

ESTUDO SOBRE MÉTODOS DE SELEÇÃO DE VARIÁVEIS EM DEA

Luis Felipe Aragão de Castro Senra

Programa de Doutorado em Administração de Empresas
IAG/PUC-Rio
Rio de Janeiro – RJ
luissenra@gmail.com

Luiz Cesar Nanci

Programa de Doutorado em Engenharia de Produção
Universidade Federal Fluminense (UFF)
Niterói – RJ
cnanci@predialnet.com.br

João Carlos Correia Baptista Soares de Mello*

Departamento de Engenharia de Produção
Universidade Federal Fluminense (UFF)
Niterói – RJ
jcsmello@producao.uff.br

Lidia Angulo Meza

Departamento Ciência dos Materiais
Universidade Federal Fluminense (UFF)
Volta Redonda – RJ
lidia@metal.eeimvr.uff.br

* *Corresponding author* / autor para quem as correspondências devem ser encaminhadas

Recebido em 08/2004; aceito em 03/2007 após 1 revisão
Received August 2004; accepted March 2007 after one revision

Resumo

Um dos pontos principais da modelagem em DEA é a escolha das variáveis a serem utilizadas. Esta escolha pode ter objetivos muitas vezes conflitantes, como aumentar a eficiência média proporcionada pelas variáveis utilizadas ou maximizar a capacidade de ordenação do modelo – uma clássica fragilidade em DEA. Neste artigo, são comparados quatro métodos de seleção de variáveis, com ênfase na ordenação das DMUs, aplicados a uma situação real de avaliação da eficiência de Operadores Logísticos na entrega domiciliar de jornais, na região metropolitana do Rio de Janeiro.

Palavras-chave: DEA; seleção de variáveis; ordenação.

Abstract

One of the main issues in DEA modeling is the variables choice. This may have conflictive objectives, like increasing the mean efficiency or maximizing the model ranking capability – a DEA classic fragility. In this paper, we compare four variable selection methods, focused on DMUs sorting. These methods are applied to a real situation of assessing the efficiency of third-party logistics in the activity of newspaper home delivery, in Rio de Janeiro.

Keywords: DEA; variable selection; ranking.

1. Introdução

A Análise Envoltória de Dados é um método não-paramétrico, utilizado para calcular a eficiência comparada de unidades de produção, chamadas DMUs (*Decision Making Units*). Sua origem data de 1978, no trabalho de E. Rhodes, sob a supervisão de W. W. Cooper. Nos últimos anos, vêm sendo desenvolvido por diversos pesquisadores de todo o mundo, e aumentado sua aceitação principalmente por se tratar de um método totalmente objetivo, sem precisar da opinião do decisor.

As variáveis necessárias para o cálculo da eficiência relativa das DMUs são divididas em *inputs* (entradas/insumos do sistema) e *outputs* (saídas/produtos do sistema). A eficiência relativa de cada DMU é definida como a razão da soma ponderada de seus produtos (*outputs*) pela soma ponderada dos insumos necessários para gerá-los (*inputs*). Em DEA, estas variáveis são ponderadas por pesos, calculados livremente ou de forma restrita através de programação linear, objetivando maximizar a eficiência de cada DMU em relação ao conjunto de referência.

Uma fragilidade clássica da DEA é a sua baixa capacidade de ordenar as DMUs, já que quanto maior o número de variáveis em relação ao número de DMUs, menor será a capacidade de ordenação pelas eficiências, já que há a tendência de muitas DMUs ficarem na fronteira (máxima eficiência). Um dos métodos usados para contornar este problema é restringir o número de variáveis usadas no modelo.

Neste artigo são revisados os métodos I-O Stepwise Exaustivo, o Método Multicritério, o Método Multicritério Combinatório Inicial e o Método Multicritério Combinatório Parcial, aqui proposto. Os quatro métodos são comparados quanto às suas principais características. Para ilustrar a aplicação dos modelos de seleção de variáveis supracitados, foi estudado o caso da entrega domiciliar de jornais na região metropolitana do Rio de Janeiro.

2. Métodos de Seleção de Variáveis em DEA

Na literatura DEA, pouco têm sido discutido sobre a problemática da seleção das variáveis utilizadas na modelagem. A grande maioria dos trabalhos publicados traz a abordagem da seleção de variáveis segundo a opinião de especialistas ou até mesmo da disponibilidade de dados. Thanassoulis (1996) ressaltou que a modificação do conjunto de variáveis selecionadas poderá ter grande impacto no resultado da avaliação. Sendo assim, torna-se extremamente relevante discutir e comparar métodos de seleção de variáveis em DEA. No entanto, antes de apresentar qualquer método deve-se atentar para o que significa realmente a seleção de variáveis.

Deve-se ter em conta que o fato de uma diferente escolha de variáveis conduzir a resultados diferentes não deve ser interpretado como uma fraqueza de DEA. Na verdade, escolher variáveis diferentes significa que se pretende levar em conta uma dimensão diferente do problema, ou seja, olhar para as DMUs segundo outro ponto de vista. Assim, por exemplo, em Soares de Mello *et al.* (2003), diferentes escolhas dos conjuntos de variáveis levaram a 3 diferentes avaliações de empresas aéreas: operacional, de vendas e total. A análise prévia de possíveis conjuntos de variáveis é uma etapa fundamental, a ser feita conjuntamente por decisores, especialistas e analistas. Só após esta escolha prévia é quem sentido pensar-se em métodos de seleção. Segundo Lins & Angulo Meza (2000), na maioria dos casos reais em que se dispõe de poucas variáveis e muitas DMUs, não se justifica a preocupação em utilizar

técnicas de seleção em variáveis. Neste caso o uso de todas as variáveis pré selecionadas não deve trazer grandes desvantagens ao modelo DEA. No entanto, caso o número de DMUs seja pequeno, o uso de um grande número de variáveis tira todo o sentido aos modelos DEA básicos. Se não se desejar usar modelos avançados, nem se consiga aumentar o número de DMUs, uma das opções é restringir as variáveis que vão entrar no modelo. Neste contexto, os métodos de seleção de variáveis devem ser vistos como instrumentos de auxílio à decisão, que orientarão a escolha final. Esta não deve ficar presa ao resultado de um modelo matemático, por mais sofisticado que seja. Sempre deve ser feita em conjunto pelos agentes de decisão, especialistas e analistas, que poderão (ou não) usar um método de seleção como ferramenta.

