DETERMINAÇÃO DOS MELHORES JOGADORES DO CAMPEONATO BRASILEIRO USANDO ANÁLISE ENVOLTÓRIA DE DADOS

Rony Walter Paiva Nolasco
rony_nolasco@hotmail.com
Celso Satoshi Sakuraba
sakuraba@ufs.br
Richard Andres Estombelo Montesco
restomb@ufs.br
Departamento de Engenharia de Produção – Universidade Federal de Sergipe
Av. Marechal Rondon, s/n, Jd. Rosa Elze São Cristóvão/SE CEP 49100-000

RESUMO

Jogadores de futebol muitas vezes acabam sendo comparados com base em informações subjetivas e tendenciosas. Para se obter um resultado menos controverso, pode haver a necessidade de métodos mais objetivos que se baseiem em ferramentas de análises de dados, tais como a Análise Envoltória de Dados (DEA). Neste trabalho, uma análise dos melhores jogadores do Campeonato Brasileiro com base no DEA foi comparada com os jogadores escolhidos para o Prêmio Brasileirão, eleitos por votação de jornalistas de todas as partes do país. O resultado mostrou que apenas 54% dos jogadores eleitos para a seleção do campeonato foram eficientes pelos cálculos do DEA. Além disso, o melhor jogador do campeonato eleito o Craque do Brasileirão, ficou apenas na 26º posição de acordo com supereficiência calculada pelo DEA. Isso sugere que uma análise de dados proporcionaria um avanço para a construção de um ranking mais imparcial e confiável.

PALAVRAS-CHAVE. DEA, Futebol, Campeonato Brasileiro.

Tópicos: Análise Envoltória de Dados, Outras aplicações em PO

ABSTRACT

Soccer players are frequently compared using subjective and biased information. To obtain a less controversial result, more objective methods based on data analysis as the Data Envelopment Analysis may be necessary. In this work, an analysis of the best players of the Brazilian League based on DEA was compared with the players chosen for the Brasileirão Award, elected by journalists all over the country. The result showed that only 54% of the players selected were efficient through DEA calculations. Besides, the best championship player elected as the Brasileirão MVP was only the 26th according to the super-efficiency calculated by DEA. The results suggest that a data analysis would provide an advance to build a more unbiased and reliable ranking.

KEYWORDS. DEA, Soccer, Brazilian League.

Paper topics: Data Envelopment Analysis, Other OR applications

Galoá { Este trabalho foi publicado utilizando o **Galoá** proceedings

1. Introdução

O futebol é o esporte mais popular do mundo, com uma estimativa de que metade da população mundial se considere fã da modalidade [World Atlas 2020]. Apesar de ser amplamente comentado em debates popularmente, há tempos ele também vem sendo alvo de estudos, como mostra o primeiro Congresso Mundial de Ciência e Futebol em 1987 [Reilly et al. 1988]. Assim como em diversos outros esportes, existem duas formas principais de se avaliar o desempenho de jogadores de futebol: baseando-se em avaliações de peritos ou através da análise de dados [Frick e Wicker 2015).

No Brasil, a Confederação Brasileira de Futebol (CBF) organiza o Prêmio Brasileirão, evento onde é apresentada a seleção dos melhores jogadores do Campeonato Brasileiro, maior campeonato nacional onde as 20 melhores equipes do país jogam entre si em turno e returno. É escolhido um jogador por posição através de uma votação entre jornalistas de todo o país, além do melhor entre todos eles, que é nomeado Craque do Brasileirão.

O problema para a escolha destes jogadores é que ela geralmente é feita de forma qualitativa, o que ocasiona diversos vieses. Listas de pontuação de revistas esportivas baseadas no julgamento de especialistas enfatizam o poder ofensivo dos jogadores sem prestar atenção ao desempenho de volantes ou defensores [Tiedemann et al. 2011]. Além disso, a distribuição de jogadores das seleções dos campeonatos dos últimos cinco anos mostrada na Tabela 1 mostra que há uma forte tendência a se escolherem jogadores da equipe campeã ou das primeiras colocadas.

