

DEA em concessionárias de energia elétrica aplicada em R

Guilherme Ventura

1 O que é o DEA?

A DEA (Data Envelopment Analysis ou Análise Envoltória de Dados), é uma metodologia de análise de eficiência que compara um grupo de DMUs (Decision Making Units) e seus planos de produção, e por programação matemática, maximiza, considerando os recursos de que se dispõe (inputs) com os resultados alcançados (outputs), e identifica aquelas empresas cujo plano de produção, dado os pesos, não pode ser superado pelo plano de nenhuma outra empresa. Dita eficiente e se tornando referência para as outras.

Todo modelo é baseado nas empresas analisadas, se caso haja uma modificação do conjunto das DMUs, modifica-se o resultado, apresentando outras empresas referências e outros scores do benchmark.

Empresas que são ditas eficientes, compõem uma chamada Fronteira de eficiência, que é a linha traçada juntando todas as empresas referências. As empresas que se situam abaixo desta linha, são empresas com planos de produção dominados. É também de feitiço deste método gerar uma medida da ineficiência para cada unidade fora da fronteira (uma distância à fronteira que representa a potencialidade de crescimento da produtividade);

Humes diz que DEA é a condição necessária para que uma empresa A seja relativamente eficiente é que sua operação seja ‘melhor’ que as demais consideradas se A tiver o poder de definir os preços.

Existem três tipos de DEA:

1.1 Modelo CCR

O modelo CCR, desenvolvido por Charnes, Cooper e Rhodes, permite uma avaliação objetiva da eficiência global e identifica as fontes e estimativas de montantes das ineficiências identificadas. Tem o objetivo minimizar insumos para produzir no mínimo o nível de produção dado.

É um modelo também conhecido como CRS (Constant Returns to Scale) que trabalha com retornos constantes de escala, ou seja, qualquer variação nos insumos leva a uma variação proporcional nos produtos.

A formulação matemática é dada por:

$$\begin{aligned}
\theta &= \underset{u,v}{Max} \sum_{i=1}^m u_i y_{i,j_0} \\
\text{subject to } &\sum_{i=1}^s v_i x_{i,j_0} = 1 \\
&\sum_{i=1}^m u_i y_{ij} - \sum_{i=1}^s v_i x_{ij} \leq 0 \forall j = 1, \dots, j_0, \dots, N \\
&u_i \geq 0 \forall i = 1, m \\
&v_i \geq 0 \forall i = 1, s
\end{aligned} \tag{1}$$

1.2 Modelo BBC

O modelo BBC (Banker, Charnes e Cooper, 1984), também conhecido como VRS (Variable Return Scale), pressupõe que as DMU's avaliadas apresentem retornos variáveis de escala e identificando se estão presentes ganhos de escala crescentes, decrescentes e constantes, para futura exploração.

A formulação matemática é dada por:

$$\begin{aligned}
\theta &= \underset{u,v}{Max} \sum_{r=1}^m u_r y_{r,k} - u_k \\
\text{subject to } &\sum_{r=1}^n v_i x_{ik} = 1 \\
&\sum_{r=1}^m u_r y_{rj} - \sum_{i=1}^n v_i x_{ij} - u_k \leq 0 \forall j = 1, \dots, j_0, \dots, N \\
&u_r \geq 0 \forall r = 1, m \\
&v_i \geq 0 \forall i = 1, s
\end{aligned} \tag{2}$$

2 Aplicação em R com exemplo

A aplicação em R é feita por meio do pacote chamado **Benchmarking**.

2.1 Eficiência

Usando um exemplo de um dataset fictício de hospitais:

	Hospital	Médicos	Enfermeiros	Pacientes alta
1	A	20	151	100
2	B	19	131	150
3	C	25	160	160
4	D	27	168	180
5	E	22	158	94
6	F	55	255	230
7	G	33	235	220
8	H	31	206	152
9	I	30	244	190
10	J	50	268	250
11	K	53	306	260
12	L	38	284	250

Então, como primeiro passo, selecionaremos quais variáveis serão os inputs e os outputs:

```
#Importando a base de dados
hospitais <- read.csv2("~/series_temporais/hospitais.csv")
```

