

Università degli Studi di Genova
DISEFIN – Series of Economic Working Papers
16126 Genova – via vivaldi 5 – Fax +39 010 209 5223



**Esemplificazione della Data Envelopment Analysis
per la valutazione di efficienza in una grande azienda
ospedaliera universitaria**

Francesco Copello Cristiana Pellicanò

**wp n. 3
July 2011**

“DISEFIN Working Papers on line”

*series of economic working papers
published online by*

*Research Doctorate in
Economics and Finance of European Union
Ph.D School in New technologies and Social Sciences
University of Genoa*

Founder:

Amedeo Fossati

Editor-in-Chief:

Marcello Montefiori

Editorial Board:

Paul De Grawe

Francesco Figari

Amedeo Fossati

Luca Gandullia

Eric Gaspérini

Andrea Monticini

Carlo Perroni

Web site:

<http://www.disefin.unige.it/>

Esemplificazione della Data Envelopment Analysis per la valutazione di efficienza in una grande azienda ospedaliera universitaria

Francesco Copello,
Controllo di Gestione, Azienda
Ospedaliera Universitaria “San
Martino” di Genova

Cristiana Pellicanò
Controllo di Gestione, Azienda
Ospedaliera Universitaria “San
Martino” di Genova

Abstract

The Data Envelopment Analysis (DEA) is applied by the management control of a huge teaching hospital to evaluate the relative efficiency of human resources. The authors show an experimental approach based on geometrical methods and simple statistical techniques that can be easily performed by basic software procedures. The results are then related to the cost of diagnostic procedures, standardized by the case-mix, and both scatterplot and cluster analysis are produced to find out related area of performance and to plan a strategy for the continuous quality improvement.

Keywords: data envelopment analysis, production frontier, cluster analysis, human resources efficiency, SWOT analysis

Jel Classification: D61, C81, C02, D24, I19

Abstract (italiano)

Il controllo di gestione di una grande azienda ospedaliera universitaria ha applicato la Data Envelopment Analysis (DEA) allo scopo di valutare l'efficienza relativa delle risorse umane. Gli autori descrivono un approccio sperimentale basato su metodi geometrici e semplici tecniche statistiche che possono essere facilmente applicate mediante procedure software di base. I risultati vengono quindi correlati con il costo della diagnostica, standardizzato per il case-mix, e vengono prodotti sia un grafico cartesiano che l'analisi di raggruppamento per individuare le aree con performance omogenea e pianificare gli interventi per il miglioramento continuo della qualità.

1. Introduzione

I risultati prodotti da tecniche di valutazione delle performance come la Data Envelopment Analysis (DEA) permettono alla Direzione Generale di un'azienda sanitaria di effettuare decisioni sulla base dei diversi livelli di efficienza che caratterizzano unità operative o specifici settori produttivi (Battese – Coelli – Rao, 1997). Ciò appare utile non solo durante la contrattazione e l'assegnazione del budget ai singoli reparti ma anche e soprattutto a livello di analisi organizzativa, partendo dall'assunto che realtà complesse e operanti in un contesto facilmente mutabile nel tempo quale quello sanitario devono essere ormai caratterizzate da ipotesi di miglioramento continuo (kaizen) (Maasaki, 1986).

L'Azienda Ospedaliera Universitaria “San Martino” di Genova, sede del Corso di laurea in Medicina e Chirurgia, rappresenta il più importante istituto di ricovero della Regione Liguria ed è suddiviso in 12 dipartimenti ad attività integrata che operano in ambito assistenziale ma anche in didattica e ricerca, per un totale di 120 unità operative sanitarie, e 2 dipartimenti gestionali composti da circa 30 unità operative.

In questi ultimi anni tale struttura sta focalizzando la propria attenzione sui processi che la caratterizzano e metodiche di tipo multidimensionale quali la DEA – oggetto di tale lavoro e attualmente implementata dal Controllo di gestione a livello sperimentale – o la tecnica di Analisi delle Componenti Principali (Copello-Pellicanò, 2009) rappresentano il nuovo approccio di osservazione del “sistema azienda” a supporto delle decisioni della Direzione.

