



Aplicação de método de seleção de variáveis em um modelo DEA na produção de bovinos de corte

Application of variables selection method in a DEA model in beef cattle production

Adriano Pinto de Freitas¹

Marcelo Xavier Guterres²

Vinicius do Nascimento Lampert³

Ana Helena Soares da Silva⁴

Júlio Otavio Jardim Barcellos⁵

Pedro Rocha Marques⁶

Resumo: Os modelos de análise de produção por estimação de fronteira não paramétrica avaliam a eficiência relativa de unidades produtivas. O método de maior abrangência na literatura é a Análise por Envoltória de Dados (DEA). A análise de eficiência realizada pela técnica de Análise Envoltória de Dados (DEA) exige especial atenção na seleção das variáveis do modelo. Este trabalho incorporou uma matriz de referência como etapa preliminar na aplicação do Método Combinatório Multicritério Inicial (MCMCI) para seleção de variáveis para a análise de eficiência de sistemas de produção de bovinos de corte. Esta matriz busca, com auxílio de especialistas, refinar um método de seleção de variáveis, para o fim de reduzir ou extinguir a seleção de variáveis inúteis em uma futura investigação de eficiência. O método proposto para seleção das variáveis do modelo DEA foi aplicado em uma problemática abordada em oito municípios da região oeste do estado do Rio Grande do Sul, com a finalidade de avaliar o desempenho de sistemas de produção de bovinos de corte, com ciclo de produção completo. As variáveis resultantes da aplicação

¹ Unipampa – Universidade Federal do Pampa

² Unipampa – Universidade Federal do Pampa

³ Embrapa – Empresa Brasileira de Pesquisa Agropecuária

⁴ UFRGS – Universidade Federal do Rio Grande do Sul

⁵ UFRGS – Universidade Federal do Rio Grande do Sul

⁶ UFRGS – Universidade Federal do Rio Grande do Sul

mostraram um alinhamento com os pontos de maior importância zootécnica para análise de eficiência de sistemas de produção de bovinos de corte que são as características reprodutivas, de crescimento e de sobrevivência dos animais. Esta técnica se mostrou importante no sentido de considerar a opinião de especialistas na etapa de pré-seleção de variáveis a fim de complementar o método clássico de redução e descarte de variáveis.

Palavras-chave: DEA; Eficiência; Pecuária de Corte.

Abstract: Nonparametric frontier estimation assesses relative efficiency of production units. The method with a larger scope in the literature is Data Envelopment Analysis (DEA). Efficiency analysis evaluated by DEA demands special attention when selecting the variables of the model. This paper incorporated a reference matrix as preliminary step in the implementation of a variable selection method, to measure the efficiency of beef cattle production systems. With the assistance of specialists, this matrix aims refine a variable selection method, to reduce or put out the selection of worthless variables to a future efficiency analysis. Proposed method for DEA variables selection was scoped in eight cities in the western region of Rio Grande do Sul state, to evaluate the performance of complete lifecycle beef cattle production systems. The variables resulting from the application demonstrated an alignment with greatest zootechnics importance points for beef cattle production systems efficiency estimation, reproductive, growth and survival characteristics of the animals in the production system. This method demonstrated to be important at supplementing the classical method of reduction and elimination of variables, by attending specialists judgment on the preliminary step.

Keywords: DEA; Efficiency; Beef Cattle.

1. Introdução

Os sistemas de produção de bovinos de corte brasileiro têm sido desafiados a evoluir em eficiência e produtividade. Esta afirmação decorre de uma série de mudanças no sentido socioeconômico e ambiental, entre estas estão o aumento dos preços das terras e a demanda global por alimento Lampert et al. (2012). Neste contexto, ferramentas matemáticas de análise de eficiência são essenciais para os atores deste sistema produtivo.

Existe uma variedade de ferramentas para a mensuração da eficiência. Estas podem ser segmentadas em dois grupos, as de modelagem econométrica e de programação matemática. Os modelos do último caso são chamados de Análise por Envoltória de Dados (*Data Envelopment Analysis – DEA*). O uso de DEA tem se destacado em diversos setores produtivos, e a sua aplicação na pecuária auxilia na identificação das fontes de ineficiência, sendo um suporte relevante aos gestores no processo de tomada de decisão (Gomes et al. 2004).

