

DEA PARA AVALIAÇÃO DA EFICIÊNCIA DENTRO DE CAMPO DAS EQUIPES PARTICIPANTES DO CAMPEONATO BRASILEIRO DE 2011

Carlos André Faccini dos Santos
Universidade Federal Fluminense (UFF)
carlosandre.faccini@hotmail.com

Lidia Angulo Meza
Universidade Federal Fluminense (UFF)
lidiaangulomeza@id.uff.br

João Carlos Correia Baptista Soares de Mello
Universidade Federal Fluminense (UFF)
jcsmello@producao.uff.br

Resumo

O futebol é conhecido por ser um esporte por produzir partidas com resultados imprevisíveis. Mesmo assim, não é estranho dizer que grandes equipes que mantêm alto nível técnico possuem na maioria das vezes condições de disputar títulos. A modelagem em DEA permite avaliar a eficiência das equipes por suas ações em campo, de maneira que equipes ineficientes identifiquem os fundamentos nos quais possuem deficiência e em que equipes podem se espelhar para melhorar seu desempenho. Neste artigo, foi aplicada a seleção de variáveis para definir quais ações em campo são mais relevantes para três modelos em DEA, que avaliam o desempenho das equipes em diferentes objetivos: Defesa, Ligação e Ataque.

Palavras-chave: DEA; seleção de variáveis; eficiência; futebol brasileiro.

Abstract

Soccer is famous for unpredictable match results. Despite that fact, it isn't odd to point that great teams that usually keep high technical level along many competitions have most of the times conditions to compete for titles. The DEA approach evaluates the efficiency analysis of the teams by their actions during the game, that way inefficient teams may identify what aspects of performance are low and find out where they could search for improvements. On this paper, variable selection was applied for choosing which aspects inside the field game are more relevant for 3 different DEA models, that evaluate teams on three different goals: Defense, Connection and Attack.

Keywords: DEA; variable selection; efficiency; Brazilian soccer.

1. Introdução

O futebol é um dos esportes mais populares do mundo e no Brasil é conhecido como a grande paixão nacional. Dentre os diversos campeonatos que acontecem ao longo do ano, o Campeonato Brasileiro tem grande destaque, sendo considerado por muitos torcedores mais emocionante que a copa do mundo. O Campeonato Brasileiro é composto das 20 melhores equipes do país e é disputado em 38 rodadas, onde cada equipe enfrenta duas vezes cada equipe adversária, uma vez como mandante e outra como visitante, em dois turnos distintos.

Mesmo com este seleto grupo de equipes, não é incomum que ocorram partidas de nível técnico baixo. Existe uma grande discrepância entre a qualidade técnica das equipes participantes, e mesmo clubes de tradição sofrem com alta oscilação de seu desempenho em campo.

Existem na literatura artigos em que foram elaborados modelos em DEA para avaliação do desempenho de equipes de futebol, porém com foco principal em avaliar a relação Investimento X Resultado final, sem oferecer resultados que permitam o apoio a decisões que viabilizem a melhoria de desempenho da equipe dentro de campo. Nestes trabalhos, a eficiência das equipes é muitas vezes impactada pelo tamanho do clube, sua torcida e capacidade de atrair investimento.

García-Sánchez (2007) analisa os clubes da liga espanhola de 2004/2005 em três aspectos: Eficiência Operacional (Ataque e Defesa), Resultados Globais e Efetividade Social. Na análise de Ataque e Defesa são consideradas ações de defesa e ataque dentro de campo com o objetivo de maximizar os gols marcados e minimizar os gols sofridos. Foram consideradas como variáveis técnicas: dribles, passes próximos a área e chutes a gols no ataque e roubadas de bola e ações do goleiro na defesa. Estas variáveis no entanto não cobrem outras importantes ações em campo, inclusive negativas, que impactam na possibilidade de marcar e sofrer gols. Não foram obtidos também alvos para melhoria destas ações em campo, ficando a equipe restrita a alvos de aumento de gols marcados e diminuição de gols sofridos.

Outros trabalhos encontrados na literatura, como Cortez (2010), avaliam resultados de investimentos contra os resultados atingidos nas competições. Estes resultados permitem avaliar a eficiência da gestão dos Clubes, porém com pouca implicação nos aspectos técnicos. Assim, seria possível trocar toda a equipe e continuar a cometer as mesmas falhas dentro de campo.

Neste trabalho, serão discutidas 18 variáveis que representam as ações dentro de campo das 20 equipes do campeonato brasileiro. São propostos 3 modelos: Defesa, Ligaç o e Ataque, com rela oes causais distintas e objetivos diferentes, onde a defici ncia de cada setor poder  ser desdobrada em fundamentos (vari veis) que as equipes precisam melhorar, e n o apenas em gols marcados e sofridos.

Com o objetivo de aumentar o poder discriminat rio dos modelos, ser  aplicado o m todo multicrit rio de sele o de vari veis, sendo mantidas apenas as vari veis mais representativas a cada modelo.

2. Objetivo

O objetivo principal deste trabalho   desenvolver um modelo que permita avaliar a efici ncia dentro de campo das equipes participantes do Campeonato Brasileiro de 2011. Esta an lise permitir   s equipes identificarem quais aspectos t cnicos devem melhorar e que equipes utilizar como Benchmark, independente do tamanho dos Clubes. Assim, ser  poss vel identificar que fundamentos precisam de mais treinamento ou at  mesmo indicar uma contrata o de um jogador que supra os gaps da equipe.

Ser  discutida a utiliza o de 18 vari veis (a oes em campo) retiradas das estat sticas de jogo contabilizadas pela "Globo.com" at  a 5ª rodada do Campeonato para constru o dos modelos de Defesa, Liga o e Ataque. Para utiliza o das vari veis que permitem melhor

discriminação das equipes mantendo a representatividade dos 3 modelos, o modelo multicritério de seleção de variáveis será utilizado.

3. Análise Envoltória de Dados

A Análise Envoltória de Dados é um método não paramétrico criado por Charnes (1978) para calcular a eficiência comparada de DMUs (*Decision Making Units*) sob a influência de fatores de produção (Inputs) e produtos (Outputs).

