

Implementação Computacional de Seleção de Variáveis em DEA: um Estudo de Caso em Avaliação Educacional

Bruno Tonioni Cunha
*João Carlos Correia Baptista Soares de Mello**
Lidia Angulo Meza
Universidade Federal Fluminense, RJ, Brasil

Resumo

O Método Multicritério de Seleção de Variáveis em DEA, embora de procedimento simples, pode tornar-se moroso na ausência de uma ferramenta computacional adequada. Dependendo do tamanho do problema faz-se necessária uma implementação computacional para garantir a rapidez de resposta na escolha das variáveis. O artigo apresenta um módulo complementar ao software SIAD, com a opção Avançada “Seleção de Variáveis”. Este desenvolvimento computacional é aplicado a uma avaliação educacional de programas de Engenharia de Produção.

Palavras-chave: *Software, Seleção de Variáveis, Análise Envoltória de dados*

1. INTRODUÇÃO

O poder discriminatório da Análise Envoltória de Dados (DEA) em problemas que contenham elevado número de variáveis pode ser muito fraco, o que dificulta o uso de DEA como uma ferramenta de Multicritério. Além das técnicas de aumento de discriminação em DEA, uma outra possibilidade é restringir o número de variáveis usadas, sem perder a relação causal mas diminuindo o número de DMUs eficientes. Um dos métodos que permite fazer essa redução é o “Método Multicritério de Seleção de variáveis em DEA” ([7] e [8]). Este método, embora de procedimento simples, pode tornar-se moroso na ausência de uma ferramenta computacional adequada, uma vez que é necessário rodar vários modelos DEA, cada um composto um número de problemas de programação linear igual ao número de DMUs.

Os softwares de DEA não estão preparados para o uso desta técnica, dado o seu recente desenvolvimento, o que obriga ao uso iterado do software, em conjunção com uma planilha de cálculo e algum trabalho manual para implementar o método. Dada a versatilidade do software SIAD, foi desenvolvida uma ferramenta que lhe permite efetuar a seleção multicritério de variáveis de forma automática sem comprometer o tempo de resposta do software.

2. SIAD - SISTEMA INTEGRADO DE APOIO À DECISÃO

* Endereço: Universidade Federal Fluminense - Departamento de Engenharia de Produção, Rua Passo da Pátria 156, 22210-240, Niterói, RJ, Brasil. E-mails: htonioni@yahoo.com.br; jcsmello@producao.uff.br ; lidia@metal.eeimvr.uff.br

O software SIAD ([1] e [2]) foi desenvolvido principalmente para resolver os problemas de programação linear da Análise Envoltória de Dados e adicionalmente métodos multicritério de apoio à decisão. Ele tem-se mostrado muito útil, pois fornece resultados completos (índices de eficiência, *benchmarks*, pesos e alvos) dos modelos DEA clássicos e ainda possui a incorporação de modelos DEA mais avançados, como fronteira invertida e restrições aos pesos. A interface do software assim como os modelos e os algoritmos de solução, foram desenvolvidos em Delphi, plataforma Windows.

Neste software foi implementado o método de seleção de variáveis que será descrito brevemente a seguir.

3.0 MÉTODO MULTICRITÉRIO DE SELEÇÃO DE VARIÁVEIS

3.1. Descrição do Método

Este método baseia-se no fato observado de que, quando usada uma grande quantidade de variáveis e uma quantidade pequena de unidades, atribui-se 100% de eficiência a um grande número de DMUs. Assim, é feita uma escolha prévia de variáveis que podem entrar na avaliação.

O método aqui utilizado foi o proposto inicialmente em [7], com variações propostas em [5] e [6], que foi denominado “Método Multicritério de Seleção de Variáveis em DEA”.

Resumidamente, este método considera duas medidas: o melhor ajuste à fronteira e a máxima discriminação. O ajuste à fronteira é medido através da eficiência média e sua normalização cria a variável S_{EF} , que descreve o impacto do critério “melhor ajuste à fronteira” e que atinge o valor 10 na eficiência máxima e zero na eficiência mínima. A máxima discriminação é medida pelo número de DMUs eficientes, que se deseja minimizar. A normalização desta medida gera a variável S_{DIS} , que descreve de impacto do critério “máxima discriminação” e que atinge valor 10 para o menor número de DMUs na fronteira e zero para o maior número.

Para selecionar a variável que produz a melhor solução de compromisso e que deve entrar no modelo, os dois critérios são agregados em uma variável S da seguinte forma $S = \omega S_{EF} + (1 - \omega) S_{DIS}$. Para não haver predominância de um dos critérios considera-se $\omega = 0,5$.

