

# ANÁLISE ENVOLTÓRIA DE DADOS COM APOIO DO SOFTWARE R: UMA APLICAÇÃO NOS INSTITUTOS FEDERAIS

Alessandro de Castro Corrêa (IFPA) E-mail: alessandro.correa4@gmail.com

Narciso Rodrigues da Costa (IFPA) E-mail: costa93y@gmail.com

Danielle Cristina Gonzaga Corrêa (FACI-Wyden) E-mail: danielle.correa1@gmail.com

Jéssica Letícia de Souza Miranda (IFPA) E-mail: jessica.miranda.gp@gmail.com

**Resumo:** O objetivo deste artigo é apresentar o Software R como uma promissora ferramenta no apoio de aplicações da Análise Envoltória de Dados (DEA), um método multicritério de análise de eficiência relativa de organizações. Analisaram-se 38 Institutos da Rede Federal de Educação, Ciência e Tecnologia (IF) por meio da DEA com rendimentos variáveis de escala (RCE) orientado a produto, no ano de 2014, utilizando-se indicadores de gestão, dois de insumo e um de produto. O estudo foi inteiramente realizado no R, realizando-se cálculos de eficiência técnica, supereficiência, coeficientes de correlação, regressão múltipla, transformação de dados e geração de gráficos personalizados. Os resultados, além de revelar o desempenho dos Institutos, evidenciaram o software R como uma vantajosa opção para aplicação da DEA, ressaltando a sua simplicidade e versatilidade.

**Palavras-chave:** Software R, DEA, Eficiência, Institutos Federais.

## DATA ENVELOPMENT ANALYSIS WITH R SOFTWARE SUPPORT: AN APPLICATION TO BRAZILIAN FEDERAL INSTITUTES

**Abstract:** The aim of this study was to demonstrate R software is a promising tool for supporting Data Envelopment Analysis (DEA) applications, a multi-criteria method used to measure relative efficiencies of organizations. The efficiencies of 38 Federal Institutes of Education, Science and Technology (IF) were assessed, in 2014, employing output-oriented variable return to scale DEA model, with two inputs and one output. The computations were entirely done applying R language, including technical efficiency, superefficiency, correlations, regressions, data transformation and customized graph. The results, in addition to revealing the Institutes performances, evidenced R software is an advantageous option for DEA application, emphasizing its simplicity and versatility.

**Keywords:** R Software, DEA, Efficiency, Federal Institutes.

### 1. Introdução

O software R (R CORE TEAM, 2018) é um ambiente computacional amplamente utilizado em análises estatísticas, mas também com forte potencial para outras aplicações. Uma possibilidade promissora reside no apoio a métodos multicritérios de tomada de decisão por sua simplicidade, versatilidade, por ser aberto e gratuito, especialmente em aplicações na Análise Envoltória de Dados (DEA) que é um método multicritério de análise de eficiência relativa destinado à análise organizacional.

Uma vez que a DEA demanda uma quantidade considerável de cálculos para resolução de um problema de programação linear para cada DMU, é indispensável o apoio de programas computacionais que permitam viabilizar a sua aplicação. Há softwares disponíveis para tal propósito. Alguns são de aplicação mais geral, destinados à programação numérica, como o MATLAB e Mathematica, ou destinados à otimização,

como o Lingo, ou o nacional PROLIN, este gratuitamente disponível. Há também os especificamente desenvolvidos para aplicações em DEA, como o DEAP e o nacional IDEAL. Existem ainda complementos para planilha eletrônica como o comercial DEA Excel Solver e o gratuito DEA-Excel.

O objetivo desse artigo é demonstrar o potencial do R como recurso computacional de apoio a DEA, apresentando uma aplicação no estudo da eficiência relativa de Institutos Federais inteiramente realizada no ambiente R.

## 2. Análise Envoltória de Dados

A DEA pode ser definida como uma técnica que emprega programação linear para estimação empírica de uma função de produção que possibilita a determinação da eficiência relativa de unidades de tomadas de decisão (DMU) que utilizam múltiplos insumos e produtos. A eficiência é denominada relativa, pois as DMU são comparadas àquelas com melhores desempenhos.

Vários modelos de DEA foram formulados para incorporar diferentes adaptações, podendo considerar tecnologias com rendimentos constantes de escala (RCE) (CHARNES *et al.*, 1978) ou tecnologias com rendimentos variáveis de escala (RVE) (BANKER *et al.*, 1984) além de poderem também serem orientados ao produto ou ao insumo. Nos modelos com orientação ao insumo, a eficiência é perseguida por variações nos insumos, mantendo os produtos constantes. Nos orientados ao produto, maximiza-se a eficiência, variando-se os produtos, sem alterar os níveis de insumos.

Os modelos RCE e RVE são apresentados nas equações dos problemas de Programação Linear a seguir, exibindo as formulações matemáticas dos modelos orientados ao produto, sendo que a diferença entre ambos é o acréscimo da restrição na Equação 5, referente ao modelo RVE.

$$\max_{\phi, \lambda} \phi \quad (1)$$

$$\text{sujeito a } \mathbf{X}\lambda \leq \mathbf{X}_o, \quad (2)$$

$$\phi \mathbf{Y}_o - \mathbf{Y}\lambda \leq 0, \quad (3)$$

$$\lambda \geq 0 \quad (4)$$

$$\mathbf{N1}'\lambda = 1 \quad (5)$$

Nas quais, para N DMU,  $\phi$  é um escalar que representa a medida de eficiência orientada ao produto, em que  $1 \leq \phi \leq \infty$ ;  $\mathbf{X}$  é a matriz da quantidade observada de insumos com dimensões  $M \times N$ , em que M é o número de insumos;  $\mathbf{Y}$  é a matriz da quantidade observada de produtos  $S \times N$ , em que S é a quantidade de produtos;  $\mathbf{N1}$  é um vetor de valores unitários com N elementos;  $\lambda$  é um vetor de constantes  $N \times 1$ , representativos dos pesos atribuídos a cada insumo e produto. Pode-se observar que a diferença entre os dois modelos é a restrição de convexidade,  $\mathbf{N1}'\lambda = 1$ , que assegura a obtenção de medidas de eficiência, considerando rendimentos variáveis de escala.

