



**Simpósio de Métodos
Numéricos em Engenharia**

25 a 27 de outubro, 2017

Análise Estatística para Avaliar o Desempenho de Atletas de Futebol Profissional

Cauê Barros Guimarães

Jair Mendes Marques

Anselmo Chaves Neto

Departamento de Métodos Numéricos em Engenharia

Universidade Federal do Paraná

Curitiba, Brasil

Resumo—O futebol profissional teve seu início em 1885 na Inglaterra com a criação das primeiras ligas oficiais. Desde então ocorreram muitas mudanças e evoluções, nos aspectos físicos, táticos e técnicos. Todavia, ainda existe muita subjetividade no meio, principalmente quando o objetivo é mensurar o desempenho de um atleta. Atualmente, nas contratações de atletas, alguns dos grandes clubes de futebol da Europa baseiam suas decisões em estudos estatísticos descritivos. Este trabalho busca aprimorar as técnicas estatísticas utilizadas para avaliar o desempenho desses atletas. Assim, este trabalho faz uma proposta de uma metodologia para classificar os atletas de futebol profissional baseada no desempenho deles. Para isso, serão utilizadas técnicas multivariadas de análise de componentes principais e análise fatorial. O estudo foi conduzido em cinco ligas, sendo as quatro principais da Europa (Alemanha, Inglaterra, Espanha e Itália) e a liga de futebol do Brasil, a partir de 91 atributos mensurados por jogador. Foram estudados os dados dos três últimos anos de cada liga. Além disso, a amostra compreendeu todos os jogadores que atuaram em pelo menos 1.000 minutos em um campeonato e que jogam na posição de meio atacante pela esquerda. O principal resultado encontrado foi a geração de uma nota para cada jogador.

Palavras-chave—futebol. Análise de desempenho. Jogador. Análise multivariada.;

I. INTRODUÇÃO

Em 2002, o diretor Billy Beane, de uma equipe pequena de *baseball*, com um dos menores orçamentos da liga,

revolucionou o mundo do desporto ao utilizar a análise estatística como base para a contratação de jogadores. Esse gestor prescindiu das habituais avaliações dos seus observadores e desafiou décadas de conhecimento sobre o jogo, em favor de milhares de números que descreviam as ações de cada jogador. Nesse ano, o time gerido por este homem chegou à final da liga com uma enorme possibilidade de conquistar o título, além de quebrar os recordes de vitórias que duravam por mais de 90 anos [10].

Beane colocou em prática os volumes mimeografados escritos por Bill James na década de 1970. James que havia começado a estudar as estatísticas de *baseball* em seu tempo livre, percebeu que algumas delas eram extremamente importantes, mas raramente mencionadas e desenvolveu toda uma base teórica sobre o assunto [8].

Fazendo um paralelo entre estatística no *baseball* para o mundo do futebol, Anderson e Sally [1] fazem uma análise interessante sobre o jogo nos dias atuais:

“Durante muito tempo, quatro palavras dominaram o futebol: Sempre foi feito assim”. Ainda, segundo Anderson e Sally [1]: “Esses homens não querem que lhes digam que há mais de um século eles estão deixando de perceber alguns fatos. Que existe um conhecimento que eles não possuem. Que o jeito como eles sempre fizeram as coisas não é como as coisas devem ser feitas”. E Anderson e Sally [1] concluem: “no cerne dessa mudança estão os números. [...] São os números que vão nos permitir ver o jogo como nunca o vimos antes.”

Tal como no *baseball*, durante décadas os observadores de jogadores de futebol influenciaram as ações dos clubes no mercado, com base na análise subjetiva de desempenho dos atletas. Em resumo, os “olheiros” - como são chamados esses observadores - observavam dezenas ou centenas de jogadores e recomendavam os que consideravam ter qualidade para serem contratados. Hoje, esse método é um risco que os clubes já não podem correr. A aquisição de um atleta implica um avultado investimento que não pode ser desperdiçado através de uma opinião, por mais reputação que tenha o avaliador. Existe, portanto, a necessidade de análises objetivas em relação à performance do jogador. A opinião já não é suficiente e a Estatística pode contribuir com suas muitas técnicas.

Atualmente, a maior parte dos grandes clubes da Europa utiliza a análise de performance de jogadores com base em estatísticas descritivas e pontuais para tomar decisões nas suas contratações. Esses clubes têm escritórios com analistas que, por meio de sistemas e ferramentas de vídeo, conseguem analisar todas as ações dos jogadores e transformá-las em números.