A variável que apresentar maior valor de S é incorporada ao modelo e, se o número de variáveis incorporadas não exceder 1/3 das DMUs, repete-se o processo para incorporar uma nova variável. Caso contrário, o procedimento está completo, ou seja todas as variáveis que devem ser incorporadas ao modelo estão selecionadas.

3.2. Implementação computacional no Software SIAD

Existem versões e melhoras posteriores para a Seleção de Variáveis [5] e [6] já mencionadas anteriormente, mas a versão escolhida para implementação no SIAD foi a apresentada em [7] e [8] que parte de um par inicial *input/output* escolhido pelo decisor ou especialista. Este método preserva, na sua fase inicial, a necessidade do conhecimento do especialista, mantendo-se assim o compromisso entre técnica e arte na modelagem.

A seguir apresenta-se o algoritmo da implementação:

Passo 1: Escolha do par inicial *Input-Output*; o decisor ou especialista escolhe qual é o melhor par de *Input-Output*. Designou-se que o par *Input-Output* deverá estar, respectivamente, no campo do *Input* 1, na primeira coluna e no campo *Output* 1, a primeira coluna dos *outputs*.

Passo 2: Identifica-se o número de DMUs, número de *Inputs*, número de *Outputs*, o modelo DEA utilizado e sua orientação (*input* ou *output*).

Passo 3: Com o par inicial *Input-Output* selecionado. Calcula-se a eficiência de cada DMU com as variáveis definidas no passo 1. Dentre as variáveis ainda não escolhidas, inclui-se dentro do modelo uma por vez e calculam-se as eficiências das DMUs para cada grupo de variáveis (inicia-se com um trio de variáveis).

Passo 4: Para cada conjunto de variáveis do passo anterior determina-se o número de DMU's na fronteira de eficiência e a eficiência média.

Passo 5: Para cada conjunto de variáveis do passo anterior, calcula-se o S_{EF} e o S_{DIS}

S_{EF} ⇒ descritor do impacto “melhor ajuste à fronteira”, que atinge valor 10(dez) na eficiência máxima normalizada e zero na eficiência mínima normalizada.

S_{DIS} ⇒ descritor de impacto do critério “máxima discriminação”, que atinge valor 10 para o menor número de DMUs na fronteira normalizada e zero para o maior número.

e calcula-se S, a soma ponderada de S_{EF} e o S_{DIS} : $S = \omega S_{EF} + (1-\omega) S_{DIS}$.

{Aqui, trata-se de dois objetivos conflitantes. O maior valor médio da eficiência e o menor número de DMU's eficientes. Quanto maior a média da eficiência melhor o ajuste à fronteira e quanto menor o numero de DMU's eficientes, melhor a discriminação. E como resultado, uma melhor ordenação.}

Passo 6: Se o número de DMUs não excede o triplo do número de variáveis **Então** deve-se reiniciar o processo do *Passo 3*, acrescentando no modelo, a variável que apresentou o maior S do *Passo 3*. **Senão** deve-se encerrar a escolha das variáveis.

4. DESCRIÇÃO DO MODO AVANÇADO “SELEÇÃO DE VARIÁVEIS”.

Após inserir os dados no SIAD, escolher a opção Avançada “Seleção de Variáveis”. Além disso, devem ter sido escolhidos o modelo (CCR ou BCC) e a orientação (*input* ou *output*) de preferência. Logo após, será apresentada uma nova tela com as seguintes informações: modelo DEA utilizado, orientação e a localização do par *Input-Output*. A figura 1 apresenta esta tela.

Figura 1: Tela de informações

Ao clicar no botão Calcular será realizado o resultado final do procedimento apresentado no algoritmo da seção anterior. Ele levará a janela de Resultados, que apresenta os resultados completos com as variáveis escolhidas pela Seleção de Variáveis.

Para saber quais das variáveis foram consideradas para análise, basta escolher a opção Pesos. A figura 2 apresenta esta opção.