Ano	Jogadores	Time	Posição	Jogadores	Time	Posição
2019	9	Flamengo	1º	2	Atlético PR	5º
2018	3	Palmeiras	1º	3	Internacional	3º
2017	4	Corinthians	1º	2	Grêmio	4º
2016	7	Palmeiras	1º	2	Flamengo	3º
2015	5	Corinthians	1º	4	Atlético MG	2º

Tabela 1 - Dois times com mais jogadores na seleção do campeonato nos últimos anos

O presente trabalho tem como objetivo avaliar o desempenho dos jogadores que foram eleitos para a seleção do campeonato pelos Prêmios Brasileirão e Craque do Brasileirão 2016 de maneira analítica. O ano de 2016 foi escolhido por ser aquele que apresentava maior número de jogadores selecionados pertencentes ao time campeão dentre os anos em que havia dados disponíveis, sugerindo um forte viés. A ferramenta escolhida para tal análise foi a Análise Envoltória de Dados (DEA, do inglês *Data Envelopment Analysis)*, por esta permitir a comparação sem a necessidade de calibração de parâmetros, eliminado possíveis tendências dos autores.

2. Revisão da literatura

A DEA foi desenvolvida primeiramente por [Farrell 1957] e consolidada por [Charnes et al. 1978] como um procedimento não-paramétrico que compara unidades tomadoras de decisão (DMUs, do inglês *Decision Making Units*) utilizando uma fronteira de eficiência de indicadores de desempenho. A definição de parâmetros é realizada através de um modelo de Programação Linear para cada DMU. [Cook e Zhu 2005] definem a DEA como uma abordagem orientada a dados, que avalia o desempenho de um conjunto de DMUs que convertem uma ou mais entradas (*inputs*) em uma ou mais saídas (*outputs*). Uma representação gráfica da fronteira de eficiência para uma entrada e duas saídas é dada na Figura 1.

Galoá { Este trabalho foi publicado utilizando o **Galoá** proceedings

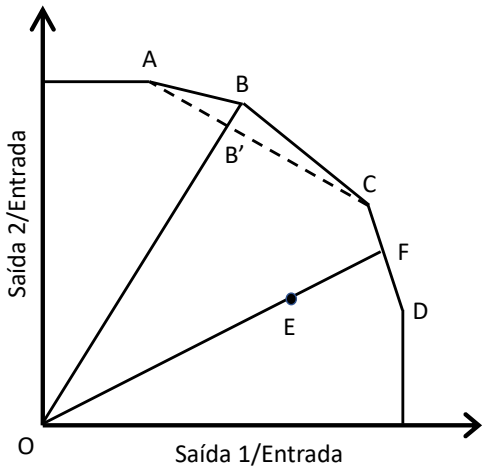


Figura 1 - Representação da fronteira de eficiência

Cada ponto no gráfico representa o desempenho de uma DMU (que neste estudo correspondem aos jogadores), sendo a fronteira de eficiência formada pelos pontos exteriores A a D. A DMU E precisariam melhorar suas saídas até atingir o ponto F para ser considerada eficiente, assim como outras DMUs que se encontrassem dentro da fronteira. A linha tracejada indica como ficaria a fronteira se a DMU B fosse excluída. A relação entre OB' e OB nos mostra a contribuição de B para a fronteira, sendo chamada de supereficiência. Valores menores de supereficiência representam DMUs com uma contribuição maior (ou jogadores com desempenhos superiores dentre os jogadores eficientes).

