

Introduzione alla D.E.A. (Data Envelopment Analysis)

La DEA è una tecnica per misurare le prestazioni di unità produttive. Introdotta nel 1978 dal lavoro di Charnes, Cooper e Rhodes, permette di valutare l'efficienza di un'unità di produzione (DMU - Decision Making Unit) relativamente a un dato insieme di unità produttive scelte per il confronto. Questa tecnica è stata utilizzata, fra l'altro, per valutare l'efficienza di banche, ospedali, sistemi di trasporto, etc. Diversamente dall'approccio "statistico" che confronta le unità produttive con qualche fantomatica unità produttiva media, DEA confronta ogni DMU con le DMU più efficienti e ne valuta l'efficienza relativa.

Le unità produttive funzionano utilizzando determinati input e producendo specifici output. Se consideriamo come esempio un'agenzia di banca, gli input possono essere, ad esempio, il numero di cassieri, di dirigenti, i locali disponibili, etc. mentre sono output il numero di clienti, di assegni cambiati, di mutui accesi, etc.

Ora, noi diremo che una data unità produttiva A è più efficiente di un'altra B se : i) con la stessa quantità di input produce maggiore output, oppure se ii) con una minore quantità di input produce (almeno) lo stesso output.

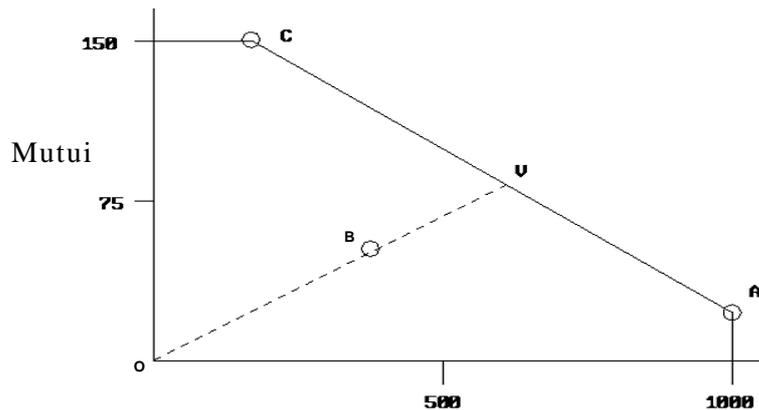
Consideriamo l'esempio di una banca con tre agenzie, denominate A, B e C. Supponiamo che ogni agenzia utilizzi un solo input, i cassieri, e produca due soli output, il numero di assegni cambiati e quello di mutui accesi. I valori di tali grandezze siano quelli espressi nella tabella

DMU	Input	Output	
	Cassieri	Assegni	Mutui
A	10	1000	20
B	10	400	50
C	10	200	150

In questo caso è facile vedere che nessuna delle agenzie è dominata da qualche altra. Come valutare l'efficienza? Supponiamo di "prelevare" 5 cassieri dell'agenzia A e sistemarli in una nuova agenzia D. Ipotizzando che questi continuino a produrre con la stessa efficienza, l'agenzia D sarà in grado di cambiare 600 assegni e di accendere 10 mutui. Ora, supponiamo di prelevare 5 cassieri dall'agenzia C; questi sono in grado di cambiare 100 assegni e accendere 75 mutui. Se trasferiamo i 5 cassieri di C nell'agenzia D, D avrà 10 cassieri e sarà in grado di cambiare 600 assegni e accendere 85 mutui.

DMU	Input	Output	
	Cassieri	Assegni	Mutui
D	10	600	85

E' evidente che D domina B. Ovviamente, D non esiste, tuttavia può essere pensata come un'agenzia potenziale, e cioè virtuale, un'agenzia che potrebbe esistere e che può essere ottenuta combinando un certo numero di agenzie. L'ipotesi fondamentale su cui DEA si basa è che è sempre possibile combinare frazioni di DMU esistenti per creare una nuova agenzia virtuale e che quantità di input e di output saranno proporzionali alle quantità delle agenzie coinvolte secondo la frazione data.