2. La Data Envelopment Analysis

La Data Envelopment Analysis (DEA) è una tecnica di misurazione delle performance utilizzata in ricerca operativa e in econometria per stimare l'efficienza delle unità produttive

all'interno del processo decisionale aziendale (DMU - decision-making units).

Tale metodo, sviluppato nel 1978 da Charnes, Cooper e Rhodes, è stato utilizzato negli anni per calcolare la funzione di produzione in vari ambiti aziendali (ospedali, banche, scuole, dipartimenti universitari, aziende manifatturiere sono solo alcune delle differenti business units ai quali tale tecnica è stata applicata).

L'efficienza che con tale tecnica si intende valutare è misura della capacità di produrre la massima quantità di risultati a partire da un dato singolo o da una molteplicità di risorse a disposizione ed è dunque un'efficienza di tipo tecnico. E' altresì un'efficienza relativa in quanto è basata sul confronto tra la combinazione di output e di input di una certa DMU e quella della DMU più efficiente, quest'ultima calcolata tra quelle a disposizione del campione osservato.

Già da queste prime caratteristiche è evidente che la DEA si fonda su alcuni concetti puramente economici, che appare utile richiamare.

Nelle *scienze economiche* l'efficienza di tipo tecnico è valutata, a partire ad esempio da un grafico come quello in figura 1, nella misura in cui una certa unità produttiva si discosta dalle combinazioni efficienti che si trovano sulla curva riportata, rappresentante la funzione di produzione:

$$x_{max} = f(\alpha, \beta)$$

dove α , β sono i fattori produttivi impiegati che consentono di avere la massimizzazione di x .

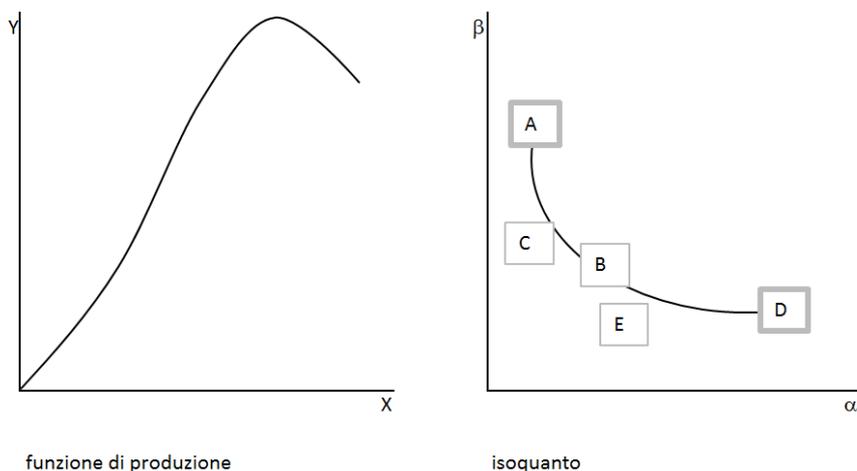


Fig. 1: Funzione di produzione e linea dell'isoquanto

La curva è la *funzione o frontiera di produzione* e, per x invariante (curva di prodotto costante), definisce i cosiddetti *isoquanti* cioè le diverse associazioni di α e di β che permettono di ottenere la stessa quantità di prodotto massimo e che sono tra loro alternative; infatti un fattore che aumenta ne sostituisce un altro che si riduce. Il criterio di scelta sarà quello che permette di ottenere il costo più basso possibile a parità di prodotto e l'ambito di scelta è rappresentato in figura dal segmento A-D, sul quale vi sono le combinazioni efficienti dei fattori produttivi impiegati.