Os próprios métodos de seleção que serão comparados permitem diferentes graus de intervenção do agente de decisão. Aqueles que exigem pouca ou nenhuma informação subjetiva são mais adequados a decisores que estejam em grande dúvida e não consigam emitir opiniões. Outros, que exigem alguma escolha, devem ser usados por decisores que queiram impor alguma direção ao método, sem no entanto serem capazes de realizar uma escolha de variáveis sozinhos. Em qualquer método usado o decisor deve ser confrontado com o resultado e verificar a sua coerência. Por exemplo, deve verificar se não foi omitida nenhuma variável que considere imprescindível, e se há relação causal entre todos os pares *input output* selecionados. Pode ainda ser recomendável fazer a seleção por mais de um método para comparar resultados e tomar a decisão com mais embasamento.

Um método de seleção de variáveis foi proposto por Norman & Stoker (1991) que aliaram a análise de correlações simples à Análise Envoltória de Dados em um procedimento iterativo que simultaneamente identifica os *inputs* e *outputs* relevantes e calcula medidas de ineficiência dos planos de operação observados.

Já Golany & Roll (1989) enunciaram três estágios do processo de seleção de variáveis: (1) *Judgmental screening*, (2) *non-DEA quantitative analysis* e (3) *DEA based analysis*. O primeiro refere-se à distinção entre variáveis determinantes de eficiência e variáveis explicativas de ineficiência, que pode ser resolvida através de análise de causalidade. O segundo estágio sugere o uso de análise de regressão para determinar se uma variável deve ser *input* ou *output*. O terceiro estágio, sugerido por Charnes *et al.* (1978) pode ser utilizado para identificar as variáveis que deverão ser excluídas, que são aquelas que não agregam eficiência significativa ao modelo.

Lins & Moreira (1999) refinaram o método do Norman & Stoker (1991) propondo várias variantes do método que chamaram de IO-Stepwise. Soares de Mello *et al.* (2002b e 2004) baseados no trabalho de Lins & Moreira (1999) propuseram o método multicritério de seleção de variáveis. No presente artigo são propostos os métodos multicritério combinatório inicial para seleção de variáveis e o método multicritério combinatório por cenários para seleção de variáveis, que serão comparados com os métodos I-O Stepwise e Multicritério.

2.1 Método I-O Stepwise Exaustivo Completo

A família de métodos I-O Stepwise parte da premissa que a seleção de variáveis deve obedecer ao princípio de máxima relação causal entre *inputs* e *outputs*. Este é um método que se preocupa em aumentar a eficiência média com um número limitado de variáveis.

Baseia-se na observação de que algumas variáveis contribuem pouco para a eficiência média do modelo. Sendo assim, uma vez identificadas podem ser retiradas no modelo. Supõe ainda que o decisor consegue identificar previamente quais das variáveis podem ser *inputs* e quais podem ser *outputs*.

Tem como elemento decisor único a eficiência média do modelo. Exige pouca intervenção do agente de decisão, a única opinião que se pede dele é avaliar se o acréscimo da eficiência média pela inclusão de uma variável extra é significativo ou não.

O método pode ser descrito da forma apresentada na Figura 1.

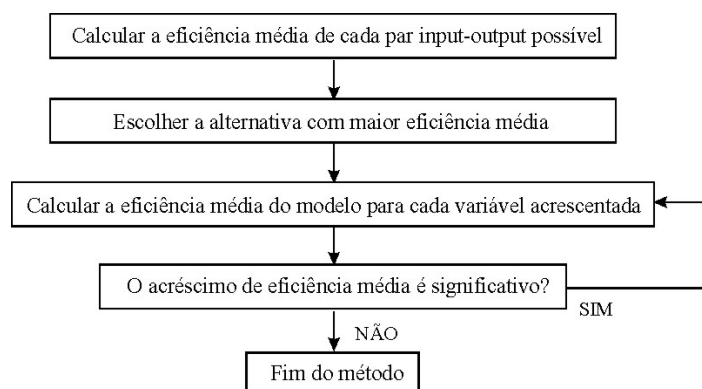


Figura 1 – Método I-O Stepwise exaustivo completo.

Brevemente, o método envolve os seguintes passos:

1. Calcular a eficiência média de cada par *input-output* possível. Nesse passo deve-se rodar $n \times m$ modelos DEA, isto é, um para cada par *input-output*. Para cada resultado calcula-se a eficiência média de todas as DMUs.
2. Escolher o par *input* e *output* inicial que gerou a maior eficiência média.
3. Uma vez de posse do par inicial, rodar modelo com mais uma variável, um para cada variável que ainda não foi incluída no modelo.
4. Calcular a eficiência média para cada variável acrescentada.
5. Escolher para entrar no modelo a variável que gerou a maior eficiência média.
6. Verificar se o aumento da eficiência foi significativo. Em caso afirmativo, repetir o passo três. Caso contrário, retirar a última variável incluída e finalizar o processo.

Tem a grande vantagem de manter relações causais, mas tem a desvantagem de não garantir poucas DMUs na fronteira e portanto corre o risco de fornecer baixa discriminação. Aplicações deste método podem ser encontradas em Lins & Moreira (1999) e Lins & Angulo-Meza (2000).

Repetir o passo 3 até que o acréscimo de eficiência não seja mais significativo. Com base no seu sentimento, o decisor deve escolher o momento em que o algoritmo pára.