As DMUs são um grupo específico e fechado de unidades que possuem uma sistemática produtiva igual, ou seja, processam o mesmo tipo de Inputs a fim de obter o mesmo tipo de Outputs. A comparação da quantidade de Inputs consumidos e Outputs gerados pelas DMUs geram o conceito de eficiência, sempre limitado pelo universo destas DMUs pela criação da chamada fronteira de eficiência.

O primeiro modelo clássico em DEA é o CCR, Charnes, Cooper e Rhodes (1978), que considera retornos constantes de escala e proporcionalidade entre as variáveis do modelo e é representado pelo PPL (1) no modelo dos multiplicadores. Banker, Charnes e Cooper (1984) criaram o BCC que considera retornos variáveis de escala e não necessita que haja proporcionalidade entre as variáveis, representado pelo PPL (2) também pelo modelo dos multiplicadores. O segundo é comumente utilizado em ordenação em competições e definição de preferências.

Modelo DEA CCR:

$$\begin{aligned} \text{Max } Eff_o &= \sum_{j=1}^s u_j y_{j0} \\ \text{sujeito a} \\ \sum_{i=1}^r v_i x_{i0} &= 1 \\ \sum_{j=1}^s u_j y_{jk} - \sum_{i=1}^r v_i x_{ik} &\leq 0, \forall k \\ u_j, v_i &\geq 0, \forall j, i \end{aligned}$$

(1)

Modelo DEA BCC:

$$\begin{aligned} \text{Max } Eff_o &= \sum_j u_j y_{j0} + u_* \\ \text{sujeito a} \\ \sum_i v_i x_{i0} &= 1 \\ - \sum_i v_i x_{ik} + \sum_j u_j y_{jk} + u_* &\leq 0, \forall k \\ u_j \geq 0, v_i \geq 0, \forall j, i \\ u_* &\in \mathfrak{R} \end{aligned}$$

(2)

A figura 1 mostra a diferente forma de envelopamento dos dados (das DMUs) entre os modelos CCR e BCC para o caso de apenas um Input e um Output. Representar graficamente o envelopamento em DEA com um número maior de variáveis é uma tarefa complexa, porém este caso simplificado é capaz de ilustrar como o BCC é mais benevolente que o CCR, já que toda DMU estará sempre mais próxima ou minimamente na mesma distância da fronteira BCC em relação a fronteira CCR. A proximidade da fronteira determina a eficiência das DMUs.

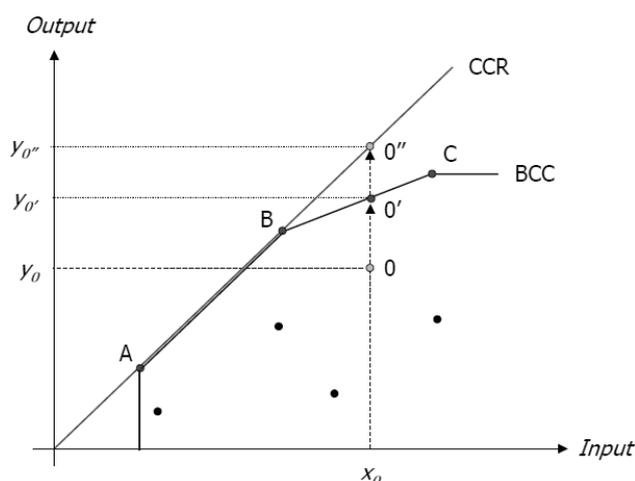


Figura 1 – Modelos CCR e BCC com par único Input-Output

A fim de realizar uma modelagem robusta neste trabalho foi necessário realizar algumas flexões e considerações especiais. Por vezes é possível que se deseje avaliar DMUs que produzam apenas Outputs ou Inputs, para solucionar este problema Lovell (1999) apresenta a utilização de Input unitário ou Output unitário respectivamente que permitem o cálculo da eficiência DEA nestes casos.

Outra necessidade particular da modelagem pode ser encontrada em Sheel (2001), que apresenta uma maneira de tratar a existência de Outputs indesejáveis na modelagem em DEA. Um Output indesejável, apesar de ser produto da operação da DMU, acaba prejudicando os resultados da mesma. Sendo assim, uma maneira de tratar na modelagem é utilizar o Output indesejável como um quociente, punindo os demais produtos gerados, porém esta aplicação pode gerar divisões por 0 quando o Output indesejável for nulo. Uma segunda maneira de tratar este problema é considerar o Output indesejável como Input, fazendo com que ele seja entendido como custos para produção da DMU.

A modelagem em DEA pode ser realizada nas mais variadas áreas fornecendo resultados consistentes desde que as variáveis sejam estabelecidas com a correta relação causal e sejam definidos os critérios de modelagem consistentes com o caso estudado, como orientação, retorno de escala, flexões no modelo e relação entre quantidade de variáveis e quantidade de DMU's (capacidade de discriminação).

O problema da capacidade de discriminação em DEA quando há um grande número de variáveis em relação ao número de DMUs é uma deficiência conhecida. Esta situação ocorre pelo fato das DMUs conseguirem encontrar pesos ótimos pelas diversas opções de variáveis existentes, formando uma configuração eficiente, porém de maneira falsa, já que ela consegue esconder seus diversos pontos fracos. Uma maneira de corrigir este problema é limitar o número de variáveis, fazendo a seleção de variáveis.

4. Seleção de Variáveis

Senra (2007) faz um estudo sobre variados métodos de seleção de variáveis por considerar importante avaliar se o método a ser utilizado se aplica a situação trabalhada, pois alerta que “a modificação do conjunto de variáveis selecionadas pode ter grande impacto no resultado da avaliação”. Senra defende ainda que a escolha de variáveis conduzirá a resultados diferentes não é necessariamente uma fraqueza, já que esta escolha pode significar a intenção de observar as DMU's por uma determinada ótica.

Os métodos de seleção de variáveis são instrumentos de auxílio a decisão, não deixando a análise do modelo restrita apenas ao modelo matemático, permitindo que os agentes tomadores de decisão e especialistas intervenham no modelo em diferentes níveis. Para isso, existem métodos de seleção de variáveis que utilizam desde nenhuma informação do decisor até aqueles

que incorporam informação subjetiva. Porém é sempre importante que se confronte os resultados para avaliar sua coerência.

