perfeita.

O nível de *multicolinearidade* pode ser indicado através do Fator de Inflação da Variância (FIV). É ele que determinará o grau de *multicolinearidade*. Se $FIV = 1$, então teremos uma ausência de *multicolinearidade*. Para valores superiores a 10, o problema é caracterizado. Os resultados encontrados para o modelo em análise são apresentados na tabela abaixo:

Tabela 13: Fator de Inflação da Variância

TESTE PARA IDENTIFICAÇÃO DE MULTICOLINEARIDADE	
Variável	VIF
<i>Idade</i>	1,3514
<i>Titular</i>	2,9664
<i>Minutos_jogados</i>	2,3165
<i>Gols_Jogos_Selecao</i>	1,5508
<i>Instagram_followers</i>	1,5179
<i>Temp_clube_years</i>	1,3708
<i>Bournemouth</i>	1,5340
<i>Brighton</i>	1,5056
<i>Burnley</i>	1,4465
<i>Cardiff</i>	1,5182
<i>Chelsea</i>	1,4139
<i>Crystal_Palace</i>	1,4521
<i>Everton</i>	1,4797
<i>Fulham</i>	1,5368
<i>Huddersfield</i>	1,5777
<i>Leicester</i>	1,5061
<i>Liverpool</i>	1,4001
<i>Manchester_City</i>	1,4213
<i>Newcastle_United</i>	1,5759
<i>Southampton</i>	1,5145
<i>Tottenham</i>	1,4674
<i>Watford</i>	1,4936
<i>West_Ham</i>	1,5442
<i>Wolverhampton_Wanderers</i>	1,4782
<i>Lateral</i>	1,6794
<i>Meia_Defensivo</i>	1,3919
<i>Meia_Ofensivo</i>	2,0336
<i>Extremo</i>	2,0294
<i>Centroavante</i>	2,3933

Elaboração própria.

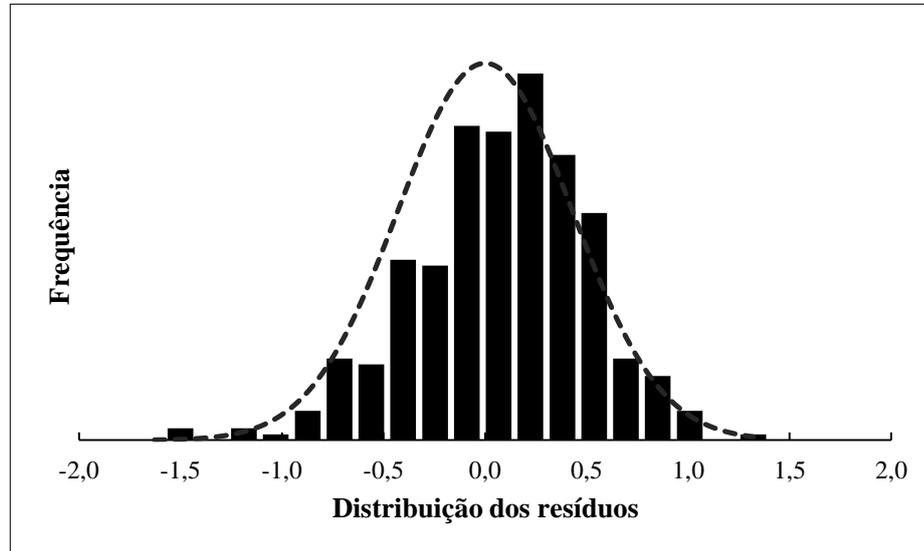
Interpretando os resultados, podemos dizer que não foram detectados problemas de *multicolinearidade* no modelo. Isso porque os VIF's encontrados se aproximam significativamente de 1 e se distanciam de 10.

4.1.6 TESTE DE NORMALIDADE DOS RESÍDUOS

A normalidade dos resíduos é um dos pressupostos da análise de regressão para os testes de hipótese sobre a significância dos coeficientes estimados. Há duas formas prevaletentes para verificar a normalidade dos resíduos. A primeira consiste em análises gráficas e a segunda

utiliza testes estatísticos. Em um primeiro momento, a fim de analisar graficamente, serão apresentados subsequentemente os resultados do gráfico de *Histograma dos resíduos* e do gráfico *Q-Q*.

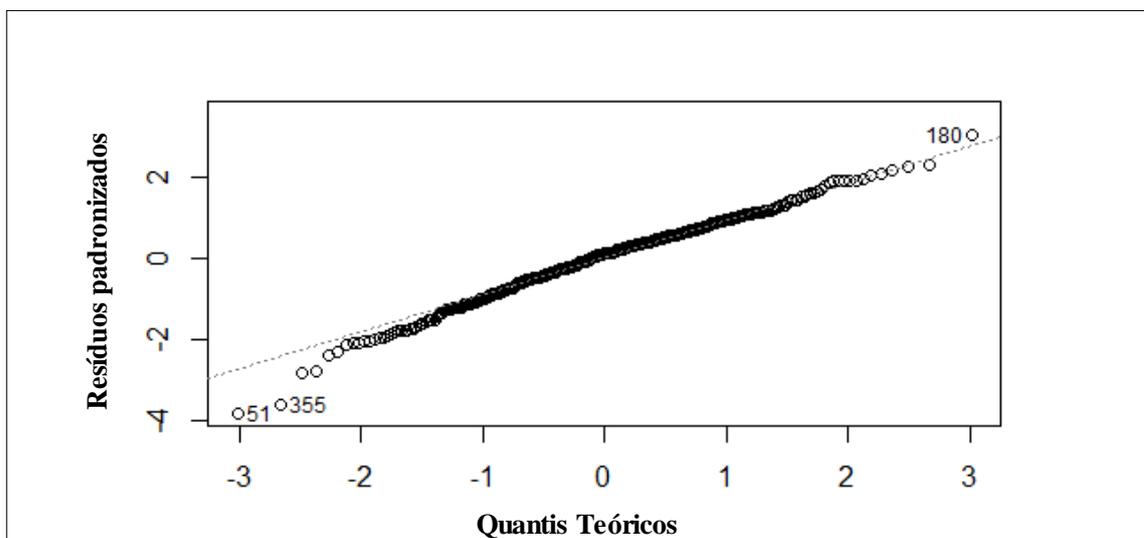
Gráfico 1: Histograma dos resíduos.