A solução ótima do problema de programação linear calcula  $\phi$  para determinada DMU, processo que se repete para cada uma individualmente. As DMU eficientes compõem uma fronteira eficiente que serve de referência para aquelas classificadas como ineficientes.

Para facilitar interpretações, o desempenho resultante é normalmente apresentado como eficiência técnica ( $E = 1/\phi$ ), em que  $0 \leq E \leq 1$ . Se  $E = 1$ , a DMU é considerada eficiente, mas

se  $E < 1$ , é considerada ineficiente. É possível também definir os parceiros de excelência (*peers*), ou *benchmarks*, que correspondem às DMU eficientes utilizadas para projetar determinada ineficiente na fronteira eficiente, de forma que as mesmas guardam mais similaridade entre si em termos de quantidades de insumos e produtos. A proporção de similaridade é dada pelo valor de  $\lambda$ , quanto maior for seu valor, maior a similaridade com a DMU eficiente.

Contudo, a restrição de que  $E$  seja limitado a 1 impede da discriminação das DMU eficientes impossibilitando a sua hierarquização uma vez que todas exibem o mesmo valor ( $E=1$ ). Uma solução foi apresentada por Andersen e Petersen (1993), denominada Modelo de Supereficiência, que consiste na extração da DMU eficiente do problema de programação, sendo comparada com as demais, mas não com ela própria, o que pode formar uma nova fronteira eficiente. A distância entre o ponto da DMU na fronteira original e na nova fronteira sem a DMU é a supereficiência (SE). Como consequência, os valores de  $\phi$  podem ser menores que 1 e  $E$  maiores do que 1, permitindo que as DMU eficientes sejam ranqueadas uma vez que não permanecem empatadas.

Todavia há o inconveniente de que, sob a suposição de RVE, a solução do problema de programação possa ser impraticável para algumas DMU, principalmente para aquelas na extremidade da fronteira. Nesse sentido, Cook (2009) propôs uma correção que têm êxito em contornar tal dificuldade e obter escores para todas as DMU.

### 3. O Software R

O R, um programa computacional aberto e gratuitamente disponível na internet ([www.R-project.org](http://www.R-project.org)), além de dispor de pacotes de funções especialmente desenvolvidas para a DEA, possui ambiente e linguagem simples e livres das limitações impostas por pacotes pelos programas especialistas para manipulação de dados, permitindo conjugações com uma vasta gama de pacotes que envolvem diversos segmentos de análise quantitativa o que o torna versátil e poderoso.

Uma busca na rede de distribuição do R (*Comprehensive R Archive Network - CRAN*), em 21/07/2018, identificou cinco pacotes dedicados a DEA, são eles o *additiveDEA*, o *Benchmarking*, o *MultiplierDEA*, o *Productivity* e o *rDEA*. Dentre os pacotes disponíveis para o R, destaca-se o **Benchmarking** (BOGETOFT e OTTO, 2018), que calcula os modelos, com diversas suposições de rendimentos de escala e orientações, além de realizar análise de fronteiras estocástica. O **additiveDEA** (SOTERIADES, 2017) é destinado ao cálculo do modelo de DEA aditivo. O **MultiplierDEA** (PUTHANPURA, 2018) calcula os modelos Multiplicativo e de Cross-efficiency. O pacote **productivity** (DAKPO *et al.*, 2018) mede os níveis e as mudanças de produtividade e lucratividade com vários índices baseados na DEA, incluindo o índice Malmquist. O **rDEA** (SIMM e BESSTREMYANNAYA, 2016) realiza a estimação robusta da DEA.

Há também o pacote **TFDEA** (SHOTT e LIM, 2015) que realiza projeções com base na DEA e oferece soluções para situações de supereficiência com soluções impraticáveis, contudo, somente pode ser obtido em versões antigas por não estar mais ativo no CRAN.

### 4. Metodologia

Na aplicação, analisou-se a eficiência relativa dos 38 Institutos da Rede Federal de Educação, Ciência e Tecnologia (IF) no ano de 2014, empregando-se o DEA com rendimentos variáveis de escala (RVE) orientado ao produto, ou DEA-RVE-Produto,

utilizando-se dois insumos, a Relação de Ingressos por Matrícula Atendida (RIM) e a Relação de Alunos por Docentes em Tempo Integral (RAD); e um produto, a Relação de Concluintes por Matrícula Atendida (RCM), selecionados a partir dos 12 indicadores de gestão, definidos no Quadro 1, obtidos do sítio do Tribunal de Contas da União.

Quadro 1 – Indicadores de gestão do TCU com suas respectivas fórmulas

I	Indicador	Fórmula
1	Relação Candidato/Vaga	$RCV = \text{Inscritos}/\text{Vagas ofertadas}$
2	Relação Ingressos/Aluno	$RIM = \text{Ingressantes}/\text{Matrículas} \times 100$
3	Relação Concluintes/Aluno	$RCM = \text{Concluintes}/\text{Matrículas} \times 100$
4	Eficiência Acadêmica dos Concluintes	$EAC = \text{Concluintes}/\text{Finalizados} \times 100$
5	Retenção do Fluxo Escolar	$RFE = \text{Retidos}/\text{Matrículas} \times 100$ , onde $\text{Retidos} = \text{Reprovação} + \text{Trancamentos}$
6	Relação de Alunos/Docente em Tempo integral	$RAD = \text{Matriculados}/\text{Docentes} \times 100$
7	Titulação do Corpo Docente	$TCD = (G \times 1 + A \times 2 + E \times 3 + M \times 4 + D \times 5) / (G + A + E + M + D)$
8	Gastos Correntes por Aluno	$GCA = \text{Gasto Total}/\text{Matriculados} \times 100$
9	Percentual de Gastos com Pessoal	$GCP = \text{Gasto com Pessoal}/\text{Gasto Total}$
10	Percentual de Gastos com outros Custeios	$GOC = \text{Gasto com outros custeios}/\text{Gasto Total} \times 100$
11	Percentual de Gastos com Investimentos	$GCP = \text{Gasto com investimentos}/\text{Gasto Total} \times 100$
12	Renda Familiar	Levantamento em questionário

O tamanho da amostra (N) é um aspecto importante no processo de análise, pois limita o número de variáveis de insumo (m) e de produto (s) que podem ser utilizados, uma vez que um número elevado de variáveis de insumo e produto reduz a capacidade do modelo discriminar as DMU eficientes das ineficientes, na verdade, conforme esclarecem Seiford e Thrall (1990), dado um número suficiente de variáveis, todas ou quase todas as DMU poderão vir a ser classificadas como eficientes. Dyson *et al.* (2001) sugerem que N seja duas vezes o produto de insumos e produtos,  $N=2 \times (m \times s)$ . Neste estudo, foram selecionados dois insumos e um produto, exigindo, no mínimo, 4 DMU; logo, a amostra possui uma dimensão confortável.

