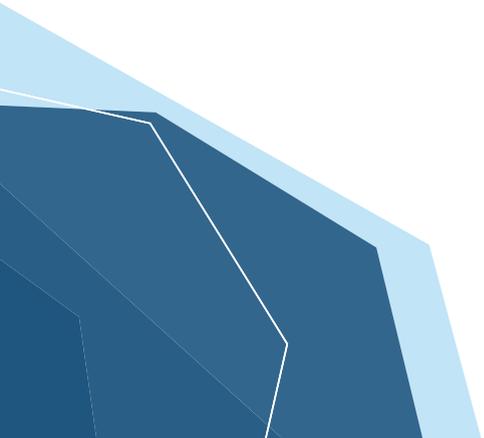


**AR
TI
GOS**

articles



EFICIÊNCIA NA LOCAÇÃO DE LOCAIS PARA A REALIZAÇÃO DE PROVAS USANDO DEA

EFFICIENCY RATING IN TEST SITES FOR PERFORMING TESTS USING DEA

ÍNDICE DE EFICIENCIA EN SITIOS DE PRUEBA PARA REALIZAR PRUEBAS UTILIZANDO DEA

Roberto Rosa da Silveira Junior¹

Daniel Lins Rodriguez²

João Carlos Félix Souza³

1 Possui graduação em Computação pela Universidade de Brasília (UnB). Especialista em Administração de Sistemas de Informação pela Universidade Federal de Lavras (UFLA). Mestre em Computação Aplicada com foco em Gestão de Riscos pela Universidade de Brasília (UnB). Atualmente é Assessor de Direção do Centro Brasileiro de Pesquisa em Avaliação e Seleção e de Promoção de Eventos (Cebbraspe). Possui certificação PMP (Project Management Professional) concedida mediante prova de habilidades pelo PMI (Project Management Institute). E-mail: noverts@hotmail.com.

2 Possui graduação em Ciências da Computação pelo Centro Universitário de João Pessoa - UNIPÊ - PB, é especialista em Tecnologia da Informação pela Coordenação de Extensão do Centro de Informática na Universidade Federal de Pernambuco - UFPE, possui formação complementar em Engenharia de Software, com ênfase em Análise de Testes promovida pela UFPE e mestre em Computação Aplicada pela Universidade de Brasília - UnB. E-mail: daniellinsr@gmail.com.

3 Possui graduação e mestrado em Estatística e Métodos Quantitativos pela Universidade de Brasília (UnB), pós-graduação em Informática pela Universidade Católica de Brasília (UCB), doutorado em Economia pela Universidade de Brasília (UnB) e Pós-Doutorado na FEUC (Faculdade de Economia da Universidade de Coimbra - Portugal). É Chefe de Departamento e Professor Associado do Departamento de Engenharia de Produção da Faculdade de Tecnologia e Engenharia da Universidade de Brasília. Tem experiência profissional como analista em avaliação de desempenho, eficiência e modelagem em Gestão de Risco de instituições financeiras. Foi gerente nas áreas de Pesquisa de Mercado e Estratégia, Controladoria e na Diretoria de Risco do Banco do Brasil. Foi Coordenador do grupo de modelagem de Risco Operacional da FEBRABAN. Atualmente trabalha nas áreas de Gestão de Riscos, Engenharia Econômica e Gestão de Projetos. Atua, principalmente, nos seguintes temas e linhas de pesquisa: sistema financeiro, gestão de riscos e risco financeiro, gestão de desempenho, eficiência e produtividade, modelos de Opções Reais e Teoria dos Jogos para avaliação de projetos de infraestrutura e controle de qualidade. Disciplinas ministradas: estatística, engenharia econômica, gestão de risco, controle estatístico de processos, projetos em sistemas de produção com utilização do PBL e PjBL. Pesquisa e projetos em andamento: aplicação dos modelos de risco em instituições financeiras e não financeiras; gestão de riscos em processos de fraudes; modelos de opções reais e teoria dos jogos com aplicação em infraestrutura de transporte. É professor

RESUMO

A medição de eficiência de unidades organizacionais constitui um fator importante nas sociedades contemporâneas. Para a obtenção de estimativas sobre eficiência, existem várias metodologias disponíveis. Um dos métodos mais conhecidos, e aqui analisado, consiste na técnica não paramétrica chamada DEA (Análise Envoltória de Dados). Dessa forma, é apresentada uma pesquisa aplicada contemplando as abordagens DEA e *Stepwise* em um banco de dados de uma instituição que realiza pesquisas em educação e concursos. Um ponto fundamental na utilização da DEA é a escolha das variáveis a serem utilizadas. No presente trabalho, a técnica *Stepwise* foi aplicada, através da execução do algoritmo *Stepwise* AIC na ferramenta R, como processo de escolha de variáveis, que, por sua vez, são utilizadas no modelo de Regressão Linear Múltipla para comprovar se o teste é confiável (através do valor de r^2) e posteriormente utilizadas na DEA. O resultado encontrado é uma relação das localidades que podem ser utilizadas para a realização de provas e sua eficiência técnica (ou locativa) relativa, podendo contribuir na gestão de locação destas localidades.

Palavras-chave: medição de eficiência; técnica não paramétrica; *stepwise*.

ABSTRACT

The efficiency measurement of organizational units is an important factor in contemporary societies. To obtain estimates of efficiency there are several methodologies available. One of the most known methods, and discussed here, is the nonparametric technique called DEA (Data Envelopment Analysis). Thus, we present a study applied with applications of DEA techniques and *Stepwise* regression in the database of an institution that conducts research in education. A key point in the use of DEA is the choice of variables to be used. In the present work, the *Stepwise* technique was applied, through the execution of the *Stepwise* AIC algorithm in the R tool, as a process for choosing variables, which in turn are used in the Multiple Linear Regression model to prove whether the test is reliable (through the r^2 value) and later used in the DEA. The results found are a list of locations and their relative technical efficiency (or allocative).