O gráfico de histograma apresenta em seu eixo horizontal os valores dos resíduos divididos em intervalos e no eixo vertical a frequência dessas classes. Podemos concluir que os resíduos parecem seguir uma distribuição normal ao observarmos as seguintes características do gráfico: simétrico, em forma de sino e em torno do zero.

Outra ferramenta gráfica utilizada para verificar a normalidade dos resíduos é o *gráfico Q-Q*:

Gráfico 2: Gráfico Q-Q.



Para o gráfico acima, o eixo vertical apresenta os valores dos resíduos do modelo e o horizontal os Quantis Teóricos. A reta pontilhada apresenta os valores esperados para os resíduos caso eles sejam normalmente distribuídos. A aparência da reta neste gráfico também sugere a normalidade. Entretanto, há de se destacar a presença de possíveis *outliers* nos primeiros *Quantis* da reta (observações 51 e 355).

A análise gráfica pode ser considerada uma ferramenta subjetiva, embora seja relevantemente informativa. Sendo assim, outra maneira para identificar se os resíduos possuem uma distribuição normal é através de testes. Neste caso, será utilizado o teste *qui-quadrado de Pearson*, cuja hipótese nula denota que os resíduos possuem distribuição normal. O resultado apresenta-se abaixo:

Tabela 14: Teste qui-quadrado de Pearson

TESTE QUI-QUADRADO DE PEARSON	
Estatística de teste	Resultado
Valor do teste qui-quadrado de Pearson	20.199
p-valor	0,3827

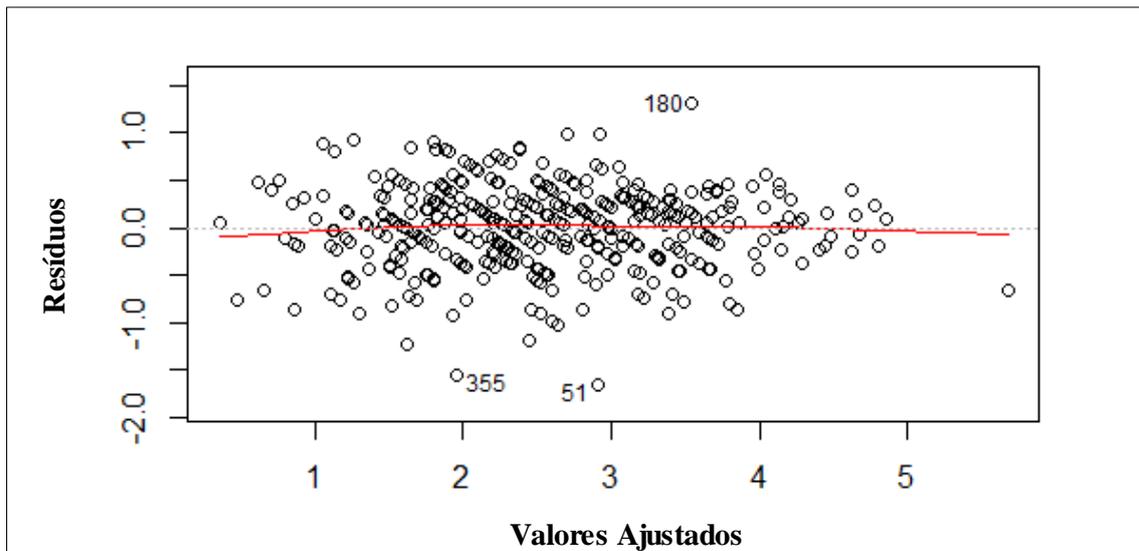
Elaboração própria.

Portanto, através do teste, podemos concluir que o erro possui distribuição normal, uma vez que a hipótese nula não pode ser rejeitada.

4.1.7 TESTE PARA HOMOCEASTICIDADE

O problema da heterocedasticidade é identificado em modelos que têm variância não-constante nos resíduos. Quando detectado, o método de Mínimos Quadrados não pode ser aplicado, pois os intervalos de confiança e testes de hipóteses apresentarão resultados inexatos. Além disso, parâmetros que parecem ser estatisticamente insignificantes podem ser significantes. Sendo assim, quando a premissa de homocedasticidade é violada, a inferência estatística do modelo deve ser invalidada.

A hipótese de homocedasticidade é provavelmente violada se os resíduos aumentam ou diminuem com os valores ajustados. Como por exemplo, se no gráfico dos Resíduos *versus* Valores Ajustados os pontos formam uma curva ao redor de zero e não estão dispostos aleatoriamente. Outra característica é que poucos pontos no gráfico ficam muito distantes dos demais. Ao analisarmos o gráfico abaixo, é possível sugerir que o modelo apresenta variância constante dos erros:

Gráfico 3: Resíduos versus Valores Ajustados.

A presença de heterocedasticidade é testada através do Teste de Goldfeld-Quandt. Para este teste, há uma exigência de que a amostra seja relativamente grande, o que se mostra aceitável no modelo em análise. A hipótese nula sugere que há homocedasticidade no modelo. Para que não seja rejeitada esta hipótese o p-valor do teste deve ser maior do que o nível de significância escolhido, de 5%. Os resultados são apresentados na tabela abaixo:

Tabela 15: Teste de Goldfeld-Quandt.

TESTE DE GOLDFELD-QUANDT	
Estatística de teste	Resultado
Valor do teste GQ	0.7902
DF1	164
DF2	163
p-valor	0,9334

Elaboração própria.

Como observado, o teste conclui que não há heterocedasticidade no modelo.