Desta forma, na tentativa de aprofundar e clarear o tema, sem pretender esgotá-lo, o presente estudo tem o objetivo de classificar os atletas de futebol de acordo com as suas performances, deixando menos subjetiva a contratação e prospecção por parte dos clubes.

II. REFERENCIAL TEÓRICO

A. O Esporte Futebol

O futebol profissional teve seu início em 1885 na Inglaterra [15] com a criação das primeiras ligas oficiais. Desde então, ocorreram muitas mudanças e evoluções, nos aspectos físicos, táticos e técnicos. Com tantas e significativas mudanças, o futebol deixou de ser uma prática empírica, realizada apenas recreacionalmente pela elite inglesa e adquiriu as grandes proporções sociais que hoje detém. Apenas para que se possa ilustrar o quanto a indústria do futebol vem gerando de movimento financeiro, nos últimos cinco anos, os valores movimentados pelos 10 maiores clubes de futebol foram na ordem de 4 bilhões e 659 milhões de euros [5]. O futebol, como uma prática de ação desportiva, é composto por um coletivo, que interagem em situações de oposição ou cooperação, com relações coerentes e consequentes, dispondo de metas e funções definidas [15]. Por tratar-se de um conjunto de ações, pode-se fragmentar as análises, visando melhoria na compreensão e na identificação de erros, bem como na performance dos atletas.

Assim, existe a necessidade do suporte científico para tal acompanhamento. Deve-se lembrar apenas que de acordo com Garganta [4] o futebol é um grupo de situações de mudanças com finais sempre em aberto, portanto, não é uma ciência exata. Além disso, é praticado por seres humanos e depende de fatores não controláveis, tais como intempéries naturais, gramados que nem sempre possuem as mesmas características, qualidade técnica (mesmo tratando-se de um esporte de alto rendimento existem diferenças técnicas entre os praticantes) ou condições psicológicas dos atletas, mas

com certeza a ciência pode agregar conhecimento à prática e melhorar o rendimento no futebol.

B. Aleatoriedade no Futebol

A aleatoriedade no futebol é um tema pouco pesquisado no meio acadêmico. Um dos primeiros estudos sobre isso foi desenvolvido por Reep e Benjamin [13]-[14] que tinham como principal objetivo responder se partidas de futebol e campeonatos são decididos por talento ou por sorte. Essa é uma questão importante sobre o assunto, pois se a resposta for de que é uma questão de talento, então existe uma lógica na competição e, portanto, é um evento passível de uma representação matemática onde os times e seus dirigentes possuem papel fundamental na obtenção de melhores resultados. Por outro lado, caso seja fruto da sorte, os jogadores, os técnicos e as torcidas possuirão somente um papel limitado dentro do espetáculo e o que acontece dentro do campo de jogo não é possível de ser controlado.

Reep e Benjamin [13]-[14] não responderam diretamente ao questionamento proposto por eles mesmos, mas encontraram alguns indícios, tais como: em média, cada gol ocorria após 10 chutes; 50% dos gols aconteciam após um passe ou nenhum (roubada de bola); 80% dos gols eram fruto de jogadas com três ou menos passes; recuperar a bola no último quarto do gramado representava 20% dos gols e, por fim, 50% dos gols aconteciam após falhas do adversário, em seu campo de defesa.

Outros estudos, mais recentes, foram de Loy [11]-[12] que se aprofundou sobre o fator da “aleatoriedade” no futebol e de que forma essa influenciava o resultado final de uma partida. De acordo com Loy [11]-[12], o gol em um jogo de futebol é muito mais influenciado por eventos aleatórios do que se pensava anteriormente - especialmente nas últimas décadas, quando o esporte se tornou mais rápido e intenso.

Loy [12] analisou 1.200 jogos da principal liga de futebol profissional da Alemanha (Bundesliga), por mais de três anos, e concluiu que 60% dos gols foram marcados como resultado da capacidade atlética ou técnica de uma equipe, enquanto os outros 40% foram marcados como um resultado do acaso. Ao estudar outras das principais ligas na Europa (Espanha, Inglaterra, Portugal, Holanda, Suíça e Bélgica), Loy [11] observou que o número de “gols ao acaso” havia subido para 46%.

Depois de avaliar uma série de jogos da Liga dos Campeões da Europa, Loy [11] apontou outro número interessante: do total de 12.405 toques na bola, que 35% foram estritamente coincidência, advindos de erros do adversário, de rebotes ou de reposições de bola em jogo. Loy concluiu que eliminar os erros não marcados eram o caminho para a vitória e quem o fizesse mais eficientemente teria garantido o caminho para títulos.