Neste contexto, Soares de Mello (2003) realizou uma pré-seleção de variáveis disponíveis para realizar 3 diferentes avaliações de empresas aéreas: operacional, vendas e total. Esta escolha permitiu então aplicar os modelos matemáticos obtendo resultados específicos para cada objetivo, todos oriundos de uma mesma base de dados global de variáveis.

Partindo deste mesmo princípio, a seleção de variáveis neste trabalho será feita em duas etapas distintas: Pré-Seleção para modelos de Defesa, Ligação e Ataque e Seleção de Variáveis pelo Método Multicritério de Seleção de Variáveis.

Na primeira etapa, serão escolhidas dentre todas as variáveis disponíveis aquelas que tenham relação com a capacidade de evitar gols (Defesa), que tenham relação com a manutenção da posse de bola (Ligação) e que tenham relação com a capacidade de marcar gols (Ataque)

Na segunda etapa, será utilizado o Método Multicritério de Seleção de Variáveis, apresentado por Soares de Mello (2002), que permite manter uma boa relação causal entre as variáveis escolhidas e fornece boa discriminação das DMUs. Este método possui alta participação do decisor, que escolhe um par inicial Input-Output do modelo, e variáveis extras são incorporadas respeitando a condição de boa relação causal e boa discriminação também com pesos definidos pelo decisor.

5. Modelagem

I. DMUs

As DMUs consideradas para o estudo são todas as equipes participantes da série A do Campeonato Brasileiro de 2011, apresentada das na Tabela 1.

AméricaMG	Bahia	Cruzeiro	Internacional
AtléticoGO	Botafogo	Figueirense	Palmeiras
AtléticoMG	Ceará	Flamengo	Santos
AtléticoPR	Corinthians	Fluminense	SãoPaulo
Avaí	Coritiba	Grêmio	Vasco

Tabela 1 – Equipes participantes da série A do Campeonato Brasileiro de 2011

II. Modelo e Orientação

Abaixo são definidos os modelos e orientações de cada uma das modelagens a serem trabalhadas. É utilizado o modelo BCC em todas as modelagens por não haver proporcionalidade entre as variáveis utilizadas, esta escolha será esclarecida a frente quando forem introduzidas as variáveis utilizadas.

O modelo de Ataque será BCC Orientado a Output. O objetivo deste modelo será avaliar a eficiência da equipe em fazer gols. Este objetivo não significa que a equipe que possui mais gols necessariamente é a mais eficiente, pois ela pode ter alcançado esta marca com uma configuração ineficiente de ações de ataque.

O modelo de Ligação será BCC Orientado a Output. O objetivo deste modelo será avaliar a eficiência da equipe em aumentar seu volume de jogo, fazendo a ligação da defesa ao ataque. Este objetivo é representado por ações em campo que influenciam diretamente na manutenção da posse de bola.

O modelo de Defesa será BCC Orientado a Input. O objetivo deste modelo será avaliar a eficiência da equipe em evitar sofrer gols. Assim como no modelo de Ataque, a equipe que sofre mais gols não é necessariamente a mais ineficiente, pois mesmo sofrendo um elevado

número de gols, a equipe pode ter sido muito atacada e ter tido sucesso em ações que evitaram sofrer mais gols.

Os modelos não avaliarão os jogadores de ataque, meio campo e defesa separadamente, mas sim a equipe como um todo, mesmo que em momentos que os defensores por vezes ataquem ou atacantes ajudem na defesa. A análise feita desta maneira permitirá avaliar se existe equilíbrio na equipe em relação aos diferentes objetivos.

III. Inputs e Outputs

Todas as ações em campo contabilizadas por jogador ,por jogo, pela Globo.com até a 5ª rodada do Campeonato Brasileiro foram consolidadas por equipe, assim são contadas todas as estatísticas independente das substituições que possam ter ocorrido. Todas as ações são variáveis produzidas pelas equipes, e desta maneira são todas Outputs.

As ações em campo contabilizadas foram as apresentadas na Tabela 3:

RB	Roubada de Bola
FC	Falta cometida
GC	Gol Contra
CA	Cartão amarelo
CV	Cartão Vermelho
SG	Jogos sem sofrer gols
DD	Defesa Díficil
DP	Defesa de pênalti
GS	Gol Sofrido
FS	Falta sofrida
PE	Passe errado
A	Assistência
FT	Finalização na trave
FD	Finalização defendida
FF	Finalização fora
G	Gol
I	Impedimento
PP	Pênalti perdido

Tabela 2 – Legenda de Variáveis

Equipe	I	PE	PP	FC	GC	CV	CA	GS	FS	FD	FT	G	FF	RB	A	DP	SG	DD
AméricaMG	17	149	0	87	0	2	16	11	97	20	2	7	30	60	5	0	0	11
AtléticoGO	6	146	0	91	0	0	11	6	57	13	2	7	26	70	4	0	6	12
AtléticoMG	7	162	0	57	0	1	9	4	72	18	2	9	31	56	9	0	5	5
AtléticoPR	4	195	0	86	1	1	9	8	98	21	2	1	31	57	0	0	0	11
Avai	5	180	0	94	1	0	16	10	84	22	1	4	21	76	3	0	0	3
Bahia	13	167	0	102	0	1	14	8	73	23	2	6	14	69	4	0	6	12
Botafogo	8	124	0	98	0	0	9	4	92	11	1	6	25	64	5	0	11	4
Ceará	11	140	1	65	0	1	7	11	77	21	1	5	21	47	4	0	6	12
Corinthians	6	120	0	75	0	0	9	10	79	15	0	7	16	61	5	0	5	9
Coritiba	12	149	1	90	0	1	9	8	84	25	3	8	25	54	6	0	0	8
Cruzeiro	13	151	0	84	1	0	10	6	90	28	3	4	32	65	3	0	0	7
Figueirense	19	161	0	86	0	0	14	2	86	12	1	5	28	65	4	0	16	12
Flamengo	12	188	0	79	0	1	9	5	83	19	1	9	26	100	7	0	10	8
Fluminense	16	127	0	91	0	0	11	6	94	25	2	3	17	63	3	0	5	10
Grêmio	10	159	1	71	0	1	10	6	74	15	0	4	17	58	3	0	12	9
Internacional	7	137	0	87	1	0	11	7	81	21	0	7	25	66	5	1	0	5
Palmeiras	12	120	0	114	1	0	13	3	115	20	2	8	26	58	7	0	13	8
Santos	6	134	0	70	0	1	9	4	60	5	1	5	14	54	4	0	0	8
São Paulo	4	138	0	74	1	0	9	1	53	16	4	8	20	70	7	1	16	8
Vasco	5	177	0	83	0	0	13	8	76	12	2	9	24	70	7	1	5	10
TOTAL	193	3024	3	1684	6	10	218	128	1625	362	32	122	469	1283	95	3	116	172

