

Implementando modelos DEA no R

José Francisco Moreira Pessanha
professorjfm@hotmail.com
UERJ

Alexandre Marinho
alexandre.marinho@ipea.gov.br
UERJ

Luiz da Costa Laurencel
llaurenc.ntg@terra.com.br
UERJ

Marcelo Rubens dos Santos do Amaral
mrubens@ime.uerj.br
UERJ

Resumo: O presente artigo tem por objetivo apresentar a implementação dos modelos clássicos da Análise Envoltória de Dados (DEA – Data Envelopment Analysis) no ambiente R, em particular o modelo com retornos constantes de escala (CRS) e o modelo com retorno variável de escala (VRS), nas versões dos multiplicadores e envelope e com orientação ao insumo. A implementação computacional é ilustrada por meio de um estudo de caso com as 18 maiores distribuidoras do setor elétrico brasileiro.

Palavras Chave: DEA - R - computação - energia elétrica -

1. INTRODUÇÃO

O presente artigo tem por objetivo apresentar a implementação dos modelos clássicos da Análise Envoltória de Dados (DEA – *Data Envelopment Analysis*) no ambiente R, em particular o modelo com retornos constantes de escala (CRS) e o modelo com retorno variável de escala (VRS), nas versões dos multiplicadores e envelope e com orientação ao insumo. A implementação computacional é ilustrada por meio de um estudo de caso com as 18 maiores distribuidoras do setor elétrico brasileiro.

Introduzida por Charnes, Cooper e Rhodes em 1978, a Análise Envoltória de Dados é uma importante área da Pesquisa Operacional e da Ciência Econômica, conforme atestado pela enorme quantidade de publicações com aplicações práticas e desenvolvimentos teóricos ao longo de pouco mais de três décadas (COOK & SEIFORD, 2009). De forma resumida, a Análise Envoltória de Dados pode ser descrita como uma técnica não paramétrica, baseada em programação linear, para a avaliação das eficiências de organizações (*Decision Making Units - DMU*) que atuam em um mesmo ramo de atividade, por exemplo, escolas públicas, agências bancárias, fábricas e concessionárias de serviços públicos.

Atualmente a Análise Envoltória de Dados conta com uma variedade de modelos que abrangem desde os modelos DEA clássicos (descritos neste artigo) e suas variações, até abordagens que combinam os modelos DEA com métodos de simulação de Monte Carlo, modelos estatísticos e lógica fuzzy.

Paralelamente aos avanços teóricos, observa-se a evolução dos programas dedicados à DEA, desde *softwares* comerciais como o DEA Solver Pro (<http://www.saitech-inc.com/products/prod-dsp.asp>), *Frontier Analyst* (<http://www.banxia.com/frontier/>), MaxDEA (<http://www.maxdea.cn/>) e PIM-DEASoft (<http://www.deasoftware.co.uk/>), até programas não comerciais como DEAP (<http://www.uq.edu.au/economics/cepa/deap.php>) e SIAD (<http://www.uff.br/decisao/>). Uma lista mais abrangente dos programas dedicados à DEA e suas funcionalidades pode ser encontrada em (ÂNGULO-MEZA et al, 2005). Destaca-se também páginas na *internet* que oferecem a possibilidade de obter resultados de diferentes modelos DEA mediante a simples introdução dos dados de entrada, por exemplo, <http://www.dea.uni-hohenheim.de/index.php>. Adicionalmente, os modelos DEA podem ser programados em planilha eletrônica por meio do Solver Excel, conforme descrito por Ragsdale (2009).

Embora os programas comerciais e livres sejam práticos e contem com diversos modelos e recursos para a aplicação da Análise Envoltória de Dados, a possibilidade de programar modelos DEA, em uma planilha eletrônica ou em qualquer outra linguagem computacional, é interessante, pois oferece ao analista uma grande flexibilidade na implementação computacional dos modelos, inclusive na investigação de inovações.

Mais recentemente, o advento do programa R (<http://www.r-project.org/>), um software livre e *open source*, altamente extensível, disponibiliza uma variedade de funções para análise de dados e rotinas gráficas por meio de funções nativas ou obtidas em pacotes (*packages*) distribuídos na *internet*. Por exemplo, o R conta com dois pacotes dedicados aos modelos DEA: o FEAR (*Frontier Efficiency Analysis with R*) (<http://www.clemson.edu/economics/faculty/wilson/Software/FEAR/fear.html>) e o *Benchmarking* (BOGETOFT & OTTO, 2011). Contudo o R é mais do que uma simples biblioteca de pacotes, pois permite que os analistas construam seus próprios programas ou *packages*. Portanto, por meio do R, os analistas podem obter soluções de baixo custo e livres das restritas opções oferecidas pelos programas dedicados apenas aos modelos DEA.