Existem diversas aplicações de diferentes modelos de DEA no futebol, tais como o trabalho de [Alp 2006] que analisou a supereficiência dos goleiros durante a Copa do Mundo de 2002. A supereficiência é uma abordagem baseada na análise da sensibilidade que analisa o impacto de se retirar uma DMU pertencente à fronteira de eficiência, permitindo comparar DMUs eficientes. [Tiedemann et al. 2011] apresentam uma revisão de diversas aplicações de DEA e FSA no futebol europeu, além de analisarem a eficiência dos jogadores da *Bundesliga* e relacioná-la aos resultados de seus times. [Santín 2014] utilizou um modelo de DEA não crescente com retornos de escala orientado às saídas (*output-oriented nonincreasing returns to scale*, NIRS) e modelos de supereficiência para determinar o jogador mais eficiente do Real Madrid de todos os tempos. No Brasil, [Calôba e Lins 2006] utilizaram o modelo de Medida de Eficiência Baseada nas Folgas (*Slacks Based Measure of Efficiency* ou SBM) para obterem um ranking dos times brasileiros e compara-lo a outros dois rankings bem conhecidos no Brasil, o da revista Placar e o da CBF.

3. Metodologia

Dois sites foram utilizados como fonte para a obtenção dos dados. O Whoscored.com (https://www.whoscored.com) possui estatísticas detalhadas de diversas ligas pelo mundo, sendo utilizado como referência pelos maiores veículos esportivos. Para os dados que não estavam disponíveis neste site, utilizou-se aqueles do *fantasy game* Cartola FC (http://www.cartolafcbrasil.com.br), que mantém dados de todos os jogos do Campeonato Brasileiro. A Tabela 2 mostra os dados utilizados e o site do qual foram obtidos.

Posição	Dados coletados	Fonte	Entrada/Saída
Goleiro	Minutos jogados	Whoscored.com	Entrada
	Quantidade de jogos sem sofrer gols	Cartola FC	Saída 1
	Defesas difíceis	Cartola FC	Saída 2
Demais posições	Minutos jogados	Whoscored.com	Entrada
	Roubadas de bola	Cartola FC	Saída 1
	Faltas sofridas	Cartola FC	Saída 2
	Assistências	Whoscored.com	Saída 3
	Gols	Whoscored.com	Saída 4
	Chutes no alvo, exceto gols	Cartola FC	Saída 5

Tabela 2 - Entradas e saídas por posição e suas fontes

Em nossas análises, a posição de cada jogador foi considerada como sendo a que ele mais vezes iniciou uma partida, pois é comum que ele mude de posição no decorrer de um jogo. O campeonato teve a participação de 736 jogadores, e análise foi realizada em dois cenários. No primeiro, foi estabelecido que apenas jogadores que houvessem entrado em campo por pelo menos 200 minutos seriam considerados. Isso evitaria vieses causados por jogadores que possam ter atuado por apenas poucos minutos e obtido saídas muito boas nesse curto período de tempo. No segundo cenário, foram considerados apenas jogadores que tenham atuado pelo menos 990 minutos ao longo do campeonato, que foi o menor tempo de jogo dentre os jogadores da seleção do campeonato. Considerou-se que apenas os jogadores que tivessem uma participação significativa em seus times compunham essa seleção, portanto apenas jogadores que tivessem atuado em uma boa parte do campeonato deveriam ser considerados nas análises. A quantidade de jogadores analisados por posição em cada cenário é mostrada na Tabela 3.

Posição	Cenário 1	Cenário 2
Goleiro	33	25
Zagueiro	80	64
Lateral direito	38	26
Lateral esquerdo	44	32
Volante	61	47
Meia	143	93
Atacante	58	32
Total	**457**	**319**

Tabela 3 – Quantidade de jogadores por posição em cada cenário

Para análise da escolha do Craque do Brasileirão, foi adotado o seguinte critério: dentre os jogadores que compunham a seleção do campeonato, seriam considerados apenas aqueles que se mostrassem eficientes em suas posições pelos cálculos do DEA. Estes jogadores seriam classificados em relação à supereficiência para determinar qual deles possuiu o melhor desempenho em relação aos outros jogadores dentro de suas posições.