2.2 Método Multicritério para Seleção de Variáveis em Modelos DEA

Baseados na observação anterior, Soares de Mello *et al.* (2002b) propuseram um método que combina a boa relação causal e boa discriminação entre as DMUs a que chamaram de método multicritério de seleção de variáveis. Este é um método com alta participação do decisor já que este deve fornecer um *input* e *output* que obrigatoriamente farão parte do modelo e ainda atribuir maior ou menor importância a cada um dos dois objetivos.

O método foi aplicado a um mesmo conjunto de dados usados por Soares de Mello *et al.* (2002a) para avaliar turmas de cálculo I em que foi usada a técnica de restrições aos pesos. Os dois resultados foram bastante semelhantes. Soares de Mello *et al.* (2004) fizeram nova aplicação do método aos dados usados por Angulo-Meza *et al.* (2003) para avaliar programas de pós-graduação da COPPE com avaliação cruzada. Novamente houve grande consistência de resultados. Estes dados viriam ainda ser usados por Soares de Mello *et al.* (2006) em um tipo de avaliação diferente mas cujos resultados mostraram coerência com os anteriores.

Neste método, a relação causal é medida através de um ajuste à fronteira, calculado pela eficiência média de todas as DMUs tal como no método I-O Stepwise. Já o poder de discriminação, leva em conta o número de DMUs com a mesma eficiência. Como é improvável que DMUs ineficientes tenham o mesmo índice de eficiência optou-se pela simplificação de considerar como descritor do poder de discriminação o número de DMUs na fronteira. Note-se que quanto maior este número, menor o poder discriminatório do modelo.

A Figura 2 mostra de uma forma simplificada o processo a seguir.

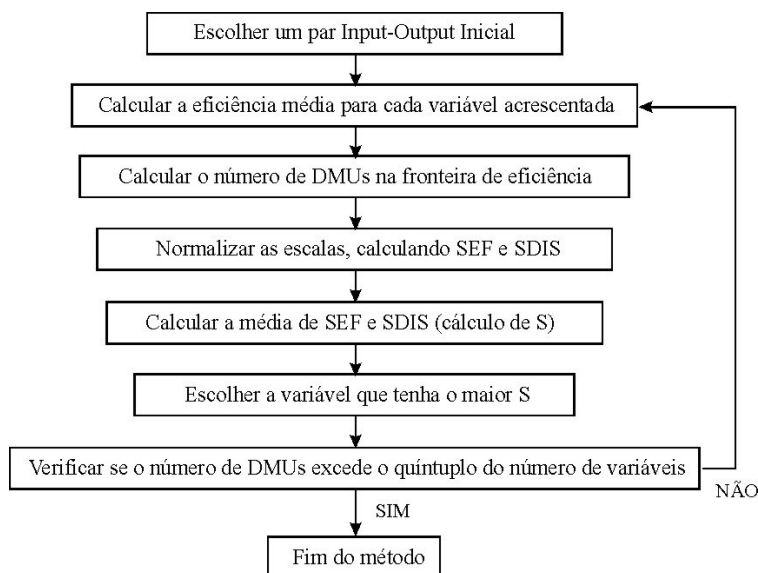


Figura 2 – Método Multicritério de Seleção de Variáveis.

Brevemente, o método envolve os seguintes passos:

1. Escolher um par *input-output* inicial. Nessa etapa o decisor escolhe qual é o melhor par *input-output* para iniciar a análise, tendo como base no seu conhecimento sobre o assunto e no tipo de medida que pretende fazer.
2. Calcular a eficiência média para cada variável acrescentada. Os valores obtidos das eficiências médias serão normalizados com o uso de uma interpolação que atribua valor 1 à maior eficiência média e 0 à menor. Este valor normalizado é chamado S_{EF} .
3. Contar o número de DMUs na fronteira para variável acrescentada. Este número será normalizado com o uso de uma interpolação que atribua valor 1 ao menor número de DMUs na fronteira e 0 ao maior número. Repare-se que esta normalização garante que quanto maior for o número de DMUs na fronteira menor será o poder discriminatório do modelo.
4. Fazer uma soma ponderada de S_{EF} e S_{DIS} . Com restrição de que a soma dos pesos seja igual a um, esta soma é equivalente a uma soma ponderada, cabe ao decisor escolher o peso de cada variável de acordo com o fator que ele julgar mais relevante. Foi assim definida a variável $S = \alpha S_{EF} + (1 - \alpha) S_{DIS}$. Repare-se que caso o decisor resolva $\alpha = 1$ o método fica quase idêntico ao método I-O Stepwise.
5. Escolher a variável que tenha maior valor de S . A alternativa que tiver o maior valor de S é considerada a alternativa que melhor concilia uma boa ordenação (alto valor de S_{DIS}) e uma boa relação causal (alto valor de S_{EF}) e será incluída no modelo.
6. Verificar se o número de DMUs excede o quádruplo do número de variáveis. Caso afirmativo deve-se encerrar a seleção de variáveis. Caso contrário, deve-se reiniciar o processo na etapa 2 incluindo mais uma variável. É importante ressaltar que a escolha de 1/5 é arbitrária e baseada em estudos de Gonzalez-Araya (2003).

Quando o método foi proposto usava-se como ponto de parada que o número de DMUs seria o triplo do número de variáveis. Na verdade, o ponto de parada deve ser deixado a critério do decisor, podendo este optar por estudar a situação com vários pontos de parada diferentes a fim de melhor embasar a sua decisão final sobre quais variáveis entraram no modelo.

2.3 Método Multicritério Combinatório Inicial para Seleção de Variáveis

Este método parte do princípio que o decisor não deseja, não é capaz ou é indiferente de emitir opiniões sobre o par inicial a ser incluído no modelo. Essa impossibilidade do decisor pode ser a indiferença entre as variáveis ou o desejo de ver que variáveis seriam selecionadas sem a sua interferência para melhor embasar a sua decisão final.

Desta forma, é introduzida uma alteração da forma de seleção do par inicial. Este é escolhido de forma semelhante ao método I-O Stepwise exaustivo completo, substituindo-se a eficiência média por $\alpha S_{EF} + (1 - \alpha) S_{DIS}$. No intuito de minimizar a quantidade de informação requerida do decisor este método já arbitra $\alpha = 0,5$.