O tratamento estatístico e cálculos dos modelos DEA foram realizados com o auxílio dos recursos do programa computacional R. Especificamente para o modelo DEA, utilizou-se do pacote Benchmarking (BOGETOFT e OTTO, 2015) e TFDEA (SHOTT e LIM, 2015) para o cálculo da supereficiência (SE) com a correção de Cook (2009), uma vez que se observaram situações em que a solução é impraticável.

Outras funções foram utilizadas para manipulação dos dados, cálculo de coeficiente correlação e análise de regressão. As principais funções do R utilizadas no estudo são listadas no Quadro 2.

Quadro 2 – Funções do R utilizadas na aplicação

Resultados	Função	Pacote
$\phi$ e $\lambda$	dea	Benchmarking
E, incremento e metas	fórmulas no programa	base
Supereficiência	SDEA	TFDEA
Figuras	sort, cut, table, sum, cumsum, round e barplot	base
Exportação para word	data.frame e write.csv2	base
Exibição de resultados	summary	base
Regressões	lm	stats
Exportação regressões para word	htmlreg	texreg
Correlações	cor e rcorr	stats e Hmisc

Fonte: Os próprios autores

O código para os cálculos dos modelos DEA e de Superficiência, além da geração de figuras e tabelas em formato para exportação para Word e Excel é apresentado no Apêndice.

## 5. Resultados e discussão

Houve dúvidas na classificação de RCV e RIM, uma vez que podem ser interpretados como insumos ou produtos. A RCV pode ser considerada como quantidade de candidatos à disposição dos IF que podem selecionar os melhores alunos para qualificação, facilitando a obtenção de melhores resultados, como maior possibilidade de conclusão e eficiência acadêmica; contudo, pode ser o produto do êxito das operações dos institutos, já que cursos bem sucedidos, normalmente gozam de prestígios junto a potenciais empregadores e candidatos, o que pode resultar em uma demanda mais intensa. A variável RIM também pode ser considerada dubiamente, a princípio, tanto como insumo quanto como produto.

No intuito de se definirem as classificações de RCV e RIM, realizou-se o procedimento proposto por Golany e Roll (1989) que consiste em rodar uma regressão entre a variável que se tem dúvida, como variável dependente (uma de cada vez), contra as demais cujas classificações como insumo (RAD, TCD e GCA) ou produto (RCM, EAC e RFEI) estavam claras.

Rodou-se, então, uma regressão da variável RIM, como variável dependente, contra os insumos RAD, TCD e GCA e contra os produtos RCM, EAC e RFEI, como variáveis independentes. Os resultados estão na última coluna da Tabela 1. Como se pode observar, a variável RIM apresenta correlação fraca com os insumos e forte com os produtos, devendo ser classificada como insumo.

Tabela 1 – Regressões entre RCV e RIM e variáveis de insumo e produto

	RCV (1)	RCV (2)	RCV (3)	RCV (4)	RIM
(Intercept)	29,75*** (10,06)	34,63*** (8,38)	6,70* (2,48)	28,77*** (7,60)	24,15 (17,76)
RAD	-0,34 (0,22)	-0,48*** (0,16)		-0,45** (0,20)	0,14 (0,39)
TCD	-2,22 (1,92)	-1,85 (1,76)			0,75 (3,39)
GCA	-0,00 (0,00)	-0,00** (0,00)		-0,00** (0,00)	0,00 (0,00)
RCM	-0,12 (0,11)		-0,17* (0,09)	-0,03 (0,09)	1,07*** (0,20)
EAC	0,05 (0,06)		0,05 (0,06)		-0,30*** (0,10)
RFEI	2,67 (6,53)		8,98 (6,09)		40,76*** (11,53)

Nota: \*\*\*p < 0,01, \*\*p < 0,05, \*p < 0,10. Os valores entre parênteses correspondem aos erros padrão.

Todavia, o comportamento de RCV não foi claro, exigindo regressões adicionais para uma melhor análise. A primeira regressão (RCV1) envolveu todas as variáveis e não revelou coeficientes significativos. Uma segunda regressão (RCV2) incluiu somente insumos, resultando em coeficientes de GCA e RAD significativos a 1% e 5%,

respectivamente. Na terceira regressão (RCV3), contendo exclusivamente produtos, somente RCM foi significativo, mas a 10%. Uma quarta regressão (RCV4) foi realizada, incluindo as variáveis que foram regressores significativos nas anteriores (GCA, RAD e RCM) e somente as variáveis de insumo foram significativas, sugerindo que RCV talvez se comporte como produto. Diante da frágil definição do indicador, poder-se-ia considerar a sua eliminação da análise, contudo, optou-se por sua manutenção tendo vista a sua importância do ponto de vista gerencial.

Relativamente aos indicadores de gastos (GCA, GCP, GOC e GCI), verificou-se que as informações neles contidas são redundantes. Primeiramente, as variáveis GCP, GOC e GCI que estão em percentuais do gasto total, foram transformadas em valores por alunos. Em seguida, realizou-se a análise de correlação de Pearson, cujo resultado é exibido na tabela a seguir e na qual se pode observar forte e significativa correlação entre as variáveis GCA e as variáveis GCP e GOCA, mas a correlação entre GCA, embora significativa, revelou-se apenas moderada.