A titolo di esempio, si consideri la figura ove sono rappresentate le tre agenzie A, B e C sul piano degli output (l'input è uguale per tutti); tutti i punti sul segmento AC rappresentano agenzie virtuali ciascuna ottenibile come combinazione convessa di A e C. Il segmento AC è detto frontiera efficiente. L'agenzia rappresentata dal punto V è una delle possibili agenzie virtuali che dominano B. In particolare, V è il punto in cui la retta passante per l'origine e per il punto B incontra il segmento AC. Questa agenzia è, fra quelle che dominano B, quella che produce output in quantità proporzionali alle quantità prodotte da B. L'efficienza di B può essere misurata come il valore del rapporto OB/OV . Infatti, affinché B fosse efficiente, il suo vettore di output dovrebbe essere moltiplicato per il fattore OV/OB .

Un modello di programmazione lineare per D.E.A.

Si è detto che un'unità produttiva è dominata se esiste una combinazione delle altre DMU che 1) produce un maggiore output con lo stesso input oppure 2) produce lo stesso output con meno input. Da ora in poi sceglieremo la definizione 2).

Torniamo all'esempio. Per decidere se B è dominata da A e C possiamo associare ad A una variabile $x_A \in \mathbb{R}$ che rappresenta la frazione di agenzia A utilizzata nell'agenzia virtuale. Analogamente, assoceremo a C la variabile x_C . Per esprimere il fatto che l'agenzia virtuale dovrà usare meno input scriviamo

$$10 x_A + 10 x_C \leq 10 - \varepsilon$$

con $\varepsilon > 0$ quantità molto piccola scelta opportunamente. Per esprimere che l'agenzia virtuale dovrà produrre almeno lo stesso numero di assegni, scriviamo

$$1000 x_A + 200 x_C \geq 400$$

Per esprimere il vincolo di output sui mutui scriveremo il vincolo:

$$20 x_A + 150 x_C \geq 50$$

Se il problema

$$\begin{aligned} 10 x_A + 10 x_C &\leq 10 - \varepsilon \\ 1000 x_A + 200 x_C &\geq 400 \\ 20 x_A + 150 x_C &\geq 50 \end{aligned} \quad (1)$$

ammette soluzione, allora B è dominato e x_A, x_C rappresentano le quantità di A e di C richieste per costruire l'agenzia virtuale che domina B.

Se (1) non ammette soluzione B non è dominato. Dunque, il programma lineare (1) risolve il problema di stabilire se B è efficiente, ma non risponde alla domanda: qual è l'efficienza di B? L'efficienza può essere definita come la frazione di input di B sufficiente ad un'agenzia virtuale a produrre almeno l'output di B. Ad esempio, se prendiamo 5 cassieri di A ($x_A = 0,5$) e 4 cassieri di C ($x_C = 0,4$) riusciamo a produrre 660 assegni e 70 mutui, ovvero un output maggiore di quello prodotto da B con 10 cassieri. Cioè, ci bastano 9 cassieri di altre agenzie per produrre almeno lo

stesso output di B e quindi l'efficienza di B è al più pari a $9/10 = 0.9$ (potrebbe in realtà essere ancora inferiore). Naturalmente, fra tutte le agenzie virtuali (si ricordi che in generale le agenzie virtuali sono infinite), vorremmo identificare quella che domina "maggiormente" B, cioè quella che, producendo almeno la stessa quantità di output di B richiede in input la minima frazione di input di B.

A questo scopo introduciamo una nuova variabile reale y che rappresenta l'efficienza di B. Con questa nuova assunzione, il problema di determinare l'efficienza di B diventa:

$$\begin{aligned} \min y \\ 10 x_A + 10 x_C &\leq 10y \\ 1000 x_A + 200 x_C &\geq 400 \\ 20 x_A + 150 x_C &\geq 50 \end{aligned} \quad (2)$$

Si osservi che la soluzione 5 cassieri di A ($x_A = 0,5$) e 4 cassieri di C ($x_C = 0,4$) e $y = 0.9$ è una soluzione ammissibile per (2), ma non necessariamente ottima. Per giustificare la funzione obiettivo, si ricordi che noi stiamo cercando la migliore agenzia virtuale, cioè quella che domina "maggiormente" B, ovvero quella che consente il più piccolo valore di y . Quindi, la soluzione ottima di (2) ci restituirà quell'agenzia virtuale che minimizza l'efficienza di B. In particolare, la soluzione ottima di (2) è $x_A = 0.342466$ $x_C = 0.287671$ $y = 0.630137$. Quindi, l'efficienza di B è 0.630137, ovvero esiste un'agenzia virtuale che con circa il 63% dell'input di B produce almeno lo stesso output di B.