Come è possibile osservare l'isoquanto ha due caratteristiche grafiche:

- ha un andamento decrescente
- è una curva convessa rispetto all'origine degli assi.

La prima caratteristica è determinata dalla sostituzione tra i fattori che si esprime matematicamente con il *saggio medio di sostituzione*, vale a dire l'indicatore che misura il numero di unità

del fattore che diminuisce sostituito in media da ogni unità del fattore che aumenta, tramite la seguente formula:

$$\frac{Pmg_A}{Pmg_B} = \left| \frac{\text{var}_\alpha}{\text{var}_\beta} \right|$$

La seconda esprime invece la difficoltà di sostituzione tra i fattori. Si prenda in considerazione il passaggio dal punto A al punto B nella figura 2; per ogni riduzione costante del fattore β , data la convessità della curva, è necessaria una quantità sempre maggiore del fattore α per sostituirlo. Matematicamente si può dire che il rapporto diminuisce gradualmente (il numeratore è costante e il denominatore aumenta sempre più).

Tutto ciò è evidente misurando geometricamente la pendenza rispetto all'asse delle ascisse della congiungente il punto A e il punto B sulla curva, che sarà negativa perché la curva è decrescente, e ciò è espressione del saggio medio negativo. Ancor più evidente se si considera il *saggio marginale di sostituzione*, che rappresenta quante unità infinitamente piccole del fattore che diminuisce vengono - nel punto A ad esempio - sostituite da un'unità infinitamente piccola del fattore che aumenta. Graficamente si ottiene tracciando la tangente alla curva in ogni punto e calcolandone la pendenza rispetto all'asse delle ascisse.

Sulla base di questi concetti economici e conoscendo un insieme di dati riguardanti ogni singola unità produttiva è possibile:

- 1) costruire a priori una frontiera di produzione utilizzando tecniche di tipo parametrico oppure
- 2) determinare l'efficienza relativa attraverso tecniche di programmazione lineare con metodi non parametrici (Cooper - Seidorf - Tone, 2002).

Nel primo caso in particolare le tecniche econometriche consentono di effettuare una stima parametrica della frontiera studiando l'equazione della retta $y = a + bx$, in cui y è la variabile

dipendente, x quella indipendente, b rappresenta la pendenza della retta e a (intercetta) la costante.

E' possibile dunque spiegare il metodo della DEA con un approccio di tipo grafico o con un approccio di tipo matematico (Seiford – Thrall, 1990). In ogni caso ciò che appare importante sottolineare è il fatto che la DEA fornisce l'efficienza relativa per un certo dato ma non l'efficienza assoluta.

A titolo di esempio è possibile considerare un certo numero di DMU (ad esempio filiali di una banca o unità operative ospedaliere), ognuna con un singolo input impiegato e un singolo output prodotto e cercare di stabilire quale risulti essere la più efficiente.

DMU	OUTPUT	INPUT	Output/Input	Eff. relativa
A	125	18	6,94	100%
B	44	16	2,75	40%
C	80	17	4,71	68%
D	23	11	2,09	30%

Tabella 1: esempio di calcolo efficienza relativa con output singolo

Per effettuare un confronto tra le stesse, occorre trasformare per ogni singola DMU gli input in output rapportandole tra loro (output/input) e poi dividere ogni singolo risultato per il rapporto che è risultato più elevato nel gruppo considerato (*DMU A*), calcolando in sostanza l'efficienza relativa. Così facendo è possibile assegnare degli obiettivi alle varie DMU che presentano variabilità (*B, C, D*) cercando di farle tendere nel tempo al raggiungimento del grado di efficienza più alto, determinato dal rapporto migliore vale a dire dall'impiego migliore di input per produrre un certo output.

L'assunto dal quale è necessario partire ai fini della valutazione risulta infatti quello che un'unità produttiva è per

definizione più efficiente di un'altra se si verificano due ipotesi alternative:

- 1) con la stessa quantità di input è prodotto un output maggiore (output target)
- 2) con una minore quantità di input viene prodotto almeno lo stesso output (input target).