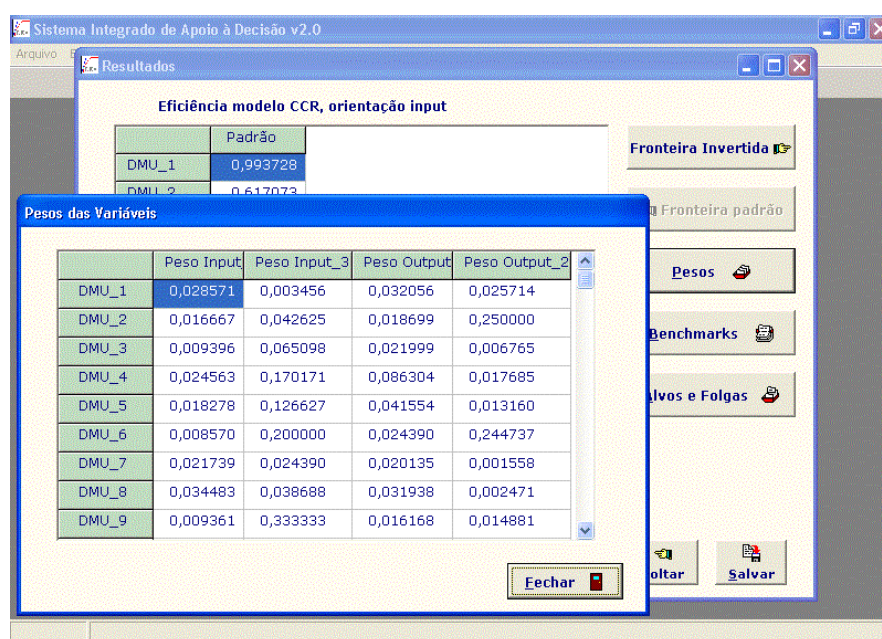


Figura 2: Apresentação das variáveis escolhidas pelo modelo.

Os resultados completos incluem a eficiência padrão, a eficiência usando fronteira invertida, o índice composto (eficiência padrão e invertida), os pesos das variáveis utilizadas, os *benchmarks* e os alvos para as DMUs (que incluem as folgas e os níveis que as variáveis devem atingir para as DMUs serem eficientes).

Computacionalmente, a seleção de variáveis não representa um esforço adicional nem de tempo nem de recursos computacionais.

5. ESTUDO DE CASO

O objetivo do Estudo de Caso é demonstrar a dificuldade de ordenação quando utiliza-se o método clássico de DEA e como se pode melhorá-la com a utilização do método multicritério de seleção de variáveis.

A base de dados estudada foi retirada do CAPES (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior [3]) que tem como objetivos: a) avaliação da pós-graduação *stricto sensu*; b) acesso e divulgação da produção científica; c) investimentos na formação de recursos de alto nível no país e exterior; d) promoção da cooperação científica internacional.

Foram utilizados dados correspondentes ao ano de 2004 e as universidades selecionadas para o estudo foram: Universidade de São Paulo, Universidade Federal do Rio de Janeiro, Universidade Federal Fluminense, Pontifícia Universidade Católica do Rio de Janeiro, Universidade Estadual do Norte Fluminense Darcy Ribeiro, Universidade Federal de Minas Gerais, Universidade Federal de São Carlos, Universidade Metodista de Piracicaba, Universidade Paulista, Universidade Federal de Itajubá, Universidade de São Paulo/São Carlos. Foram escolhidas universidades da região do sudeste devido à disponibilidade dos dados das mesmas e no modelo cada universidade recebeu uma numeração, porque o intuito do artigo não é por em conflito a capacidade das mesmas. O cenário propõe a análise das eficiências dos cursos de pós-graduação *stricto sensu* das universidades no que diz respeito a produção bibliográfica de cada curso.

Assim, este estudo de caso possui onze DMU's. Propõe-se o uso de 3(três) *inputs*: total de docentes, linhas de pesquisas e projetos de pesquisa; e 6 (seis) *outputs*: artigos em periódicos, trabalhos em anais (completos e resumos), livros (textos e capítulos) e artigos em jornais.

O modelo DEA utilizado foi o CCR [4] orientado ao *output*, pois deseja-se aumentar o nível de publicações, em que será difundido o conhecimento técnico e a diversidade de conceitos, enquanto são mantidos os *inputs* constantes.

6. ANÁLISE DOS RESULTADOS

6.1 Análise pelo método clássico de DEA

A Tabela 1 apresenta os resultados utilizando o modelo CCR orientação *output* utilizando-se o SIAD [1] e [2] sem a utilização da opção “Seleção de Variáveis”.

Tabela 1: Resposta do modelo DEA utilizando todas as variáveis

<i>DMU</i>	<i>Eficiências</i>
IES- 01	1,00
IES- 02	1,00
IES- 03	0,58
IES- 04	1,00
IES- 05	0,57
IES- 06	1,00
IES- 07	1,00
IES- 08	1,00
IES- 09	1,00
IES- 10	1,00
IES- 11	1,00

Nota-se que, tal como relatado na introdução, a fragilidade de ordenação do método DEA clássico quando há múltiplas variáveis em relação ao número de DMU's. Verifica-se que nove das onze DMU's são eficientes o que impossibilita uma ordenação e a determinação da melhor e da pior unidade.