A Figura 3 mostra o processo a ser seguido.

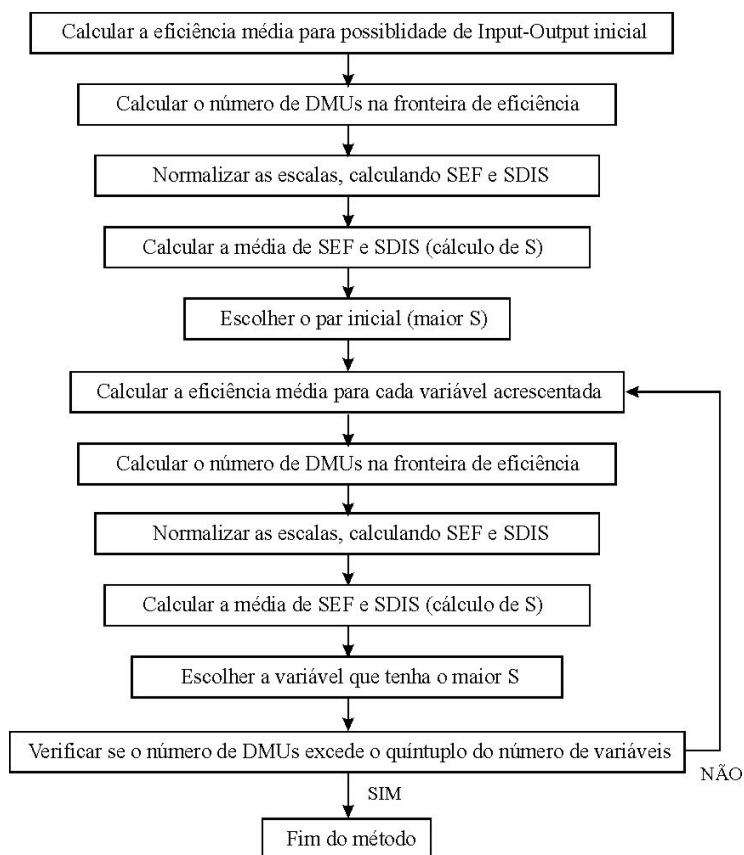


Figura 3 – Método Multicritério Combinatório Inicial de Seleção de Variáveis.

Os passos envolvidos neste método podem ser resumidos da seguinte forma:

1. Calcular a eficiência média para cada par *input-output* inicial.
2. Contar o número de *DMUs* na fronteira de eficiência.
3. Fazer uma normalização como descrito no método anterior obtendo as variáveis SEF e SDIS.
4. Fazer a média aritmética de SEF e SDIS obtendo a variável S.
5. Escolher como par inicial o de maior valor de S. Em caso de empate, considera-se a que tem maior S_{DIS} .

Os passos seguintes são análogos ao método anterior com $\alpha = 0,5$.

2.4 Método Multicritério Combinatório por Cenários para Seleção de Variáveis

Este método é uma evolução na linha de exigir menos informação ao decisor. Para isso, não há critério de parada do algoritmo, devendo ser todas as variáveis incluídas para em seguida ser feita a comparação entre os modelos com diferentes número de variáveis.

Este é um método em duas fases. A primeira fase constrói cenários que serão analisados na segunda. Chama-se cenário da primeira fase a o melhor modelo com duas, três, quatro, etc, variáveis. A segunda fase limita-se a escolha do melhor cenário.

O método pode ser descrito pelas Figuras 4 e 5.

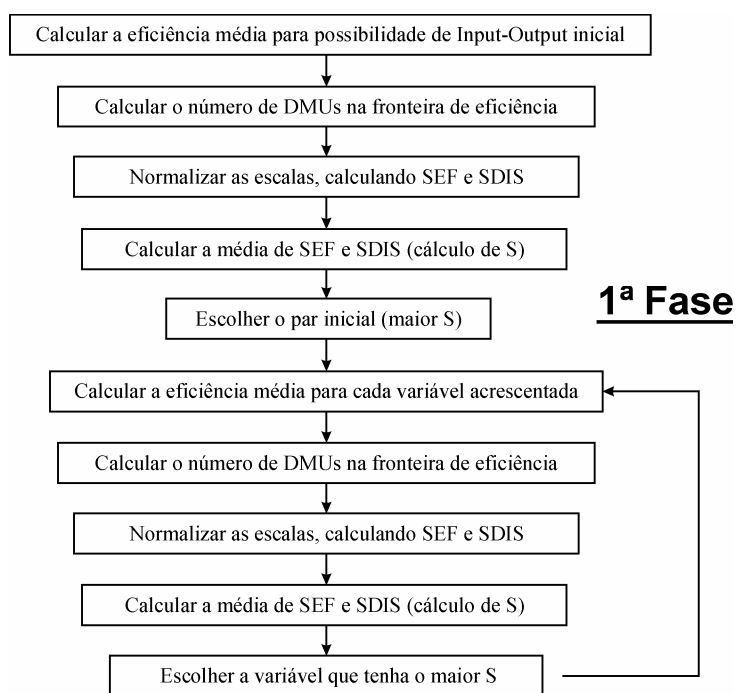


Figura 4 – 1ª fase do Método Multicritério Combinatório por Cenários de Seleção de Variáveis.

Esta fase é idêntica ao método anterior tendo como ponto de parada a inclusão de todas as variáveis no modelo.

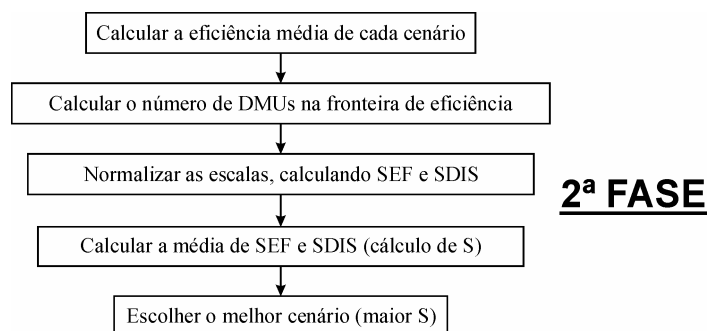


Figura 5 – 2ª fase do Método Multicritério Combinatório por Cenários de Seleção de Variáveis.