Anderson e Sally [1] visitaram casas de apostas, laboratórios e clubes de futebol. Examinaram milhares de partidas de campeonatos europeus e de copas, e concluíram que 50% dos gols eram devido ao acaso e à sorte e, os outros 50% eram devido ao talento ou à habilidade.

Lames [9] assistiu junto com sua equipe a vídeos de mais de 2.500 gols, buscando em cada um - durante a jogada inteira, desde a saída da bola até o gol -

circunstâncias de sorte. Encontrou que em 44,4% das vezes durante a jogada que resultou em gol, houve um fato do acaso que se não tivesse acontecido, não teria saído o gol. Esse percentual apresentou pequenas variações de um país para outro e de uma competição para a outra. Também concluiu que esses gols, frutos da sorte, eram mais comuns quando o placar foi zero a zero, pois ambos os clubes ainda estavam seguindo rigorosamente o perfil tático desenhado.

Nos quatro estudos, [11]-[12], [1] e [9] verificou-se que o percentual de gols resultante do aleatório foi em torno de 40% a 50%. Isso mostra que, por outro lado, 50% a 60% dos gols foram oriundos de talento, técnica e tática. Logo, se os clubes puderem melhorar seus processos de seleção de jogadores, existe uma possibilidade de que o controle do jogo, ou de que o evento mais importante do esporte – o gol – possa ser manipulado em até 60% das ocasiões.

C. Análise de Componentes Principais (ACP)

A Análise de Componentes Principais é um procedimento matemático que utiliza uma transformação ortogonal para converter um conjunto de observações de variáveis possivelmente correlacionadas em um conjunto de valores de variáveis linearmente não-correlacionadas chamadas “componentes principais”. O número de componentes principais é menor ou igual ao número de variáveis originais. Esta transformação é definida de forma que o primeiro componente principal tem a maior variância possível (ou seja, é responsável pelo máximo de variabilidade nos dados), e cada componente seguinte, por sua vez, tem a máxima variância sob a restrição de ser ortogonal a aos componentes anteriores, isto é, não-correlacionado com eles [7]. Na prática, os parâmetros $\underline{\mu}$ e Σ são desconhecidos e, portanto, devem ser estimados. Logo a estimativa de $\underline{\mu}$ é dada pelo vetor $\underline{\hat{x}}$ e a estimativa de Σ pela matriz S .

Para o caso de variáveis padronizadas de um vetor z , a matriz de covariância amostral será:

$$S_z = R = \begin{bmatrix} \mathbf{1} & \hat{p}_{12} & \dots & \hat{p}_{1p} \\ \hat{p}_{21} & 1 & \dots & \hat{p}_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \hat{p}_{p1} & \hat{p}_{p2} & \dots & 1 \end{bmatrix} \quad (1)$$

E a j -ésima componente principal amostral padronizada é dada por:

$$\hat{y}_j = \hat{e}_j' z = \hat{e}_{j1} z_1 + \hat{e}_{j2} z_2 + \dots + \hat{e}_{jp} z_p, \quad j=1, 2, \dots, p \quad (2)$$

onde: $(\hat{\lambda}_1, \hat{e}_1), (\hat{\lambda}_2, \hat{e}_2), \dots, (\hat{\lambda}_p, \hat{e}_p)$ são os autovalores e autovetores de R com $\hat{\lambda}_1 \geq \hat{\lambda}_2 \geq \dots \geq \hat{\lambda}_p \geq 0$. Além disso, tem-se que:

$$V(\hat{y}_j) = \hat{\lambda}_j, \quad j = 1, 2, \dots, p \quad (3)$$

$$e \quad \text{COV}(\hat{y}_i, \hat{y}_j) = 0, \quad \text{para } i \neq j \quad (4)$$

A variância total pode ser definida como:

$$\text{tr}(R) = p = \sum_{j=1}^p \hat{\lambda}_j = \hat{\lambda}_1 + \hat{\lambda}_2 + \dots + \hat{\lambda}_p \quad (5)$$

Já a proporção da variância total devida à j -ésima componente principal:

$$\frac{\hat{\lambda}_j}{p}, \quad j = 1, 2, \dots, p \quad (6)$$

Logo, o coeficiente de correlação amostral entre as componentes principais \hat{y}_i e as variáveis padronizadas z_i é:

$$r_{\hat{y}_i z_i} = \hat{e}_{ij} \sqrt{\hat{\lambda}_j}, \quad j = 1, 2, \dots, p \quad (7)$$

D. Análise Fatorial

Segundo [7] a Análise Fatorial é uma técnica da estatística destinada a representar um processo aleatório multivariado, por meio da criação de novas variáveis, decorrentes das variáveis originais e, geralmente, em menor número, que elas. Em outras palavras, é uma técnica que tem por objetivo explicar as correlações entre um conjunto de variáveis em termos de um número menor de fatores.