O artigo está organizado em quatro seções. A seguir, na seção 2, tem-se a formulação matemática dos modelos DEA clássicos (modelos CRS e VRS). Na sequência, na seção 3 são apresentados os códigos para execução dos modelos DEA clássicos no R. Por fim, na seção 4 são apresentadas as principais conclusões do artigo.

2. MODELOS DEA CLÁSSICOS

Considere uma tecnologia de produção que transforma um vetor com s tipos de *inputs* $X = \{x_1, \dots, x_s\} \in R_+^s$ em um vetor com m tipos de *outputs* $Y = \{y_1, \dots, y_m\} \in R_+^m$. A tecnologia de produção pode ser representada pelo conjunto de possibilidades de produção, definido como o conjunto formado por todas as possíveis formas de produzir o vetor de produtos Y a partir de diferentes combinações dos insumos no vetor X :

$$T(X, Y) = \{ (X, Y) \mid \text{é viável produzir } Y \text{ a partir de } X \}$$

Sob o enfoque de conservação de recursos (orientação ao insumo), define-se a medida de eficiência técnica (θ) de uma DMU (X, Y) como sendo a máxima contração radial do vetor de insumos X que permite produzir a mesma quantidade de produtos Y , ou seja:

$$\text{Eficiência} = \text{Min} \{ \theta \mid (\theta X, Y) \in T(X, Y) \}$$

onde θ pode assumir um valor menor ou igual a unidade.

Um valor unitário para θ indica que não é possível reduzir a quantidade de insumos e manter a mesma produção. Neste caso a DMU é tecnicamente eficiente. Caso contrário, quando $\theta < 1$, significa que há um excesso de insumos e a DMU é tecnicamente ineficiente.

Com base nos resultados acima e admitindo as hipóteses de rendimentos constantes de escala e tecnologia convexa Charnes, Cooper e Rhodes (1978) propuseram o primeiro modelo DEA, conhecido como CCR ou CRS (*Constant Return of Scale*). Neste modelo a eficiência é formulada como um Problema de Programação Linear (PPL), onde a função objetivo é a máxima contração dos insumos (orientação ao insumo) e as restrições do problema representam o conjunto de possibilidades de produção.

Posteriormente, Banker, Charnes e Cooper (1984) adicionaram uma restrição de combinação convexa no modelo CRS, criando um modelo que contempla a hipótese de rendimentos variáveis de escala, conhecido como BCC ou VRS (*Variable Return of Scale*).

A seguir, na Tabela 1 são apresentados dois modelos DEA com orientação ao insumo e na formulação envelope, onde N é o total de DMUs analisadas e o par (X_j, Y_j) representa os vetores de insumos e produtos da j -ésima DMU, $j=1, N$.

Denotando a solução ótima dos modelos (1) e (2) por $(\theta^*, \lambda^*_1, \dots, \lambda^*_N)$, a DMU avaliada (DMU j_0) é eficiente, se e somente se, $\theta^*=1$ e todas as folgas nas restrições são nulas na solução ótima. Caso contrário, quando $\theta^* < 1$ ou $\theta^*=1$, porém com folgas positivas, a DMU j_0 é ineficiente. Os coeficientes $\lambda^*_j > 0$ $j=1, N$ indicam as DMU que compõem o conjunto de referência (*peer set*) da DMU j_0 .

Tabela 1: Modelos DEA com orientação ao insumo na formulação envelope.

Modelo CRS	Modelo VRS
$efici\ênc\i\i\ a\ DMU_{j_0} = Min\ \theta \quad (1)$ <p>s.a.</p> $\theta X_{j_0} \geq \sum_{j=1}^N \lambda_j X_j$ $Y_{j_0} \leq \sum_{j=1}^N \lambda_j Y_j$ $\lambda_j \geq 0 \forall j = 1, \dots, j_0, \dots, N$ $\theta \geq 0$	$efici\ênc\i\i\ a\ DMU_{j_0} = Min\ \theta \quad (2)$ <p>s.a.</p> $\theta X_{j_0} \geq \sum_{j=1}^N \lambda_j X_j$ $Y_{j_0} \leq \sum_{j=1}^N \lambda_j Y_j$ $\sum_{j=1}^N \lambda_j = 1$ $\lambda_j \geq 0 \forall j = 1, \dots, j_0, \dots, N$
<p>$m+s$ restrições $N+1$ variáveis</p>	<p>$m+s+1$ restrições $N+1$ variáveis</p>