A análise de eficiência realizada pela técnica DEA exige especial atenção na seleção das variáveis à serem utilizadas no processo de modelagem matemática. Segundo Gomes et al. (2009), a etapa de seleção das variáveis em DEA é de grande importância, pois dela depende a qualidade dos resultados.

Assim, a literatura indica que diferentes pesquisas vêm sendo realizadas para avaliar e propor métodos de seleção de variáveis para aplicação em modelos DEA. Destaca-se Lins et al. (1999) que propuseram o método chamado *IO-Stepwise*, bem como o Método Multicritério Combinatório Inicial (MMCI), aplicado no presente estudo, proposto em Senra et al. (2007).

Mediante o cenário apresentado, é proposto neste estudo a seleção das variáveis para mensuração da eficiência de sistemas de produção de bovino de corte na região oeste do estado do Rio Grande do Sul. Para tal seleção foi aplicado o Método Combinatório Multicritério Inicial (MCMI) adicionado à sugestão de uma etapa preliminar de seleção. Esta etapa, se dá com a incorporação de uma Matriz de Referência inicial antes à aplicação do MCMI. A utilização desta matriz se mostra importante, pois, busca refinar o processo com o fim de reduzir ou extinguir a seleção de variáveis inúteis em uma futura investigação de eficiência.

2. Materiais e métodos

A grande maioria dos trabalhos publicados na literatura, de análise de eficiência por meio da DEA, não aborda a temática da seleção de variáveis. Nesta perspectiva, o rumo que tem se tomado é o de seleção segundo a opinião do aplicador ou simplesmente da disponibilidade de dados. Conforme Thanassoulis (1996) a alteração do conjunto de variáveis afeta diretamente o

resultado final da DEA. Isto demonstra necessário discutir metodologias que definam modos de seleção de variáveis.

Os resultados dos escores de eficiência obtidos por meio da metodologia DEA, estão estreitamente ligados ao conjunto de variáveis utilizadas na análise. Contrapondo isto, é relativamente pequena a atenção despendida a como estas variáveis devem ser selecionadas em situações reais, conforme destaca Wagner e Shimshak (2007).

Norman e Stoker (1991) propuseram um método de seleção dos *inputs* e *outputs* por intermédio da adição de variáveis ao modelo DEA uma a uma progressivamente. Iniciaram com um modelo simples de um output e um input. Então as eficiências de todas DMU's (*Decision Making Units*) foram calculadas. Todas as variáveis não presentes no modelo foram correlacionadas com a eficiência calculada. Eles afirmaram que a alta correlação estatística era um indicador de que aquela variável em particular influenciaria no nível da performance. Logo, uma nova variável foi adicionada, utilizando os valores da correlação como critério. O processo foi repetido até que não restasse variáveis que pudessem influenciar no modelo.

Em outra abordagem Jenkins e Anderson (2003), utilizaram de uma metodologia baseada em análises de correlação e regressão para rastrear as variáveis que futuramente não contribuiriam com o modelo DEA, e assim não fazer o uso das mesmas. Também buscaram neste rastreamento minimizar a perda de informações. Semelhante a outros estudos prévios, os autores perceberam que os resultados obtidos com a metodologia DEA podem variar radicalmente, de acordo com o grupo de variáveis selecionadas para o modelo.

O método proposto para seleção das variáveis do modelo DEA foi empregado na problemática abordada em Marques et al. (2011). O qual desenvolveu um estudo em oito municípios da região oeste do estado do Rio Grande do Sul, com a finalidade de avaliar o desempenho de sistemas de produção de bovinos de corte, com ciclo de produção completo. O banco de dados inicial é composto por 48 propriedades (DMU's) e 51 variáveis, sendo 28 inputs e 23 outputs. Foram descartados deste, 13 propriedades por não possuírem dados em algumas variáveis em análise. Logo, o banco de dados inicial era composto por 35 DMU's.