Tabela 3 – Ações contabilizadas por equipe até a 5ª rodada

O Insumo das equipes para a produção das ações em campo é sua mera existência, que não é diferenciada pelo tamanho do clube ou mesmo o número total de jogadores inscritos no campeonato, pois participam do jogo e da produção de ações sempre 11 jogadores por partida em cada equipe. Assim, o Input a ser considerado é o número 1, que representa existência da DMU. Esta aplicação foi utilizada por Soares de Melo (2008) ao considerar a existência de tornos mecânicos como output unitário para avaliação da eficiência dos tornos, onde os desvios geométricos eram o preço a se pagar (Inputs) pela fabricação de peças.

No entanto, para cada um dos três modelos propostos, os Outputs gerados podem ser desejáveis ou indesejáveis. Outro trabalho sobre outputs indesejáveis pode ser encontrado em Nanci (2004). Da mesma forma, para cada modelo proposto os Outputs indesejáveis serão considerados como Inputs que devem ser minimizados.

Nenhuma destas variáveis possui proporcionalidade entre elas, o que justifica a utilização do BCC em todos os modelos. Como exemplo, podemos afirmar que o aumento de Roubadas de Bola (RB) não causa um aumento proporcional na ocorrência de Passes Errados (PE), da mesma maneira o aumento de Passes Errados (PE) não causa um aumento proporcional na ocorrência de uma Defesa Difícil (DD), apesar de estas ações impactarem na possibilidade de ocorrência da outra.

Dado o grande número de variáveis em relação ao número de DMU's, é necessário realizar a pré-seleção das variáveis, identificando as que pouco podem agregar aos modelos e aquelas que são candidatas a participar de cada modelo. É importante frisar que uma mesma variável pode ser candidata e ser selecionada para diferentes objetivos.

IV. Pré-Seleção de Variáveis

Nesta etapa são analisadas quais as variáveis com informação que contribuem na eficiência de uma DMU em cada modelo.

Variáveis Descartáveis:

Pode-se dizer que Pênalti Perdido (PP), Gol Contra (GC) e Defesa de Pênalti (DP) são redundantes, pois são representados diretamente pelas variáveis Gol (G) e Gol Sofrido (GS). Além disso, variaram pouco entre os times até a 5ª rodada.

Jogos sem Sofrer Gols (SG) é uma variável contabilizada por jogador de defesa quando uma equipe não sofre gols em um jogo. Assim, esta variável é impactada pelo número de defensores no time, inclusive levando-se em consideração substituições que acontecem ao longo do jogo. Desta forma a variável é impactada pela formação tática, prejudicando a unicidade das DMU's e afetando a confiabilidade da modelagem em DEA.

Os cartões amarelos e vermelhos (CA, CV) possuem problema de causalidade com as demais ações em campo, por não possuir influência na possibilidade de ocorrência das demais variáveis. Isto ocorre pois um jogador que é suspenso será substituído por outro no jogo seguinte, e este outro jogador poderá realizar as ações em campo da mesma maneira. Além disso, estas variáveis são representadas em campo pelo volume de faltas cometidas (FC).

Assim, com as variáveis restantes propõe-se as seguintes modelagens resumidas na Tabela 4:

MODELO	BCC/CCR	ORIENTAÇÃO	OBJETIVO	IMPUTS(OUTPUTS INDESEJÁVEIS)	OUTPUTS DESEJÁVEIS
Defesa	BCC	Output	Evitar sofrer Gols	Passe Errado (PE) Falta Cometida (FC) Gol Sofrido (GS)	Roubada de Bola (RB) Defesa Difícil (DD)
Ligação	BCC	Output	Intensificar posse de bola	Passe Errado (PE) Falta Cometida (FC)	Falta Sofrida (FS) Roubada de Bola (RB) Assistência (A)
Ataque	BCC	Imput	Fazer Gols	Impedimento (I) Finalização na Trave (FT) Finalização para Fora (FF)	Falta Sofrida (FS) Finalização Defendida (FD) Gol (G) Assistência (A)

Tabela 4 – Resumo da Modelagem Inicial

Causalidade e Alocação das Variáveis nos Modelos Propostos

Nesta sessão será discutida a relação causal existente entre as variáveis pré-selecionadas para cada modelo, justificando a alocação de cada variável em cada modelo. É importante ressaltar que uma mesma variável pode fazer parte de modelos diferentes, desde que possua influência em ambos os objetivos destes diferentes modelos.

No Ataque:

Todos Outputs indesejáveis reduzem as chances de produzir Outputs desejáveis, pois todos eles resultam no encerramento de uma ação de ataque que poderia resultar em Gol, que é o objetivo do modelo. Todos Outputs desejáveis por sua vez representam ações que permitem a continuidade da ação de ataque ou mesmo representam seu sucesso diretamente.

O Output Falta Sofrida (FS) aumenta as chances de produzir Outputs desejáveis por finalizações de falta ou mesmo jogada aérea. É importante frisar que as Finalizações Defendidas (FD) são desejáveis pois representam o acerto ao alvo, não significando que se deve buscar finalizar para defesa, mas prioritariamente no gol ao invés de Finalizar para Fora (FF) ou na Trave (FT), que são Outputs indesejáveis.