Keywords: measuring efficiency; nonparametric technique; *stepwise*.

e pesquisador no Programa de Pós-graduação em Computação Aplicada (PPCA) da UnB na linha de pesquisa de Gestão de Risco. E-mail: jocafs@unb.br.

RESUMEN

La medición de la eficiencia de las unidades de la organización es un factor importante en las sociedades contemporáneas. Para obtener estimaciones de la eficiencia, son varias metodologías disponibles. Uno de los métodos más conocidos, y discutidos aquí, es la técnica no paramétrica llamada DEA (Análisis Envoltente de Datos). De esta forma, se presenta la investigación aplicada cubriendo los enfoques DEA y *Stepwise* en la base de datos de una institución que realiza investigación en educación. Un punto fundamental en el uso de DEA es la elección de variables a utilizar. En el presente trabajo, se aplicó la técnica *Stepwise*, mediante la ejecución del algoritmo *Stepwise* AIC en la herramienta R, como proceso de elección de variables, las cuales a su vez son utilizadas en el modelo de Regresión Lineal Múltiple para comprobar si la prueba es confiable (a través del valor de r^2) y luego utilizado en la DEA. Los resultados encontrados se muestran una lista de los lugares y su eficiencia técnica relativa (o la asignación de recursos).

Palabras clave: eficiencia de medición; técnica no paramétrica; *stepwise*.

1. Introdução

Segundo Neumann (2013), os processos de gestão são extremamente importantes para o sucesso das empresas ou indústrias, visto que são exercidos de forma contínua e respondem pelo alcance de seis indicadores de desempenho: eficiência, eficácia, produtividade, lucratividade, efetividade e competitividade.

O presente trabalho apresenta uma pesquisa aplicada com a utilização da abordagem Análise Envoltória de Dados (na nomenclatura anglo-saxônica: *Data Envelopment Analysis* - DEA) e *Stepwise* em um banco de dados cedidos por uma instituição ligada a pesquisas em educação, especificamente aos dados que se referem aos locais de prova, composta por características, com suas respectivas pontuações, e visando obter como resultado a relação de eficiência destas localidades. **A eficiência no contexto considerado se refere aos locais de provas mais adequados, de acordo com suas características e valor pago na locação, para aplicação de provas de seleção.**

A instituição considerada, que cedeu as informações utilizadas no presente trabalho, realiza pesquisas na área educacional e aplica provas diversas em

todo o território nacional há mais de 20 anos. Entre 2011 e 2013, realizou mais de 160 concursos, com mais de 4 milhões de inscritos em todos os estados do país, nas mais diferentes áreas de atuação: jurídica, administrativa, educacional, executiva, entre outras.

2. Revisão bibliográfica

Os sistemas de medição de desempenho constituem os elos entre os objetivos e a execução prática das atividades nas empresas. Por um lado, são os guias básicos que suportam a tomada de decisão e, por outro, são a lógica dos critérios de avaliação e controle dos resultados através dos indicadores de desempenho (Neuman, 2013).

É importante ressaltar alguns estudos correlatos realizados sobre o tema. No caso de Oliveira et al. (2015), o DEA foi utilizado para a avaliação dos níveis de eficiência dos hotéis de 4 e 5 estrelas da região do Algarve, em Portugal, com o objetivo de aferir o desempenho de 28 (vinte e oito) hotéis que operaram no período de 2005 a 2007. A modelagem permitiu indicar que as diferenças de eficiência se prendem com a gestão, o fraco uso de infraestrutura (época baixa), a sazonalidade e o ambiente institucional e contextual.

Já Sameie & Arvan (2015) efetuaram uma simulação utilizando DEA para avaliar a localização de instalações eólicas no Irã. Esse estudo propõe um modelo para decidir qual é a melhor localização para um parque eólico. Logo, a técnica DEA realiza a seleção da melhor área, dentre todas as áreas candidatas a parque eólico relacionadas na simulação.

Entre os métodos mais aplicados, a técnica DEA e os números índices são considerados métodos não paramétricos, ao passo que as fronteiras estocásticas (na nomenclatura anglo-saxônica: *Stochastic Frontier Analysis* - SFA) e os modelos de regressão, baseados nos quadrados mínimos ordinários (na nomenclatura anglo-saxônica: *Ordinary Least Squares* - OLS) e nos quadrados mínimos ordinários corrigidos (na nomenclatura anglo-saxônica: *Corrected Ordinary Least Squares* - COLS), são alguns exemplos da utilização dos métodos paramétricos.

Os métodos paramétricos ou não paramétricos podem ainda ser classificados em não fronteira ou em fronteira, devido a pressupostos de que as unidades de decisão (na nomenclatura anglo-saxônica: *Decision Making Unit* - DMU) sejam tecnicamente eficientes ou não, caso os benchmarks sejam assentados nas aproximações médias ou nas melhores práticas (Coelli et al., 1998).

Soares de Mello et al. (2005) afirmam que, em contraste com as aproximações paramétricas, que otimizam um plano de regressão a partir das observações, DEA otimiza cada observação individual com o objetivo de calcular uma fronteira de eficiência, determinada pelas unidades que são paretos eficientes. Esse mesmo autor menciona que uma unidade é pareto eficiente se, e somente se, ela não conseguir melhorar alguma de suas características sem piorar as demais.