O vetor aleatório \underline{X} com p componentes, tem média $\underline{\mu}$ e matriz de covariância Σ . No modelo fatorial, \underline{X} é linearmente dependente sobre algumas variáveis aleatórias não observáveis F_1, F_2, \dots, F_m , chamadas de “fatores comuns” e p “fontes adicionais de variação” $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_p$, chamados “erros” ou “fatores específicos”. O modelo de análise fatorial é representado por:

$$\begin{aligned} X_1 - \mu_1 &= l_{11} F_1 + l_{12} F_2 + \dots + l_{1m} F_m + \varepsilon_1 \\ X_2 - \mu_2 &= l_{21} F_1 + l_{22} F_2 + \dots + l_{2m} F_m + \varepsilon_2 \\ &\vdots \\ X_p - \mu_p &= l_{p1} F_1 + l_{p2} F_2 + \dots + l_{pm} F_m + \varepsilon_p \end{aligned} \quad (8)$$

ou, na notação matricial,

$$\underline{X} - \underline{\mu} = \underline{L}\underline{F} + \underline{\varepsilon} \quad (9)$$

onde μ_i é a média da variável i ; ε_i representa o i -ésimo fator específico; F_j é o j -ésimo fator comum; e l_{ij} é o peso ou carregamento na i -ésima variável do j -ésimo fator. Assumindo que:

$$E(\underline{F}) = \mathbf{0}, \quad V(\underline{F}) = E(\underline{F}\underline{F}') = {}_m I_m \quad (10)$$

e

$$E(\underline{\varepsilon}) = \mathbf{0}_{(px1)}, \quad V(\underline{\varepsilon}) = E(\underline{\varepsilon}\underline{\varepsilon}') = {}_p \Psi_p = \begin{bmatrix} \Psi_1 & 0 & \dots & 0 \\ 0 & \Psi_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \Psi_p \end{bmatrix} \quad (11)$$

Além disso, \underline{F} e $\underline{\varepsilon}$ são independentes, ou seja, a covariância entre elas é zero. No modelo fatorial ortogonal a covariância tem a seguinte estrutura:

$$\Sigma = \underline{L}\underline{L}' + \Psi \quad (12)$$

e

$$\text{COV}(X_i, X_k) = l_{i1}l_{k1} + \dots + l_{im}l_{km} \quad (13)$$

sendo que

$$\text{COV}(X_i, F_j) = l_{ij} \quad (14)$$

Desta forma, são extraídas as communalidades h_i^2 – porção da variância da variável X_i que é distribuída pelos m fatores comuns – e as especificidades Ψ_i que são a porção da variância relativa aos fatores específicos.

Sejam os pares de autovalores e autovetores de S : $(\hat{\lambda}_1, \hat{e}_1), (\hat{\lambda}_2, \hat{e}_2), \dots, (\hat{\lambda}_p, \hat{e}_p)$ onde $\hat{\lambda}_1 \geq \hat{\lambda}_2 \geq \dots \geq \hat{\lambda}_p \geq 0$.

A matriz dos pesos ou carregamentos estimados dos fatores \hat{I}_{ij} é dada por

$$\hat{L} = \hat{C}\hat{D}_\lambda^{1/2} \quad (15)$$

onde

$$\hat{C} = \begin{bmatrix} \hat{e}_{11} & \hat{e}_{12} & \dots & \hat{e}_{1p} \\ \hat{e}_{21} & \hat{e}_{22} & \dots & \hat{e}_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \hat{e}_{p1} & \hat{e}_{p2} & \dots & \hat{e}_{pp} \end{bmatrix} \quad (16)$$

e

$$\hat{D}_\lambda^{1/2} = \begin{bmatrix} \sqrt{\hat{\lambda}_1} & 0 & \dots & 0 \\ 0 & \sqrt{\hat{\lambda}_2} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sqrt{\hat{\lambda}_p} \end{bmatrix} \quad (17)$$