Esta segunda fase pode ser descrita da seguinte forma:

1. Calcular a eficiência média para cada cenário.
2. Contar o número de *DMUs* na fronteira de eficiência em cada cenário.
3. Normalizar as escalas, calculando S_{EF} e S_{DIS} , como no método multicritério.
4. Fazer a média aritmética de S_{EF} e S_{DIS} , obtendo o valor S .
5. Escolher o cenário que tiver o maior valor de S . No caso de empate considera-se a que tem maior S_{DIS} .

3. Estudo de Caso

A entrega domiciliar de jornais representa, do ponto de vista logístico, um enorme desafio, tanto pela particularidade desta atividade como pela dificuldade em se identificar as variáveis que impactam diretamente na eficiência dos operadores logísticos responsáveis pela entrega. Segundo Nanci *et al.* (2004), dentre os fatores considerados explicativos para índices de eficiência, pode-se citar o índice de verticalização das áreas e o grau de dispersão da entrega.

Neste estudo, aplicou-se os métodos de seleção de variáveis abordados neste artigo ao conjunto de 09 (nove) operadores logísticos (DMUs), responsáveis pela entrega domiciliar de jornais de uma empresa de comunicação do Rio de Janeiro, com foco em sua região metropolitana.

Dado o pequeno número de DMUs e a necessidade de identificar variáveis relevantes para o cálculo da eficiência, torna-se ainda mais relevante a utilização de técnicas de seleção de variáveis.

3.1 Modelo Utilizado

Apesar das empresas operarem no mesmo ambiente de negócios, as características geográficas das regiões onde atuam são muito diferentes, o que justificou a utilização do modelo BCC, que apresenta retornos variáveis de escala (Banker *et al.*, 1984).

Orientou-se o modelo a *input*, ou seja, deseja-se minimizar os recursos envolvidos na operação de entrega domiciliar, mantendo-se os *outputs* constantes.

Nanci *et al.* (2004) avaliaram a eficiência de operadores logísticos na entrega domiciliar, utilizando apenas variáveis indicadas pelos especialistas (decisores): número de entregadores, número de reclamações e número de jornais entregues. Neste estudo, além das variáveis supracitadas, outras foram incluídas com o objetivo de possibilitar a seleção de variáveis e determinar aquelas que melhor representam a modelagem da atividade de entrega domiciliar.

Tabela 1 – Variáveis identificadas na modelagem do Caso Prático.

Sigla	Inputs
EN	Entregadores
RO	Roteiros
JT	Janela de Tempo
RE	Reclamações
Sigla	Outputs
JO	Jornais entregues
PE	Pontos de Entrega

3.2 Aplicação do Método I/O Stepwise Exaustivo Completo

Aplicou-se a modelagem DEA para cada par *input-output* possível obtendo-se os resultados apresentados na Tabela 2.

Selecionou-se o par inicial **JT – JO**, pois foi o que apresentou a maior eficiência média (neste caso, obteve-se um empate com o par JT – PE). Recorreu-se à opinião do decisor, que considerou mais importante a variável JO em detrimento a PE. Dando seqüência ao método, comparou-se o acréscimo de eficiência média que cada variável agregou ao par inicial JT – JO anteriormente definido, conforme Tabela 3.

Tabela 2 – Comparação das opções do Par Inicial – Método I-O Stepwise.

<i>Input</i>	EN	RO	JT	RE	EN	RO	JT	RE
<i>Output</i>	PE	PE	PE	PE	JO	JO	JO	JO
DMU 1	0,6215	0,6240	1,0000	0,7923	0,5930	0,5964	1,0000	0,7721
DMU 2	0,7320	0,7404	0,7407	0,8633	0,6877	0,6981	0,7407	0,8401
DMU 3	0,5604	0,6186	0,8000	0,7693	0,5046	0,5598	0,8000	0,7314
DMU 4	0,5646	0,5422	0,9524	0,5280	0,5925	0,5680	0,9524	0,5427
DMU 5	0,4677	0,5333	0,8696	0,5888	0,4032	0,4627	0,8696	0,5453
DMU 6	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000
DMU 7	0,8184	0,6702	1,0000	0,7617	0,8935	0,7298	1,0000	0,8040
DMU 8	1,0000	1,0000	0,8333	1,0000	1,0000	1,0000	0,8333	1,0000
DMU 9	0,2823	0,5609	0,7407	0,5308	0,2538	0,5068	0,7407	0,5038
Eficiência Média	0,6719	0,6988	0,8819	0,7594	0,6587	0,6802	0,8819	0,7488

Tabela 3 – Escolha da terceira e quarta variáveis – Método I-O Stepwise.

<i>Input</i>	-	EN	RO	RE	<i>Input</i>	-	EN	RE
<i>Output</i>	PE	-	-	-	<i>Output</i>	PE	-	-
DMU 1	1,0000	1,0000	1,0000	1,0000	DMU 1	1,0000	1,0000	1,0000
DMU 2	0,7407	0,8330	0,8430	0,8567	DMU 2	0,8447	0,8430	0,8567
DMU 3	0,8000	0,8418	0,8608	0,8669	DMU 3	0,8638	0,8608	0,8669
DMU 4	0,9524	0,9524	0,9524	0,9524	DMU 4	0,9524	0,9524	0,9524
DMU 5	0,8696	0,8696	0,8784	0,8696	DMU 5	0,8784	0,8784	0,8784
DMU 6	1,0000	1,0000	1,0000	1,0000	DMU 6	1,0000	1,0000	1,0000
DMU 7	1,0000	1,0000	1,0000	1,0000	DMU 7	1,0000	1,0000	1,0000
DMU 8	0,8333	1,0000	1,0000	1,0000	DMU 8	1,0000	1,0000	1,0000
DMU 9	0,7407	0,7407	0,7908	0,7416	DMU 9	0,7936	0,7908	0,7908
Eficiência Média	0,8819	0,9153	0,9250	0,9208	Eficiência Média	0,9259	0,9250	0,9272
Acréscimo	0,0000	0,0334	0,0432	0,0389	Acréscimo	0,0008	0,0000	0,0022