Na Ligação:

Os Outputs desejáveis Roubada de Bola (RB) e Falta Sofrida (FS) implicam diretamente no aumento da posse de bola e os Outputs indesejáveis Passe Errado (PE) e Falta Cometida (FC) implicam diretamente na diminuição da posse de bola. A posse de bola por sua vez, indica a estabilidade no time e o quanto ele está mantendo seu volume de jogo. As chances de produzir o Output desejável Assistência (A) é produto da posse de bola e representa o efeito do sucesso do modelo de Ligação em apoiar o ataque.

Na Defesa:

Os Outputs indesejáveis Passe Errado (PE) e Falta Cometida (FC) implicam nas chances de se produzir o Output Indesejável Gol Sofrido (GS), pois permitem que o adversário realize mais ações de ataque. O Output desejável Roubada de Bola (RB) por sua vez interrompe a movimentação de ataque do adversário, reduzindo as chances de Sofrer Gol (GS). A Defesa Difícil (DD) possui relação direta com a variável (GS) por representar a interrupção de uma ação bem sucedida de ataque do adversário.

7. Seleção de Variáveis – Método Multicritério

Nesta segunda etapa será utilizado o Método Multicritério de Seleção de Variáveis (Soares de Mello (2002)). O primeiro passo deste método é a escolha pelo decisor do par inicial Input-Output, pertencente às variáveis pré-selecionadas, que farão parte do modelo. Variáveis adicionais serão incluídas uma a uma até que se atinja um número de variáveis igual a um quinto do número de DMUs, que são 4 variáveis no caso estudado.

No segundo passo, cada variável ainda não pertencente ao modelo é incorporada no cálculo da eficiência DEA padrão com o par inicial isoladamente. Para escolha desta variável são calculados os índices SEF (eficiência média) e SDIS (nº DMUs na fronteira) obtidos com o acréscimo de cada variável candidata.

SEF e SDIS são normalizados e utilizados para o cálculo de S por uma média ponderada de acordo com a importância dada pelo decisor a cada critério. O maior S então indicará a variável que deve ser incorporada no modelo. O processo é repetido até que se atinja o nº de variáveis indicado.

O método foi executado com as variáveis pré-selecionadas para os modelos de Ataque, Ligação e Defesa conforme as tabelas 5, 6 e 7:

V. ATAQUE

Para o modelo de ataque foram Selecionados como par inicial o Input Finalização para Fora (FF) e o Output Gols (G). Dado o objetivo do modelo de ataque, estas duas variáveis são as que melhor representam o sucesso da equipe em fazer gols, pois identificarão aquela que possui uma boa relação entre as finalizações no alvo e de sucesso (gols) e a quantidade de tentativas. Desta maneira uma equipe pode ter muitos gols e ser ineficiente no objetivo de fazer gols por finalizar muito para fora, o que significa possuir uma pontaria ruim.

As variáveis adicionadas ao modelo foram Assistência (A) e Falta Sofrida (FS) por garantirem ao modelo o melhor conjunto entre poder de discriminação (SDIS) e alta eficiência média (SEF).

Par Inicial: FF e G		1ª Variável: A					2ª Variável: FS					
DMU	Padrão	IMP OUTPUT	-	-	-	I	FT	IMP OUTPUT	-	-	FT	I
			A	FD	FS	-	-		FD	FS	-	-
AméricaMG	78%	AméricaMG	78%	85%	85%	78%	78%	AméricaMG	85%	87%	78%	78%
AtléticoGO	78%	AtléticoGO	78%	78%	78%	78%	78%	AtléticoGO	78%	78%	78%	78%
AtléticoMG	100%	AtléticoMG	100%	100%	100%	100%	100%	AtléticoMG	100%	100%	100%	100%
AtléticoPR	11%	AtléticoPR	11%	76%	76%	13%	11%	AtléticoPR	76%	85%	11%	13%
Avaí	48%	Avaí	48%	87%	87%	48%	49%	Avaí	87%	81%	49%	48%
Bahia	100%	Bahia	100%	100%	100%	100%	100%	Bahia	100%	100%	100%	100%
Botafogo	67%	Botafogo	67%	67%	67%	67%	68%	Botafogo	68%	82%	71%	67%
Ceará	61%	Ceará	61%	85%	85%	61%	61%	Ceará	85%	77%	63%	61%
Corinthians	100%	Corinthians	100%	100%	100%	100%	100%	Corinthians	100%	100%	100%	100%
Coritiba	89%	Coritiba	89%	100%	100%	89%	89%	Coritiba	100%	93%	89%	89%
Cruzeiro	44%	Cruzeiro	44%	100%	100%	44%	44%	Cruzeiro	100%	78%	44%	44%
Figueirense	56%	Figueirense	56%	58%	58%	56%	56%	Figueirense	58%	75%	57%	56%
Flamengo	100%	Flamengo	100%	100%	100%	100%	100%	Flamengo	100%	100%	100%	100%
Fluminense	41%	Fluminense	55%	100%	100%	41%	41%	Fluminense	100%	100%	55%	55%
Grêmio	55%	Grêmio	55%	63%	63%	55%	57%	Grêmio	65%	83%	60%	55%
Internacional	78%	Internacional	78%	86%	86%	78%	100%	Internacional	86%	83%	100%	78%
Palmeiras	89%	Palmeiras	90%	93%	93%	89%	89%	Palmeiras	97%	100%	91%	90%
Santos	83%	Santos	100%	83%	83%	100%	100%	Santos	100%	100%	100%	100%
São Paulo	100%	São Paulo	100%	100%	100%	100%	100%	São Paulo	100%	100%	100%	100%
Vasco	100%	Vasco	100%	100%	100%	100%	100%	Vasco	100%	100%	100%	100%
E. Média	74%	E. Média	76%	88%	88%	75%	76%	E. Média	89%	90%	77%	76%
DMU's Front	6	DMU's Front	7	9	9	7	8	DMU's Front	10	9	8	7
		SEF	0,053	1	1	0	0,095	SEF	0,938	1	0,121	0
		SDIS	1	0	0	1	0,5	SDIS	0	0,333	0,667	1
		S	0,527	0,5	0,5	0,5	0,298	S	0,469	0,667	0,394	0,5

Tabela 5 – Seleção de Variáveis para o Modelo de Ataque

VI. LIGAÇÃO

O par inicial selecionado foi Passe Errado (PE) e Roubada de Bola (RB). Estas duas variáveis são as ações típicas que representam diretamente na perda e na recuperação da posse de bola. Equipes que erram muitos passes sofrem para manter volume de jogo, enquanto as que roubam muitas bolas são enaltecidas por dominarem as ações do jogo proativamente.