As variâncias específicas estimadas são fornecidas pelos elementos diagonais da matriz $\hat{\Psi} = S - \hat{L}\hat{L}'$, assim:

$$\hat{\Psi} = \begin{bmatrix} \hat{\Psi}_1 & 0 & \dots & 0 \\ 0 & \hat{\Psi}_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \hat{\Psi}_p \end{bmatrix} \quad (18)$$

com

$$\hat{\Psi}_i = s_i^2 - \sum_{j=1}^m l_{ij}^2 \quad (19)$$

e as communalidades são estimadas como:

$$\hat{h}_i^2 = \hat{l}_{i1}^2 + \dots + \hat{l}_{im}^2 \quad (20)$$

Em algumas aplicações, se faz necessário estimar o valor de cada um dos fatores para uma observação individual. Estes valores dos fatores são conhecidos como “escores fatoriais”. Os escores fatoriais estimados para variáveis padronizadas é dado pela equação:

$$\hat{\underline{F}} = (\hat{L}\hat{L}')^{-1}\hat{L}'\underline{z} \quad (21)$$

e os escores fatoriais estimados para as variáveis originais:

$$\hat{\underline{F}} = (\underline{L}'\underline{\Psi}^{-1}\underline{L})^{-1}\underline{L}'\underline{\Psi}^{-1}(\underline{x} - \underline{\bar{x}}) \quad (22)$$

III. METODOLOGIA

A. População e Amostra

Para a análise de desempenho dos atletas foram utilizadas 134 variáveis, sendo 85 referentes à performance dos atletas nos campeonatos, quatro relativas ao perfil/características dos atletas e mais 45 criadas a partir das 85 variáveis originais. Foram analisados 140 times e 468 atletas, gerando mais de 1.000 observações de atletas.

As ligas selecionadas para análise foram: a inglesa, a espanhola, a italiana, a alemã e a brasileira. Em todos os casos, foi utilizada a principal liga de futebol do país. Para os campeonatos europeus, foram estudadas as temporadas de 2013/2014, 2014/2015 e 2015/2016, pois no referido continente, os torneios têm início em setembro de um ano e terminam em abril do ano posterior. Na liga brasileira, a análise ocorreu sobre as temporadas de 2014, 2015 e 2016.

Fizeram parte da amostra todos os jogadores que tiveram ao menos 1.000 minutos jogados no mesmo campeonato e que atuam na posição de meia-atacantes pelo lado esquerdo do campo.

B. Coleta e Tratamento dos Dados

Segundo Fachin [2], a coleta de dados preocupa-se com a observação de fatos colhidos no contexto natural, sem que haja intervenção do pesquisador, por meio de métodos e técnicas específicas.

Neste sentido, foi utilizado o banco de dados de uma empresa inglesa que disponibiliza em seu *site* um rol de estatísticas individuais e coletivas de 18 ligas de futebol profissional mundial. O *site* é denominado WhoScored.com e conta com informações detalhadas desde 2009. A coleta dos dados foi executada após o término de todos os torneios.

A escolha das quatro ligas europeias (Inglaterra, Alemanha, Espanha e Itália) ocorreu por serem os principais núcleos do futebol mundial. Essa conclusão pode ser constatada pela representatividade financeira e de mídia desses países. Os principais jogadores do mundo atuam ou estão migrando para lá. Além do mais, as pesquisas sobre o desempenho de jogadores e clubes no futebol são, em boa parte, de pesquisadores europeus. No caso da liga brasileira, o motivo de selecioná-la foi o fato de ser considerado um dos campeonatos mais equilibrados do mundo, uma vez que desde a última mudança no formato de disputa do torneio – em 2003 – houve seis diferentes clubes vencedores em 13 campeonatos nacionais disputados.

C. Procedimentos da Análise Fatorial

Alguns procedimentos da análise fatorial são importantes de serem mencionados. O primeiro é relativo aos testes de adequação. No estudo foi utilizado o teste de esfericidade de Bartlett, pois é um teste que avalia a presença de correlações entre as variáveis [6]. A outra medida de adequação utilizada, foi a estatística Kaiser-Meyer-Olkin (KMO) que fornece a proporção da variância considerada comum a todas as variáveis na amostra [3].

No estudo se optou por trabalhar com os fatores rotacionados. O método de rotação utilizado foi o EQUIMAX pelo motivo de ser um meio-termo entre os métodos QUARTIMAX e VARIMAX [6].