Procede-se agora a utilizar a seleção de variáveis para melhorar a discriminação em DEA e possibilidade de se fazer uma ordenação das unidades. Após a execução do software com a opção Seleção de Variáveis, foi obtido o seguinte resultado. Como *inputs* as variáveis Projeto de Pesquisa e Linhas de Pesquisa; como *outputs*, Artigos em Periódicos e Artigos em Jornais e Revistas. A Tabela 2 mostra o resultado das eficiências usando a implementação computacional realizada no SIAD.

Tabela 2: Resultado do modelo DEA após seleção de variáveis

<i>DMU</i>	<i>Eficiências</i>
IES- 01	0,99
IES- 02	0,61
IES- 03	0,19
IES- 04	0,17
IES- 05	0,41
IES- 06	0,87
IES- 07	1,00
IES- 08	0,69
IES- 09	0,41

IES- 10	0,67
IES- 11	0,48

Aqui pôde-se observar que a melhor unidade foi a IES 07 que corresponde a Universidade de São Paulo e a IES de pior desempenho foi a 04, Universidade Estadual do Norte Fluminense Darcy Ribeiro. Nota-se que com a seleção de variáveis foram escolhidas variáveis tal que permitem fazer uma ordenação das IES de melhor para pior.

7. CONSIDERAÇÕES FINAIS

Por ser um método em que há necessidade de várias verificações e ciclos, ocorre uma lentidão dos cálculos. Com a implementação conseguiu-se uma maior rapidez e confiabilidade nas respostas do método.

O artigo também abre precedentes para desenvolvimentos futuros, tais como:

- Conjunção do método com as outras opções do SIAD, como por exemplo, utilizar o método seleção de variáveis com o método Avaliação Cruzada
- Diferentes abordagens quanto à escolha do par *Input-Output*, permitindo ao usuário opções de entrada das variáveis (automática ou manual).

Pode-se notar a facilidade de ordenação, após a utilização do método, das IES, permitindo uma classificação entre as mesmas mais claramente.

8. AGRADECIMENTOS:

Ao CNPq, pelo apoio financeiro, por meio do Edital CNPq 19/2004 – Universal, processo n.º 480628/2004-1 e n.º 502885/2004-1

9. BIBLIOGRAFIA

- [1] L. Angulo-Meza.; L. Biondi Neto; J. C. C. B. Soares de Mello; E. G. Gomes & P. H. G. Coelho. *FSDA – Free Software for Decision Analysis (SLAD – Software Livre de Apoio a Decisão)*, Rio de Janeiro, 2004.
- [2] L. Angulo-Meza; M. P. E. Lins.: “Data envelopment analysis na determinação da eficiência dos programas de pós-graduação da coppe/ufrrj”. Anais do XXXI Simpósio Brasileiro de Pesquisa Operacional – Minas Gerais, 1999.
- [3] CAPES - Coordenação de Aperfeiçoamento de Pessoal de Nível Superior; www.capes.org.br, 2005
- [4] A. Charnes, W. W. Cooper, E. Rhodes, E., *Measuring the Efficiency of Decision-Making Units*, European Journal of Operational Research, vol. 2, pp. 429-444, 1978
- [5] L.F.A.C Senra, J.C.C.B. Soares De Mello. Uso de Técnicas de Seleção de Variáveis em DEA para Analisar o Setor Elétrico, Rio de Janeiro, 2003.
- [6] L.F.A.C Senra, L.C. Nanci, L. Angulo-Meza. “Revisão dos métodos Totais de Seleção de Variáveis em DEA”. Anais do XXXVII SBPO – Simpósio Brasileiro de Pesquisa Operacional, Gramado, 2005.
- [7] J. C. C. B. Soares de Mello; E. G. Gomes; M. H. C. Soares De Mello; M. P. Estellita Lins; *Método Multicritério para Seleção de Variáveis em Modelos DEA*, Revista de Pesquisa Naval, Rio de Janeiro, v.15, p. 55-66, 2002.
- [8] J. C. C. B. Soares de Mello; E. G. Gomes; M. P. Estellita Lins, L. Angulo-Meza. “Selección de variables para el incremento del poder de discriminación de los modelos DEA”. Revista EPIO, Escuela de Perfeccionamiento En Investigación Operativa. Buenos Aires: , n. 24, pp.40-52, 2004.