Quando uma DMU transforma x unidades de um tipo de insumo em y unidades de um tipo de produto, a sua eficiência pode ser avaliada pelo quociente de produtividade total y/x . A generalização, para o caso com múltiplos insumos e múltiplos produtos, consiste em calcular o seguinte quociente:

$$efici\ênc\i\i\ a = \frac{u_1 y_1 + u_2 y_2 + \dots + u_m y_m}{v_1 x_1 + v_2 x_2 + \dots + v_s x_s} = \frac{U \cdot Y}{V \cdot X} \quad (3)$$

onde os vetores $V=(v_1, v_2, \dots, v_s)$ e $U=(u_1, u_2, \dots, u_m)$ denotam os pesos atribuídos às quantidades de *inputs* e *outputs* respectivamente.

Charnes, Cooper e Rhodes (1978) sugerem que os vetores U e V sejam determinados por meio do problema de programação linear (4) indicado na Tabela 2 e denominado modelo DEA/CRS com orientação ao insumo na versão dos multiplicadores. Ainda na Tabela 2, o PPL (5) é o modelo DEA/VRS com orientação ao insumo na versão dos multiplicadores. Prova-se que o modelo (4) é o dual do modelo (1) e que o modelo (5) é o dual do modelo (2).

Tabela 2: Modelos DEA com orientação ao insumo na formulação dos multiplicadores.

Modelo CRS	Modelo VRS
$\theta = Max \sum_{i=1}^m u_i y_{i,j_0} \quad (4)$ <p>s.a.</p> $-\sum_{i=1}^s v_i x_{ij} + \sum_{i=1}^m u_i y_{ij} \leq 0 \quad \forall j = 1, \dots, j_0, \dots, N$ $\sum_{i=1}^s v_i x_{i,j_0} = 1$ $u_i \geq 0 \quad \forall i = 1, m$ $v_i \geq 0 \quad \forall i = 1, s$	$\theta = Max \sum_{i=1}^m u_i y_{i,j_0} + u_0 \quad (5)$ <p>s.a.</p> $-\sum_{i=1}^s v_i x_{ij} + \sum_{i=1}^m u_i y_{ij} + u_0 \leq 0 \quad \forall j = 1, \dots, j_0, \dots, N$ $\sum_{i=1}^s v_i x_{i,j_0} = 1$ $u_i \geq 0 \quad \forall i = 1, m$ $v_i \geq 0 \quad \forall i = 1, s$
<p>$N+1$ restrições $m+s$ variáveis</p>	<p>$N+1$ restrições $m+s+1$ variáveis</p>

No PPL (5) a variável u_0 é irrestrita em sinal e indica se o rendimento é decrescente ($u_0 < 0$), constante ($u_0 = 0$) ou crescente ($u_0 > 0$) em escala. Denotando a solução ótima de (4) e

(5) por (θ^*, u^*, v^*) , a DMU j_0 é considerada eficiente, se e somente se, $\theta^*=1$ e todos os elementos de u^* e v^* são positivos. Caso contrário, quando $\theta^* < 1$ ou quando $\theta^*=1$, porém com elementos nulos em u^* e v^* , a DMU j_0 é considerada ineficiente. O *peer set* de uma DMU ineficiente é formado pelas DMUs associadas com as restrições ativas.

3. IMPLEMENTANDO O MODELO DEA NO R

O programa pode ser organizado em três partes: entrada de dados, processamento e gravação dos resultados. Na maioria das vezes os dados de entrada de um modelo DEA encontram-se armazenados em uma planilha MS Excel, por exemplo, a planilha ilustrada na Figura 1 contendo dados de 2009 das 18 maiores distribuidoras de energia elétrica que atuam no Brasil, onde são informados o custo operacional total em R\$ (OPEX), a extensão total (km) da rede de distribuição (NETWORK), o volume de energia transacionada ao longo do ano (MWh) e o número de unidades consumidoras (CUSTOMERS).