O método escolhido para comparação foi o modelo clássico CCR [Charnes, Cooper e Rhodes 1978] voltado às saídas, que pode ser linearizado produzindo o modelo a seguir, no qual a eficiência da m-ésima DMU E_m é maximizada. No modelo, u_{im} e v_{jm} são os pesos da i-ésima entrada (i = 1,2,...,I) e da j-ésima saída (j = 1, 2,...,J) e as variáveis y_{jn} e x_{in} são, respectivamente, saída e a entrada da n-ésima DMU (n = 1, 2, ..., N), onde m, $n \leq$ N. Resolvendo-se o modelo para cada uma das N DMUs, obtém-se a eficiência máxima de cada uma delas.

$$\max E_m = \sum_{j=1}^{J} v_{jm} y_{jm}$$

sujeito a

$$\sum_{j=1}^{J} v_{jm} y_{jn} - \sum_{i=1}^{I} u_{im} x_{in} \leq 0 \,; n = 1, 2, \ldots, N$$

$$\sum_{i=1}^{I} u_{im} x_{im} = 1$$

$v_{jm}, u_{im} \geq 0$; $i = 1, 2, ..., I$; $j = 1, 2, ..., J$

Foram utilizados o *Open Source DEA* (http://opensourcedea.org) para os cálculos gerais e o *Efficiency Measurement System* (http://www.holger-scheel.de/ems) para o cálculo da supereficiência, que não estava disponível no primeiro programa.

4. Resultados

4.1. Primeiro cenário

A Tabela 4 exibe a eficiência dos jogadores que compõem a seleção do campeonato segundo o método CCR considerando apenas jogadores que participaram por pelo menos 200 minutos ao longo do campeonato.

Jogador	Minutos jogados	Posição na seleção	Classific. do time	Posição na DEA	Eficiência
Jaílson	1710	Goleiro	Campeão	Goleiro	78,95%
Jean	2990	Lateral direito	Campeão	Lateral direito	100,00%
Pedro Geromel	2310	Zagueiro	9º	Zagueiro	100,00%
Yerry Mina	990	Zagueiro	Campeão	Zagueiro	100,00%
Jorge	2876	Lateral esquerdo	3º	Lateral esquerdo	70,22%
Tchê Tchê	3270	Volante	Campeão	Volante	44,92%
Moisés	1464	Volante	Campeão	Meia	87,80%
Dudu	2582	Meia esquerda	Campeão	Atacante	83,21%
Diego	2139	Meia direita	3º	Meia	100,00%
Gabriel Jesus	2690	Atacante	Campeão	Atacante	100,00%
Robinho	2319	Atacante	4º	Meia	100,00%

Tabela 4 - Análise dos jogadores da seleção com jogadores que atuaram pelo menos 200 minutos

Podemos ver que dos onze jogadores que compõem a seleção, cinco não estão na fronteira de eficiência. Quatro destes cinco jogadores pertencem ao time campeão, mostrando o possível viés em sua escolha. Além disso, o jogador com menor eficiência (Tchê Tchê) é aquele com mais minutos jogados, que corresponde a mais do que o triplo dos minutos jogados pelo jogador da seleção com menos minutos, Yerry Mina. Sabe-se que o desgaste causado por atuações demasiadas influencia o desempenho dos atletas, portanto pode ser injusto comparar suas saídas com outros jogadores que atuaram por apenas 200 minutos (que corresponde a menos de 10% do seu tempo

de jogo). Deve-se notar que os jogadores Dudu e Robinho foram analisados pela DEA em posições diferentes daquelas para as quais foram escolhidos dentro da seleção.

Os seis jogadores da seleção do campeonato que obtiveram 100% de eficiência na DEA foram analisados em relação à sua supereficiência. A Tabela 5 apresenta os resultados destes cálculos, mostrando além da classificação de cada um desses seis jogadores em relação à supereficiência dentro de sua posição, os dados relativos ao jogador que obteve o melhor índice de supereficiência em cada uma de suas posições e os minutos jogados por eles no campeonato.