Tabela 2 – Matriz de correlações entre as GCA e os demais indicadores de gastos por aluno

	GCA	GCIA	GCPA	GOCA
GCA	1			
GCIA	0,4928***	1		
GCPA	0,9374***	0,2095	1	
GOCA	0,8525***	0,5842***	0,6875***	1

Nota: \*\*\*p < 0,01, \*\*p < 0,05, \*p < 0,10. GCIA = Gastos com Investimentos; GCPA = Gastos com Pessoal por aluno e GOCA= Gastos com outros custeios, deduzindo benefícios e Pasesp

No intuito de dirimir dúvidas com relação à redundância entre as variáveis, especialmente entre GCA e GCIA, rodou-se uma regressão de GCA como variável dependente e GCIA, GCPA e GOCA, conforme sugestão de Golany e Roll (1989). Todos os coeficientes são altamente significantes, de modo que se optou por manter GCA por ser a variável com maiores correlações entre as variáveis e por ser um indicador de gastos totais contendo as informações dos demais indicadores de gastos.

A última etapa da seleção das variáveis que compuseram o modelo DEA foi realizada por meio do procedimento iterativo de Norman e Stoker (1991). Os resultados do procedimento são exibidos nas Tabelas 3 e 4.

Primeiramente, buscou-se a maior correlação entre insumo e produto. De todas as correlações entre pares de insumo e produto, a maior foi entre RIM e RCM (r = 0,6794) que por essa razão passou a ser considerado o par inicial.

Tabela 3 – Resultados do método iterativo de seleção de variáveis

Variáveis	Insumos				Produtos		
	RIM	RAD	TCD	GCA	RCM	EAC	RFEI
RIM	1						
RAD	0,3914	1					
TCD	-0,2347	-0,1712	1				
GCA	-0,2395	-0,7775	0,3215	1			
RCM	0,6794	0,6393	-0,1704	-0,3588	1		
EAC	0,2600	0,4171	-0,0044	-0,2836	0,6740	1	
RFEI	0,3243	-0,1865	-0,2581	-0,0408	-0,0375	0,0476	1
Passo 2		0,5918	-0,1219	-0,394		0,6812	-0,1256
Passo 3			-0,0244	0,1192	-0,0245	0,5824	-0,0273
Passo 4			-0,0309			0,6289	-0,0054

Passo 5	-0,1208	0,5602	0,1346
---------	---------	--------	--------

Nota: As primeiras quatro variáveis são indicadores de insumos e as demais, indicadores de produtos. Os valores em destaque são as correlações que justificaram incorporações das variáveis no modelo a cada passo do processo.

A partir do segundo passo, busca-se a maior correlação entre os indicadores de desempenho calculado pelo par RIM e RCM e as variáveis restantes. Devem ser consideradas as correlações positivas entre a eficiência e insumos e as correlações negativas entre eficiência e produtos. Apesar da maior correlação observada ter sido com o produto EAC, o sinal foi trocado; de modo que foi considerada a variável RAD que apresentou a maior correlação com o sinal correto, insumo com correlação positiva ( $r = 0,5918$ ). Calculam-se os escores de eficiência com RIM e RAD como insumos e RCM como produto.

No terceiro passo, selecionou-se o insumo GCA com maior correlação positiva ( $r = 0,1192$ ). Novamente um produto com correlação maior foi observado (EAC), contudo com sinal trocado. Calcularam-se os escores de eficiência com RIM, RAD e GCA como insumos e RCM como produto. Neste ponto convém observar a evolução da eficiência média na Tabela 4. A inclusão da variável GCA ao modelo provocou um acréscimo de menos de 0,05 à eficiência média, ou seja, a sua presença no modelo pode não ser justificada do ponto de vista de análise de eficiência técnica, contudo, considerando a importância dos custos em questões gerenciais, optou-se por manter o indicador para avaliações adicionais.

No quarto passo, selecionou-se o produto RFEI com maior correlação negativa ( $r = -0,0054$ ). Calcularam-se os escores de eficiência com RIM, RAD e GCA como insumos e RCM e RFEI como produtos. Novamente, a inclusão dessa nova variável pouco acrescentou à eficiência média, somente 0,0217, de forma que sua inclusão no modelo não pode ser justificada. Finalmente, no quinto passo, tendo em vista que todas as variáveis restantes exibiram coeficientes de correlação com sinais trocados, encerrou-se o procedimento.

Tabela 4 – Evolução da eficiência média com incorporação de variáveis ao modelo

Passo	Insumos	Produtos	Ef. Média	Var.	Var. relativa	n	perc.
1	RIM	RCM	0,3920			3	0,0789
2	RIM e RAD	RCM	0,6303	0,2383	0,6079	9	0,2368
3	RIM, RAD e GCA	RCM	0,6780	0,0478	0,0758	12	0,3158
4	RIM, RAD e GCA	RCM e RFEI	0,6997	0,0217	0,0320	14	0,3684

Pode-se observar na Tabela 4 que o acréscimo relativo da inclusão das variáveis à eficiência média nos passos 3 e 4 foram pequenos. Pode ser que tais variáveis pouco agreguem ao modelo de análise em termos de avaliação da eficiência. Uma forma de avaliar as variáveis num modelo DEA é pelo procedimento Jackknifing, sugerido por Huguenin (2015), que envolve a extração de uma das variáveis do modelo completo de cada vez com reposição e o cálculo da correlação entre as eficiências das DMU obtidas entre os escores de eficiência. Elevadas correlações indicam as variáveis cujas eliminações não afetam o modelo. A Figura 1 exhibe as correlações entre as eficiências de cada modelo incompleto com completo e, conforme se observa, o modelo completo é robusto à extração das variáveis GCA e RFEI, confirmando que ambas não se justificam para a análise da eficiência das IF da amostra.

O modelo completo incluiu os insumos RIM e RAD e o produto RCM. Cinco modelos

de eficiência com uma das variáveis extraída foram aplicados, isto é, com três variáveis às quais a eficiência média é mais sensível. Analisou-se, portanto, a eficiência dos 38 IF por meio da DEA com retornos variáveis de escala (RVE) orientado ao produto ou DEA-RVE-Produto, utilizando-se dois insumos (RIM, RAD) e um produto (RCM). Os resultados das eficiências dos IF são apresentados na Figura 2 e nas Tabelas de 5 a 7.