4.1.8 OBSERVAÇÕES NÃO USUAIS

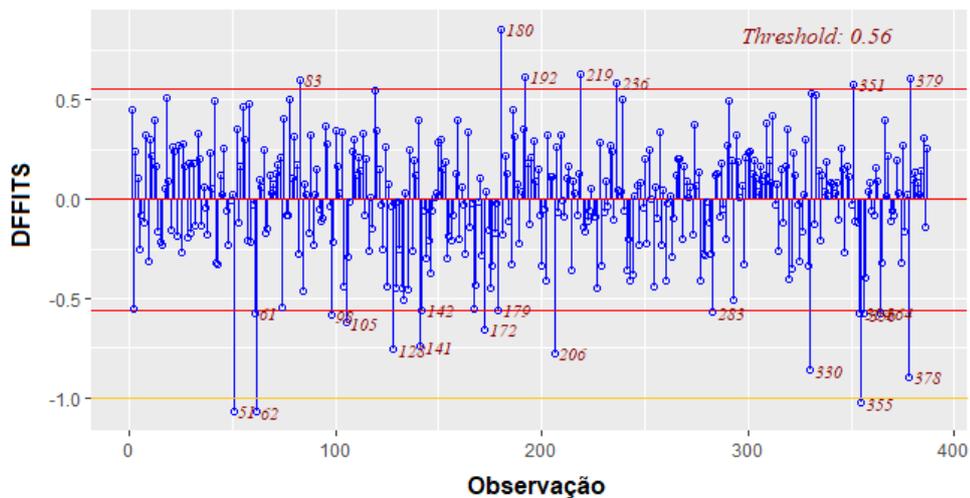
Em análises de regressão, um ponto é considerado influente quando sua retirada da amostra causa mudanças substanciais nos resultados do modelo. Para o estudo em questão, foi selecionado o critério de DFFITS para checagem de possíveis *outliers*.

O critério de análise de DFFITS mede o quanto a exclusão da observação i impacta os valores ajustados da regressão. Diz-se que uma observação é um ponto influente, segundo a medida DFFITS, se:

- a) $DFFITs > 1$ para amostras pequenas ou médias;
 b) $DFFITs > \frac{2 \cdot (p+1)}{\sqrt{(n-p-1)}}$ para amostras grandes.

Para este estudo, onde $n = 387$, a amostra será considerada de tamanho médio e, portanto, apenas observações com $DFFITs > 1$ serão retiradas do modelo. O gráfico abaixo apresenta os resultados do teste:

Gráfico 4: Análise de DFFITS, n = 387



Sendo assim, foram retiradas do modelo as observações cujo DFFITS foi superior ao limite estabelecido. A exclusão se deu de forma individual, seguindo sempre a ordem do maior valor de DFFITS, até o momento em que não foram mais registradas observações com DFFITS superior a 1. Como resultado, sete atletas foram excluídos da análise:

Tabela 16: Observações não usuais

Nome	Clube	Posição	Valor de Mercado (em € mi)	Idade
Mohamed Salah	Liverpool	Extremo Direito	150,0	26,9
Paul Pogba	Manchester United	Médio Centro	100,0	26,2
David Luiz	Chelsea	Defesa Central	20,0	32,1
Mesut Özil	Arsenal	Médio Ofensivo	25,0	30,6
Fernando Llorente	Tottenham	Ponta de Lança	3,5	34,2
Kyle Walker-Peters	Tottenham	Lateral Esquerdo	3,5	22,1
Jack Simpson	Bournemouth	Defesa Central	1,5	22,4

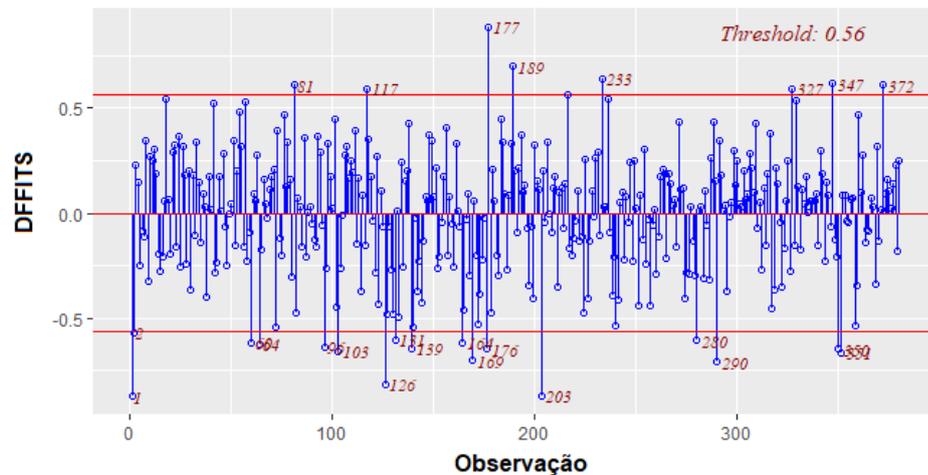
Elaboração própria.

De qualquer forma, é possível observar que estes são atletas de suma importância para a amostra, uma vez que têm notória importância no futebol e, principalmente, na liga estudada. A exclusão destes jogadores pode prejudicar a análise, ocultando informações essenciais e

fatídicas. O futebol, por si só, é reconhecido por recompensar, com exorbitantes cifras, jogadores que nem sempre devolvem o investimento ao clube, o que corrobora a numerosa quantidade de *outliers* que podem ser observados em estudos desse tipo.

Os DFFITS obtidos com a nova amostra, com $n = 380$, estão apresentados abaixo:

Gráfico 5: Análise de DFFITS, n = 380



Assim, o modelo encontrado após a exclusão dos pontos influentes da amostra é dado por:

$$\begin{aligned} \text{Log}(\text{Val_mercado}) = & 4,4449 - 0,0933.\text{Idade} + 0,7611.\text{Titular} + 0,0003.\text{Minutos_jogados} + \\ & 0,7329.\text{Gols_Jogos_Selecao} + 0,0599.\text{Instagram_followers} - 0,0403.\text{Temp_clube_years} - 0,3340.\text{Bournemouth} \\ & - 0,3489.\text{Brighton} - 0,2273.\text{Burnley} - 0,3283.\text{Cardiff} + 0,0421.\text{Chelsea} - 0,0956.\text{Crystal_Palace} - \\ & 0,0392.\text{Everton} - 0,1108.\text{Fulham} - 0,1523.\text{Huddersfield} - 0,0459.\text{Leicester} + 0,0346.\text{Liverpool} + \\ & 0,0384.\text{Manchester_City} - 0,0567.\text{Newcastle_United} - 0,0386.\text{Southampton} + 0,0276.\text{Tottenham} - \\ & 0,0518.\text{Watford} - 0,0271.\text{West_Ham} - 0,0342.\text{Wolverhampton_Wanderers} - 0,2306.\text{Lateral} + \\ & 0,2129.\text{Meia_Ofensivo} + 0,3031.\text{Extremo} + 0,3990.\text{Centroavante} + \varepsilon \end{aligned}$$

O ajuste nos pontos influentes impactou a perda de significância da variável *Meia_Defensivo*, que foi retirada da regressão. Além disso, modificou significativamente os coeficientes das demais *Dummies* de posição. A variável *Centroavante*, por exemplo, que indicava um coeficiente de 0,5388 antes do ajuste, passou a registrar 0,3990. A variável *Instagram_followers* também sofreu impacto significativo, como é possível verificar na Tabela 17:

Tabela 17: Comparação entre os modelos.