	A	B	C	D	E
1	UTILITY	OPEX	NETWORK	MWH	CUSTOMERS
2	ELETROPAULO	1,249,143,613	45,213	39,922,710	5,987,873
3	CEMIG	1,682,334,644	460,219	37,476,802	6,832,546
4	CPFL - Paulista	497,290,782	89,879	25,267,579	3,502,793
5	COPEL	1,018,866,491	224,817	23,525,040	3,628,209
6	LIGHT	557,206,112	58,074	22,902,552	3,640,182
7	CELESC	721,455,274	144,896	18,105,811	2,237,127
8	COELBA	436,436,014	215,001	14,286,757	4,622,046
9	ELEKTRO	414,602,018	107,116	13,398,558	2,123,670
10	CPFL - Piratininga	195,789,961	22,236	13,013,378	1,367,488
11	BANDEIRANTE	286,832,273	27,496	12,536,237	1,482,518
12	CELPE	350,651,684	120,428	10,001,560	2,994,259
13	AMPLA	436,532,756	51,050	9,506,961	2,365,558
14	CELG	691,472,253	199,494	9,344,291	2,213,198
15	RGE	186,357,415	84,997	7,993,103	1,226,079
16	COELCE	316,166,876	120,300	7,929,212	2,744,830
17	ESCELSA	241,433,335	56,960	7,897,969	1,185,432
18	AES SUL	206,122,962	76,133	7,616,460	1,150,518
19	CEEE	385,990,997	71,892	7,277,929	1,438,072

Figura 1: Matriz de Dados (Fonte: Agência Nacional de Energia Elétrica - Aneel)

O serviço prestado por uma distribuidora tem duas dimensões básicas: o volume de energia distribuída e o número de unidades consumidoras atendidas e que juntos constituem os principais *drivers* do custo operacional. Adicionalmente, os custos operacionais também são influenciados por fatores não controláveis pela distribuidora, por exemplo, a dispersão das unidades consumidoras um aspecto associado com o tamanho e a cobertura da rede de distribuição. Para um determinado nível de serviço, a distribuidora deve provê-lo com o menor custo possível. Com base nas considerações acima e tendo como objetivo a obtenção de uma medida de eficiência que quantifique o potencial de redução dos custos operacionais, em cada concessionária, a Agência Nacional de Energia Elétrica (ANEEL) propõe um modelo DEA com orientação ao insumo, no qual o OPEX é o único *input* e com três *outputs*: volume de energia distribuída, número de unidades consumidoras e extensão da rede de distribuição.

A importação dos dados pode ser realizada pelo comando *read.xls*, uma função disponibilizada no pacote *xlsReadWrite* (<http://cran.r-project.org/web/packages/xlsReadWrite/xlsReadWrite.pdf>). Admitindo que os dados estejam armazenados em uma planilha denominada *dados.xls*, localizada no diretório *c:\exemplo*, a importação dos dados da planilha para o R pode ser realizada pelos seguintes comandos (comentário após #):

```
require(lpSolve) # carrega pacote lpSolve previamente instalado
require(xlsReadWrite) # carrega pacote xlsReadWrite previamente instalado
setwd('c:/exemplo') # define o diretório de trabalho
```

```
data <- read.xls('dados.xls',colNames =TRUE,sheet=1, type='data.frame',from=1) # importa os dados
```

No código R acima, `colNames=TRUE` indica que a primeira linha contém os rótulos das variáveis, `sheet=1` informa que os dados estão na pasta 1 do arquivo e que devem ser lidos a partir da célula A1 da planilha (`from=1`), enquanto a opção `type` informa ao R que o objeto `data` (que recebe os dados da planilha) é do tipo `data.frame` (VERZANI, 2005).

```
> data
      UTILITY      OPEX      NETWORK      MWH CUSTOMERS
1      ELETROPAULO 1249143613  45212.99 39922710  5987873
2      CEMIG      1682334644  460219.00 37476802  6832546
3      CPFL - Paulista 497290782  89879.00 25267579  3502793
4      COPEL      1018866491  224817.29 23525040  3628209
5      LIGHT      557206112   58074.00 22902552  3640182
6      CELESC      721455274  144896.32 18105811  2237127
7      COELBA      436436014  215001.47 14286757  4622046
8      ELEKTRO      414602018  107115.75 13398558  2123670
9      CPFL - Piratininga 195789961  22235.63 13013378  1367488
10     BANDEIRANTE 286832273  27496.38 12536237  1482518
11     CELPE      350651684  120427.84 10001560  2994259
12     AMPLA      436532756   51050.29 9506961  2365558
13     CELG      691472253  199494.10 9344291  2213198
14     RGE      186357415   84996.52 7993103  1226079
15     COELCE      316166876  120299.97 7929212  2744830
16     ESCELSA      241433335   56959.90 7897969  1185432
17     AES SUL      206122962   76133.22 7616460  1150518
18     CEEE      385990997   71892.26 7277929  1438072
```