IV. ANÁLISE DE DADOS

No primeiro momento, munido de mais de 130 variáveis, sendo diversas referentes à aspectos ofensivos do jogo, outras relacionadas à forma defensiva, algumas ligadas à parte disciplinar e, por fim, variáveis cadastrais, a dificuldade foi estabelecer quais eram importantes para o jogador de futebol que atua como meia-atacante pelo lado esquerdo. Para contornar esta dificuldade, foi usada a técnica de regressão linear simples, considerando como variável dependente as notas médias, dadas por jornalistas e comentaristas esportivos, para os jogadores da amostra.

A partir dessa etapa inicial, foram selecionadas as variáveis com p-valor inferior a 5% através do processo *stepwise*. No modelo final restaram 22 variáveis, sendo um valor de R² de 0,7924 e com valor de R² ajustado de 0,7866. Além disso, o modelo foi significativo com um p-valor inferior à 0,01%.

Na fig. 1, abaixo, encontram-se as variáveis selecionadas de acordo com seu grau de explicação. As seis principais foram: o número de gols que cada atleta faz por hora (GH); em segundo, o número total de dribles por hora (DTH); em terceiro, o número de passes-chave por hora (KEYH); em quarto, o número de assistências por hora (AH); em quinto, o total de passes curtos por hora (TPHC); e, por fim, em sexto, a taxa de desarmes corretos (TX_D).

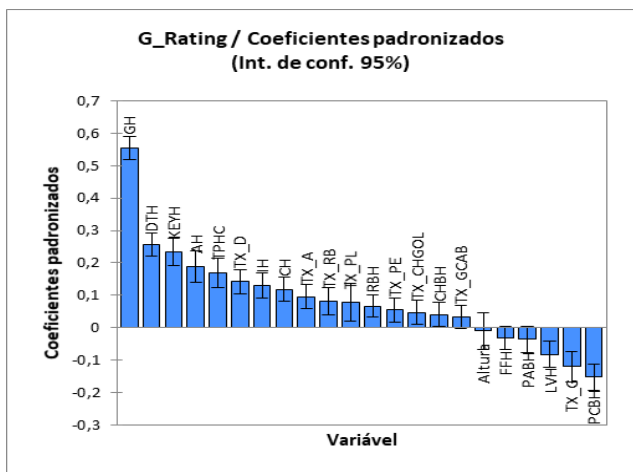


Figura 1: Coeficientes padronizados do modelo de regressão

Prosseguindo com a análise dos dados, visando encontrar um indicador sintético que descrevesse o comportamento dos jogadores, mas sem considerar o viés subjetivo dos jornalistas esportivos, foi utilizada a técnica de análise fatorial nos dados onde a estimação dos pesos fatoriais e das variâncias específicas ocorreu por meio da análise de componentes principais.

Para constatar a condição de aplicação da análise fatorial foram realizados os testes KMO e Bartlett. Os resultados estão divulgados na Tabela I. Como o p-valor foi inferior a 0,01% no teste de Bartlett e o valor do teste KMO foi maior que 0,5 é possível utilizar a análise fatorial.

TABELA I: TESTE DE KMO E BARTLETT

Teste de KMO e Bartlett		
Medida Kaiser-Meyer-Olkin de adequação de amostragem.		0,6788
Teste de esfericidade de Bartlett	Qui-quadrado aproximado	3857,448
	Graus de liberdade	190
	p-valor	0,0000

A Tabela II, a seguir, apresenta os resultados iniciais da análise fatorial. A partir dela, foi definido 12 como o número de fatores para este estudo, uma vez que o percentual da variância explicado foi de 82,76.

TABELA II: ANÁLISE FATORIAL

Análise Fatorial			
Fator	Autovalor	% da Variância	% Acumulado
1	3,19424	15,97	15,97
2	2,58414	12,92	28,89
3	1,81859	9,09	37,99
4	1,37094	6,86	44,84
5	1,29858	6,49	51,33
6	1,17665	5,88	57,22
7	1,01828	5,09	62,31
8	0,948837	4,74	67,05
9	0,869281	4,35	71,40
10	0,825006	4,13	75,52
11	0,744566	3,72	79,25
12	0,701848	3,51	82,76
13	0,650186	3,25	86,01
14	0,576812	2,88	88,89
15	0,533618	2,67	91,56
16	0,465817	2,33	93,89
17	0,413697	2,07	95,96
18	0,321507	1,61	97,56
19	0,300389	1,50	99,07

Na Tabela III, adiante, são apresentadas as variâncias específicas e as comunalidades após a rotação Equimax. Além disso, é possível observar que duas variáveis pré-selecionadas não se encontram na tabela, pois suas comunalidades foram inferiores a 0,5. Estas variáveis são o total de faltas cometidas por hora (FFH) e o percentual de chutes corretos na direção do gol (TX_CHGOL).