O método utilizado na seleção das variáveis no presente estudo foi o MCMI. Porém, antes da aplicação do mesmo foi realizada uma pré-seleção das variáveis candidatas, assim, de maneira geral o processo de seleção das variáveis empregado no estudo em descrição, pode ser resumido conforme o fluxo indicado pela Figura 01.



Figura 1: Fluxo do processo de seleção de variáveis

2.1. Matriz de indicadores de referência

Como critério inicial foi utilizado uma matriz de indicadores do sistema de pecuária de corte. Esta classifica previamente as variáveis na pré-seleção do processo de seleção de variáveis. A matriz de indicadores de referência considerou os aspectos de importância, precisão, dificuldade de coleta, tipo de coleta e quem pode coletar as variáveis.

O indicador importância busca avaliar a relevância das variáveis do banco de dados, para uma análise de eficiência futura. Precisão avalia o banco de dados na ótica da qualidade dos dados obtidos. E por fim, a dificuldade de coleta, avalia o esforço a ser despendido para a coleta das informações.

O banco de dados também foi avaliado em tipo de coleta, sendo automática, manual ou ambas. E quem pode realizar a coleta se produtor, técnico ou ambos.

A construção desta matriz foi realizada com a participação de empresas de consultoria e técnicos de campo. Após a construção da Matriz cada variável do banco de dados foi julgada usando os critérios estabelecidos na mesma. Na Tabela 1 estão expostos e descritos os indicadores, assim como a escala de pontuação dos mesmos.

A construção desta matriz surgiu a partir da necessidade de filtrar as variáveis do banco de dados inicial. Visto que, a utilização do método de seleção em sua forma bruta, absorvia para o modelo variáveis com irrelevância na aplicação real. A inclusão desta etapa resguardou o processo sentido de não selecionar variáveis apenas pela característica matemática dos dados levantados.

Tabela 1: Matriz de indicadores de referência

Indicadores	Escala
Importância	1 à 5
Precisão	1 à 5
Dificuldade de Coleta	1 à 5
Tipo de Coleta	1: Automática; 2:Manual; 3:Ambas
Quem pode coletar?	1:Produtor; 2: Técnicos; 3: Ambos

Fonte: Autores

As variáveis com escore total dos indicadores superior a 9 foram pré-selecionadas, e neste trabalho foram desconsiderados os indicadores tipo de coleta e quem pode coletar.

Dando seguimento ao fluxo do processo, é finalizada a seleção das variáveis a serem utilizadas no modelo DEA ao aplicar o Método Combinatório Multicritério Inicial (MCMI).

2.2. Método combinatório multicritério inicial

Este método tem por princípio combinar uma boa relação causal entre as variáveis e a discriminação entre as DMU's. A relação causal é medida por meio do ajuste a fronteira de eficiência, obtido pela eficiência média das DMU's. Já a capacidade de discriminação é feita via análise da quantidade de DMU's sobre a fronteira (100 por cento eficientes).

A eficiência média é calculada e normalizada em uma escala onde a maior eficiência média vale um e a menor zero. Então o valor é armazenado em variável chamada S_{ef} . O mesmo é feito com o número de DMU's na fronteira, onde a escala utilizada é o maior número de DMU's na fronteira vale zero e o menor vale um. A variável que recebe este é chamada S_{dis} . Obtido esses dois valores calcula-se o índice S , que é a média aritmética de S_{ef} e S_{dis} . Este processo é realizado inicialmente entre todos os possíveis pares Input/Output, e o par com maior valor de S é selecionado. Após vai se repetindo a cada inclusão individual de variável, até que o número de variáveis desejado seja obtido. Na Figura 2 é apresentado o método em forma de fluxograma.