Figura 2: The data object in R and the data import code

O objeto `data` contém a lista de variáveis candidatas aos insumos e produtos do modelo DEA. Assim, é necessário indicar quais são as variáveis consideradas no modelo e classificá-las entre insumos (*inputs*) e produtos (*outputs*). Conforme ilustrado na Figura 2, a variável *input* (OPEX) está na segunda coluna da matriz de dados, enquanto os *outputs* estão nas colunas 3 (NETWORK), 4 (MWH) e 5 (CUSTOMERS). A seleção dos *inputs* e *outputs* pode ser feita pelos seguintes comandos do R:

```
inputs<- data.frame(data[,2]) # seleciona coluna 2 do objeto data
```

```
outputs <- data.frame(data[,c(3,4,5)]) # seleciona colunas 3, 4 e 5 do objeto data
```

```
N<- dim(data)[1] # número de DMUs
```

```
s <- dim(inputs)[2] # número de inputs
```

```
m <- dim(outputs)[2] # número de outputs
```

O processamento de modelo DEA consiste em resolver um PPL para cada uma das N DMUs analisadas. No caso de um modelo DEA/CRS orientado ao insumo na versão dos multiplicadores, para cada DMU é resolvido um modelo com a mesma estrutura do PPL (4). A solução do PPL no R pode ser obtida por meio da função `lp` do pacote `lpSolve` (<http://cran.r-project.org/web/packages/lpSolve/lpSolve.pdf>), cujos argumentos de entrada são os coeficientes da função objetivo, a orientação do problema (maximização ou minimização), os coeficientes tecnológicos das restrições e os respectivos valores do RHS e o tipo de cada restrição (\leq , $=$, \geq). O PPL (4) pode ser expresso da seguinte maneira:

$$\begin{aligned}
 & \text{Max} && c^T \cdot z \\
 & \text{s.t.} && [-\text{inputs} \quad \text{outputs}] \cdot z \leq 0 \\
 & && b^T \cdot z = 1 \\
 & && z \geq 0
 \end{aligned} \tag{6}$$

Em (6) os *inputs* e *outputs* são os objetos do R definidos pelos códigos imediatamente acima do parágrafo anterior, $z^T = [v \ u_1 \ u_2 \ u_3]$ é o vetor com as variáveis de decisão, b^T é o vetor com o *input* da DMU avaliada (DMU_{j0}) e c^T é o vetor com os coeficientes da função objetivo.

A eficiência de cada DMU pode ser avaliada pelo código R indicado a seguir, onde *i* é o índice da DMU avaliada (DMU_{j0}) e os vetores *b* e *c* são modificados automaticamente para cada DMU:

```
f.rhs <- c(rep(0,1,N),1) # RHS
f.dir <- c(rep("<=" ,1,N), "=") # direção das restrições
aux <- cbind(-1*inputs,outputs) # matriz com os coeficientes tecnológicos em (6)
for (i in 1:N) {
  f.obj <- c(0*rep(1,s),as.numeric(outputs[i,])) # coeficientes da função objetivo
  f.con <- rbind(aux ,c(as.numeric(inputs[i,]) , + rep(0,1,m))) # adiciona a restrição  $b^T z = 1$ 
  results <- lp("max",as.numeric(f.obj),f.con,f.dir,f.rhs,scale=1,compute.sens=TRUE) # resolve PPL
  multipliers <- results$solution # multiplicadores
  efficiency <- results$objval # índices de eficiência
  duals <- results$duals #  $\lambda$ 
  if (i==1) {
    weights <- multipliers
    effcrs <- efficiency
    lambdas <- duals [seq(1,N)]
  } else {
    weights <- rbind(weights,multipliers)
    effcrs <- rbind(effcrs , efficiency)
    lambdas <- rbind(lambdas,duals[seq(1,N)])
  }
}
```

Note que na função *lp* são informados todos os elementos de um PPL: a orientação do problema (max ou min), os coeficientes da função objetivo (f.obj), os coeficientes tecnológicos das restrições (f.con), as direções das restrições <=, = ou >= (f.dir) e o RHS (f.rhs) das restrições. Os resultados gerados pela função *lp* são armazenados no objeto *results*, por exemplo, os multiplicadores (results\$solution), as variáveis duais (results\$duals) e a eficiência (results\$objval).