TABELA III: COMUNALIDADES APÓS ROTAÇÃO

Comunalidades Após Rotação		
Variáveis	Comunalidades estimada	Variância Específica
Altura	0,8949	0,1051
GH	0,8909	0,1091
KEYH	0,8009	0,1991
AH	0,7957	0,2043
DTH	0,9463	0,0537
TX_D	0,8547	0,1453
TX_A	0,7312	0,2688
TPHC	0,7744	0,2256
PCBH	0,7979	0,2021
IH	0,7159	0,2841
CH	0,7411	0,2589
TX_RB	0,9360	0,0640
TX_PL	0,7091	0,2909
TX_G	0,8829	0,1171
TX_PE	0,8740	0,1260
TX_GCAB	0,9786	0,0214
LVH	0,8289	0,1711
PABH	0,8119	0,1881
RBH	0,7374	0,2626
CHBH	0,8483	0,1517

Após a aplicação dos escores fatoriais com o percentual da variância explicada de cada fator foi possível elaborar um indicador sintético que explicasse a classificação dos jogadores de acordo com seu desempenho em cada temporada. Para finalizar o indicador, este foi colocado em uma escala de zero a dez, onde dez representa o melhor valor possível, ou seja, o melhor jogador e zero o seu pior. O indicador pode ser observado na Tabela IV, através da coluna “Nota”. Uma observação importante é relativa aos dois melhores jogadores do mundo, eleitos pela FIFA, nos últimos anos, Lionel Messi e Cristiano Ronaldo. O jogador Lionel Messi não aparece nos resultados, pois não atua como meia-atacante pelo esquerdo do campo de jogo, e sim, como atacante ou meia-atacante pelo lado direito. Já Cristiano Ronaldo não consta na tabela a seguir para as temporadas de 2013/2014 e 2015/2016. Neste caso, o mesmo encontra-se em 40° e 63°, respectivamente para cada temporada.

TABELA IV: CLASSIFICAÇÃO DOS JOGADORES

Classificação dos Jogadores				
Ordem	Nome	Equipe	Temporada	Nota
1	Franck Ribéry	Bayern Munich	2013/2014	8,88
2	Gareth Bale	Real Madrid	2015/2016	8,66
3	Arjen Robben	Bayern Munich	2014/2015	8,58
4	James Rodríguez	Real Madrid	2014/2015	8,34
5	Cristiano Ronaldo	Real Madrid	2014/2015	8,29
6	Marco Reus	Borussia Dortmund	2013/2014	7,99
7	James Rodríguez	Real Madrid	2015/2016	7,70
8	Isco	Real Madrid	2013/2014	7,67
9	Henrikh Mkhitaryan	Borussia Dortmund	2015/2016	7,65
10	Sejad Salihovic	Hoffenheim	2013/2014	7,62
11	Gareth Bale	Real Madrid	2013/2014	7,62
12	Kevin De Bruyne	Wolfsburg	2014/2015	7,54
13	David Silva	Manchester City	2013/2014	7,51
14	Neymar	Barcelona	2015/2016	7,36
15	Roberto Firmino	Hoffenheim	2013/2014	7,31
16	Lucas Vázquez	Real Madrid	2015/2016	7,30
17	David Silva	Manchester City	2014/2015	7,28
18	Riyad Mahrez	Leicester	2015/2016	7,25
19	Neymar	Barcelona	2014/2015	7,23
20	James Milner	Manchester City	2014/2015	7,21

V. CONSIDERAÇÕES FINAIS

A análise de atletas de futebol profissional por meio de estatística multivariada se mostrou eficiente. Foi possível classificar os jogadores de acordo com sua produtividade nos campeonatos que disputaram e os resultados são próximos ao de que especialistas no assunto divulgam na mídia. Para estudos subsequentes sugere-se a utilização de mais variáveis, mais campeonatos e também a análise de atletas de outras posições. Uma outra variável que pode agregar aos resultados é o preço de cada atleta.

Uma limitação do estudo é a falta de informações psicológicas relacionadas aos atletas, uma vez que a pressão aos quais estes jogadores são submetidos é enorme e pode influenciar no desempenho dos mesmos.