Ciò appare semplice nel caso appena visto di un singolo input e un singolo output.

Nel caso in cui gli output risultassero più di uno, stabilire quale sia l'unità produttiva più efficiente tra quelle indicate, appare un po' più complicato in quanto è possibile, ad esempio, che una DMU presenti il rapporto migliore se si considera singolarmente il primo output, ma che ciò non si verifichi invece prendendo a riferimento il secondo output.

Considerando ad esempio due output e un input, come nell'esempio sotto indicato (Beasley, 2009), l'unità produttiva *A* è più efficiente nel primo caso ma nel secondo appare migliore *C*. Inoltre se volessimo giudicare le altre due (*B*, *D*) meno efficienti comunque, come sarebbe possibile giudicare il loro rapporto reciproco?

DMU	OUTPUT ₁	OUTPUT ₂	INPUT	Output ₁ /Input	Output ₂ /Input
A	125	50	18	6,94	2,78
B	44	20	16	2,75	1,25
C	80	55	17	4,71	3,24
D	23	12	11	2,09	1,09

Tabella 2: esempio di calcolo efficienza relativa con doppio output

Quanto detto complica il giudizio complessivo sull'analisi dello scenario ed è chiaro che in tal caso risulta alquanto difficile affidarsi solamente ad una valutazione di tipo numerico.

Per superare dunque l'ostacolo intercorso, appare utile affidarsi ad una rappresentazione grafica plottando i due rapporti

per ogni DMU rispettivamente sull'asse delle ordinate e sull'asse delle ascisse, individuando così come punti nello spazio bidimensionale le varie unità produttive tramite le coordinate relative (figura 2).

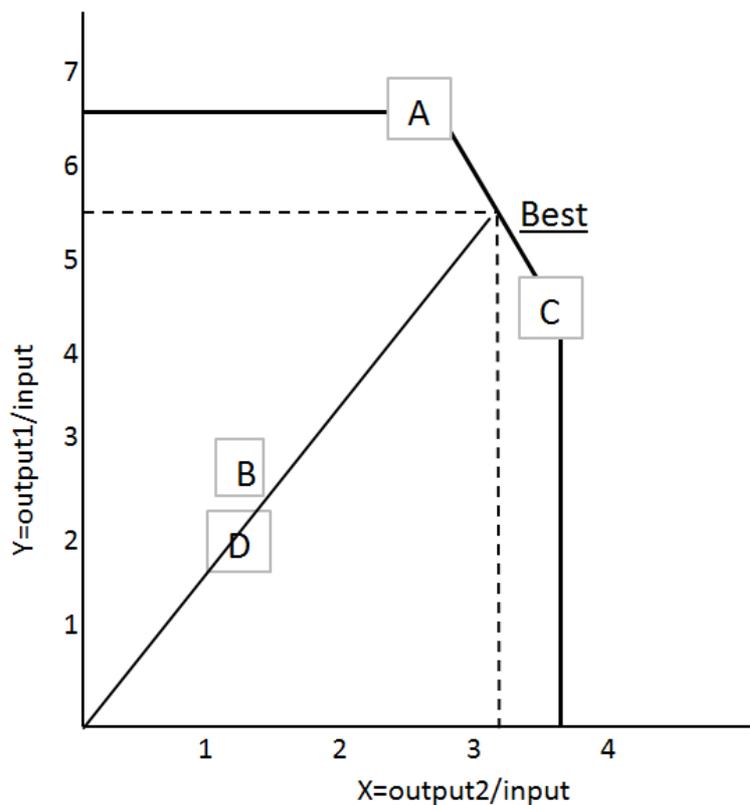


Fig. 2: Rappresentazione grafica dell'efficienza relativa

Appare subito evidente che A e C presentano un livello di performance raggiunto più elevato rispetto alle altre due e disegnando una linea che unisce i due punti e gli stessi agli assi cartesiani è possibile costruire la frontiera efficiente o di efficienza (efficiency frontier). Essa si presenta graficamente come uno scafo concavo verso il basso, sotto

la tangente) e rappresenta i valori che hanno un'efficienza del 100% con diversa combinazione dei due output (A , C) e allo stesso tempo lo standard che le altre unità produttive non sulla linea potrebbero provare a raggiungere (da qui il termine Data Envelopment Analysis).