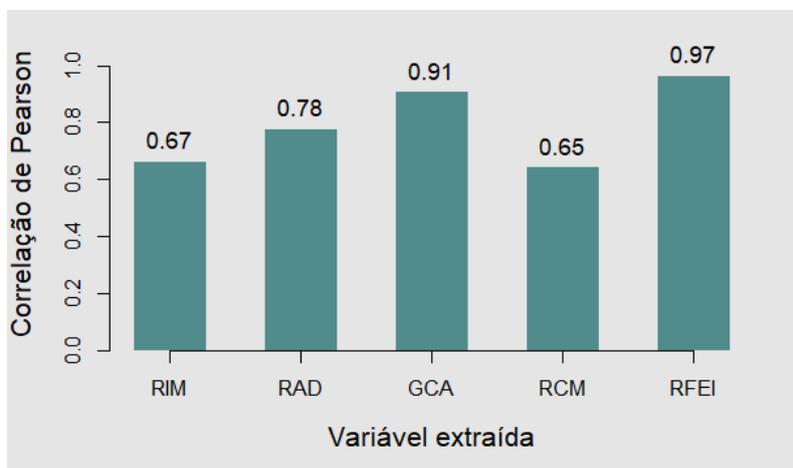


Figura 1 – Análise de sensibilidade das variáveis do modelo com produtos RIM, RAD e GCA e os insumos RCM e RFEI.

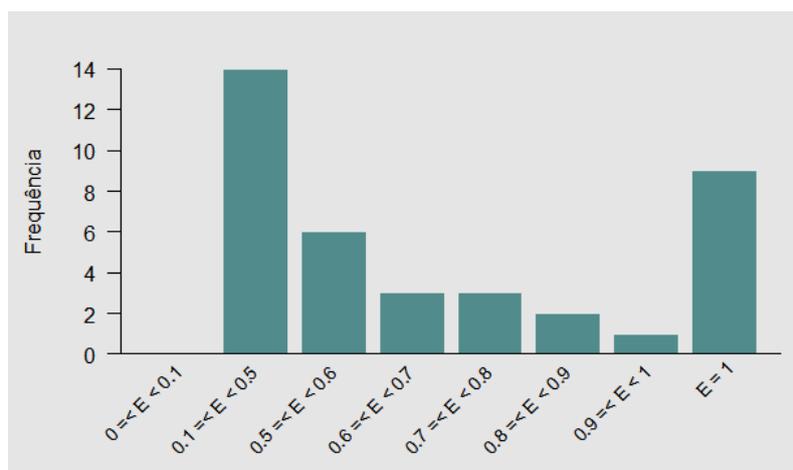


Figura 2 – Distribuição das Eficiências relativas dos IF.

De acordo com os resultados apresentados na Figura 2, 9 IF, correspondentes a 23,68% do total, foram classificados como eficientes ( $E=1$ ); 20 não alcançaram 60% de eficiência, sendo que o pior desempenho foi o do IFMA ( $E_{IFMA}=0,16$ ).

Os IF puderam ser estudados individualmente com maior detalhamento. O IFAP, por exemplo, com  $E=0,78$ , foi classificado como ineficiente, com produção 22% aquém do possível, de modo que, para se tornar eficiente, deve atingir a meta de  $RCM = 52,38$ , o que implica num incremento de seu indicador de produto na ordem de 11,56 sobre o valor atualmente observado (40,82). Tal incremento é justamente 22% da meta (11,56/52,38). Por essa razão, é que se diz que IFAP é 22% ineficiente.

No caso do IFAP, o seu conjunto de referência é formado por dois institutos,  $E_{IFAP}=\{\lambda_{IFRN}=0,041, \lambda_{IFSULDEMINAS}=0,959\}$ , com maior similaridade com o IFSuldeminas, em virtude da maior magnitude de seu  $\lambda$  associado.

Tabela 5 – Eficiência relativa, incrementos potenciais e metas para os produtos dos institutos, com base no modelo DEA-RCE-Produto em 2014.

i	DMU	$\phi$	$E=1/\phi$	RCM	Incremento	Meta	u1	u2	v1
1	IFAC	2,07	0,48	7,74	8,31	16,05	0,000	0,220	0,129
2	IFAL	1,00	1,00	3,31	0,00	3,31	2,262	0,240	0,302
3	IFAP	1,28	0,78	40,82	11,56	52,38	0,000	0,030	0,024
4	IFAM	1,00	1,00	41,15	0,00	41,15	0,034	0,000	0,024
5	IFBA	2,92	0,34	5,34	10,28	15,62	0,573	0,102	0,187
6	IFBaiano	1,16	0,86	13,75	2,20	15,95	0,009	0,104	0,073
7	IFB	2,03	0,49	16,55	17,09	33,64	0,001	0,073	0,060
8	IFC	1,01	0,99	15,87	0,09	15,96	0,000	0,107	0,063
9	IFCE	3,49	0,29	8,70	21,68	30,38	0,014	0,164	0,115
10	IFES	1,92	0,52	11,34	10,40	21,74	0,011	0,126	0,088
11	IFFarroupilha	1,11	0,90	15,75	1,81	17,56	0,008	0,091	0,063
12	IFF	4,00	0,25	4,77	14,29	19,06	0,483	0,186	0,210
13	IFGoiano	4,17	0,24	5,50	17,42	22,92	0,022	0,260	0,182
14	IFG	1,00	1,00	11,41	0,00	11,41	0,000	0,149	0,088
15	IFMA	6,16	0,16	3,35	17,29	20,64	1,088	0,020	0,299
16	IFMT	1,77	0,57	9,80	7,53	17,33	0,065	0,144	0,102
17	IFMS	2,21	0,45	7,31	8,86	16,17	0,000	0,232	0,137
18	IFMG	1,49	0,67	15,15	7,42	22,57	0,008	0,094	0,066
19	IFNMG	2,57	0,39	9,62	15,06	24,68	0,013	0,149	0,104
20	IFPA	1,00	1,00	15,04	0,00	15,04	0,203	0,036	0,066
21	IFPB	5,71	0,18	4,19	19,73	23,92	0,550	0,211	0,239
22	IFPR	2,82	0,35	8,89	16,15	25,04	0,410	0,008	0,112
23	IFPE	1,00	1,00	8,08	0,00	8,08	0,454	0,000	0,124
24	IFPI	2,71	0,37	7,24	12,42	19,66	0,088	0,195	0,138
25	IFRJ	1,00	1,00	9,81	0,00	9,81	0,235	0,090	0,102
26	IFRN	1,00	1,00	28,35	0,00	28,35	0,000	0,045	0,035
27	IFRS	1,79	0,56	14,22	11,22	25,44	0,009	0,101	0,070
28	IFRO	2,00	0,50	10,30	10,28	20,58	0,012	0,139	0,097
29	IFRR	2,95	0,34	5,05	9,83	14,88	0,722	0,013	0,198
30	IFSC	1,64	0,61	19,76	12,65	32,41	0,001	0,061	0,051
31	IFSP	1,29	0,78	14,00	4,04	18,04	0,009	0,102	0,071
32	IFS	1,34	0,75	9,06	3,06	12,12	0,070	0,156	0,110
33	IFSertão	1,81	0,55	13,32	10,73	24,05	0,009	0,107	0,075
34	IFSudesteMG	1,00	1,00	14,29	0,00	14,29	0,044	0,099	0,070
35	IFSULDEMINAS	1,00	1,00	53,41	0,00	53,41	0,006	0,000	0,019
36	IFSul	2,00	0,50	9,83	9,81	19,64	0,013	0,146	0,102
37	IFTO	2,37	0,42	11,38	15,61	26,99	0,056	0,124	0,088
38	IFTM	1,52	0,66	9,60	5,02	14,62	0,000	0,177	0,104