Selecionou-se a variável **RO**, por apresentar o maior ganho de eficiência ao modelo. Dado que o acréscimo na eficiência média foi significativo (0,0432), prosseguimos com o método, fazendo a análise da inserção da quarta variável. Como se pode notar, a inserção da quarta variável representa apenas um pequeno acréscimo de eficiência média (0,0022), portanto o modelo pode ser bem representado apenas por 3 variáveis (**JT – RO – JO**). É importante citar o empate de 4 DMUs na fronteira de eficiência, o que demonstra baixa capacidade de ordenação do método.

3.3 Aplicação do Método Multicritério para Seleção de Variáveis em Modelos *DEA*

Escolheu-se o par inicial **EN-JO**, com a ajuda de um especialista da área, dando início ao método. Vale a pena ressaltar que a eficiência média deste par inicial foi de apenas 0,6587, considerada relativamente baixa. Em seguida comparou-se a adição da terceira variável ao par inicial, conforme Tabela 4.

Tabela 4 – Escolha da terceira variável – Método Multicritério.

<i>Input</i>	-	RO	JT	RE
<i>Output</i>	PE	-	-	-
DMU 1	0,6215	0,5964	1,0000	0,7721
DMU 2	0,7320	0,6981	0,8330	0,8401
DMU 3	0,5604	0,5598	0,8418	0,7314
DMU 4	0,5925	0,5925	0,9524	0,5925
DMU 5	0,4677	0,4627	0,8696	0,5453
DMU 6	1,0000	1,0000	1,0000	1,0000
DMU 7	0,8935	0,8935	1,0000	0,8935
DMU 8	1,0000	1,0000	1,0000	1,0000
DMU 9	0,2823	0,5068	0,7407	0,5038
Eficiência Média	0,6833	0,7011	0,9153	0,7643
DMUs na fronteira	2	2	4	2
SEF	0,0000	0,0766	1,0000	0,3492
SDIS	1	1	0	1
S	0,5000	0,5383	0,5000	0,6746

Escolheu-se a variável **RE**, que apresenta o maior valor de S, para fazer parte do modelo. Como o quádruplo do número de variáveis (15) excede o número de DMUs (9), encerra-se a aplicação do método. Nota-se que apenas duas DMUs encontram-se na fronteira de eficiência, evidenciando a boa capacidade de ordenação conseguida pelo método utilizado.

3.4 Aplicação do Método Multicritério Combinatório Inicial para Seleção de Variáveis em *DEA*

Inicia-se o método comparando todos os possíveis pares iniciais de variáveis, de forma normalizada (S), conforme Tabela 5.

Tabela 5 – Escolha do Par Inicial – Método Multicritério Combinatório Inicial.

<i>Input</i>	EN	RO	JT	RE	EN	RO	JT	RE
<i>Output</i>	PE	PE	PE	PE	JO	JO	JO	JO
DMU 1	0,6215	0,6240	1,0000	0,7923	0,5930	0,5964	1,0000	0,7721
DMU 2	0,7320	0,7404	0,7407	0,8633	0,6877	0,6981	0,7407	0,8401
DMU 3	0,5604	0,6186	0,8000	0,7693	0,5046	0,5598	0,8000	0,7314
DMU 4	0,5646	0,5422	0,9524	0,5280	0,5925	0,5680	0,9524	0,5427
DMU 5	0,4677	0,5333	0,8696	0,5888	0,4032	0,4627	0,8696	0,5453
DMU 6	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000
DMU 7	0,8184	0,6702	1,0000	0,7617	0,8935	0,7298	1,0000	0,8040
DMU 8	1,0000	1,0000	0,8333	1,0000	1,0000	1,0000	0,8333	1,0000
DMU 9	0,2823	0,5609	0,7407	0,5308	0,2538	0,5068	0,7407	0,5038
Eficiência Média	0,6719	0,6988	0,8819	0,7594	0,6587	0,6802	0,8819	0,7488
DMUs na fronteira	2	2	3	2	2	2	3	2
SEF	0,0591	0,1799	1,0000	0,4511	0,0000	0,0963	1,0000	0,4039
SDIS	1	1	0	1	1	1	0	1
S	0,5295	0,5899	0,5000	0,7255	0,5000	0,5481	0,5000	0,7019

Escolheu-se o par inicial **RE-PE** por apresentar o maior valor de S. É interessante destacar que o par inicial escolhido difere do escolhido pelo especialista no método anterior, apresentando uma eficiência média consideravelmente maior. Continuando com a aplicação do método, comparou-se a adição da terceira variável, apresentada na Tabela 6.

Tabela 6 – Escolha da terceira variável – Método Multicritério Combinatório Inicial.

<i>Input</i>	EN	RO	JT	-
<i>Output</i>	-	-	-	JO
DMU 1	0,7923	0,7923	1,0000	0,7923
DMU 2	0,8633	0,8633	0,8633	0,8633
DMU 3	0,7693	0,7693	0,8714	0,7693
DMU 4	0,5646	0,5422	0,9524	0,5427
DMU 5	0,5888	0,5888	0,8696	0,5888
DMU 6	1,0000	1,0000	1,0000	1,0000
DMU 7	0,8184	0,7617	1,0000	0,8040
DMU 8	1,0000	1,0000	1,0000	1,0000
DMU 9	0,5308	0,5609	0,7416	0,5308
Eficiência Média	0,7697	0,7643	0,9220	0,7657
DMUs na fronteira	2	2	4	2
SEF	0,0346	0,0000	1,0000	0,0090
SDIS	1	1	0	1
S	0,5173	0,5000	0,5000	0,5045

Escolheu-se a variável **EN** por apresentar o maior valor de *S*, encerrando assim o método, já que o quádruplo do número de variáveis (15) excede o número de DMUs (9). É interessante ressaltar a boa capacidade de ordenação alcançada, assim como no método anterior.