Por meio do código R apresentado abaixo, os índices de eficiência e os multiplicadores (*U* e *V*) podem ser exportados para a planilha MS Excel ilustrada na Figura 3.

```
planilha <- cbind(effcrs,weights)
rownames(planilha) <- data[,1] # nomeas das linhas (empresas)
colnames(planilha) <- c('efficiency', names(inputs),names(outputs)) # nomes das colunas
write.xls(planilha,"resultscrs.xls",colNames=TRUE,sheet = 1,from = 1) # grava arquivo
```

	A	B	C	D	E	F
1		efficiency	OPEX	NETWORK	MWH	CUSTOMERS
2	ELETROPAULO	0.5827	8.0055E-10	0	6.0740E-09	5.6817E-08
3	CEMIG	0.5795	5.9441E-10	9.7382E-07	3.5032E-09	0
4	CPFL - Paulista	0.8854	2.0109E-09	0	1.5257E-08	1.4272E-07
5	COPEL	0.5186	9.8148E-10	7.9522E-07	1.1191E-08	2.1093E-08
6	LIGHT	0.7755	1.7947E-09	0	1.3617E-08	1.2737E-07
7	CELESC	0.5239	1.3861E-09	1.2843E-06	1.8660E-08	0
8	COELBA	1.0000	2.2913E-09	0	1.7385E-08	1.6262E-07
9	ELEKTRO	0.6879	2.4120E-09	1.9542E-06	2.7502E-08	5.1834E-08
10	CPFL - Piratininga	1.0000	5.1075E-09	0	7.6844E-08	0
11	BANDEIRANTE	0.6984	3.4864E-09	0	2.6452E-08	2.4744E-07
12	CELPE	0.8225	2.8518E-09	0	2.1638E-08	2.0240E-07
13	AMPLA	0.5498	2.2908E-09	0	1.7381E-08	1.6258E-07
14	CELG	0.5856	1.4462E-09	2.9357E-06	0	0
15	RGE	1.0000	5.3660E-09	4.3477E-06	6.1187E-08	1.1532E-07
16	COELCE	0.8198	3.1629E-09	0	0	2.9866E-07
17	ESCELSA	0.6697	4.1419E-09	3.3559E-06	4.7229E-08	8.9013E-08
18	AES SUL	0.8406	4.8515E-09	3.9308E-06	5.5319E-08	1.0426E-07
19	CEEE	0.4460	2.5907E-09	2.0991E-06	2.9541E-08	5.5677E-08

Figura 3: Resultados do modelo DEA CRS/M/I

Da mesma forma, o seguinte código R exporta as variáveis duais (λ) para a planilha MS Excel apresentada na Figura 4.

planilha <- lambdas

rownames(planilha) <- data[,1] # escreve nomes das empresas nas linhas

colnames(planilha) <- data[,1] # escreve nomes das empresas nas colunas

write.xls(planilha,"dualscrs.xls",colNames = TRUE,sheet = 1,from = 1) # grava arquivo

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S
1		ELETROPAULO	CEMIG	CPFL - Paulista	COPEL	LIGHT	CELESC	COELBA	ELEKTRO	CPFL - Piratininga	BANDEIR/CELPE	AMPLA	CELG	RGE	COELCE	ESCELSA	AES SUL	CEEE	
2	ELETROPAULO	0	0	0	0	0	0.5744	0	2.4372	0	0	0	0	0	0	0	0	0	
3	CEMIG	0	0	0	0	0	0.9781	0	0	0	0	0	0	0	2.9403	0	0	0	
4	CPFL - Paulista	0	0	0	0	0	0.2716	0	1.6435	0	0	0	0	0	0	0	0	0	
5	COPEL	0	0	0	0	0	0.0751	0	0.2590	0	0	0	0	0	2.3874	0	0	0	
6	LIGHT	0	0	0	0	0	0.3953	0	1.3260	0	0	0	0	0	0	0	0	0	
7	CELESC	0	0	0	0	0	0	0	0.4101	0	0	0	0	0	1.5974	0	0	0	
8	COELBA	0	0	0	0	0	1.0000	0	0	0	0	0	0	0	0	0	0	0	
9	ELEKTRO	0	0	0	0	0	0.1244	0	0.3720	0	0	0	0	0	0.8483	0	0	0	
10	CPFL - Piratininga	0	0	0	0	0	0	0	1.0000	0	0	0	0	0	0	0	0	0	
11	BANDEIRANTE	0	0	0	0	0	0.0529	0	0.9052	0	0	0	0	0	0	0	0	0	
12	CELPE	0	0	0	0	0	0.6227	0	0.0849	0	0	0	0	0	0	0	0	0	
13	AMPLA	0	0	0	0	0	0.4379	0	0.2498	0	0	0	0	0	0	0	0	0	
14	CELG	0	0	0	0	0	0.9279	0	0	0	0	0	0	0	0	0	0	0	
15	RGE	0	0	0	0	0	0	0	0	0	0	0	0	0	1.0000	0	0	0	
16	COELCE	0	0	0	0	0	0.5939	0	0	0	0	0	0	0	0	0	0	0	
17	ESCELSA	0	0	0	0	0	0.0575	0	0.2639	0	0	0	0	0	0.4555	0	0	0	
18	AES SUL	0	0	0	0	0	0.0041	0	0.0440	0	0	0	0	0	0.8739	0	0	0	
19	CEEE	0	0	0	0	0	0.1682	0	0.1387	0	0	0	0	0	0.3840	0	0	0	