Nota: Os produtos são referentes aos valores atuais dos indicadores de produtos; Incremento são referentes aos incrementos potenciais dos indicadores de produto; e Meta são os valores que os IF devem atingir para se tornarem eficientes.

Uma vez que 23,68% da amostra foi classificada como eficiente e para aprofundamento do estudo, foi realizada ainda uma análise utilizando o Modelo de Supereficiência entre os 9 Institutos considerados eficientes, obtendo-se um ranking entre os eficientes, tendo o IFPE obtido o melhor desempenho entre os melhores, seguido por IFG e IFAM. Deve-se notar também que a supereficiência não altera os valores originais das DMU ineficientes.

Finalmente, vale mencionar a simplicidade do R na aplicação da DEA, uma vez que os valores do modelo, neste estudo,  $\phi$  e  $\lambda$ , foram obtidos com a execução de uma única função, a DEA do pacote Benchmarking, e o ajuste adequado de seus argumentos. As demais medidas derivadas, quais sejam  $E$ , incremento potencial dos produtos e as suas

quantidades metas, foram calculadas com operações matemáticas simples. A segunda função, **SDEA** do pacote TFDEA, foi utilizada em virtude da necessidade de cálculo da

Tabela 6 – Lambdas dos IF produzidos.

i	DMU	L1	L2	L3	L4	L5	L6	L7	L8	L9	L10	L11	L12	L13	L14	L15	L16	L17	L18	L19	L20	L21	L22	L23	L24	L25	L26	L27	L28	L29	L30	L31	L32	L33	L34	L35	
1	IFAC	0	0	0	0	0	0	0	0	0	0	0	0	0	0,7262	0	0	0	0	0	0	0	0	0	0	0	0,2738	0	0	0	0	0	0	0	0	0	0
2	IFAL	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3	IFAP	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0,041	0	0	0	0	0	0	0	0	0	0,959	0
4	IFAM	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
5	IFBA	0	0	0	0,1117	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0,4405	0	0	0	0	0,4478	0	0	0	0	0	0	0	0	0	0	
6	IFBaiano	0	0	0	0,0674	0	0	0	0	0	0	0	0	0	0,7828	0	0	0	0	0	0	0	0	0	0	0	0,1498	0	0	0	0	0	0	0	0	0	
7	IFB	0	0	0	0,0406	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0,7692	0	0	0	0	0	0	0	0	0	0,1902	
8	IFC	0	0	0	0	0	0	0	0	0	0	0	0	0	0,7312	0	0	0	0	0	0	0	0	0	0	0	0,2688	0	0	0	0	0	0	0	0	0	
9	IFCE	0	0	0	0,5956	0	0	0	0	0	0	0	0	0	0,3303	0	0	0	0	0	0	0	0	0	0	0	0,0741	0	0	0	0	0	0	0	0	0	
10	IFES	0	0	0	0,2548	0	0	0	0	0	0	0	0	0	0,5826	0	0	0	0	0	0	0	0	0	0	0	0,1626	0	0	0	0	0	0	0	0	0	
11	IFFarroupilha	0	0	0	0,0289	0	0	0	0	0	0	0	0	0	0,6591	0	0	0	0	0	0	0	0	0	0	0	0,312	0	0	0	0	0	0	0	0	0	
12	IFF	0	0	0	0,2699	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0,5527	0	0	0	0	0	0	0	0	0	0,1774	
13	IFGoiano	0	0	0	0,2912	0	0	0	0	0	0	0	0	0	0,5406	0	0	0	0	0	0	0	0	0	0	0	0,1682	0	0	0	0	0	0	0	0	0	
14	IFG	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
15	IFMA	0	0	0	0,2335	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0,6949	0	0	0,0716	0	0	0	0	0	0	0	0	0	0	0	0	0
16	IFMT	0	0	0	0,1762	0	0	0	0	0	0	0	0	0	0,5874	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0,2364	
17	IFMS	0	0	0	0	0	0	0	0	0	0	0	0	0	0,7192	0	0	0	0	0	0	0	0	0	0	0	0,2808	0	0	0	0	0	0	0	0	0	
18	IFMG	0	0	0	0,0883	0	0	0	0	0	0	0	0	0	0,4081	0	0	0	0	0	0	0	0	0	0	0	0,5036	0	0	0	0	0	0	0	0	0	
19	IFNMG	0	0	0	0,1065	0	0	0	0	0	0	0	0	0	0,2974	0	0	0	0	0	0	0	0	0	0	0	0,5961	0	0	0	0	0	0	0	0	0	
20	IFPA	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
21	IFPB	0	0	0	0,4287	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0,4207	0	0	0	0	0	0	0	0	0	0,1505	0
22	IFPR	0	0	0	0,4904	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0,1071	0	0	0,4024	0	0	0	0	0	0	0	0	0	0	0	0	0
23	IFPE	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
24	IFPI	0	0	0	0,2213	0	0	0	0	0	0	0	0	0	0,2013	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0,5774	0
25	IFRJ	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	
26	IFRN	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	
27	IFRS	0	0	0	0,386	0	0	0	0	0	0	0	0	0	0,4634	0	0	0	0	0	0	0	0	0	0	0	0,1506	0	0	0	0	0	0	0	0	0	
28	IFRO	0	0	0	0,1782	0	0	0	0	0	0	0	0	0	0,5934	0	0	0	0	0	0	0	0	0	0	0	0,2284	0	0	0	0	0	0	0	0	0	
29	IFRR	0	0	0	0,0042	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0,9566	0	0	0,0392	0	0	0	0	0	0	0	0	0	0	0	0	0
30	IFSC	0	0	0	0,2102	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0,7351	0	0	0	0	0	0	0	0	0	0,0547	
31	IFSP	0	0	0	0,0648	0	0	0	0	0	0	0	0	0	0,6576	0	0	0	0	0	0	0	0	0	0	0	0,2776	0	0	0	0	0	0	0	0	0	
32	IFS	0	0	0	0,0046	0	0	0	0	0	0	0	0	0	0,7946	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0,2008	0	
33	IFSertão	0	0	0	0,3421	0	0	0	0	0	0	0	0	0	0,5123	0	0	0	0	0	0	0	0	0	0	0	0,1457	0	0	0	0	0	0	0	0	0	
34	IFSudesteMG	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
35	IFSULDEMINAS	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
36	IFSul	0	0	0	0,2377	0	0	0	0	0	0	0	0	0	0,6936	0	0	0	0	0	0	0	0	0	0	0	0,0687	0	0	0	0	0	0	0	0	0	0
37	IFTO	0	0	0	0,5161	0	0	0	0	0	0	0	0	0	0,4034	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0,0805	0
38	IFTM	0	0	0	0	0	0	0	0	0	0	0	0	0	0,8104	0	0	0	0	0	0	0	0	0	0	0	0,1896	0	0	0	0	0	0	0	0	0	0