Variável	Modelo 3 (n = 387)		Modelo 4 (n = 380)	
	Coefficiente	P-valor	Coefficiente	P-valor
<i>Intercept</i>	4,3724	< 2e-16 ***	4,4449	< 2e-16 ***
<i>Idade</i>	-0,1015	< 2e-16 ***	-0,0933	< 2e-16 ***
<i>Altura</i>	-	-	-	-
<i>Titular</i>	0,8767	8,72e-07 ***	0,7611	5.10e-06 ***
<i>Gols_assist_90minutos</i>	-	-	-	-
<i>Minutos_jogados</i>	0,0003	< 2e-16 ***	0,0003	< 2e-16 ***
<i>Melhor</i>	-	-	-	-
<i>Nota_Media_jogo</i>	-	-	-	-
<i>Jogos_Selecao</i>	-	-	-	-
<i>Gols_Jogos_Selecao</i>	0,8267	1,33e-05 ***	0,7329	5.28e-05 ***
<i>Titulos_idade</i>	-	-	-	-
<i>copas</i>	-	-	-	-
<i>champions</i>	-	-	-	-
<i>Premios_idade</i>	-	-	-	-
<i>Instagram_followers</i>	0,0222	0,00385 **	0,0599	2.09e-06 ***
<i>Temp_clube_years</i>	-0,0262	0,01951 *	-0,0403	0.000169 ***
<i>Bournemouth</i>	-0,4144	1,09e-10 ***	-0,3340	6.11e-08 ***
<i>Brighton</i>	-0,3706	1,30e-15 ***	-0,3489	2.17e-15 ***
<i>Burnley</i>	-0,2510	1,97e-12 ***	-0,2273	2.01e-11 ***
<i>Cardiff</i>	-0,3436	< 2e-16 ***	-0,3283	< 2e-16 ***
<i>Chelsea</i>	0,0464	0,03101 *	0,0421	0.043661 *
<i>Crystal_Palace</i>	-0,1065	5,20e-08 ***	-0,0956	2.43e-07 ***
<i>Everton</i>	-0,0452	0,00412 **	-0,0392	0.008396 **
<i>Fulham</i>	-0,1174	2,52e-15 ***	-0,1108	3.08e-15 ***
<i>Huddersfield</i>	-0,1588	< 2e-16 ***	-0,1523	< 2e-16 ***
<i>Leicester</i>	-0,0528	5,98e-06 ***	-0,0459	3.24e-05 ***
<i>Liverpool</i>	0,0344	0,00140 **	0,0346	0.000820 ***
<i>Manchester_City</i>	0,0416	1,38e-05 ***	0,0384	2.29e-05 ***
<i>Manchester_United</i>	-	-	-	-
<i>Newcastle_United</i>	-0,0607	1,18e-12 ***	-0,0567	2.88e-12 ***
<i>Southampton</i>	-0,0432	9,27e-08 ***	-0,0386	5.03e-07 ***
<i>Tottenham</i>	0,0165	0,02217 *	0,0276	9.10e-05 ***
<i>Watford</i>	-0,0551	5,26e-13 ***	-0,0518	8.07e-13 ***
<i>West_Ham</i>	-0,0304	5,47e-06 ***	-0,0271	1.87e-05 ***
<i>Wolverhampton_Wanderers</i>	-0,0380	8,93e-08 ***	-0,0342	3.38e-07 ***
<i>Lateral</i>	-0,1682	0,02497 *	-0,2306	0.000455 ***
<i>Meia_Defensivo</i>	0,1944	0,04720 *	-	-
<i>Meia_Ofensivo</i>	0,3153	4,24e-05 ***	0,2129	0.001330 **
<i>Extremo</i>	0,4495	8,54e-07 ***	0,3031	0.000166 ***
<i>Centroavante</i>	0,5388	1,69e-07 ***	0,3990	1.48e-05 ***
<i>Canhoto</i>	-	-	-	-
<i>Ambidestro</i>	-	-	-	-

<i>Nac_africa</i>	-	-	-	-
<i>Nac_america_sul</i>	-	-	-	-
<i>Nac_america_norte</i>	-	-	-	-
<i>Nac_america_central</i>	-	-	-	-
<i>Nac_asia</i>	-	-	-	-
<i>Nac_oceania</i>	-	-	-	-
Residual standard error	0,4483 on 357 DF		0,42 on 351 DF	
Multiple R-squared	0,8173		0,8344	
Adjusted R-squared	0,8024		0,8211	

Nota: Códigos para a significância: p -valor < 0,001 ***; $0,001 < p$ -valor < 0,01 **; $0,01 < p$ -valor < 0,05 *; $0,05 < p$ -valor < 0,1 .?.

Como observado, o R^2 ajustado foi superior no novo modelo. No que se refere aos testes de regressão, todos mantiveram resultados estáveis, conforme apresentado na tabela abaixo.

Tabela 18: Modelos e resultados, n = 380.

	Teste	Valor	P-valor
R² Ajustado		0,8231	-
Teste F		63,14	< 2.2e-16***
Reset de Ramsey (má especificação)			
Termos quadráticos e cúbicos		2,679	0,0700
Teste Qui-quadrado de Pearson (normalidade)		25,611	0,1414
Teste de Goldfeld-Quandt (homocedasticidade)		0,7335	0,9750
FIV (multicolinearidade)		N/A variáveis com VIF > 10	

Elaboração própria.