Quindi, DEA permette di determinare l'efficienza di una DMU specifica rispetto a tutte le altre. Possiamo ora generalizzare quanto descritto nell'esempio. Sia quindi dato un insieme di n DMU.

Ognuna delle n DMU riceve in ingresso (al più) p input distinti e fornisce in uscita (al più) q output distinti. Avremo:

I = matrice reale non negativa di dimensione $p \times n$ degli input delle DMU
 O = matrice reale non-negativa di dimensione $q \times n$ degli output delle DMU

Inoltre, indichiamo con e l'indice della DMU della quale si voglia determinare l'efficienza (si osservi che e è una delle n unità). Infine, indichiamo con $x = (x_1, \dots, x_n)$ il vettore reale non negativo la cui j -esima componente rappresenta la frazione della j -esima DMU nella DMU virtuale e con y l'efficienza della DMU di riferimento e . Il seguente programma lineare valuta l'efficienza della DMU e :

$$\begin{aligned} \min y \\ \sum_{j=1, \dots, n} I_{ij} x_j &\leq I_{i,e} y & i = 1, \dots, p & \quad (3.a) \\ \sum_{j=1, \dots, n} O_{ij} x_j &\geq O_{i,e} & i = 1, \dots, q & \quad (3.b) \\ x, y &\geq 0 \end{aligned}$$

I vincoli (3a) rappresentano il fatto che l'unità virtuale dovrà usare al più la frazione y dell'input I_e della DMU di riferimento, mentre i vincoli (3b) assicurano che l'output prodotto sia almeno pari a O_e . Si osservi che, diversamente dall'esempio (2), in questo caso anche la variabile x_e corrispondente alla DMU e (cioè la DMU della quale si vuole determinare l'efficienza) appare nei vincoli di input e output. Questo vuol dire che la DMU e può essere utilizzata per "generare" se stessa. In altri termini, la soluzione $y = 1$, $x_e = 1$ e $x_j = 0$ per $j \neq e$, è una soluzione ammissibile. Inoltre, indicando con (y^*, x^*) il valore della soluzione ottima, si può dimostrare che se la DMU e è dominata (e cioè $y^* < 1$) allora $x_e^* = 0$.

Per valutare tutte le DMU efficienti del nostro insieme di unità, dovremmo risolvere il programma (3) n volte, ogni volta cambiando la DMU di riferimento.

Soluzione del modello scelto attraverso AMPL

Il modello che rappresenta tutti e 3 i problemi descritti sopra è:

```
set INPUT;           # insieme degli input
set OUTPUT;          # insieme degli output
set DMU;             # insieme delle Decision Making Unit

param I{INPUT,DMU};  # coefficienti dei vincoli sull'uso degli input
param O{OUTPUT,DMU}; # coefficienti dei vincoli sulla produzione degli output
param E symbolic in DMU; # DMU in esame

var x{i in DMU} >= 0; # frazioni delle DMU reali usate per la DMU virtuale
var y{i in DMU} >= 0; #efficienza della DMU in esame rispetto alla DMU virtuale

minimize obj: y[E];  # cerco la DMU virtuale che minimizza l'efficienza
                  # della DMU in esame

subject to uso{j in INPUT}: sum{i in DMU} I[j,i] * x[i] <= I[j,E]*y[E];
subject to prod{j in OUTPUT}: sum{i in DMU} O[j,i] * x[i] >= O[j,E];
```

Si noti l'uso del parametro simbolico E per rappresentare la DMU in esame.
Il file di dati corrispondente al problema di valutare A è:

```
set INPUT := CASSIERI;
set OUTPUT := ASSEGNI, MUTUI;
set DMU := A, B, C;

param I:      A      B      C :=
  CASSIERI  10     10     10;

param O:      A      B      C :=
  ASSEGNI   1000   400    200
  MUTUI     20     50     150;

param E:= A; # porlo pari a A, B, C per calcolare l'efficienza delle varie DMU
```

Per valutare invece l'efficienza di B e di C basta cambiare il valore del parametro simbolico.