As variáveis adicionadas ao modelo foram Falta Sofrida (FS) e Assistência (A). Para a seleção da assistência foi necessário priorizar um dos objetivos da seleção, e o objetivo priorizado foi o aumento da discriminação garantido pela variável.

Par Inicial: PE e RB		1ª Variável: FS				2ª Variável: A		
DMU	Padrão	IMPUT	-	-	FC	IMPUT	-	FC
		OUTPUT	A	FS	-	OUTPUT	A	-
AméricaMG	77%	AméricaMG	78%	90%	77%	AméricaMG	90%	100%
AtléticoGO	92%	AtléticoGO	92%	92%	92%	AtléticoGO	92%	92%
AtléticoMG	66%	AtléticoMG	100%	72%	100%	AtléticoMG	100%	100%
AtléticoPR	57%	AtléticoPR	57%	89%	57%	AtléticoPR	89%	100%
Avai	80%	Avai	80%	89%	80%	Avai	89%	89%
Bahia	78%	Bahia	78%	79%	78%	Bahia	79%	79%
Botafogo	100%	Botafogo	100%	100%	100%	Botafogo	100%	100%
Ceará	64%	Ceará	65%	71%	100%	Ceará	71%	100%
Corinthians	100%	Corinthians	100%	100%	100%	Corinthians	100%	100%
Coritiba	69%	Coritiba	79%	79%	69%	Coritiba	82%	85%
Cruzeiro	82%	Cruzeiro	82%	88%	82%	Cruzeiro	88%	98%
Figueirense	77%	Figueirense	77%	85%	77%	Figueirense	85%	92%
Flamengo	100%	Flamengo	100%	100%	100%	Flamengo	100%	100%
Fluminense	96%	Fluminense	96%	97%	96%	Fluminense	97%	100%
Grêmio	69%	Grêmio	69%	74%	77%	Grêmio	74%	90%
Internacional	93%	Internacional	93%	93%	93%	Internacional	93%	93%
Palmeiras	95%	Palmeiras	100%	100%	95%	Palmeiras	100%	100%
Santos	78%	Santos	78%	78%	89%	Santos	78%	89%
SãoPaulo	97%	SãoPaulo	100%	97%	100%	SãoPaulo	100%	100%
Vasco	75%	Vasco	88%	81%	75%	Vasco	91%	85%
Ef. Média	82%	Ef. Média	86%	88%	87%	Ef. Média	90%	95%
DMU's Front	3	DMU's Front	6	4	6	DMU's Front	6	9
		SEF	0	1	0	SEF	0	1
		SDIS	0	1	0,60214	SDIS	1	0
		S	0	1	0,30107	S	0,6	0,4

Tabela 6 – Seleção de Variáveis para o Modelo de Ligação

VII. DEFESA

O par inicial selecionado foi Gol Sofrido (GS) e Defesa Difícil (DD). Estas duas variáveis não significam apenas o aumento ou diminuição da chance de sofrer gols, que é o objetivo do modelo. Estas ações representam diretamente o sucesso da DMU em alcançar o objetivo proposto.

As variáveis adicionadas ao modelo foram Falta Cometida (FC) e Roubada de Bola (RB). Da mesma forma que o modelo de Ligação foi priorizada a discriminação proporcionada para escolha da variável Roubada de Bola (RB)

Par Inicial: GSe DD

DMU	Padrão
AméricaMG	16%
AtléticoGO	33%
AtléticoMG	25%
AtléticoPR	22%
Avai	10%
Bahia	25%
Botafogo	25%
Ceará	18%
Corinthians	13%
Coritiba	13%
Cruzeiro	17%
Figueirense	100%
Flamengo	20%
Fluminense	25%
Grêmio	21%
Internacional	14%
Palmeiras	33%
Santos	25%
SãoPaulo	100%
Vasco	19%
Ef. Média	29%
DMU's Front	2

1ª Variável: FC

INPUT	-	FC	PE
OUTPUT	RB	-	-
AméricaMG	16%	77%	90%
AtléticoGO	100%	86%	100%
AtléticoMG	25%	100%	74%
AtléticoPR	22%	83%	71%
Avai	18%	61%	67%
Bahia	65%	75%	87%
Botafogo	25%	66%	97%
Ceará	18%	100%	100%
Corinthians	13%	82%	100%
Coritiba	13%	71%	81%
Cruzeiro	17%	75%	79%
Figueirense	100%	100%	100%
Flamengo	100%	84%	64%
Fluminense	25%	80%	100%
Grêmio	21%	94%	78%
Internacional	14%	66%	88%
Palmeiras	33%	63%	100%
Santos	25%	96%	90%
SãoPaulo	100%	100%	100%
Vasco	23%	82%	72%
Ef. Média	39%	82%	87%
DMU's Front	4	4	7
SF	0	0,90241	1
SDIS	1	1	0
S	0,5	0,95121	0,5

2ª Variável: RB

INPUT	-	PE
OUTPUT	RB	-
AméricaMG	80%	90%
AtléticoGO	100%	100%
AtléticoMG	100%	100%
AtléticoPR	83%	83%
Avai	71%	74%
Bahia	88%	87%
Botafogo	67%	100%
Ceará	100%	100%
Corinthians	87%	100%
Coritiba	71%	85%
Cruzeiro	77%	86%
Figueirense	100%	100%
Flamengo	100%	84%
Fluminense	80%	100%
Grêmio	94%	94%
Internacional	71%	92%
Palmeiras	63%	100%
Santos	96%	100%
SãoPaulo	100%	100%
Vasco	87%	82%
Ef. Média	86%	93%
DMU's Front	6	9
SF	0	1
SDIS	1	0
S	0,6	0,4

Tabela 7 – Seleção de Variáveis para o Modelo de Defesa

A Tabela 8 possui o resumo do resultado da seleção de variáveis para os diferentes modelos. Pode-se observar que as variáveis Roubada de Bola (RB), Falta Sofrida (FS) e Assistência (A) fazem parte de 2 modelos cada uma, porém para cada um destes modelos elas foram selecionadas em momentos diferentes, o que explicita a diferença entre os modelos proporcionada pelos diferentes objetivos e diferentes variáveis utilizadas e suas correlações.