3.5 Aplicação do Método Multicritério Combinatório por Cenários para Seleção de Variáveis

Como a primeira fase é igual ao método combinatório inicial, mostrar-se-á aqui somente a segunda fase do método.

Considerando esta dinâmica tem-se ao final da primeira fase os seguintes cenários:

Tabela 7 – Cenários ao final da primeira fase – Método Multicritério Combinatório por Cenários.

Variáveis	2	3	4	5	6
<i>Inputs</i>	RE	RE EN	RE EN	RE EN RO	RE EN RO JT
<i>Outputs</i>	PE	PE	PE JO	PE JO	PE JO

Sendo assim dá-se início a segunda fase, de comparação entre os cenários, apresentada na Tabela 8.

Portanto, ao final deste método conclui-se que o melhor, que possibilita uma eficiência média alta e uma boa capacidade discriminatória é o cenário com cinco variáveis.

Tabela 8 – Segunda fase – Método Multicritério Combinatório por Cenários.

<i>Input</i>	Cenários – Número de Variáveis				
<i>Output</i>	2	3	4	5	6
DMU 1	0,7923	0,7923	0,7923	0,7923	1,0000
DMU 2	0,8633	0,8633	0,8633	0,8633	0,8633
DMU 3	0,7693	0,7693	0,7693	0,7693	0,8714
DMU 4	0,5280	0,5646	0,5925	0,5925	0,9524
DMU 5	0,5888	0,5888	0,5888	0,5888	0,8784
DMU 6	1,0000	1,0000	1,0000	1,0000	1,0000
DMU 7	0,7617	0,8184	0,8935	0,8935	1,0000
DMU 8	1,0000	1,0000	1,0000	1,0000	1,0000
DMU 9	0,5308	0,5308	0,5308	0,5609	0,7936
Eficiência Média	0,7594	0,7697	0,7812	0,7845	0,9288
DMUs na fronteira	2	2	2	2	4
SEF	0,0000	0,0612	0,1287	0,1484	1,0000
SDIS	1	1	1	1	0
S	0,5000	0,5306	0,5644	0,5742	0,5000

4. Análise dos Resultados

A Tabela 9 mostra as variáveis selecionadas em cada método, a eficiência média alcançada e a ordenação obtida.

Tabela 9 – Comparação de Resultados.

		I/O Stepwise	Multicritério	Combinatório Inicial	Combinatório por Cenários
Variáveis Escolhidas	Inputs	JT, RO	EN, RE	EN, RE	RE EN RO
	Outputs	JO	JO	PE	PE JO
Eficiência média		0,925	0,7643	0,7697	0,7845
DMUs na fronteira		1, 6, 7, 8	6, 8	6, 8	6, 8
Ordenação das demais DMUs		4, 5, 3, 2, 9	7, 2, 1, 3, 4, 5, 9	2, 7, 1, 3, 5, 4, 9	2, 7, 1, 3, 5, 4, 9

Destaca-se que os métodos atingiram seus propósitos. O primeiro atingiu uma alta eficiência média, e os outros três uma boa ordenação das DMUs, sem muito prejuízo na relação causal.

Observa-se o fato de que a variável JO, escolhida pelo decisor, não se faz presente no rol de variáveis selecionadas pelo método combinatório. Este fato deve ser levado ao decisor para que ele verifique se mantém a opinião inicial ou se deseja insistir na entrada da variável JO.

O método por cenários selecionou um cenário com cinco das seis variáveis do modelo completo com apenas duas DMUs na fronteira, possibilitando um modelo com mais variáveis (portanto mais explicativo), sem perda de discriminação.

5. Conclusão

Observando os resultados obtidos e tendo em vista o propósito de cada modelo, verificaram-se as seguintes vantagens e desvantagens da utilização de cada método:

Método I-O Stepwise exaustivo completo

Vantagens:

1. Chega-se rapidamente a uma eficiência média elevada;
2. Em modelos com número de variáveis superior ao recomendado (1 variável para cada 5 DMUs), é um método aconselhado para descartar variáveis que agregam pouca ou nenhuma eficiência média ao modelo (menos representativas) e destacar as variáveis que agregam maior eficiência média (mais representativas).

Desvantagens:

1. Não é adequado para aumentar a capacidade discriminatória do modelo;
2. Por não considerar a capacidade discriminatória, pode gerar um número elevado de DMUs na fronteira;
3. Para instâncias com grande número de DMUs e variáveis, exige um custo de cálculo elevado.

Método Multicritério

Vantagens:

1. Possibilita uma eficiência média relativamente alta com boa capacidade de ordenação, através de uma ponderação entre a capacidade discriminatória e a eficiência média do modelo;
2. Por utilizar a opinião do decisor para a escolha do par inicial, apresenta custo de cálculo inferior aos demais métodos apresentados;
3. Possibilita incorporar a opinião do decisor (o que pode também ser visto como desvantagem, como mostrado a seguir).

Desvantagens:

1. Pode não chegar a uma eficiência média elevada;
2. Exige opinião do decisor (especialista) para iniciar o processo de escolha das variáveis, o que pode ser tendencioso e levar a resultados distintos, dependendo do decisor;
3. Pode apresentar baixa eficiência média no par inicial.

Método Multicritério Combinatório Inicial

Vantagens:

1. Possibilita uma eficiência média relativamente alta com boa capacidade de ordenação, através de uma ponderação entre a capacidade discriminatória e a eficiência média do modelo;
2. Independe da opinião do decisor para escolha do par inicial, que é obtido pelo mesmo processo de ponderação supracitado;
3. Desde o par inicial, apresenta eficiência média relativamente alta.