Posição	**Seleção do campeonato**			**Melhor da posição na DEA**		
	Jogador	**Superef.**	**Clas.**	**Jogador**	**Min.**	**Superef.**
Lateral direito	Jean	84,20%	5º	Léo Príncipe	435	62,18%
Zagueiro	Pedro Geromel	97,91%	9º	Yerry Mina	990	49,50%
	Yerry Mina	49,50%	1º			
Meia	Robinho	96,63%	11º	Marinho	2267	72,26%
Atacante	Dudu	81,37%	5º	Magno Alves	802	58,91%
	Gabriel Jesus	83,90%	6º			

Tabela 5 - Supereficiência dos jogadores da seleção com jogadores que atuaram mais de 200 minutos

Podemos notar que a maioria dos jogadores da seleção do campeonato não são aqueles com os melhores índices de supereficiência em suas posições. A excessão é Yerry Mina, que é aquele com melhor índice de eficiência entre os zagueiros. Seu índice de supereficiência mostra que ele ainda seria considerado um jogador eficiente mesmo que produzisse apenas 49,5% de suas saídas. Não devemos esquecer que ele é o jogador com menos minutos jogados entre aqueles que compõem a seleção do campeonato, tendo em teoria mais tempo de recuperação entre partidas. A Tabela 6 mostra como seria a seleção do campeonato caso ela fosse determinada pelos primeiros jogadores classificados pela supereficiência.

Posição	**Jogador**	**Supereficiência na Posição**	**Minutos Jogados**	**Classificação do time**
Goleiro	Paulo Victor	76,53%	450	3º
Lateral direito	Léo Príncipe	62,18%	435	7º
Zagueiro	Yerry Mina	49,50%	990	1º
Zagueiro	Felipe	84,85%	630	7º
Lateral esquerdo	Pará	58,72%	721	18º
Volante	Eduardo Henrique	56,82%	427	4º
Volante	Marcílio	58,97%	345	19º
Meia	Marinho	72,26%	2267	16º
Meia	Arthur Maia	73,23%	668	11º
Atacante	Magno Alves	58,91%	802	13º
Atacante	Leandro Pereira	67,37%	519	1º

Tabela 6 – Seleção do campeonato segundo a supereficiência de cada posição

Podemos observar que há uma distribuição maior em relação aos times aos quais os jogadores pertencem. O time campeão é o único que possui mais de um jogador na lista, mas ela contém até mesmo jogadores dos times que ficaram nas quatro últimas colocações, ou seja, que foram rebaixados naquele ano. Devemos destacar o meia Marinho, que além de ser o primeiro colocado segundo a supereficiência dentre os meias, é o único que atuou mais do que o dobro do

tempo do que todos os outros, e o zagueiro Yerry Mina, que é o único com supereficiência menor do que 50% e o segundo da lista com mais minutos jogados.

4.2. Segundo cenário

A Tabela 7 apresenta os resultados da DEA em relação aos jogadores que atuaram por pelo menos 990 minutos ao longo do campeonato. Para cada posição descrita na primeira coluna, as três colunas seguintes mostram o jogador escolhido para a seleção do campeonato, seu índice e classificação em relação à supereficiência (jogadores com valores maiores que 100% não são considerados eficientes); as quatro colunas seguintes mostram os jogadores com melhores índices de supereficiência calculados pelo DEA para cada posição, a quantidade de minutos jogados por eles e a classificação final de seus times.

Posição	Jogador	Superef.	Cl.	Jogador	Superef.	Min.	Time
Goleiro	Jaílson	108,71%	7º	M. Grohe	86,88%	2476	9º
L. direito	Jean	66,00%	2º	M. Rocha	61,83%	1112	4º
Zagueiro	P. Geromel	97,22%	5º	Y. Mina	40,11%	990	1º
Zagueiro	Y. Mina	40,11%	1º	Ernando	83,79%	3318	17º
L. esquerdo	Jorge	106,03%	9º	Juan Jr.	66,29%	2210	15º
Volante	Tchê Tchê	171,93%	42º	L. Cittadini	47,00%	1027	2º
Volante	Moisés	102,57%	13º	Cícero	74,53%	3141	13º
Meia (esq.)	Dudu	92,07%	7º	Marinho	62,00%	2267	16º
Meia (dir.)	Diego	113,89%	24º	Marcelo	74,14%	1188	3º
Atacante	G. Jesus	78,27%	2º	Sassá	77,37%	1306	5º
Atacante	Robinho	81,00%	3º	Dudu	78,27%	2690	1º