A superioridade dos métodos de **fronteira** em relação aos **não fronteira** parece ser evidente e relativamente consensual na aferição de eficiência. O mesmo não ocorre, porém, dentro dos métodos de fronteira, em relação aos não paramétricos e aos paramétricos ou vice-versa, nomeadamente entre a DEA e SFA. Não obstante, o argumento de maior simplicidade na aplicabilidade da técnica DEA pode sustentar a sua supremacia (Tavares, 2002).

A eficiência econômica, numa abordagem por fronteira, pode ser analisada por duas ópticas: a óptica da orientação por insumo (na nomenclatura anglo-saxônica: *inputs*) e da orientação por produto (na nomenclatura anglo-saxônica: *outputs*). Segundo Coelli (1994), a abordagem da orientação por insumo busca responder a questão de quanto se pode reduzir as quantidades de insumo de modo que as quantidades de produto feitas não sofram alteração, ao passo que a abordagem da orientação por produto mede quanto do produto pode ser aumentado sem que altere as quantidades de insumos utilizadas.

Em termos de tecnologia de produção, existem um vetor com (s) tipos de entradas e um vetor com (m) tipos de saída. A medida de eficiência técnica (E) de uma DMU deve ser a máxima contração radial do vetor de insumos que permite produzir a mesma quantidade de produtos (considerando a óptica da orientação por insumo). Um valor unitário para (E) indica que não

é possível reduzir a quantidade de insumos e manter a mesma produção. Nesse caso, a DMU é tecnicamente eficiente. Caso contrário, quando $(E) < 1$, significa que há um excesso de insumos e a DMU é tecnicamente ineficiente.

Pode-se referenciar, como mais citados na literatura sobre DEA, os seguintes modelos:

- a) **Modelo CCR** - desenvolvido por Charnes, Cooper e Rhodes (1978), permite uma avaliação objetiva da eficiência global e identifica as fontes e estimativas de montantes das ineficiências identificadas. A correspondência com o CCR é o retorno constante de escala (na nomenclatura anglo saxônica: *Constant Returns to Scale* - CRS).
- b) **Modelo BCC** - criado por Banker, Charnes e Cooper (1984), distingue entre ineficiências técnicas e de escala, estimando a eficiência técnica pura, a uma dada escala de operações, identificando se estão presentes ganhos de escala crescentes, decrescentes e constantes, para futura exploração. A correspondência com o BCC é o retorno variável de escala (na nomenclatura anglo-saxônica: *Variable Returns to Scale* - VRS).

A técnica DEA foi definida como uma abordagem baseada em programação matemática para medir a eficiência relativa das DMUs, considerando as entradas e saídas (Charnes et al., 1978). Na tabela 1, está representado o modelo denominado CRS/VRS com orientação ao insumo na versão dos multiplicadores. A DEA é taxada como uma técnica de simples aplicação, mas é considerada uma poderosa ferramenta para medir a eficiência relativa.

TABELA 1. MODELOS DEA COM ORIENTAÇÃO AO INSUMO NA VERSÃO DOS MULTIPLICADORES

MODELO CRS	MODELO VRS
$\theta = \text{Max} \sum_{i=1}^m u_i y_{i,j_0}$ <p>s.a.</p> $-\sum_{i=1}^s v_i x_{ij} + \sum_{i=1}^m u_i y_{ij} \leq 0 \forall j$ $= 1, \dots, j_0, \dots, N$ $\sum_{i=1}^s v_i x_{i,j_0} = 1$ $u_i \geq 0 \forall i = 1, m$ $v_i \geq 0 \forall i = 1, s$	$\theta = \text{Max} \sum_{i=1}^m u_i y_{i,j_0} + u_0$ <p>s.a.</p> $-\sum_{i=1}^s v_i x_{ij} + \sum_{i=1}^m u_i y_{ij} + u_0 \leq 0 \forall j$ $= 1, \dots, j_0, \dots, N$ $\sum_{i=1}^s v_i x_{i,j_0} = 1$ $u_i \geq 0 \forall i = 1, m$ $v_i \geq 0 \forall i = 1, s$
<p>N + 1 restrições m + s variáveis</p>	<p>N + 1 restrições m + s + 1 variáveis</p>

Ressalta-se, ainda, que, nos dias atuais, a técnica DEA conta com uma variedade de modelos que abrange desde os modelos clássicos (utilizados neste artigo) e suas variações até abordagens que combinam DEA com lógica *fuzzy*, simulação de Monte Carlo e *bootstrap*.

A utilização da DEA possui algumas vantagens, das quais se destacam (Cunha & Silva, 2006): identificação para cada DMU ineficiente de um conjunto de DMU (*peers*) eficientes com combinação de *inputs* e *outputs* semelhante; facilidade em lidar com múltiplos *inputs* e *outputs*; adoção dos melhores resultados como elementos de comparação; não admissão de uma forma paramétrica para a fronteira ou para a ineficiência quando

associada ao erro; natureza conservativa das avaliações e decomposição da natureza da eficiência em vários componentes.

Entretanto, é necessário ter alguns cuidados na utilização da DEA, a saber: a sensibilidade elevada aos *outliers*, a exigência em termos de informação requerida e a dificuldade de medir o erro associado ou de testar estatisticamente os resultados e os modelos adotados. Dessa forma, também, na DEA, a análise dos fatores explanatórios é complexa, dependendo da correlação existente.