Fonte: Os próprios autores. L(i) refere-se ao λ do IF com relação ao seu parceiro de excelência i.

supereficiência com a correção de Cook, uma vez que houve DMU com soluções impraticáveis, caso contrário, o pacote Benchmarking oferece a função **sdea** para a supereficiência, no entanto, sem a opção de correção, que pode ser útil para os modelos com RCE, ou mesmo RVE se não apresentarem soluções impraticáveis.

Tabela 7 – Rank dos IF, com base na Supereficiência (SE).

i	DMU	SE	i	DMU	SE
1	IFPE	3,4565	4	IFRJ	1,9079
2	IFG	2,3322	5	IFAL	1,3973
3	IFAM	1,9483	6	IFSULDEMINAS	1,2979
			7	IFSudesteMG	1,2142
			8	IFPA	1,1197
			9	IFRN	1,0868

### 6. Considerações Finais

A aplicação neste artigo evidenciou que o software R constitui uma vantajosa opção para aplicação da DEA, não somente por dispor de pacotes capazes de resolver os seus principais modelos, mas por permitir ampla versatilidade e liberdade na manipulação de dados, nos cálculos e na análise gráfica.

O estudo foi inteiramente realizado no R, sendo realizados cálculos de coeficientes de correlação, regressão múltipla, transformação e manipulação de dados, geração de gráficos e tabelas especialmente elaborados para as necessidades do estudo e exportação

dos resultados para o processador de texto. Podem-se realizar análises estatísticas tradicionais e mais aprimoradas, como *bootstrap*, tratamentos especificamente desenvolvidos para DEA, ou mesmo personalizados sem a necessidade de se recorrer a outro programa.

Um obstáculo para a sua utilização do R é a sua curva de aprendizagem relativamente mais longa. Acredita-se que a familiarização com a sua linguagem pode ser facilitada se for usada frequentemente no ensino de várias disciplinas quantitativas na academia que podem usufruir de seus pacotes, as quais citamos a administração financeira, engenharia econômica, pesquisa operacional, estatística, matemática, economia, teoria da decisão, entre outras.

A produtividade de estudos com os recursos do R pode ser ainda mais potencializada, considerando a existência de interfaces mais amigáveis, como a interface RStudio, que facilitam a sua utilização; a sua capacidade de exportar resultados e dados para planilhas eletrônicas e para o popular processador de texto Word, bem como a sua interação com os programas de marcação para a editoração de documentos Latex e Markdown, permitindo a geração quase instantânea de relatórios em diversos formatos, inclusive HTML. Acrescentando-se o fato de ser aberto e livre, conclui-se que o R é uma ferramenta promissora como apoio para a DEA e para as diversas áreas de análise quantitativa.