Tabela 7 - Análise dos jogadores da seleção com jogadores que atuaram pelo menos 990 minutos

Novamente, apenas seis dos 11 jogadores da seleção do campeonato se mostraram eficientes. É interessante notar que ao contrário do que acontecia no cenário anterior, nesse novo cenário o meia Dudu é considerado eficiente e o meia Diego não, o que demonstra que o critério de corte de jogadores possui influência no resultado final. A lista de jogadores mais eficientes calculada pela DEA também apresenta maior diversidade de times, contendo ainda dois jogadores do time campeão, mas também um de um time rebaixado. Podemos destacar os jogadores Yerry Mina e Léo Cittadini por seu índice de supereficiência, e também o meia Marinho, que possui o menor índice dentre aqueles que atuaram por mais de 1.710 minutos, que representa metade do tempo que cada time atuou no campeonato desconsiderando acréscimos.

5. Conclusões

A análise da DEA revelou que nem todos os jogadores eleitos para a seleção do campeonato de 2016 foram considerados eficientes. Levando em consideração os dados utilizados neste trabalho, apenas 54% (6 dos 11 jogadores) eleitos para a seleção do campeonato foram eficientes, sendo que Jaílson, Jorge, Tchê Tchê, Moisés e Diego precisariam aumentar as suas saídas para se igualarem a jogadores eficientes em suas respectivas posições. Com exceção de Yerry Mina, nenhum dos jogadores da seleção do campeonato conseguiram obter a máxima eficiência em sua posição através do método de supereficiência. Gabriel Jesus, escolhido como Craque do Brasileirão 2016, ficou em terceiro (sexto, se considerarmos jogadores que atuaram no mínimo 200 minutos) na classificação de supereficiência de sua posição, sendo que o segundo colocado Dudu atou mais minutos que ele ao longo do campeonato.

Galoá { Este trabalho foi publicado utilizando o **Galoá** proceedings

A seleção definida pelos cálculos de supereficiência mostrou que existe uma grande tendência em se escolher jogadores dos times mais bem colocados, de forma que jogadores de times que terminaram o campeonato na segunda metade da tabela dificilmente são escolhidos por jornalistas e outros especialistas para comporem a seleção do campeonato. Jogadores como Marinho deveriam receber uma maior importância no ranking mostrado no Prêmio Brasileirão, uma vez que ele jogou durante a maior parte do campeonato e conseguiu uma alta eficiência, mesmo que seu time tenha ficado próximo da zona de rebaixamento.

Referências

ALP, I. (2006) Performance evaluation of goalkeepers of the World Cup. *Journal of Science*, 19:119–125.

Calôba, G. M. e Lins, M. P. E. (2006) Performance assessment of the soccer teams in Brazil using DEA. *Pesquisa Operacional*, 26:521–536.

Charnes, A., Cooper, W. W. e Rhodes, E. (1978) Measuring the efficiency of decision making units. *European Journal of Operational Research*, 2:429–444.

Farrell, M. J. (1957) The measurement of productive efficiency. *Journal of the Royal Statistical Society*, 120:253–290.

Frick, B. e Wicker, P. (2015) Football experts versus sports economists: Whose forecasts ate better? *European Journal of Sport Science*, 16:603–608.

Reilly, T., Lees, A., Davids, K. e Murphy, W. J. (1988) Science and Football. E & FN Spon, Nova Iorque.

Santín, D. (2014) Measuring the technical efficiency of football legends: Who were Real Madrid's all-time most efficient players? International Transactions in Operational *Research*, 21: 439–452.

Tiedemann, T., Francksen, T. e Latacz-Lohmann, U. (2011) Assessing the performance of German Bundesliga football players: A non-parametric metafrontier approach. *Central European Journal of Operations Research*, 19: 571–587.

World Atlas (2020) Web page. https://www.worldatlas.com/articles/what-are-the-most-popular-sports-in-the-world.html. Acessado: 2020-05-18.

Galoá { Este trabalho foi publicado utilizando o **Galoá** proceedings