D'altronde anche le stesse performance di A e C potrebbero essere migliorate, ma fino a che punto? E i punti B e D sono meno efficienti del 100% ma di quanto?

Per rispondere a quest'ultima domande occorre prendere a riferimento B e D e calcolare per ciascuna di esse il rapporto tra i due output, stabilendo in tal modo quanto del primo è parte del secondo (mix di produzione). Nella figura 2 tale risultato è lo stesso su ogni punto della linea che parte dall'origine dei dati e arriva al punto D e costituisce il suo gradiente o, in altre parole, il suo particolare mix dei due diversi output prodotti. Prolungando tale linea fino alla frontiera di efficienza è possibile definire un punto *Best* che la stessa DMU potrebbe cercare di raggiungere per migliorarsi e arrivare con lo stesso business mix a un'efficienza del 100%. Dunque il punto *Best* è il punto in cui la retta che parte dall'origine e attraversa il punto D incontra la frontiera, vale a dire la linea ideale tracciata partendo dalla combinazione dei dati reali a disposizione e dunque tale punto è potenzialmente raggiungibile. Inoltre tutti i punti che appaiono sulla retta che congiunge A e C rappresentano DMU "virtuali" ciascuna ottenibile dalla combinazione degli stessi.

Graficamente l'efficienza di D è calcolata dal rapporto tra la distanza origine-punto D e la distanza origine-punto *Best*.

Per migliorare la propria performance D potrebbe utilizzare, come in precedenza sottolineato, un input target, minimizzando gli input, o un output target, massimizzando gli output oppure una combinazione delle due modalità.

Aggiungendo altre DMU alle precedenti ognuna con coordinate y/x , e soprattutto aumentando il numero di input utilizzati e di output, la sola rappresentazione grafica appare poco

chiara nella sua definizione. In tal caso è necessario procedere ad utilizzare calcoli matematici, in particolare per calcolare l'efficienza:

$$E = \frac{\sum_{i=1}^n (output_i \cdot peso_i)}{\sum_{i=1}^n (input_i \cdot peso_i)}$$

posto che: $0 \leq E \leq 1$ e $output \geq 0$

dove E = efficienza,

e utilizzare la programmazione lineare per risolvere il problema. (Beasley, 2009), posto che: (output pesati) – (input pesati) = working (≤ 0) e che l'input pesato (o l'output pesato) della DMU della quale si vuole mantenere come obiettivo per gli altri = 1.

3. L'efficienza nell'utilizzo delle risorse umane

Le risorse umane rappresentano una percentuale variabile tra il 65% ed il 70% dei costi nelle Aziende Sanitarie; la loro corretta allocazione costituisce pertanto un fattore strategico ai fini dell'efficienza e dell'efficacia del processo attraverso il quale vengono erogate le prestazioni. Un primo passo può essere rappresentato dall'analisi della matrice organizzativa di ciascuna struttura complessa nella quale attività, tempi e personale dedicato vengono chiaramente esplicitati; è così possibile un confronto tra diversi modelli gestionali e l'elaborazione di standard ai quali ricondurre tutte le attività omogenee.