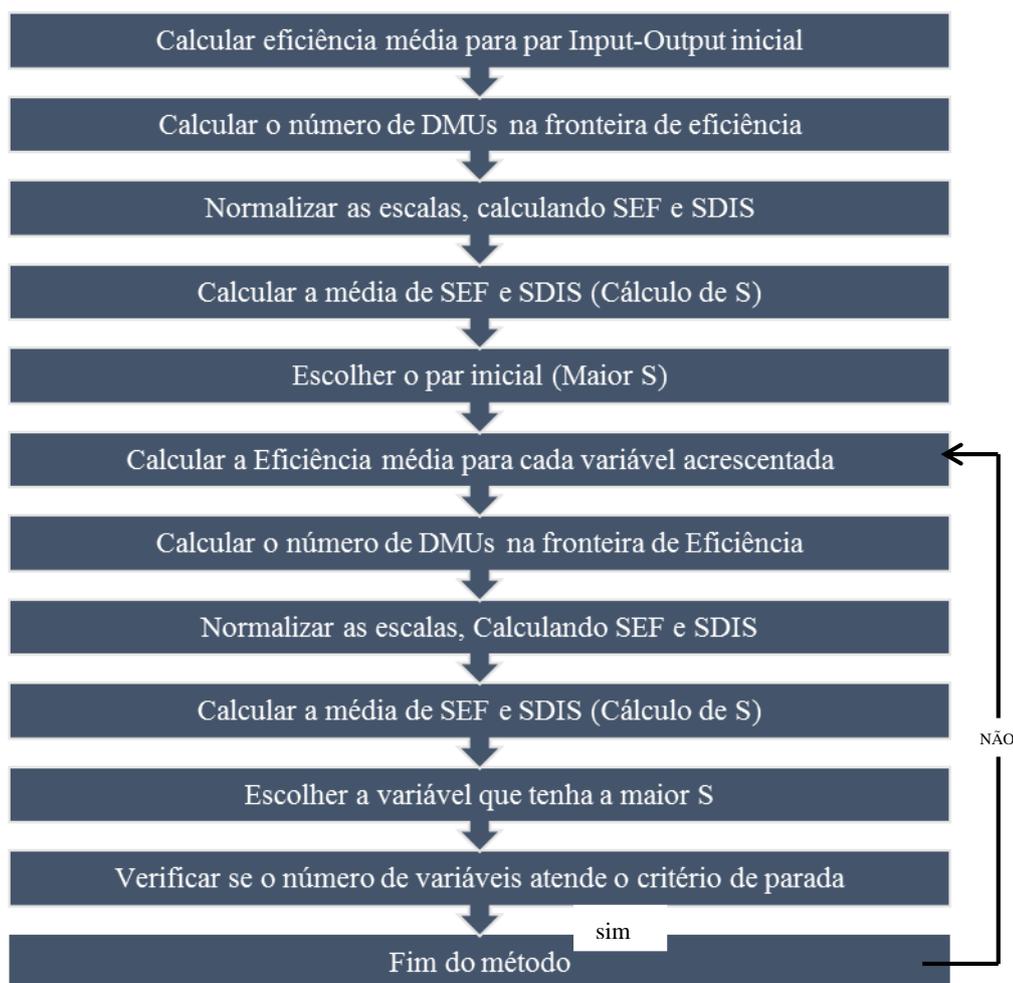


Figura 2: Método Combinatório Multicritério Inicial

3. Resultados e discussões

Ao final da pré-seleção foram eliminadas 22 das 51 variáveis do banco de dados inicial. Entretanto esta eliminação não se destaca apenas no quesito de redução numérica das variáveis, como também na eliminação das variáveis que não seriam úteis em um futuro modelo de investigação da eficiência das propriedades rurais.

Após o transcurso total do fluxo de seleção das variáveis resultou-se numa redução do número de variáveis do banco de dados inicial de 51 para 7 variáveis finais. Reduziu-se ainda de 86% para 14% o número de DMU's sobre a fronteira. Isto representa um aumento da capacidade de ordenamento do modelo DEA, pois menor a quantidade de DMU's sobre a fronteira menores as possibilidades de empate nos escores de eficiência relativa calculados. A Tabela 2 resume as características do banco de dados inicial, e após a aplicação do método de seleção de variáveis.

Tabela 2: Resumo dos resultados

Descrição	Antes	Após
Total de Variáveis	51	7
Relação Variáveis/DMU	1,06	0,14
Proporção de DMU's na fronteira	86%	14%

Fonte: Autores

A redução do número de variáveis a 0,14, ou seja, 1/7 do número de DMU's, vai de encontro com Gomes et al. (2004), o qual recomendam que o número de DMU's deve ser de no mínimo 4 a 5 vezes o número de variáveis. No Quadro 1 são apresentadas as variáveis resultantes ao final do método.