A tabela abaixo detalha a interpretação adequada para cada um dos coeficientes do melhor modelo encontrado, dadas as demais variáveis constantes:

Tabela 19: Interpretação das variáveis independentes do Modelo.

Variável	Coefficiente	Interpretação
<i>Idade</i>	-0,0933	Para cada variação de um ano de idade, o valor de mercado do atleta é reduzido em 9,33%.
<i>Titular</i>	0,7611	Para cada variação de 1 p.p. na taxa de partidas jogadas como titular na liga, o valor de mercado do atleta é aumentado em 0,76%
<i>Minutos_jogados</i>	0,0003	Para cada minuto jogado a mais na liga, o valor de mercado do atleta é aumentado em 0,03%.
<i>Gols_Jogos_Selecao</i>	0,7329	Para cada variação de 1 p.p. na taxa de gols feitos por jogo disputado com a camisa da seleção do país, o valor de mercado do atleta é aumentado em 0,73%.
<i>Instagram_followers</i>	0,0599	Para cada 1 milhão de seguidores a mais no Instagram, o atleta tem seu valor de mercado aumentado em 5,99%.
<i>Temp_clube_years</i>	-0,0403	Para cada ano adicional que o jogador mantém seu contrato vigente com o clube, o seu valor de mercado é reduzido em 4,03%.
<i>Bournemouth</i>	-0,3340	O valor de mercado de atletas do Bournemouth se reduz em 33,40% quando comparado ao valor de mercado de atletas do Arsenal.
<i>Brighton</i>	-0,3489	O valor de mercado de atletas do Brighton se reduz em 34,89% quando comparado ao valor de mercado de atletas do Arsenal.
<i>Burnley</i>	-0,2273	O valor de mercado de atletas do Burnley se reduz em 22,73% quando comparado ao valor de mercado de atletas do Arsenal.
<i>Cardiff</i>	-0,3283	O valor de mercado de atletas do Cardiff se reduz em 32,83% quando comparado ao valor de mercado de atletas do Arsenal.
<i>Chelsea</i>	0,0421	O valor de mercado de atletas do Chelsea aumenta em 4,21% quando comparado ao valor de mercado de atletas do Arsenal.
<i>Crystal_Palace</i>	-0,0956	O valor de mercado de atletas do Crystal Palace se reduz em 9,56% quando comparado ao valor de mercado de atletas do Arsenal.
<i>Everton</i>	-0,0392	O valor de mercado de atletas do Everton se reduz em 3,92% quando comparado ao valor de mercado de atletas do Arsenal.
<i>Fulham</i>	-0,1108	O valor de mercado de atletas do Fulham se reduz em 11,08% quando comparado ao valor de mercado de atletas do Arsenal.
<i>Huddersfield</i>	-0,1523	O valor de mercado de atletas do Huddersfield se reduz em 15,23% quando comparado ao valor de mercado de atletas do Arsenal.
<i>Leicester</i>	-0,0459	O valor de mercado de atletas do Leicester se reduz em 4,59% quando comparado ao valor de mercado de atletas do Arsenal.
<i>Liverpool</i>	0,0346	O valor de mercado de atletas do Liverpool aumenta em 3,46% quando comparado ao valor de mercado de atletas do Arsenal.

<i>Manchester_City</i>	0,0384	O valor de mercado de atletas do Manchester City aumenta em 3,84% quando comparado ao valor de mercado de atletas do Arsenal.
<i>Newcastle_United</i>	-0,0567	O valor de mercado de atletas do Newcastle United se reduz em 5,67% quando comparado ao valor de mercado de atletas do Arsenal.
<i>Southampton</i>	-0,0386	O valor de mercado de atletas do Southampton se reduz em 3,86% quando comparado ao valor de mercado de atletas do Arsenal.
<i>Tottenham</i>	0,0276	O valor de mercado de atletas do Tottenham aumenta em 2,76% quando comparado ao valor de mercado de atletas do Arsenal.
<i>Watford</i>	-0,0518	O valor de mercado de atletas do Watford se reduz em 5,18% quando comparado ao valor de mercado de atletas do Arsenal.
<i>West_Ham</i>	-0,0271	O valor de mercado de atletas do West Ham se reduz em 2,71% quando comparado ao valor de mercado de atletas do Arsenal.
<i>Wolverhampton_Wanderers</i>	-0,0342	O valor de mercado de atletas do Wolverhampton Wanderers se reduz em 3,42% quando comparado ao valor de mercado de atletas do Arsenal.
<i>Lateral</i>	-0,2306	O valor de mercado de jogadores da posição Lateral é 23,06% inferior ao valor de mercado de atletas Zagueiros.
<i>Meia_Ofensivo</i>	0,2129	O valor de mercado de jogadores da posição Meia Ofensivo é 21,29% superior ao valor de mercado de atletas Zagueiros.
<i>Extremo</i>	0,3031	O valor de mercado de jogadores da posição Extremo é 30,31% superior ao valor de mercado de atletas Zagueiros.
<i>Centroavante</i>	0,3990	O valor de mercado de jogadores da posição Centroavante é 39,90% superior ao valor de mercado de atletas Zagueiros.

Fonte: Elaboração própria.

Há muitas semelhanças entre os resultados que foram encontrados no presente trabalho e o que foi abordado previamente na literatura. Além disso, muitos dos resultados expostos neste capítulo reforçam as teses que já eram esperadas quando o estudo foi proposto. É o caso da relação entre o aumento marginal da variável *Idade* e a perda de valor de mercado dos atletas, ou do maior valor de mercado daqueles atletas que jogam em posições mais ofensivas ou que permanecem por mais tempo como titulares. Outro efeito interessante que pode ser observado no estudo é o da relação positiva entre seguidores no *Instagram* e o ganho de valor no mercado de transferências. Tal hipótese sustenta as ideias previamente abordadas do “apelo de bilheteria” levantadas por Rosen (1981), do “efeito de superestrelas”, de Lucifora e Simmons (2003), e do “fenômeno da superestrela”, de Bali, R. (2012), já que há um ganho com a imagem do atleta que não necessariamente tem relação com o seu desempenho esportivo. Existem também alguns resultados inesperados, como é o caso da fraca relação entre as variáveis de desempenho e o valor de mercado dos atletas. No entanto, parte da explicação se dá pela limitação do estudo em analisar exclusivamente o desempenho dos atletas nas partidas da Liga Inglesa, diante da diversidade de competições que são disputadas em paralelo a esta em uma temporada e que não foram abordadas neste trabalho.