Per la valutazione di tali standard è peraltro necessaria una misurazione di efficienza relativa dei singoli modelli rispetto ad un benchmark, misurata in termini di rapporto tra volume della produzione e teste pesate (full time equivalent – FTE) cioè numero

di operatori ponderato per la tipologia di orario contrattuale. Tale misurazione, per essere concretamente utilizzabile, deve tenere conto del contesto in cui l'attività è svolta e del diverso mix di prestazioni erogate, non soltanto per diversa complessità e intensità di cure ma anche per differente setting assistenziale. Si pensi alle molteplici attività di una Azienda Ospedaliera quali ricoveri per acuti, riabilitazione, day hospital e day surgery, prestazioni diagnostiche di laboratorio e strumentali; spesso componenti multiple, in varia percentuale, della produzione di uno stesso centro di responsabilità. (Rebba – Rizzi, 2000)

Particolarmente utile a questo scopo, per le considerazioni espresse nel capitolo precedente, appare la Data Envelopment Analysis che consente di valutare come output tutte le tipologie di prestazioni variandone continuamente l'importanza relativa al fine di massimizzare l'efficienza di ciascuna struttura e costruire di conseguenza la frontiera di produzione quale linea ideale che unisce i best performers aziendali.

3.1. La scelta del modello di analisi

Poiché l'approccio seguito è stato di tipo sperimentale si è deciso di non investire inizialmente risorse per l'implementazione delle procedure della ricerca operativa mediante algoritmo del sinflesso e basare l'analisi su un modello a due dimensioni nel quale la costruzione della frontiera di produzione ed il calcolo dell'efficienza relativa di ciascuna struttura potessero rispondere ad un approccio di tipo geometrico, di più facile ed economica implementazione.

Le due dimensioni sono rappresentate dal rapporto tra diverse unità di prodotto e FTE di personale. Considerato che le risorse umane possono grossolanamente suddividersi in due macro aree, dirigenza e professioni sanitarie, e che gli output sono invece rappresentati da molteplici prestazioni (ricoveri ordinari, ricoveri diurni, prestazioni specialistiche ambulatoriali, accessi di pronto

soccorso, episodi di osservazione breve intensiva (OBI), sedute di sala operatoria, ecc..) si è deciso di restringere l'elaborazione in questa prima fase al personale dirigente medico delle unità operative degenziali e di suddividere le prestazioni in due gruppi: gli episodi di ricovero, misurati attraverso il metodo dei punti DRG e tutte le altre, valorizzate applicando le tariffe del nomenclatore aziendale o regionale (prestazioni ambulatoriali, episodi di OBI) o i risultati dell'activity based costing (sedute di sala operatoria).

Sono state quindi prodotte le procedure software in ambiente SAS per la costruzione della matrice degli output secondo quanto descritto di seguito.

Episodi di ricovero

Poiché l'analisi ha lo scopo di verificare l'appropriatezza allocativa delle risorse umane nella gestione delle attività e non l'efficienza nella gestione del caso clinico si è ritenuto opportuno valorizzare sia il singolo episodio di ricovero nel suo complesso sia le giornate di degenza prodotte da ciascuna struttura che possono presentare differenze notevoli a parità di entità nosologica. Per il primo aspetto sono stati utilizzati i punti DRG derivanti dalla somma dei pesi DRG specifici di tutti gli episodi di ricovero, tenendo conto anche dei pazienti trasferiti ad altre unità operative aziendali e non solo dei dimessi; nel secondo le giornate di degenza sono state trasformate in punti DRG equivalenti utilizzando il costo standard della singola giornata, così come risultante dall'activity based costing, stratificato per area di intensità assistenziale (acuzie, intensiva, riabilitazione) rapportato al valore medio del punto DRG dei ricoveri ordinari standard mediante le formule seguenti:

$$\begin{aligned}
 \text{puntiDRG}_{CDR} &= \sum_{DRG=pz1}^{pzN} \text{peso}_{DRG} & \text{valore_puntoDRG} &= \frac{\sum_{ricord=1}^N \text{importo}_{ricord}}{\sum_{ricord=1}^N \text{peso}_{ricord}} \\
 \text{puntiDRGequiv}_{CDR} &= \frac{\sum_{gg=1}^N \text{importo}_{gg}}{\text{valore_puntoDRG}}
 \end{aligned}$$

dove *CDR*=centro di responsabilità; *ricord*=ricovero ordinario standard; *gg*=giornata di degenza.