Figura 4: Variáveis duais

Conforme ilustrado na Tabela 2, a diferença entre os modelos CRS e VRS reside na variável irrestrita u_0 presente apenas neste último modelo. A variável irrestrita em sinal é modelada pela diferença de duas variáveis não negativas ($u_0 = u^+ - u^-$, $u^+ \geq 0$ e $u^- \geq 0$), conforme ilustrado no programa abaixo para um modelo VRS orientado ao insumo na versão dos multiplicadores (5):

f.rhs <- c(rep(0,1,N),1) # RHS

f.dir <- c(rep("<=",1,N), "=") # direção das restrições

aux <- cbind(-1*inputs,outputs,1,-1) # matriz de coeficientes tecnológicos

for (i in 1:N) {

f.obj <- c(rep(0,1,s),as.numeric(outputs[i,]),1,-1) # 1 e -1 representam u^+ - u^-

f.con <- rbind(aux,c(as.numeric(inputs[i,]), rep(0,1,m),0,0))

results <- lp("max",as.numeric(f.obj),f.con,f.dir,f.rhs,scale=1,compute.sens=TRUE) # resolve o PPL



```
multipliers <- results$solution
efficiency <- results$objval
duals <- results$duals
u0 <- multipliers[s+m+1]-multipliers[s+m+2]
if (i==1) {
  weights <- c(multipliers[seq(1,s+m)],u0)
  effvrs <- efficiency
  lambdas <- duals [seq(1,N)]
} else {
  weights <- rbind(weights,c(multipliers[seq(1,s+m)],u0))
  effvrs <- rbind(effvrs , efficiency)
  lambdas <- rbind(lambdas,duals[seq(1,N)])
}
}
```

Os índices de eficiência oriundos dos modelos CRS e VRS e a eficiência de escala (COOPER et al, 2002) são ilustrados na Figura 5 gerada pelo seguinte código R.

```
par(mar=c(10,5,1, 8), xpd=TRUE)
scale <- effcrs/effvrs
spreadsheet <- cbind(effcrs,effvrs,scale)
rownames(spreadsheet) <- data[,1]
colnames(spreadsheet) <- c ("CRS", "VRS", "SE")
barplot(t(spreadsheet),col=palette()[c(1,4,7)], ylab="Efficiency",beside=TRUE,las=3)
legend("topright",inset=c(-0.2,0),colnames(spreadsheet),fill=palette()[c(1,4,7)],bty="n")
```

A decomposição apresentada na Figura 5 permite identificar se a fonte de ineficiência reside na operação, na escala ou em ambas (COOPER et al, 2002).