Desvantagens:

1. Pode não chegar a uma eficiência média elevada;
2. Para instâncias com grande número de DMUs e variáveis, exige um custo de cálculo elevado;
3. Não considera nenhuma opinião do decisor, podendo levar a um modelo irreal.

Método Multicritério Combinatório por Cenários

Vantagens:

1. Possibilita uma eficiência média relativamente alta com boa capacidade de ordenação, através de uma ponderação entre a capacidade discriminatória e a eficiência média do modelo;
2. Independe da opinião do decisor para escolha do par inicial, que é obtido pelo mesmo processo de ponderação supracitado;
3. Desde o par inicial, apresenta eficiência média relativamente alta;
4. Seleciona um cenário próximo ao modelo completo.

Desvantagens:

1. Pode não chegar a uma eficiência média elevada;
2. Para instâncias com grande número de DMUs e variáveis, exige um custo de cálculo elevado.

Mesmo com a boa performance do método Combinatório por Cenários, aqui proposto, ele é apenas uma simplificação de um método completo, denominado Método Multicritério Combinatório Total por Cenários (Senra, 2004). Este método faria a comparação de todas as possíveis combinações de variáveis. A escolha seria feita pela mesma variável S do método multicritério. No entanto este método apresenta alto esforço computacional, razão que motivou a sua exclusão deste artigo.

A aplicação dos métodos a um caso prático foi de extrema valia, pois foi possível validar os resultados alcançados com especialistas da área e reafirmar a metodologia DEA como ferramenta de gestão.

Deve-se destacar que, caso se opte por usar todas as variáveis iniciais, sem nenhuma seleção, podem ser usados outros métodos de aumento de discriminação. Revisões destes métodos podem ser visto em Adler *et al.* (2002) ou Angulo Meza & Lins (2002).

Referências Bibliográficas

- (1) Adler, N.; Friedman, L. & Sinuany-Stern, Z. (2002). Review of ranking methods in the data envelopment analysis context. *European Journal of the Operational Research*, **140**(2), 249-265.
- (2) Angulo-Meza, L.; Gomes, E.G.; Biondi Neto, L. & Coelho, P.H.G. (2003). Avaliação do Ensino nos cursos de pós-graduação em engenharia: um enfoque quantitativo de avaliação em conjunto. *Engevista*, **5**(9), 41-49.
- (3) 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.
- (4) Banker, R.D.; Charnes, A. & Cooper, W.W. (1984). Some models for estimating Technical and Scale inefficiencies in Data Envelopment Analysis. *Management Science*, **30**(9), 1078-1092.
- (5) Charnes, A.; Cooper, W.W. & Rhodes, E. (1978). Measuring the efficiency of decision-making units. *European Journal of Operational Research*, **2**, 429-444.
- (6) Golany, B. & Roll, Y. (1989). An application procedure for DEA. *Omega International Journal of Management Science*, **17**(3), 237-1250.
- (7) Gonzalez-Araya, M.C. (2003). Projeções Não Radiais em regiões fortemente eficientes da fronteira DEA – Algoritmos e aplicações. Tese – Programa de Engenharia de Produção, Universidade Federal do Rio de Janeiro/COPPE.
- (8) Lins, M.P.E. & Angulo Meza, L. (2000). *Análise Envoltória de Dados e perspectivas de integração no ambiente do Apoio à Decisão*. Editora da COOPE/UFRJ, Rio de Janeiro.
- (9) Lins, M.P.E. & Moreira, M.C.B. (1999). Método I-O Stepwise para Seleção de Variáveis em Modelos de Análise Envoltória de Dados. *Pesquisa Operacional*, **19**(1), 39-50.
- (10) Nanci, L.C.; Azevedo, S.M. & Soares de Mello, J.C.C.B. (2004). Estudo da Eficiência de Empresas Distribuidoras de Jornais Usando Análise Envoltória de Dados. *Produto & Produção*, **7**(3), 27-35.

- (11) Norman, M. & Stoker, B. (1981). *Data Envelopment Analysis: The Assessment of Performance*. J. Wiley and Sons.
- (12) Senra, L.F.A.C. (2004). Métodos de Seleção de Variáveis em DEA: Estudo de caso no setor elétrico brasileiro. Dissertação de mestrado, Programa de Engenharia de Produção, Universidade Federal Fluminense.
- (13) Senra, L.F.A.C. & Soares de Mello, J.C.C.B. (2004) Uso de Técnicas de Seleção de Variáveis em DEA para analisar o Setor Elétrico. *Relatórios de Pesquisa em Engenharia de Produção*, **4**(4).
- (14) 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.
- (15) Soares de Mello, J.C.C.B.; Estellita Lins, M.P.; Soares de Mello, M.H.C. & Gomes, E.G. (2002a). Evaluating the performance of calculus classes using operational research tools. *European Journal of Engineering Education*, **27**(2), 209-218.
- (16) Soares de Mello, J.C.C.B.; Gomes, E.G.; Angulo Meza, L. & Lins, M.P.E. (2004). Selección de variables para el incremento del poder de discriminación de los modelos DEA. *Revista de la Escuela de Perfeccionamiento En Investigación Operativa*, **24**, 40-52.
- (17) Soares de Mello, J.C.C.B.; Gomes, E.G.; Angulo Meza, L.; Soares de Mello, M.H.C. & Soares de Mello, A.J.R. (2006). Engineering post-graduate programmes: a quality and productivity analysis. *Studies in Educational Evaluation*, **32**(2), 136-152.
- (18) Soares de Mello, J.C.C.B.; Gomes, E.G.; Soares de Mello, M.H.C. & Lins, M.P.E. (2002b). Método Multicritério para Seleção de Variáveis em Modelos DEA. *Pesquisa Naval*, **15**, 55-66.
- (19) Thanassoulis, E. (1996). Assessing the efficiency of schools with pupils of different ability using Data Envelopment Analysis. *Journal of the Operational Research Society*, **47**(1), 84-97.