MODELO	INPUT INICIAL	OUTPUT INICIAL	1ª VARIÁVEL ADICIONADA	2ª VARIÁVEL ADICIONADA
Defesa	Gol Sofrido (GS)	Defesa Difícil (DD)	Falta Cometida (FC)	Roubada de Bola (RB)
Ligação	Passe Errado (PE)	Roubada de Bola (RB)	Falta Sofrida (FS)	Assistência (A)
Ataque	Finalização para Fora (FF)	Gol (G)	Assistência (A)	Falta Sofrida (FS)

Tabela 8 – Variáveis selecionadas para cada modelo

8. Resultados

Uma vez definidas as variáveis mais relevantes para cada modelo, foi possível obter uma ordenação das equipes que possuem as melhores práticas em cada modelo. No intuito de apontar as equipes falsamente eficientes, foi calculada também para cada modelo a eficiência em relação a fronteira invertida, que representa as equipes que possuem práticas extremamente ineficientes. Esta aplicação consiste em realizar os cálculos considerando os Inputs como Outputs e os Outputs como Inputs, o que permite identificar as DMUs que estão tendo sucesso em realizar práticas indesejáveis.

A fim de apontar as equipes que possuem um bom equilíbrio em cada setor e também no conjunto de todos os setores, foi construída uma ordenação pela Eficiência Conjunta, que é a média simples da eficiência composta pelas fronteiras padrão e invertida de cada modelo. Os resultados estão representados na Tabela 9.

DMU	DEFESA			LIGAÇÃO			ATAQUE			EFICIÊNCIA CONJUNTA			Ordenação
	Padrão	Invertida	Composta*	Padrão	Invertida	Composta*	Padrão	Invertida	Composta*	Padrão	Invertida	Composta*	
Flamengo	100%	76%	92%	100%	100%	82%	100%	72%	92%	100%	83%	89%	1
SãoPaulo	100%	65%	100%	100%	100%	82%	100%	100%	72%	100%	88%	85%	2
Corinthians	87%	93%	70%	100%	84%	95%	100%	77%	89%	96%	85%	84%	3
Palmeiras	63%	100%	47%	100%	81%	97%	100%	61%	100%	88%	81%	81%	4
Fluminense	80%	88%	68%	97%	84%	92%	100%	88%	81%	92%	87%	80%	5
Vasco	87%	86%	75%	91%	100%	74%	100%	75%	90%	93%	87%	80%	6
Figueirense	100%	75%	92%	85%	87%	80%	75%	87%	63%	87%	83%	78%	7
AtléticoMG	100%	100%	74%	100%	100%	82%	100%	100%	72%	100%	100%	76%	8
Internacional	71%	91%	59%	93%	80%	92%	83%	80%	74%	82%	84%	75%	9
Botafogo	67%	100%	50%	100%	78%	100%	82%	77%	75%	83%	85%	75%	10
AméricaMG	80%	100%	59%	90%	82%	88%	87%	80%	77%	85%	87%	75%	11
Santos	96%	78%	88%	78%	100%	63%	100%	100%	72%	91%	93%	74%	12
AtléticoGO	100%	87%	84%	92%	100%	75%	78%	100%	56%	90%	96%	72%	13
Grêmio	94%	75%	88%	74%	99%	61%	83%	96%	63%	84%	90%	71%	14
Bahia	88%	100%	65%	79%	99%	65%	100%	87%	81%	89%	95%	71%	15
Cruzeiro	77%	83%	70%	88%	84%	85%	78%	100%	56%	81%	89%	70%	16
Coritiba	71%	100%	52%	82%	92%	74%	93%	75%	85%	82%	89%	70%	17
AtléticoPR	83%	93%	67%	89%	100%	73%	85%	100%	61%	86%	98%	67%	18
Ceará	100%	100%	74%	71%	100%	58%	77%	89%	63%	83%	96%	65%	19
Avai	71%	100%	53%	89%	99%	74%	81%	89%	67%	81%	96%	64%	20

Tabela 9 – Resultados

9. Interpretação dos Resultados

Pode-se observar que o Flamengo possui a maior Eficiência Conjunta, mesmo sem possuir a melhor Eficiência Composta em nenhum dos 3 modelos, o que mostra que a equipe possui equilíbrio em campo.

O São Paulo é a melhor equipe na Defesa, e mesmo sendo 100% eficiente na fronteira padrão da Ligaç o e Ataque,   tamb m 100% eficiente nas fronteiras invertidas, sendo um falso eficiente.

Palmeiras e Botafogo possuem a maior efici ncia conjunta no Ataque e Liga o respectivamente, mas ambos possuem baixa efici ncia na Defesa.   poss vel observar que ambos cometem muitas faltas e roubam poucas bolas, o que prejudica a estabilidade defensiva.

Este trabalho n o busca identificar qual equipe deve ser campe , mas permite que seja identificada a consist ncia das equipes em campo, e assim apontar aquelas que possuem mais chances de produzir resultados positivos com o tempo e assumirem o topo da tabela se forem eficientes, ou a parte inferior se possuem a t cnica em campo pouco apurada.

Com a medi o das vari veis apenas at  a 5  rodada,   plaus vel afirmar que uma equipe possa ter tido um bom desempenho em campo e obtido resultados ruins, ou o contr rio tamb m   verdade, quando equipes possuem um mau rendimento, mas acabam vencendo a partida. Resultados como estes s o comuns e fazem parte da imprevisibilidade do futebol. Mesmo assim, seria poss vel afirmar que as equipes tendem a obter resultados que condigam com seu desempenho m dio.

Na tabela 10 foi poss vel avaliar a consist ncia do modelo pelo comportamento do deslocamento das DMU's ao longo da tabela da 5  at  a 9  rodada. Ou seja, com os dados obtidos at  a 5  rodada, p de-se tra ar a tend ncia de evolu o da maioria das DMU's pela posi o que elas mereciam estar por seu desempenho em campo, que   justamente a Efici ncia Composta calculada.