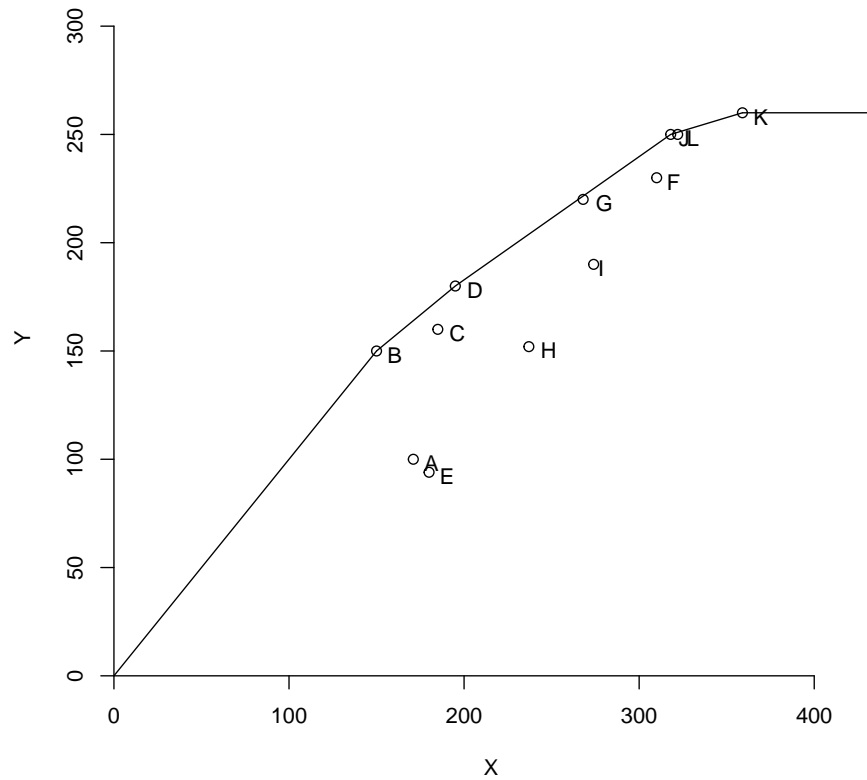
```
#Setando inputs e outputs
inp <- as.matrix(hospitais[,2:3])
out <- hospitais[,4]
```

Logo depois disto, já conseguimos visualizar a fronteira de eficiência, usaremos então:

```
library(Benchmarking)

## Loading required package: lpSolveAPI
## Loading required package: ucminf

#Plor da fronteira de eficiência
dea.plot.frontier(inp, out, RTS = "drs", txt=hospitais$Hospital)
```



Podemos perceber que há seis empresas em cima da fronteira de eficiência. Para sabermos exatamente quais são e os seus scores, useremos os comandos:

```
#Resultado
result_dea <- dea(inp, out)
summary(result_dea)

## Summary of efficiencies
## VRS technology and input orientated efficiency
## Number of firms with efficiency==1 are 6 out of 12
## Mean efficiency: 0.932
## ---
##   Eff range      #    %
##   0.6<= E <0.7   1   8.3
##   0.7<= E <0.8   0   0.0
##   0.8<= E <0.9   3  25.0
##   0.9<= E <1     2  16.7
```

```
##          E ==1      6 50.0
##      Min. 1st Qu.  Median      Mean 3rd Qu.      Max.
## 0.6479 0.8935 0.9750 0.9319 1.0000 1.0000

eff(result_dea)

## [1] 0.9500000 1.0000000 0.8958333 1.0000000 0.8636364 0.9389356 1.0000000
## [8] 0.6478964 0.8866667 1.0000000 1.0000000 1.0000000
```

Assim fica o resultado da eficiência de cada hospital:

	Hospital	Médicos	Enfermeiros	Pacientes alta	Eficiência
1	A	20	151	100	0.95
2	B	19	131	150	1.00
3	C	25	160	160	0.90
4	D	27	168	180	1.00
5	E	22	158	94	0.86
6	F	55	255	230	0.94
7	G	33	235	220	1.00
8	H	31	206	152	0.65
9	I	30	244	190	0.89
10	J	50	268	250	1.00
11	K	53	306	260	1.00
12	L	38	284	250	1.00

Ou seja, podemos ver que o hospital B, D, G, J, K e L são eficientes, sendo assim, dado os recursos que utilizam, conseguem ter o máximo de pacientes com alta. Se tornando hospitais de referência para os outros.