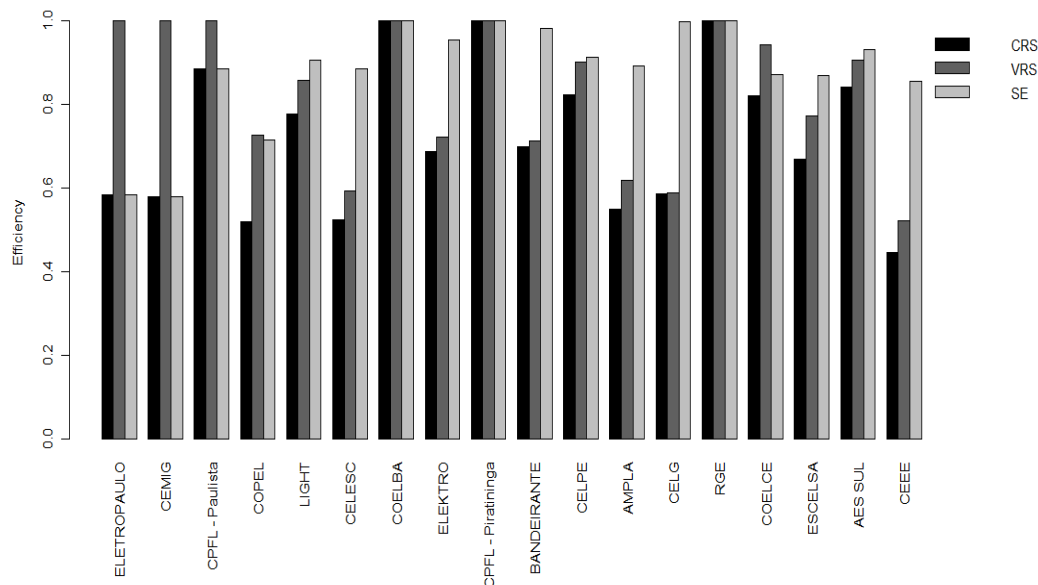


Figura 5: Índices de eficiência CRS, VRS e de escala (SE – Scale Efficiency)

4. CONCLUSÕES

O R é um programa livre e *open source*, altamente extensível e largamente difundido na comunidade acadêmica oferece um ambiente computacional interessante para atividades de pesquisa e desenvolvimento de aplicações de métodos quantitativos. Embora a Análise Envoltória de Dados (DEA) conte com uma ampla disponibilidade de *softwares* comerciais e não comerciais que oferecem aos analistas uma variedade de modelos DEA e sofisticados recursos gráficos, as opções oferecidas pelos softwares são restritas.

A possibilidade de implementar os modelos DEA no ambiente R torna o analista livre das limitações dos *softwares* dedicados e, principalmente, facilita a integração da Análise Envoltória de Dados com outras técnicas quantitativas, um aspecto importante quando se considera uma visão mais ampla do processo de tomada de decisão.

Os programas apresentados no presente artigo constituem apenas um exemplo de como implementar modelos DEA no ambiente R. Contudo, destaca-se que os códigos descritos no artigo podem ser facilmente adaptados para versões mais sofisticadas do modelo DEA, por exemplo, modelos com esquemas de restrições aos pesos, avaliação cruzada, DEA em dois estágios, modelos para alocação de recursos e até mesmo modelos Fuzzy-DEA. Adicionalmente o R conta com funções que permitem elaborar interfaces gráficas amigáveis (pacote RGtk2 <http://www.ggobi.org/rgtk2/>), além de poder ser facilmente integrado ao Excel (pacote RExcel <http://rcom.univie.ac.at/>).

5. REFERÊNCIAS

ANGULO-MEZA, L., BIONDI NETO, L., SOARES DE MELLO, J.C.C.B., GOMES, E.G. ISYDS – Integrated System for Decision Support (SIAD – Sistema Integrado de Apoio à Decisão): A software package for Data Envelopment Analysis Model, *Pesquisa Operacional*, v.25, n.3, pp. 493-503, Setembro a Dezembro, 2005.

BANKER, R.D., CHARNES A., COOPER, W.W. Some models for estimating technical and scale inefficiencies in data envelopment analysis, *Management Science* (30): 1078-1092, 1984.



BOGETOFT, P., OTTO, L. Benchmarking with DEA, SFA and R, Springer Science, 2011.

CHARNES, A., COOPER, W.W., RHODES, E. Measuring the Efficiency of Decision Making Units, European Journal of Operational Research: 429-444, 1978.

COOK, W.D., SEIFORD, L.M. Data Envelopment Analysis (DEA) – Thirty years on, European Journal of Operational Research, 192: 1-17, 2009.

COOPER, W.W., SEIFORD, L.M., TONE, K. Data Envelopment Analysis: A comprehensive text with models, applications, references and DEA-Solver Software, Kluwer Academic Publishers, 2002.

R DEVELOPMENT CORE TEAM R. A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <http://www.R-project.org/>, 2011.

VERZANI, J. Using R for introductory statistics, Chapman & Hall/ CRC Press, 2005.