AGRADECIMENTOS

À Fundação de Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) pelo apoio financeiro dado ao primeiro autor, por meio da bolsa de doutorado.

REFERÊNCIAS

- [1] C. Anderson and D. Sally, “The numbers game: why everything you know about soccer is wrong,” Schwarzcz, 2013.
- [2] O. Fachin, “Fundamentos de metodologia,” São Paulo. Editora Saraiva, 2004.
- [3] L. P. Fávero and P. Belfiore, “Análise de dados: técnicas multivariadas exploratórias”. 1ª edição. Elsevier, 2015.
- [4] J. Garganta, “A análise da performance nos jogos desportivos: revisão acerca da análise do jogo,” Revista Portuguesa de Ciências do Desporto, 1(1): 57-64, 2001.
- [5] M. Gonçalves, “Negócio futebol: onde a grana está. Blog dos manos,” disponível em: <<http://www.blogdosmanos.com/negocio-futebol-onde-a-grana-esta/>>. Acesso em: 27 julho 2013.
- [6] J. F. Hair, W. C. Black, B. J. Babin, R. E. Anderson and R. L. Tatham, “Análise multivariada de dados,” 6ª edição. Bookman, 2009.
- [7] R. A. Johnson and D. W. Wichern, “Applied multivariate statistical analysis,” 6th ed. UpperSaddle River, New Jersey: Prentice-Hall, 2007.
- [8] S. Kuper and S. Szymanski, “Soccernomics: por que a Inglaterra perde, a Alemanha e o Brasil ganham, e os Estados Unidos, o Japão, a Austrália, a Turquia – e até mesmo o Iraque – podem se tornar os reis do esporte mais popular do mundo,” Tinta Negra Bazar Editorial, 2010.
- [9] M. Lames, “Glücksspiel Fußball – Zufallseinflüsse beim Zustandekommen von Toren,” In B. Halberschmidt & B. Strauß (Hrsg.), Elf Freunde sollt ihr sein!? 38. Jahrestagung der asp. Hamburg, Abstracts, p. 92, 2006.
- [10] M. Lewis, “Moneyball: the art of winning an unfair game,” New York. W. W. Norton & Company, 2003.
- [11] R. Loy, “Das lexikon der fußballirrtümer. Über die flügel zum erfolg,” Verlag C. Bertelsmann, 2008.
- [12] R. Loy, “Fußball - taktik und analyse,” (2). Hamburg: Czwalina, 2006.
- [13] C. Reep and B. Benjamin, “Skill and chance in association football,” Journal of the Royal Statistical Society, 131, 581-585, 1968.
- [14] C. Reep, R. Pollard and B. Benjamin, “Skill and chance in ball games,” Journal of the Royal Statistical Society, 134, 623-629, 1971.
- [15] S. Viera and A. Freitas, “O que é futebol? - histórias, regras, curiosidades - coleção o que é?,” Casa da Palavra, 2006.

APÊNDICE

TABELA V: DESCRIÇÃO DAS VARIÁVEIS UTILIZADAS NO TRABALHO

Variável	Descrição
Altura	altura em centímetros do atleta
GH	número de gols por hora marcados pelo atleta
KEYH	número de passes-chave por hora efetuados pelo atleta
AH	número de assistências por hora efetuadas pelo atleta
DTH	número de vezes que o atleta tentou driblar o adversário por hora
TX_D	taxa de dribles bem sucedidos pelo atleta
TX_A	taxa de duelos aéreos vencidos pelo atleta
TPHC	total de passes curtos por hora efetuados pelo atleta
PCBH	número de vezes que o atleta não dominou a bola e a perdeu por hora
IH	número de interceptações por hora que o atleta efetuou
CH	número de chutes para longe (chutões) por hora efetuados pelo atleta
TX_RB	taxa de roubadas de bola bem sucedidas executadas pelo atleta
TX_PL	taxa de passes longos corretos efetuados pelo atleta
TX_G	taxa de gols por hora marcados pelo atleta
TX_PE	taxa de escanteios corretos efetuados pelo atleta
TX_GCAB	taxa de gols com a cabeça por hora marcados pelo atleta
LVH	número de cartões vermelhos por hora que o atleta levou
PABH	número de passes que o atleta conseguiu bloquear por hora
RBH	número de tentativas de roubar a bola por hora efetuadas pelo atleta
CHBH	número de chutes que o atleta conseguiu bloquear por hora