DMU	Colocação 5ª		Desloc.	Colocação	Desloc.	Previsão
	Rodada	Tendência	Previsto	9ª Rodada	Real	
Flamengo	10	1	9	2	8	Tendência Correta
São Paulo	1	2	-1	3	-2	Tendência Correta
Corinthians	3	3	0	1	2	Tendência Correta
Palmeiras	2	4	-2	4	-2	Tendência Correta
Fluminense	11	5	6	9	2	Tendência Correta
Vasco	7	6	1	7	0	Tendência Correta
Figueirense	4	7	-3	8	-4	Tendência Correta
Atlético MG	5	8	-3	13	-8	Tendência Correta
Internacional	12	9	3	6	6	Tendência Correta
Botafogo	6	10	-4	5	1	Previsão Equivocada
América MG	15	11	4	18	-3	Previsão Equivocada
Santos	13	12	1	16	-3	Previsão Equivocada
Atlético GO	8	13	-5	17	-9	Tendência Correta
Grêmio	9	14	-5	11	-2	Tendência Correta
Bahia	14	15	-1	15	-1	Tendência Correta
Cruzeiro	18	16	2	10	8	Tendência Correta
Coritiba	16	17	-1	14	2	Tendência Correta
Atlético PR	19	18	1	20	-1	Tendência Correta
Ceará	17	19	-2	12	5	Previsão Equivocada
Avai	20	20	0	19	1	Tendência Correta

Tabela 10 – Análise de Consistência

Para 11 DMUs o deslocamento real da equipe ocorreu no sentido previsto, de melhora ou piora da posição na tabela. Como exemplo, o Flamengo encontrava-se apenas na 10ª colocação mesmo sendo a DMU melhor avaliada, e até a 9ª rodada produziu resultados com seu desempenho assumindo a 2ª colocação. O Atlético GO por sua vez estava na 9ª colocação mesmo sendo a 14ª DMU na ordenação do modelo, e na 9ª rodada caiu com seu desempenho para a 17ª colocação.

5 DMUs não tiveram o deslocamento previsto da maneira correta, porém seus deslocamentos foram pouco significativos e mesmo assim o modelo foi capaz de apontar se estas equipes deveriam estar no topo ou na parte de baixo da tabela do campeonato. Como exemplo pode ser apontado o Atlético PR que desceu uma colocação, enquanto o modelo previu sua subida em uma colocação. No entanto sua presença nas últimas colocações foi confirmada pelo modelo.

Apenas 4 DMUs tiveram uma previsão equivocada da tendência de evolução, o que pode ser explicado por eventos novos que ocorreram na equipe no decorrer destas 4 rodadas, como o Santos, que não dispunha de 3 de seus principais jogadores por estarem convocados para a seleção brasileira e assim não tiveram o aumento de rendimento previsto.

10. Considerações Finais

A utilização de DEA como método de avaliação do desempenho dentro de campo das equipes mostrou-se eficiente, por produzir resultados coerentes com os acontecimentos reais, podendo até mesmo prever tendências que não seriam óbvias com a mera avaliação da colocação das equipes na tabela.

O método multicritério de seleção de variáveis foi capaz de auxiliar a construção dos modelos garantindo uma boa discriminação sem comprometer a representatividade do modelo, o que ficou comprovado pela análise de consistência do modelo e dos resultados.

A aplicação recorrente do modelo com os dados de rodadas consecutivas pode auxiliar na identificação das lacunas de cada uma, permitindo a evolução contínua de uma equipe em busca de um desempenho técnico que a permita disputar títulos.

Para avaliar esta possibilidade, é necessária a realização de uma análise temporal em momentos distintos do campeonato com o cruzamento das ações reais aplicadas nas equipes entre estes períodos, como alterações táticas ou aquisições e vendas de jogadores.

11. Bibliografia

Angulo Meza L., Lins M. P. E., (2002). Review of Methods for Increasing Discrimination in Data Envelopment Analysis. *Annals of Operations Research*, 116, 225–242.

Banker, R.D., A. Charnes e W.W. Cooper (1984). Some models for estimating technical scale inefficiencies in Data Envelopment Analysis. *Management Science*, v.30, n.9, 1078-1092.

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

Cortez, L. C. S. (2010). Estudo da Eficiência dos times de futebol que mais investiram em jogadores para temporada 2008-2009. *Relatórios de Pesquisa em Engenharia de Produção*, V. 10, n. 02.

Lovell, C. A. K., Pastor, J. T. (1999). Radial DEA models without inputs or without outputs. *European Journal of Operational Research*, Vol. 118, 46-51.

Nanci, L. C., Senra, L. F. A., Soares De Mello J. C. C. B., Caldas, M. A. F. (2004). Uso de análise envoltória de dados para avaliar a distribuição de jornais. *Pesquisa Naval*, n. 17, 113-118.

Sánchez, I. M. G. (2007). Efficiency and effectiveness of Spanish football teams: a three-stage-DEA approach. *CEJOR*, 15:21-45.

Scheel, H. (2001). Undesirable outputs in efficiency evaluations. *European Journal of Operational Research*, v. 132, 400-410.

Senra, L. F. A., Nanci, L. C., Soares de Mello, J. C. C. B., Angulo Meza, L. (2007) Estudo sobre métodos de seleção de variáveis em DEA. *Pesquisa Operacional*, v. 27, n. 2, 191-207.

Soares de Mello, J.C.C.B., Angulo Meza, L., Gomes, E.G., Serapião, B.P., Estellita Lins, M.P. (2003). Análise de envoltória de dados no estudo da eficiência e dos benchmarks para companhias aéreas brasileiras. *Pesquisa Operacional*, 23(2), 325-345.

Soares de Mello, J.C.C.B., Gomes, E.G., Soares de Mello, M.H.C., Lins, M.P.E. (2002). Método Multicritério para Seleção de Variáveis em Modelos DEA. *Pesquisa Naval*, 15, 55-66.

Soares De Mello J. C. C. B., Gomes E. G., Angulo Meza L., Leta F. R. (2008). DEA Advanced Models for Geometric Evaluation of used Lathes. *WSEAS TRANSACTIONS on SYSTEMS*, v. 7.