Quanto à hipótese de haver ou não qualquer tipo de discriminação étnica ou racial na determinação do valor de mercado dos jogadores da liga, os resultados se assemelham ao que concluiu Frick (2007). As variáveis *Dummy* que indicavam o continente de origem dos atletas não se mostraram significativas no modelo, demonstrando que não existem evidências suficientes para apontar a presença de discriminação para jogadores nascidos em diferentes locais do planeta. No entanto, é importante que novos estudos sejam feitos para aprofundar esta hipótese, já que John Goddard e John O. S. Wilson (2009) encontraram, em seus resultados, uma maior dificuldade para jogadores negros chegarem ao futebol profissional do que jogadores brancos com as mesmas ou semelhantes características.

5 CONSIDERAÇÕES FINAIS

O futebol é definitivamente um esporte de expressiva importância para a sociedade brasileira. Não à toa, o Brasil é conhecido como “o país do futebol”. Este trabalho buscou identificar quais são as variáveis que mais impactam o valor de mercado de um atleta profissional de futebol da liga inglesa, bem como quais são as relações existentes entre estas, a partir de métodos estatísticos e econômicos. Com isso, foi possível fornecer novas ferramentas de análise e, principalmente, trazer novos incentivos ao debate da economia do esporte no âmbito científico nacional, em decorrência de uma necessidade urgente de profissionalização das análises esportivas no país.

Muitos dos resultados encontrados neste estudo assemelham-se aos que foram abordados na revisão da literatura. Da mesma forma, grande parte dos resultados confirmaram o que já era esperado quando da proposição do trabalho. É o caso da relação negativa entre a variável *Idade* e o valor de mercado, indicando que quanto mais próximo do final da carreira um atleta está, menos valioso ele passa a ser, já que a possibilidade de o revender futuramente e recuperar parte do dinheiro investido se reduz. No mesmo sentido, o coeficiente negativo da variável *Temp_clube_years* também reforça a hipótese de relação inversa entre tempo restante de carreira e valor de mercado. Outro ponto interessante confirmado pelo modelo é a relação positiva entre a variável *Instagram_followers* e a variável dependente, indicando que, as demais variáveis constantes, quanto maior o número de seguidores na rede social *Instagram*, isto é, quanto maior for o ganho com imagem do atleta, maior é o seu valor de mercado, corroborando as hipóteses do “apelo de bilheteria” levantadas por Rosen (1981), do “efeito de superestrelas”, de Lucifora e Simmons (2003), e do “fenômeno da superestrela”, de Bali, R. (2012).

Dentre os regressores de desempenho, o único que se mostrou relevante para o modelo é o que indica o número de gols marcados para cada partida realizada defendendo a seleção nacional, dada por *Gols_Jogos_Selecao*. Como não era de se esperar, boa parte das variáveis de desempenho não se mostraram significativas. É o caso do número de gols e assistências para cada 90 minutos jogados (*Gols_assist_90minutos*), número de vezes eleito o melhor em campo proporcional ao número de aparições (*Melhor*), nota média nas partidas do campeonato (*Nota_Media_jogo*), número de partidas jogadas pela seleção que defende historicamente (*Jogos_Selecao*), número de títulos conquistados pela idade (*Titulos_idade*), número de vezes que foi campeão da Copa do Mundo da FIFA (*copas*), número de vezes que foi campeão da UEFA Champions League (*champions*) e número de prêmios individuais conquistados pela

idade (*Premios_idade*). Imaginava-se que o desempenho do atleta na liga, cuja variável que melhor descreve é a *Nota_Media_jogo* impactasse positivamente o valor de mercado. Entretanto, por essa variável se restringir a abordar o desempenho do atleta apenas na Barclays Premier League e excluir os demais jogos da temporada (jogos pela seleção e outros campeonatos), é possível que por isso a relação não tenha se mostrado significativa. Já os regressores *Titular* e *Minutos_jogados* mostraram-se altamente significativos e impactantes para o modelo, reforçando a hipótese de que os jogadores mais valiosos são os que mais vezes são escalados nas partidas.

Os resultados encontrados para as variáveis *Dummy* dos clubes da elite inglesa indicam que apenas Chelsea, Manchester City, Liverpool e Tottenham acrescentam (significativamente) maior valor aos atletas quando comparados aos atletas do Arsenal. Também se esperava uma relação significativa para os atletas do Manchester United, mas o nível de significância não foi suficiente para que essa variável integrasse o modelo final. Para as variáveis *Dummy* de posição, o resultado denota que, quanto mais ofensiva for a posição do atleta, maior será o seu valor de mercado, o que também já era esperado. No entanto, é de se destacar que o coeficiente da posição Lateral indica que estes têm menor valor de mercado do que atletas da posição Zagueiro. O modelo também não encontra relação significativa para a perna preferencial do atleta. Esperava-se, entretanto, que atletas ambidestros pudessem ser mais valiosos.

No que se refere ao continente de origem do atleta, utilizado para testar a hipótese de discriminação racial ou étnica na precificação dos atletas, as variáveis *Dummy* não se mostraram significativas. Dessa forma, assim como Frick (2007), é possível concluir que não existem evidências suficientes neste trabalho para apontar a presença de discriminação, esta diretamente relacionada com o continente de origem, nos valores de mercado dos atletas analisados. Entretanto, novos estudos são necessários para aprofundar esta hipótese, já que John Goddard e John O. S. Wilson (2009) apontaram uma maior dificuldade para jogadores negros chegarem ao futebol profissional do que jogadores brancos com as mesmas ou semelhantes características.

Dentre as dificuldades encontradas nesse estudo, está a necessidade de analisar o desempenho dos atletas em todos os jogos da temporada. Restringir o desempenho de um atleta aos jogos realizados em apenas uma competição pode camuflar o real desempenho deste, já que o valor de mercado não é definido exclusivamente pelos jogos realizados na Barclays Premier League.