Esta pesquisa visa avaliar as localidades que podem ser utilizadas para a realização de provas em termos de eficiência técnica, considerando variáveis independentes (características dos locais de prova) e uma variável de *input* (custo do aluguel). Dessa forma, por um lado, se um local apresenta um nível de eficiência máximo (igual a 1), então não há desperdício. Se, por outro lado, o nível de eficiência não é máximo, há de se considerar a existência de um certo desperdício (o local poderia ter melhores características e/ou um custo de aluguel menor). A partir dessa avaliação, é possível calcular o desperdício de cada local com ineficiência relativa através da seguinte fórmula:

$$\sum_{k=1}^n (1 - Eff(k)) * input(k)$$

Em que n é o número de locais de prova, k representa um local de prova (uma DMU). A função $Eff(k)$ retorna o valor da eficiência da DMU k . $Input(k)$ retorna o custo de aluguel para alocação de cada DMU.

3. METODOLOGIA DA PESQUISA

De acordo com Jung (2003), “A pesquisa aplicada (tecnológica) tem como objetivo alcançar a inovação em um produto ou processo, frente a uma demanda ou necessidade preestabelecida”, ou seja, busca solucionar um problema concreto. Dessa forma, o artigo pode ser visto, quanto à

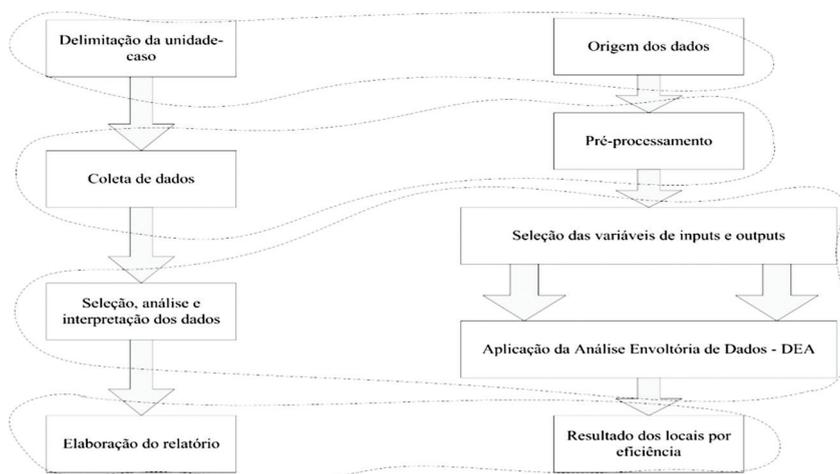
natureza, como uma pesquisa aplicada.

Quanto ao procedimento, a pesquisa ocorreu através de um estudo de caso em uma instituição de pesquisa em educação para revelar os locais de provas mais eficientes. Segundo Bruyne et al. (1977), o estudo de caso justifica sua importância por reunir informações numerosas e detalhadas que possibilitam apreender a totalidade de uma situação. A riqueza das informações detalhadas auxilia o pesquisador a adquirir maior conhecimento e solucionar problemas relacionados ao assunto estudado.

Segundo Gil (1999), o estudo de caso não aceita um roteiro rígido para a sua delimitação, mas é possível definir quatro fases que mostram o seu delineamento, tais como: a) delimitação da unidade-caso; b) coleta de dados; c) seleção, análise e interpretação dos dados; d) elaboração do relatório.

Do ponto de vista da estruturação da pesquisa, foi elaborada uma correlação entre as fases definidas por Gil (1999) e as fases elaboradas pelo autor, como representado na Figura 1. Cada uma dessas fases será detalhada no desenvolvimento do trabalho.

FIGURA 1. AS 4 FASES DEFINIDAS POR GIL (1999) EM ANALOGIA ÀS 5 FASES DEFINIDAS PELO AUTOR



4. DESENVOLVIMENTO

Para a aplicação de técnicas estatísticas, com foco na análise de eficiência, é possível e produtivo utilizar um ambiente de desenvolvimento de *software*. Para a presente pesquisa, o ambiente de programação R foi escolhido por contar com um conjunto de ferramentas para manipulação de dados, cálculos e apresentação gráfica (Venables, 2015). O R é mais do que uma simples biblioteca de pacotes, pois permite aos usuários a construção de seus próprios programas ou *packages*. Portanto, por meio do R, os usuários podem obter soluções de baixo custo e livres das restritas opções oferecidas pelos programas dedicados apenas aos modelos DEA (Pessanha et al, 2013). O pacote *Benchmarking* contém métodos de análise de fronteira. Ele suporta a técnica DEA, abordando diferentes tecnologias (*fdh*, *vrs*, *drs*, *crs*, *irs*, *add/frh* e *fdh*) e utilizando diferentes meios de eficiência (baseados em entrada e saída, gráfico hiperbólico, aditivo, superior e eficiência direcional). Possui também funcionalidades para que sejam gerados gráficos que facilitem a análise dos dados que estão sendo manipulados. Esse pacote é um complemento do livro do Bogetoft & Otto (2011). Este estudo está baseado na utilização desse pacote. A seguir serão detalhadas as cinco fases estabelecidas na metodologia.

4.1 Origem dos dados

Os dados utilizados na análise foram concedidos por uma instituição de pesquisa em educação que aplica provas diversas. A instituição em questão possui um banco de dados com informações de locais de avaliações da região do Distrito Federal, capital do Brasil. Esse banco de dados é oriundo de um sistema específico para administração cadastral das instalações dessas localidades. Os dados a seguir compõem a lista de possíveis *inputs* e *outputs* para a modelagem.