Nello stesso modo sono stati trattati gli episodi di ricovero diurni sostituendo le giornate di degenza con il numero degli accessi e il loro relativo costo standard.

Le basi dati utilizzate per l'elaborazione sono costituite dal flusso del debito informativo SDO dell'anno 2009.

Altre prestazioni

Le altre prestazioni sono state valorizzate applicando il tariffario del nomenclatore aziendale o regionale, dove disponibile. Sono state così trattate le prestazioni di specialistica ambulatoriale per esterni, le prestazioni intermedie (consulenze, diagnostica di laboratorio, per immagini o strumentale effettuate su pazienti degenti), gli accessi di pronto soccorso, gli episodi di osservazione breve intensiva (OBI).

Per il solo personale anestesista sono state inoltre valorizzate le sedute di sala operatoria mediante l'utilizzo del costo standard della specifica risorsa umana; i chirurghi non sono stati ricompresi in tale elaborazione in quanto si è considerato che l'attività operatoria di ogni caso clinico fosse già stata valutata dal relativo peso DRG imputato all'unità operativa che ha avuto in carico il paziente.

In Regione Liguria esiste inoltre una modalità alternativa al ricovero diurno per l'erogazione di prestazioni integrate diagnostico-terapeutiche costituita dai Pacchetti Ambulatoriali Complessi (PAC) che presentano una tariffa unitaria omnicomprensiva. Anch'essi sono stati pertanto inseriti in questo capitolo utilizzando tale tariffa per la loro valorizzazione.

Le basi dati utilizzate sono costituite dai flussi del debito informativo regionale 2009 della specialistica ambulatoriale, dei

PAC, dell'attività di pronto soccorso nonché dai dati 2009 estratti dagli applicativi aziendali per la gestione delle prestazioni intermedie e dei blocchi operatori. Questi ultimi, unitamente ai flussi del debito informativo, rappresentano parte integrante del datawarehouse del controllo di gestione.

Risorse umane

Per le risorse umane sono stati utilizzati i dati dell'anagrafica del personale considerando in particolare il campo relativo alla tempistica del contratto individuale anche se il riscontro di dirigenti medici a tempo parziale è risultato assolutamente sporadico. A tutto il personale universitario è stata invece applicata la percentuale del 60% per l'elaborazione del dato FTE ricavata dall'ultima convenzione Regione/Università che fissa a tale livello la quota di orario dedicata all'assistenza. Sono stati inoltre presi in considerazione gli specializzandi assegnati a ciascuna unità operativa per i quali gli FTE sono stati calcolati applicando una percentuale stimata nella misura del 50%. La base dati utilizzata per quest'ultima informazione è stata fornita direttamente dagli uffici didattici dell'Ateneo.

3.2. Costruzione degli indicatori

A partire dai dati di input (risorse umane) e output stratificati per unità operativa (CDR) sono stati quindi elaborati gli indicatori, espressione delle due dimensioni da inserire nell'analisi.

Allo scopo di consentire una maggiore omogeneità di valutazione si è deciso innanzitutto di procedere alla costruzione di due frontiere di produzione: la prima sulle strutture caratterizzate da tutte le attività di ricovero, sia ordinario che diurno e la seconda sulle strutture con sola degenza diurna. Ambedue i gruppi sviluppano inoltre la loro produzione in almeno una delle altre aree sopra descritte (ambulatoriale, prestazioni intermedie, ecc.) il che consente la costruzione delle due dimensioni necessarie.

L'indicatore espressione della prima dimensione è stato calcolato dal rapporto tra punti DRG, intesi come sommatoria delle due componenti relative rispettivamente alla gestione del caso clinico ed alle giornate di degenza, ed FTE del personale dirigente medico.