Ampliar o banco de dados para outras ligas europeias, sul-americanas e de todos os continentes onde o futebol é jogado profissionalmente certamente agregaria uma maior diversidade aos dados e resultaria em uma amostra mais completa. Variáveis econômicas e financeiras dos clubes compradores e vendedores também são interessantes pontos a serem analisados, já que, como levantado por Carmichael e Thomas (1993), é possível haver relação entre o grau de riqueza de um clube e o valor de mercado do atleta no processo de barganha.

Propõe-se, com este trabalho, uma ampliação do debate sobre o futebol no país, levando os conceitos e as metodologias mais utilizadas na ciência econômica, e na economia do esporte em particular, aos clubes e atletas. Assim, é possível fazer com que estes agentes consigam tomar decisões cada vez mais racionais em meio a uma profissionalização iminente das relações neste meio. É preciso, entretanto, uma disponibilização maior e mais assertiva de dados dos campeonatos nacionais para que o debate possa agregar maior valor científico ao esporte do país.

REFERÊNCIAS

- Bali, R. (2012). **“An Analysis of the Link between Football Player Wages and the Degree of Transfer Market Competition between Football Clubs”**.
https://www.academia.edu/7745547/Analysis_Football_Transfer_Market_Competition
- Carmichael, Fiona & Thomas, Dennis. (1993). **“Bargaining in the Transfer Market: Theory and Evidence”**, Applied Economics. 25. 1467-76.
<https://doi.org/10.1080/00036849300000150>
- Frick, B. (2007). **“The Football Players’ Labor Market: Empirical Evidence from the Major European Leagues”**. Scottish Journal of Political Economy, 54: 422-446.
<https://doi:10.1111/j.1467-9485.2007.00423.x>
- Geurts, J. (2016). **“Football Players’ Transfer Price Determination Based on Performance in the Big Five European Leagues”** <http://hdl.handle.net/10362/18646>
- Kahn, L. (2000). **“The Sports Business as a Labor Market Laboratory”**. Journal of Economic Perspectives, 14, issue 3, p. 75-94. http://unionstats.gsu.edu/8220/Kahn_2000.pdf
- Lucifora, C., & Simmons, R. (2003). **“Superstar Effects in Sport: Evidence From Italian Soccer”**. Journal of Sports Economics, 4(1), 35–55.
<https://doi.org/10.1177/1527002502239657>
- Majewski S. (2016). **"Identification of Factors Determining Market Value of the Most Valuable Football Players"**. Journal of Management and Business Administration. Central Europe, Sciendo, vol. 24(3), pages 91-104.
<https://ideas.repec.org/a/vrs/jmbace/v24y2016i3p91-104n5.html>
- Nash, J. F. (1950). **“The Bargaining Problem”**. Econometrica, 21, 128-40.
- Rosen, Sherwin. (1981). **“The Economics of Superstars”**. The American Economic Review, vol. 71, no. 5, 1981, pp. 845–858. JSTOR, www.jstor.org/stable/1803469.
- Rottenberg, Simon. (1956). **“The Baseball Players' Labor Market”**. Journal of Political Economy, vol. 64, no. 3, 1956, pp. 242–258. JSTOR, <http://www.jstor.org/stable/1825886>
- Tunaru, R., & Viney, H. (2010). **“Valuations of Soccer Players from Statistical Performance Data”**. Journal of Quantitative Analysis in Sports, 6(2).
<https://doi.org/10.2202/1559-0410.1238>
- Preston, Ian & Szymanski, Stefan. (2000). **“Racial Discrimination in English Football”**. Scottish Journal of Political Economy. <https://doi.org/10.1111/1467-9485.00168>
- Marques, Jader & Faria da Silva, Maurício. (2013). **“O Direito e a Copa do Mundo de Futebol”**. Livraria do Advogado Editora.

Wicker, P., Prinz, J., Weimar, D., Deutcher, Ch. and Upmann, T. (2013). **“No Pain, No Gain? Effort and Productivity in Professional Soccer”**. International Journal of Sport Finance, 8.2 (May 2013): 124–139.

Montgomery, D. C.; Peck, E. A. e Vining, G. G. (2001). **“Introduction to Linear Regression Analysis”**.

Bozdogan, H. (2000). **“Akaike's Information Criterion and Recent Developments in Information Complexity”**.

<https://doi.org/10.1006/jmps.1999>

Akaike, H. (1974). **“A new look at the statistical model identification. IEEE Transactions on Automatic Control”**. Boston, v.19, n.6, p.716-723.

Emiliano, P.C. (2009). **“Fundamentos e aplicações dos critérios de informação: Akaike e Bayesiano”**. Universidade Federal de Lavras. Dissertação de mestrado.

http://repositorio.ufla.br/jspui/bitstream/1/3636/1/DISSERTA%C3%87%C3%83O_Fundamentos%20e%20Aplica%C3%A7%C3%B5es%20dos%20Crit%C3%A9rios%20de%20Informa%C3%A7%C3%A3o%20Akaike%20e%20Bayesiano.pdf