TABELA 2. - IDENTIFICAÇÃO DAS VARIÁVEIS DAS INFORMAÇÕES OBTIDAS PARA A REALIZAÇÃO DOS ESTUDOS

CAMPO DE IDENTIFICAÇÃO DOS DADOS (VARIÁVEIS)	DESCRIÇÃO DA INFORMAÇÃO
TotalPorInstituicao	Representa o valor do aluguel do espaço físico.
Qt_Capacidade	Representa a capacidade total que a localidade possui para alocação de pessoas.
Qt_Salas	Representa a quantidade de salas disponíveis na localidade.
Nu_NotaConservacaoBanheiro	Representa a nota de avaliação (na escala de zero a dez) efetuada pela organizadora para determinar o estado de conservação do banheiro.
Qt_CapacidadeMediaSalas	Representa a capacidade média das salas disponíveis na localidade.
Qt_Banheiro	Representa a quantidade total de banheiros disponíveis na localidade.
Nu_NotaVentilacaoSala	Representa a nota de avaliação (na escala de zero a dez) efetuada pela organizadora para determinar a ventilação das salas.
Nu_NotaConservacaoCarteira	Representa a nota de avaliação (na escala de zero a dez) efetuada pela organizadora para determinar o estado de conservação das carteiras.

As notas referentes às avaliações dos locais de provas, variáveis de nota da Tabela 2, foram lançadas no sistema por membros da área de logística da instituição. Tratam-se de especialistas em alocação de espaços para a realização de provas que definem as notas após visita *in loco* a cada estabelecimento. A descrição e identificação dos locais de aplicação foram omitidas. As variáveis na tabela 2 foram as utilizadas no modelo.

4.2 Pré-processamento

O pré-processamento compreende as funções que se relacionam à captação, à organização e ao tratamento de dados, cujo objetivo é preparar os dados (Goldschmidt & Passos, 2005). Nessa etapa, analisaram-se os dados de forma a assegurar a qualidade (completude, veracidade e integridade) dos fatos por eles representados. As variáveis com ausência de informações foram descartadas. **Do total de 431 registros recebidos, foram considerados 401.** Para a importação das informações, checagens e manipulações necessárias nesta fase de pré-processamento, foi utilizado o sistema de gerenciamento de banco de dados (SGBD) SQL Server Express (*Microsoft*[®]).

4.3 Seleção das variáveis de *inputs* e *outputs*

Em estudos de eficiência, costuma-se utilizar como variável de entrada ou insumo (também chamada de *input*) um custo qualquer de produção ou serviço, ou diversas variáveis de custo quando estas existem. Isso pode ser verificado em diversos trabalhos (Pessanha et al., 2013, Sousa & Ramos, 1999). Nesta pesquisa, será considerado, como variável de entrada, o custo de aluguel das localidades de realização das avaliações (a variável Total-PorInstituicao). As demais variáveis são candidatas a produtos (também chamadas de *outputs*) no contexto do modelo DEA. A determinação das melhores variáveis candidatas pode ser feita através da regressão linear múltipla. Tal modelo foi escolhido porque é capaz de abarcar mais de uma variável explicativa. Este modelo matemático que estabelece a relação funcional entre as variáveis é definido como:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k + \varepsilon$$

y é a variável dependente (*input*);

$\beta_0, \beta_1, \dots, \beta_k$ são parâmetros a serem estimados;

x_1, x_2, \dots, x_k são as variáveis independentes (*outputs*);

ε é o erro aleatório referente à variabilidade que não pode ser explicada pelas variáveis independentes.

Sob o modelo exposto, utiliza-se o método dos mínimos quadrados para estimação dos parâmetros.

Existem alguns critérios que permitem a escolha da melhor função de regressão. A técnica *Stepwise* utiliza um algoritmo que possibilita escolher um modelo em que todos os parâmetros sejam significativamente diferentes de 0 (zero) (Draper & Smith, 1998). É possível, ainda, determinar qual é o modelo mais plausível para representar um fenômeno dentro de um conjunto de alternativas. A opção é pelo Critério de Informação de Akaike (*AIC*) (Lindsey, 1996). O critério de informação de Akaike (Tsay, 2005) é dado por:

$$AIC = -\ln(L) + 2k$$

L é o máximo da função verossimilhança e k é o número de parâmetros estimado. Nesse contexto, o melhor modelo será aquele que apresentar o menor valor utilizando o *AIC*.

A técnica *Stepwise* foi escolhida porque é adequada para estudos exploratórios. Com essa técnica, a seleção da sequência dos preditores na equação é feita estatisticamente, sem um modelo empírico a ser seguido. É importante ressaltar que, segundo Thanassoulis et al. (1996), a modificação do conjunto de variáveis selecionadas poderá ter grande impacto no resultado da avaliação. A família de métodos *Stepwise* parte da premissa de que a seleção de variáveis deve obedecer ao princípio de máxima relação causal entre *inputs* e *outputs*. Esse é um método que se preocupa em aumentar a eficiência média com um número limitado de variáveis.

A técnica *Stepwise* é sintetizada pelos seguintes passos (Kittelsen, 1998):

- a) definição de um modelo base que contenha todas as variáveis consideradas essenciais por razões empíricas ou teóricas;
- b) obtenção da estimativa da eficiência para o modelo
- c) cálculo da estimativa da eficiência com uma variável candidata ou com as variáveis desagregadas;
- d) determinar os 4 testes (F^m - meia normal, F^e - exponencial, K-S - *Kolmogorov-Smirnov* e T - comparação das médias) e aceitar, ou não, o novo modelo.

Caso existam mais variáveis a desagregar ou novas variáveis candidatas, é preciso voltar ao item (b), embora a técnica *Stepwise* exija, do ponto de vista teórico, requisitos que, em geral, não se verificam, como a dimensão da amostra ou a independência (Simar & Wilson, 2002). A técnica *Stepwise* foi utilizada neste estudo em conjunto com o Critério de Seleção de *Akaike*. A tabela 3 sintetiza o resultado da aplicação da técnica *Stepwise*, por meio do algoritmo *Stepwise AIC*, com auxílio computacional da ferramenta R e função *stepAIC* do pacote *MASS*, que selecionam o modelo através do método *Stepwise* com opções *backward*, *forward* e *both*.