Conseguimos também calcular a folga para cada hospital, ou seja, o quanto precisa para cada empresa não eficiente ser empurrada para eficiência:

```
#Resultado
dea_folga <- dea(inp,out,SLACK=TRUE)
dea_folga_table <- data.frame(dea_folga$sx,dea_folga$sy)
names(dea_folga_table) <- c("Folga Insumo 1", "Folga Insumo 2", "Folga Output")
```

	Folga Insumo 1	Folga Insumo 2	Folga Output
1	0.00	12.45	50.00
2	0.00	0.00	0.00
3	0.73	0.00	0.00
4	0.00	0.00	0.00
5	0.00	5.45	56.00
6	8.21	0.00	0.00
7	0.00	0.00	0.00
8	0.55	0.00	0.00
9	0.00	24.15	0.00
10	0.00	0.00	0.00
11	0.00	0.00	0.00
12	0.00	0.00	0.00

2.2 Distância Euclidiana e Clusters

Podemos também recuperar a distância euclidiana e com isso separar os hospitais por clusters, precisamos apenas de alguns comandos do R base para isto:

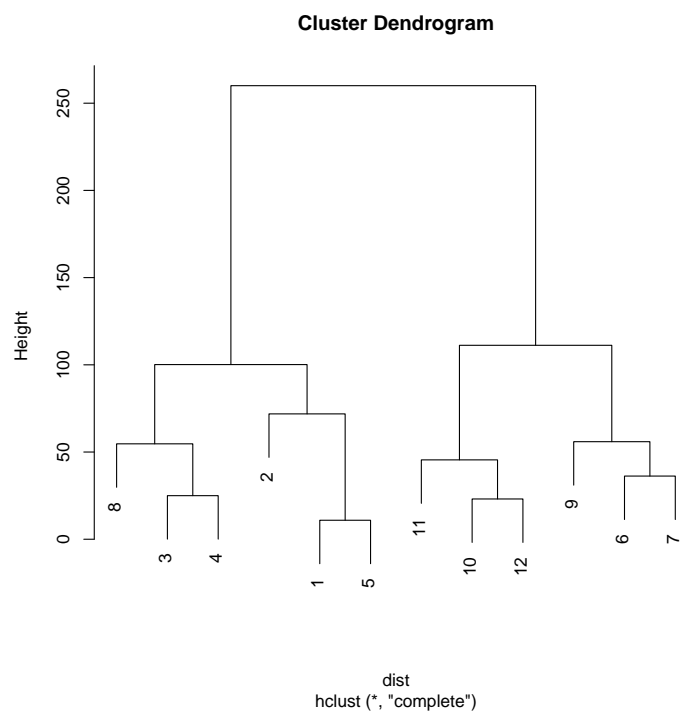
```
dist <- round(dist(hospitais), 2)

## Warning in dist(hospitais): NAs introduzidos por coerção

dist
##      1      2      3      4      5      6      7      8      9     10
## 2  62.19
## 3  70.29  36.09
## 4  94.78  55.77  24.98
## 5  10.89  71.87  76.32 100.14
## 6 196.44 175.39 140.59 120.29 196.62
## 7 169.80 145.66 111.29  90.37 170.98  36.22
## 8  88.32  87.73  54.36  54.70  87.55 109.92  85.39
```

```
## 9  149.88 139.00 103.16  88.58 149.11  55.93  36.33  62.06
## 10 222.38 199.10 164.88 143.43 222.77  28.14  55.11 135.69  78.11
## 11 260.04 241.88 206.88 186.62 259.28  68.36  96.89 171.84 111.19  45.50
## 12 232.42 212.19 177.56 156.96 232.29  45.17  66.59 144.85  83.78  23.09
##      11
## 2
## 3
## 4
## 5
## 6
## 7
## 8
## 9
## 10
## 11
## 12  32.84
```

E para o cluster:



Sendo que para para 1, seria equivalente ao hospital A.