APÊNDICE A – ROTINA UTILIZADA NO SOFTWARE R STUDIO.

```
# PACOTE PARA LEITURA DE DADOS
```

```
require("readxl")
```

```
require("MASS")
```

```
require("lmtest")
```

```
require("het.test")
```

```
require("olsrr")
```

```
require("faraway")
```

```
require("tseries")
```

```
require("nortest")
```

```
require("vars")
```

```
require("outliers")
```

```
#####
```

```
setwd("C:/Users/henri/OneDrive/Área de Trabalho/Henrique - UFRGS/TCC")
```

```
dados <- read_excel("BANCO DE DADOS.xlsx")
```

```
dados <- dados[!(dados$Val_mercado==0),]
```

```
# TRANSFORMA COPA DO MUNDO EM DUMMY, PARECE FAZER MAIS SENTIDO
```

```
dados$copas = 1
```

```
for (i in 1:nrow(dados)){
```

```
  if (dados$Copa_mundo[i] == 0){
```

```
    dados$copas[i] = 0
```

```
  }
```

```
}
```

```
# TRANSFORMA CHAMPIONS EM DUMMY, PARECE FAZER MAIS SENTIDO
```

```
dados$champions = 1
```

```
for (i in 1:nrow(dados)){
```

```
  if (dados$Champions_League[i] == 0){
```

```
    dados$champions[i] = 0
```

```
  }
```

```
}  
reg = lm(log(dados$Val_mercado) ~  
dados$Idade  
+ dados$Altura  
+ dados$Titular  
+ dados$Gols_assist_90minutos  
+ dados$Minutos_jogados  
# + dados$assist_minuto  
# + dados$gols_minuto  
#+ dados$Cart_Amar_Minuto  
#+ dados$Cart_Verm_Minuto  
#+ dados$Chutes_jogo  
#+ dados$Perc_Acert_Pass  
#+ dados$Duelos_Aereos_Jogo  
+ dados$Melhor  
#+ dados$Carrinhos_jogo  
#+ dados$Interceptac_jogo  
#+ dados$Faltas_comet_jogo  
#+ dados$Impedim_ganho_jogo  
#+ dados$Bolas_afast_jogo  
#+ dados$Dribles_sofr_jogo  
#+ dados$Chutes_bloquead_jogo  
#+ dados$Gols_c_minuto  
#+ dados$Passes_chave_jogo  
#+ dados$Passes_jogo  
#+ dados$Cruzament_jogo  
#+ dados$Bolas_longas_jogo  
#+ dados$Lancamento_jogo  
#+ dados$Faltas_sofr_jogo
```

```
#+ dados$Vezes_impedido_jogo
#+ dados$Desarmado_jogo
#+ dados$Mal_controlebola_jogo
+ dados$Nota_Media_jogo
+ dados$Jogos_Selecao
+ dados$Gols_Jogos_Selecao
+ dados$Titulos_idade
+ dados$copas #adicionada
+ dados$champions #adicionada
+ dados$Premios_idade #adicionada
+ dados$Instagram_followers
+ dados$Temp_clube_years

#DUMMIES

#CLUBE, ARSENAL É O PADRÃO
+ dados$`Bournemouth(2)`
+ dados$`Brighton(3)`
+ dados$`Burnley(4)`
+ dados$`Cardiff(5)`
+ dados$`Chelsea(6)`
+ dados$`Crystal_Palace(7)`
+ dados$`Everton(8)`
+ dados$`Fulham(9)`
+ dados$`Huddersfield(10)`
+ dados$`Leicester(11)`
+ dados$`Liverpool(12)`
+ dados$`Manchester_City(13)`
+ dados$`Manchester_United(14)`
+ dados$`Newcastle_United(15)`
+ dados$`Southampton(16)`
```

```

+ dados$`Tottenham(17)`
+ dados$`Watford(18)`
+ dados$`West_Ham(19)`
+ dados$`Wolverhampton_Wanderers(20)`
#POSIÇÃO, ZAGUEIRO É O PADRÃO
+ dados$`Lateral(2)`
+ dados$`Meia_Defensivo(3)`
+ dados$`Meia_Ofensivo(4)`
+ dados$`Extremo(5)`
+ dados$`Centroavante(6)`
#PERNA, DESTRO É O PADRÃO
+ dados$`Canhoto(2)`
+ dados$`Ambidestro(3)`
#NACIONALIDADE, EUROPEU É O PADRÃO
+ dados$`Nac_africa(2)`
+ dados$`Nac_america_sul(3)`
+ dados$`Nac_america_norte(4)`
+ dados$`Nac_america_central(5)`
+ dados$`Nac_asia(6)`
+ dados$`Nac_oceania(7)`
)
summary(reg)
step(reg, direction = "both")
reg = lm(log(dados$Val_mercado) ~
dados$Idade +
dados$Titular +
dados$Minutos_jogados +
dados$Gols_Jogos_Selecao +
#dados$copas +

```

```

#dados$Premios_idade +
dados$Instagram_followers +
dados$Temp_clube_years +
dados$`Bournemouth(2)` +
dados$`Brighton(3)` +
dados$`Burnley(4)` +
dados$`Cardiff(5)` +
dados$`Chelsea(6)` +
dados$`Crystal_Palace(7)` +
dados$`Everton(8)` +
dados$`Fulham(9)` +
dados$`Huddersfield(10)` +
dados$`Leicester(11)` +
dados$`Liverpool(12)` +
dados$`Manchester_City(13)` +
dados$`Newcastle_United(15)` +
dados$`Southampton(16)` +
dados$`Tottenham(17)` +
dados$`Watford(18)` +
dados$`West_Ham(19)` +
dados$`Wolverhampton_Wanderers(20)` +
dados$`Lateral(2)` +
#dados$`Meia_Defensivo(3)` +
dados$`Meia_Ofensivo(4)` +
dados$`Extremo(5)` +
dados$`Centroavante(6)`
)
summary(reg)

```

```
#VARIANCIA
plot(reg, which = 1)
gqtest(reg)
ols_test_score(reg)
ols_test_f(reg)
ols_test_breusch_pagan(reg)
bptest(reg)

#NORMALIDADE
hist(reg$residuals)
ols_plot_resid_hist(reg)
#lillie.test(reg$residuals)
#jarque.bera.test(reg$residuals)
#cvm.test(reg$residuals)
#ad.test(reg$residuals)
#sf.test(reg$residuals)
#shapiro.test(reg$residuals)
pearson.test(reg$residuals)
#ols_test_normality(reg)
plot(reg, which = 2)
ols_plot_resid_qq(reg)

#TESTE RESET
resettest(reg,power=2,type="fitted",data=dados)

#MULTICOLINEARIDADE
vif(reg)

#OUTLIERS
boxplot(reg$residuals)
ols_plot_cooksd_bar(reg)
ols_plot_cooksd_chart(reg)
```

```
#ols_plot_dfbetas(reg)
ols_plot_dffits(reg)
ols_plot_resid_stand(reg)
ols_plot_resid_lev(reg)
ols_plot_resid_stud(reg)
ols_plot_resid_stud_fit(reg)
ols_plot_hadi(reg)
ols_plot_resid_pot(reg)
#DEMAIS GRAFICOS
plot(reg, which = 3)
plot(reg, which = 4)
plot(reg, which = 5)
plot(reg, which = 6)
```