Quadro 1: Variáveis Resultantes

Variáveis	Input/Output
Lotação Geral (UA/ha)	Input
Lotação Pastagens (UA/ha)	Input
GMD em Pastagens (Gramas)	Output
GMD em Suplementação (Gramas)	Output
NDT do Suplemento (%)	Input
Taxa de Sobrevivência Geral (%)	Output
Taxa de Desmama (%)	Output

Fonte: Autores

4. Conclusão

As variáveis resultantes da aplicação do MCMI mostraram um alinhamento com o conhecimento de especialistas da área. Destaca-se que o método de seleção atingiu os pontos de maior importância para análise de eficiência de sistemas de produção de bovinos de corte que são as características reprodutivas, de crescimento e de sobrevivência.

A incorporação da Matriz de referência teve essencial relevância na seleção das variáveis do estudo ao descartar variáveis que não teriam as características necessárias para futura análise da eficiência das propriedades.

Visto que os objetivos do método para seleção das variáveis eram reduzir os números das variáveis, mantendo uma eficiência média elevada das DMU's e a capacidade de ordená-las, os resultados foram satisfatórios. Esta técnica mostra-se também importante no sentido de considerar a opinião de especialistas na etapa de pré-seleção de variáveis a fim de complementar o método clássico de redução e descarte de variáveis.

5. Referências

GOMES, E. G., MANGABEIRA, JA de C. Uso de análise de envoltória de dados em agricultura: o caso de Holambra. *Engevista*, v. 6, n. 1, p. 19-27, 2004.

GOMES, E. G.; GREGO, C. R.; DE MELLO, J. C. C. B. S.; VALLADARES, G. S.; MANGABEIRA, J. A. C.; DE MIRANDA, E. E. Dependência espacial da eficiência do uso da terra em assentamento rural na Amazônia. *Produção*, v. 19, n. 2. p. 417-432, 2009.

JENKINS, Larry, ANDERSON, Murray. A multivariate statistical approach to reducing the number of variables in data envelopment analysis. *European Journal of Operational Research*, v. 147, n. 1, p. 51-61, 2003.

LAMPERT, V.N., BARCELLOS, J.O.J., KLIEMANN NETO, F.J., CANELLAS, L.C., DILL, M.D. and CANOZZI, M.E.A., 2012. Development and application of a bioeconomic efficiency index for beef cattle production in Rio Grande do Sul, Brazil, *Revista Brasileira de Zootecnia*, 41, 775–782.

LINS, M. P. E.; MOREIRA, M. C. B. Método IO stepwise para seleção de variáveis em modelos de análise envoltória de dados. *Pesquisa Operacional*, v. 19, n. 1, p. 39-50, 1999.

MARQUES, P. R., BARCELLOS, J. O. J., MCMANUS, C., OAIGEN, R. P., COLLARES, F. C., CANOZZI, M. E. A., LAMPERT, V. N. Competitiveness of beef farming in Rio Grande do Sul State, Brazil. *Agricultural Systems*, v. 104, n. 9, p. 689-693, 2011.

NORMAN, Michael; STOKER, Barry. Data envelopment analysis: the assessment of performance. *John Wiley & Sons, Inc.*, 1991.

SENRA, L. F. A. D. C., NANJI, L. C., MELLO, J. C. C. B. S., MEZA, L. A. Estudo sobre métodos de seleção de variáveis em DEA. *Pesquisa Operacional*, v. 27, n. 2, p. 191-207, 2007.

THANASSOULIS, Emmanuel; BOUSSOFIANE, A., DYSON, R. G. A comparison of data envelopment analysis and ratio analysis as tools for performance assessment. *Omega*, v. 24, n. 3, p. 229-244, 1996.

WAGNER, Janet M., SHIMSHAK, Daniel G. Stepwise selection of variables in data envelopment analysis: Procedures and managerial perspectives. *European Journal of Operational Research*, v. 180, n. 1, p. 57-67, 2007.