TABELA 3. RESUMO DO RESULTADO DA APLICAÇÃO DA TÉCNICA *STEPWISE* COM O AUXÍLIO DA FERRAMENTA R E A FUNÇÃO *STEPAIC* DO PACOTE MASS

MODELO INICIAL					
TotalPorInstituicao ~ qt_Capacidade + qt_Salas + qt_CapacidadeMediaSalas + nu_NotaConservacaoBanheiro + nu_NotaVentilacaoSala + nuNotaConservacaoCarteira + qt_Bebedouros + qt_banheiro					
MODELO FINAL					
TotalPorInstituicao ~ qt_Capacidade + qt_Salas + qt_Bebedouros + qt_banheiro.					
Passo	Df	Deviance Resid.	Df	Resid. Dev	AIC
1			392	2.211.365.507	6.242,689
2 NuNotaConservacao- Carteira	1	503.284,4	393	2.211.868.791	6.240,780
3 Nu_NotaVentilacaoSala	1	569.478,1	394	2.212.438.269	6.238,883
4 Nu_NotaConservacao- Banheiro	1	153.1250,1	395	2.213.969.520	6.237,161
5 Qt_CapacidadeMedia- Salas	1	3.709.070,4	396	2.217.678.590	6.235,832

As variáveis preditoras que melhor explicam o critério (variável dependente), obtidas com a aplicação da técnica *Stepwise*, foram as variáveis: qt_Capacidade, qt_Salas, qt_Bebedouros e qt_banheiro. Segundo a informação de *Akaike*, o melhor modelo é o que possui menor valor de *AIC*. A tabela 4 apresenta os dados obtidos através da utilização de mais uma regressão, desta vez aplicada somente com as variáveis selecionadas no pelo *Stepwise*, que constituem o modelo final, no sentido de verificar se o teste é confiável:

TABELA 4. RESULTADO DA REGRESSÃO UTILIZANDO O MODELO FINAL

RESIDUAIS					
Min	1Q	Mediana	3Q	Máx	
-14474,9	-5471	-68,1	430,1	16565,0	
COEFICIENTES					
	Estimativa	Erro Padrão	Valor t	Pr (> t)	
(INTERCEPTO)	-1838,3319	247,0768	-7,440	6,30e-13	***
QT_CAPACIDADE	1,4326	0,5849	2,449	0,0147	*
QT_SALAS	150,7830	31,0786	4,852	1,76e-06	***
QT_BANHEIRO	52,2240	26,9580	1,937	0,0534	.
QT_BEBEDOUTOS	64,1511	34,0954	1,882	0,0606	.
NÍVEL DE SIGNIFICÂNCIA	0 '***' 0,001 '**' 0,01 '*' 0,05 '.' 0,1 ' ' 1				
ERRO RESIDUAL PADRÃO	2.366 EM 396 GRAUS DE LIBERDADE				
R-QUADRADO MÚLTIPLO	0,7864	R-QUADRADO AJUSTADO		0,7842	
ESTATÍSTICA F	364,4 EM 4 E 396 GRAUS DE LIBERDADE		P-VALOR	<2,2e-16	

Observa-se, pela tabela 4 e pelo valor de t (*t value*), que a variável preditora que possui maior correlação com a variável dependente, TotalPorInstituição, é qt_Salas, e posteriormente destaca-se a variável qt_Capacidade. Entretanto, conforme resultado da aplicação da técnica *Stepwise*, todas essas variáveis independentes do modelo final serão utilizadas na aplicação do modelo de eficiência (DEA). Essa decisão ocorre porque é sabido também que esse modelo de eficiência tem alta sensibilidade a cada uma das variáveis independentes, mesmo que estas apresentem níveis de significância praticamente irrelevantes no modelo final. Ressalta-se, ainda, através de informações da tabela 4, que a regressão utilizada com as variáveis independentes, obtidas pela aplicação da técnica *Stepwise*, mostra-nos um teste confiável (o valor ajustado de r^2 é 0,7842, *p*-valor é baixo e o teste *F*-statistic é 364,4).

parceria pontual com uma universidade, somente possível em eventos de um cliente específico, ou devido a uma necessidade de aplicação em presídio, cadastrado como local de avaliação e utilizado ocasionalmente em aplicações a detentos como o Exame Nacional do Ensino Médio (Enem), e que, por isso, poderiam ser considerados *outliers* e conseqüentemente desprezados.

É possível ainda que se identifiquem quais DMUs deverão servir como *benchmark* para cada uma das unidades ineficientes. As DMUs observadas que aparecem com maior frequência como padrão de eficiência ou *benchmark* das observações ineficientes são candidatas a serem mais eficientes dentro do conjunto, ou seja, são as referências ou *peers* (melhores práticas) obtidas pela técnica DEA. Ao definir a fronteira, a técnica DEA possibilita identificar os *peers* para cada operador, bem como os seus valores-alvo (*targets*) para os diferentes *inputs* e *outputs*. Assim, a tabela 5 apresenta uma amostra de DMUs com a identificação de suas referências para ilustração.