### Referências

- ANDERSEN, P.; PETERSEN, N. C. (1993), *A procedure for ranking efficient units in data envelopment analysis*, *Management science*, Vol. 39, No. 10, pp. 1261-1264, disponível em: <https://pubsonline.informs.org/doi/abs/10.1287/mnsc.39.10.1261> (acesso em 09 ago. 2018).
- BANKER, R.D.; CHARNES, R.F.; COOPER, W.W. (1984), *Some models for estimating technical and scale inefficiencies in Data Envelopment Analysis*, *Management Science*. Vol. 30, No. 9, pp. 1078–1092, 1984, disponível em <https://doi.org/10.1287/mnsc.30.9.1078> (acesso em 09 ago. 2018).
- BOGETOFT, P.; OTTO, L. (2018), *Benchmarking with DEA and SFA*, R package version 0.27, disponível em: <https://cran.r-project.org/web/packages/Benchmarking/index.html> (acesso em 30 jul. 2018).
- CHARNES, A.; COOPER, W.W.; RHODES, E. (1978), *Measuring the efficiency of decision making units*. *European journal of operational research*, Vol. 2, No. 6, pp. 429-444, disponível em: <https://www.sciencedirect.com/science/article/pii/0377221778901388> (acesso em 30 jul. 2018).
- COOK, W. D.; LIANG, L.; ZHA, Y.; ZHU, J. (2009), *A modified super-efficiency DEA model for infeasibility*, *Journal of the Operational Research Society*, Vol. 60, No. 2, pp. 276-281, disponível em: <https://www.tandfonline.com/doi/abs/10.1057/palgrave.jors.2602544> (acesso em 30 jul. 2018).
- DAKPO, H.; DESJEUX, Y.; LATRUFFE, L. (2018). *productivity: Indices of Productivity and Profitability Using Data Envelopment Analysis (DEA)*, R package version 1.1.0, disponível em: <https://CRAN.R-project.org/package=productivity/> (acesso em 30 jul. 2018).
- DYSON, R.G.; ALLEN, R.; CAMANHO, A.S.; PODINOVSKI, V.V.; SARRICO, C.S.; SHALE, E.A. (2001), *Pitfalls and protocols in DEA*, *European Journal of operational research*, Vol. 132, No. 2, pp. 245-259, disponível em: [https://doi.org/10.1016/S0377-2217\(00\)00149-1](https://doi.org/10.1016/S0377-2217(00)00149-1) (acesso em 30 jul. 2018).
- GOLANY, B.; ROLL, Y. (1989), *An application procedure for DEA*, *Omega*, Vol. 17, No. 3, pp. 237-250, disponível em: [https://doi.org/10.1016/0305-0483\(89\)90029-7](https://doi.org/10.1016/0305-0483(89)90029-7) (acesso em 1 jul. 2018).
- NORMAN, M.; STOCKER, B. (1991) *Data Envelopment Analysis: the assessment of performance*. John Wiley and Sons Inc., New York.
- PUTHANPURA, A.K. (2018). *MultiplierDEA: Multiplier Data Envelopment Analysis and Cross Efficiency*, R package version 0.1.18, disponível em: <https://CRAN.R-project.org/package=MultiplierDEA> (acesso em 30 jul. 2018).

**R CORE TEAM (2018)**, *R: a language and environment for statistical computing*, Vienna, Austria, disponível em: <https://www.R-project.org/> (acesso em 1 jul. 2018).

**SEIFORD, L. M.; THRALL, R. M.** (1990), *Recent developments in DEA: the mathematical programming approach to frontier analysis*. *Journal of econometrics*, Vol. 46, No. 1, pp. 7-38, disponível em: [https://doi.org/10.1016/0304-4076\(90\)90045-U](https://doi.org/10.1016/0304-4076(90)90045-U) (acesso em 1 jul. 2018).

**SEIFORD, L. M.; ZHU, J.** (1999), *Infeasibility of super-efficiency data envelopment analysis models*, *INFOR: Information Systems and Operational Research*, Vol. 37, No. 2, pp. 174-187, disponível em: <https://doi.org/10.1080/03155986.1999.11732379> (acesso em 1 jul. 2018).

**SHOTT, T.; LIM, D. J.** (2015), *TFDEA: Technology Forecasting using DEA (Data Envelopment Analysis)*, R package version 0.9.8.3, disponível em: <https://CRAN.R-project.org/package=TFDEA> (acesso em 30 jul. 2018).

**SIMM, J.; BESSTREMYANNAYA, G.** (2016), *rDEA: Robust Data, Envelopment Analysis (DEA) for R*, R package version 1.2-5, disponível em: <https://CRAN.R-project.org/package=rDEA> (acesso em 30 jul. 2018).

**SOTERIADES, A.** (2017), *additiveDEA: Additive Data Envelopment Analysis Models*, R package version 1.1, disponível em: <https://CRAN.R-project.org/package=additiveDEA> (acesso em 30 jul. 2018).

#### APENDICE – Código em R para cálculo do modelo DEA-RCE-Produto.

```
x <- matrix(c(dados$rim,dados$rad),nrow=38, ncol=2,
             dimnames=list(dados$dmu, (c("RIM","RAD"))))
y <- matrix(c(dados$rcm), nrow=38, ncol=1,
            dimnames=list(dados$dmu, (c("RCM"))))
install.packages("Benchmarking") # instala pacote
library(Benchmarking)           # carrega pacote
e <- dea(x, y, RTS="vrs", ORIENTATION="out")
e$eff # Phi
1/e$eff # Eficiência técnica
peers(e, NAMES=TRUE) # benchmarks
e$lambda
e.dual <- dea.dual(x, y, RTS="vrs", ORIENTATION="out")
#Relatório
incremento<-data.frame(y*(e$eff-1))
meta<-data.frame(y+y*(e$eff-1))
et<-data.frame(et)
e.tab <- data.frame("PHI"=e$eff, "E"=et$et,
                  "Incremento"=incremento$RCM, "Meta"=meta$RCM,e.dual$u, e.dual$v)
print(round(e.tab,4))
write.csv2(round(e.tab,4))# para excel/word, separação com";"
# Gráfico de eficiências
et.sort<-sort(1/e$eff)
scores <- cut(et.sort, breaks = bins,right = F)
par(mar=c(5, 4.5, 1.5, 1))
par(bg="gray90")
mp <- barplot(table(scores), xaxt="n", las=2,ylim=c(0,15),
              ylab="Frequência", col="darkslategray4", border=F)
lablist.x<-as.vector(c("0 =< E < 0.1","0.1 =< E < 0.5",
                      "0.5 =< E < 0.6",
                      "0.6 =< E < 0.7", "0.7 =< E < 0.8",
                      "0.8 =< E < 0.9","0.9 =< E < 1","E = 1"))
text(mp, par("usr")[3], labels = lablist.x, srt = 45, adj =
c(1.1,1.1), xpd = TRUE, cex=.9)
abline(h=0)
install.packages("TFDEA") # instala pacote
library(TFDEA)           # carrega pacote
super<-SDEA(x, y, rts="vrs", orientation="output", slack=T,
            dual = F,cook=T, second="none", z=0, round=F,debug=1)
```

```
e.super<-data.frame("DMU"=sort(1/super$eff,decreasing = T))
```