TABELA 5. IDENTIFICAÇÃO DOS PEERS DE ALGUMAS DMUS PARA ILUSTRAÇÃO

LOCAL	PEER1	PEER2	PEER3	PEER4	PEER5
1	105	194	344		
2	88	187	356		
..
49	31	140	194	344	
..
144	194				

4.5 Resultado dos locais por eficiência

O A tabela 6 demonstra um resumo da aplicação da técnica DEA nos registros de locais de prova (DMUs). Verifica-se que foram encontrados 25 locais eficientes dentre os 401 analisados. Isso quer dizer que somente 6,2% dos locais estão na margem de eficiência. A tabela 7 relaciona alguns

dos locais com as eficiências encontradas. A listagem completa foi omitida sem prejuízo algum do entendimento, haja vista que os locais estão sem identificação.

TABELA 6. RESUMOS DAS INFORMAÇÕES DE EFICIÊNCIAS DOS LOCAIS DE REALIZAÇÃO DE PROVAS

<i>RESUMO DE EFICIÊNCIAS</i>					
<i>A tecnologia é VRS e eficiência com orientação a insumos</i>					
<i>Número de locais com eficiência igual a 1 (por isso eficientes): 25</i>					
<i>Eficiência média: 0,434</i>					

<i>FAIXA DE EFICIÊNCIA</i>	<i>QTDE</i>	<i>%</i>			
<i>0 <= E < 0,1</i>	2	0,5			
<i>0,1 <= E < 0,2</i>	9	2,2			
<i>0,2 <= E < 0,3</i>	41	10,2			
<i>0,3 <= E < 0,4</i>	209	52,1			
<i>0,4 <= E < 0,5</i>	41	10,2			
<i>0,5 <= E < 0,6</i>	29	7,2			
<i>0,6 <= E < 0,7</i>	21	5,2			
<i>0,7 <= E < 0,8</i>	11	2,7			
<i>0,8 <= E < 0,9</i>	8	2,0			
<i>0,9 <= E < 1</i>	5	1,2			
<i>E == 1</i>	25	6,2			
<i>MIN.</i>	<i>1º. Q.</i>	<i>MEDIANA</i>	<i>MÉDIA</i>	<i>3º. Q.</i>	<i>MÁX.</i>
0,07794	0,33330	0,33330	0,43400	0,48970	1,00000

TABELA 7. EXEMPLOS DE LOCAIS DE PROVA COM SUAS RESPECTIVAS EFICIÊNCIAS

LOCAL	TOTALPOR INSTITUICAO	QT_ CAPACIDADE	QT_ SALAS	QT_ BANHEIRO	QT_ BEBE- DOUROS	EFICIÊNCIA
1	1935	645	20	16	1	0,33
...
23	655	655	16	2	1	1
...
30	695	695	16	6	6	1
31	6900	3065	73	6	36	1
32	6900	1760	39	6	36	1
...
42	492	492	15	4	3	1
...
44	50820	10164	197	40	41	1
45	50465	10093	193	40	41	1
...
353	450	150	6	6	3	1
354	945	315	9	4	9	0,83
355	945	315	9	4	9	0,83
356	450	450	20	8	1	1
...
401	1461	487	13	8	6	0,39

As eficiências com valor unitário (destacados na tabela 7) indicam que não é possível reduzir a quantidade de insumos e manter a mesma produção. Assim, as DMUs com eficiência unitária são tecnicamente eficientes. Caso contrário, quando é menor que 1, significa que há um excesso de insumos e a DMU é tecnicamente ineficiente. O resultado do modelo nos mostra que os valores para alocação destes locais de prova ineficientes são elevados pelo que eles oferecem. Para entendimento do volume de recursos

financeiros que poderiam ser economizados, consideraremos um cenário em que a utilização dos 401 locais poderia ocorrer numa única aplicação de provas, de modo que é possível calcular o desperdício de cada local com ineficiência relativa através da fórmula citada no item de revisão bibliográfica e reproduzida a seguir:

$$\sum_{k=1}^n (1 - Eff(k)) * input(k)$$

Considerando um local ($k=2$) com nível de eficiência $Eff(2)$ equivalente a 0,52295, aproximadamente, e custo de aluguel equivalente a R\$ 900,00, teríamos: $(1 - 0,52295) * 900$. Assim, o desperdício de escolher este local ($k=2$) seria de R\$ 429,35. Um local com eficiência 1 não traria, dessa forma, desperdício (o resultado obtido com a aplicação da fórmula seria zero).

O desperdício total, obtido através da soma dos desperdícios dos locais não eficientes, não será mencionado por ser uma informação estratégica e talvez sensível da instituição que forneceu os dados. Foi possível verificar, entretanto, que as variáveis que mais impactaram os níveis de eficiência, $Eff(k)$, foram a capacidade total e ainda a quantidade de salas de determinado espaço físico. Obviamente que isso pode variar de acordo com a amostra de dados utilizada. Comparando determinadas localidades com as suas referências ou *peers* (melhores práticas), sugere-se um ajuste do valor do aluguel das localidades não eficientes para torná-las eficientes, ou uma ampliação das suas capacidades e/ou aumento na quantidade de salas.

5. CONCLUSÃO

Os objetivos do presente trabalho consistiram no tratamento de dados e na seleção de variáveis para a aplicação da técnica DEA clássica (com orientação por insumo), referente ao cálculo da eficiência de locais de aplicações de provas.

O uso da DEA apresentou vantagens, das quais se destacam: facilidade na aplicação e na identificação de um conjunto de DMU (*peers*) eficientes com combinação de *input* e *outputs*. Os locais tidos como eficientes constituem

benchmark (referências) para outros não eficientes, ou seja, um local não eficiente precisa melhorar suas características (capacidade, quantidade de salas, etc.) para se tornar eficiente, considerando o seu custo de aluguel.

Para selecionar as variáveis que possuem maior correlação com a variável dependente (custo de aluguel das localidades de realização das provas), foi utilizada a técnica *Stepwise/AIC*, que se baseia em regressão.

Ressalta-se que a técnica DEA não permitiu, com facilidade, a inferência estatística dos resultados obtidos, uma consequência óbvia da escolha de modelo não paramétrico.

Como sugestão para estudos futuros, vislumbrou-se a elaboração de um modelo que, de acordo com a localização geográfica relativa de prova e outras variáveis de *output*, seria possível prever o preço eficiente de alocação. Uma possibilidade para isso seria estabelecer índices para cada localização (considerando bairros, por exemplo). Fica como sugestão, ainda, a inclusão de outras variáveis no cadastro das informações dos locais com características de qualidade, como aspectos mais detalhados sobre a ventilação das salas e ainda aspectos de acessibilidade a PcDs (pessoas com deficiência).

Referências

Bogetoft, P., & Otto, L. (2011). *Benchmarking with DEA, SFA, and R*. Springer, New York.

Bruyne, P., Herman, J., & Schoutheete, M. (1977). *Dinâmica da pesquisa em ciências sociais*. Rio de Janeiro: Francisco Alves Editora.

Charnes, A., Cooper, W., & Rhodes, E. (1978). Measuring the efficiency of decision making units. *European Journal of Operational Research*.

Coelli, T., Prasada, R., & Battese, G. (1998). *An Introduction to Efficiency and Productivity Analysis*, Kluwer Academic Publishers, USA.

Coelli, T. (1994). A Guide to DEAP Version 2.1.: a data envelopment analysis (computer) program. CEPA Working Papers, Recuperado de: <http://www.une.edu.au/economet/rics/cepawp.htm>.

Cunha Marques, R., & Silva, D. (2006). Inferência Estatística dos Estimadores de Eficiência obtidos com a Técnica Fronteira Não Paramétrica de DEA. Uma Metodologia de Bootstrap. *Investigação Operacional*.

Draper, N. R., & Smith, H. (1998). *Applied regression analysis*. New York: Wiley-Interscience publication.

Dunlap, W. P., & Landis, R.S. (1998). Interpretations of multiple regression borrowed from factor analysis and canonical correlation. *The Journal of General Psychology*.

Kittelsen, S. (1998). Stepwise DEA: Choosing Variables for Measuring Technical Efficiency in Norwegian Electricity Distribution. Memorandum, Nº. 6, Oslo University.

Gil, A., & Métodos, C. (1999). *Técnicas de pesquisa social*. São Paulo. Editora Atlas.

Goldschmidt, R. R., & Passos, E. (2005). *Data Mining: Um Guia Prático*. Rio de Janeiro: Campus.

Jung, F. C. (2003). *Metodologia Científica Ênfase em Pesquisa Tecnológica*. Recuperado de: <http://www.jung.pro.br/moodle/>.

Lindsey, J. K. (1996). *Parametric Statistical Inference*. New York. Oxford science publications.

Lins, M. E., Lobo, M. S. C., Silva, A. C. M., Fiszman, R., & Ribeiro, V. J. P. (2007). O uso da análise envoltória de dados (DEA) para avaliação de hospitais universitários brasileiros. *Ciência & Saúde Coletiva*.

Neumann, C. (2013). *Gestão de Sistemas de Produção e Operações*. Rio de Janeiro. Elsevier.

Pessanha, J. F. M., Marinho, A., Laurencel, L. C., & Amaral M. R. S. (2013). Implementando modelos DEA no R. X Simpósio de Excelência em Gestão de Tecnologia (SEGeT).

Oliveira, R. S. L. P., Pedro, M. I. C. & Marques, R. D. R. da C. Avaliação da eficiência das empresas hoteleiras do Algarve pelo metodologia Análise Envoltória de Dados (DEA). RGBN 17, 788–805 (2015).

Sameie, H., & Arvan, M. (2015). A simulation-based Data Envelopment Analysis (DEA) model to evaluate wind plants locations. Decision Science Letters.

Senra, L. F. A. C, Nanci, L. C, Mello, J. C. C. B. S, & Meza, L. A. (2007). Estudo sobre métodos de seleção de variáveis em DEA. Rio de Janeiro. Pesquisa Operacional, vol. 27, nº 2.

Simar, L., & Wilson, P. W. (2002) Non-parametric Tests of Returns to Scale, European Journal of Operational Research. Vol 139, No 1.

Soares de Mello, J. C. B. S., Angulo, M. L.; Gomes, E. G., & Biondi Neto, L. (2005) Curso de análise de envoltória de dados. Rio Grande do Sul: Gramado. Simpósio Brasileiro de Pesquisa Operacional.

Sousa, M. C. S., & Ramos, F. S. (1999). Eficiência técnica e retornos de escala na produção de serviços públicos municipais: o caso do nordeste e do sudeste brasileiros. Rio de Janeiro. Revista Brasileira de Economia, vol. 53, nº 54.

Tavares, G. (2002). A bibliography of data envelopment analysis (1978-2001). RUTCOR Research Report, Rutgers University.

Thanassoulis, E., Boussofiene, A., & Dyson, R. G. (1996). A comparison of data envelopment analysis and ratio analysis as tools for performance assessment. Omega - International Journal of Management Science, 24 (3).

Tsay, R. S., (2005). Analysis of financial Time Series. 2ª ed. Wiley & Sons, Inc., Hoboken, New Jersey.

Venables W.N., Smith D.M., & R Core Team. (2015) An Introduction to R. A Programming Environment for Data Analysis and Graphics. Version: 3.2.1.

Wagner, J. M., & Shimshak, D. G., (2007). Stepwise selection of variables in data envelopment analysis: Procedures and managerial perspectives. European Journal of Operational Research, Volume 180.