La seconda dimensione è invece rappresentata dall'indicatore ricavato dal rapporto tra valore economico di tutte le altre prestazioni, non di ricovero, espresso con le modalità descritte nel sottocapitolo 3.1, ed FTE del personale dirigente medico.

Si ricavano dunque due variabili quantitative per ciascuna struttura che partecipa al modello di valutazione che consentono la produzione di un grafico di dispersione (scatterplot) nel quale ogni punto rappresenta una singola unità operativa (CDR). Sull'asse delle ascisse è posto il valore del primo indicatore (attività di ricovero/FTE) e sull'asse delle ordinate il valore del secondo indicatore (altra attività/FTE). Quanto più il punto è lontano dall'origine degli assi tanto più è elevato il rapporto tra prodotto ed FTE di quel CDR. Nell'arco compreso tra i gradi da 0° a 30° del quadrante risultano plottate le strutture con una attività orientata prevalentemente nel settore alternativo al ricovero; nell'arco tra i gradi da 60° a 90° troviamo le strutture con attività di ricovero prevalente sulle altre; nell'arco tra i gradi intermedi si localizzano invece le strutture con un mix di attività bilanciato tra ricoveri ed altre prestazioni. (figura 3)

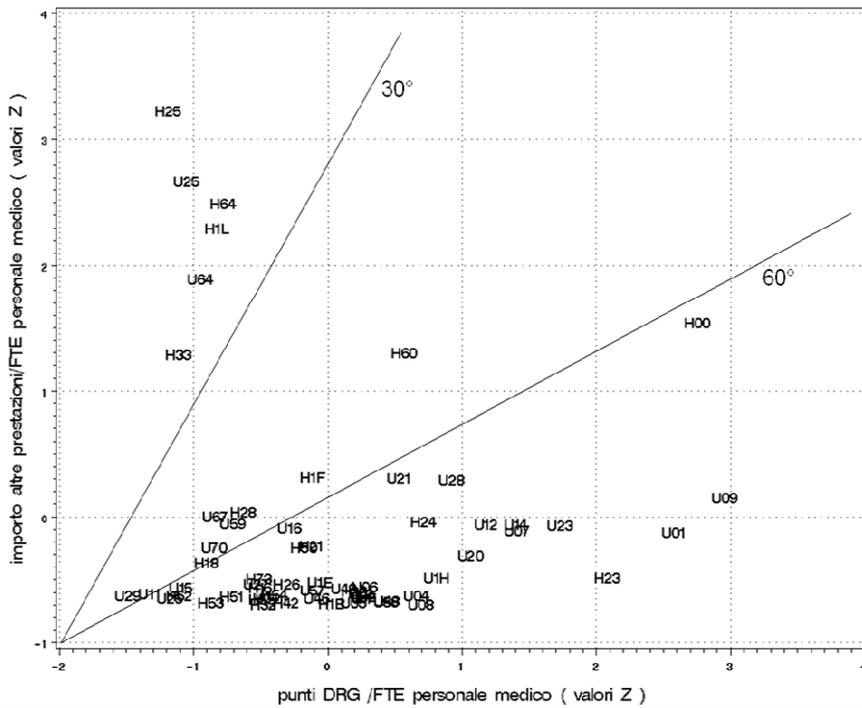


Fig.3: Distribuzione delle unità operative per volumi di attività procapite

3.3. Modalità di calcolo della frontiera di produzione

La frontiera di produzione è stata individuata mediante lo sviluppo ciclico di equazioni di regressione.

Le procedure software, sviluppate in ambiente SAS comprendono i seguenti passaggi:

- a) individuazione dei best performers delle due aree esterne del quadrante (archi 0° - 30° e 60° - 90° gradi) cioè di quelle due strutture rispettivamente con il massimo valore della variabile in ordinata e della variabile in ascissa ed elaborazione della equazione di regressione, espressione della retta congiungente le due strutture (figura 4);

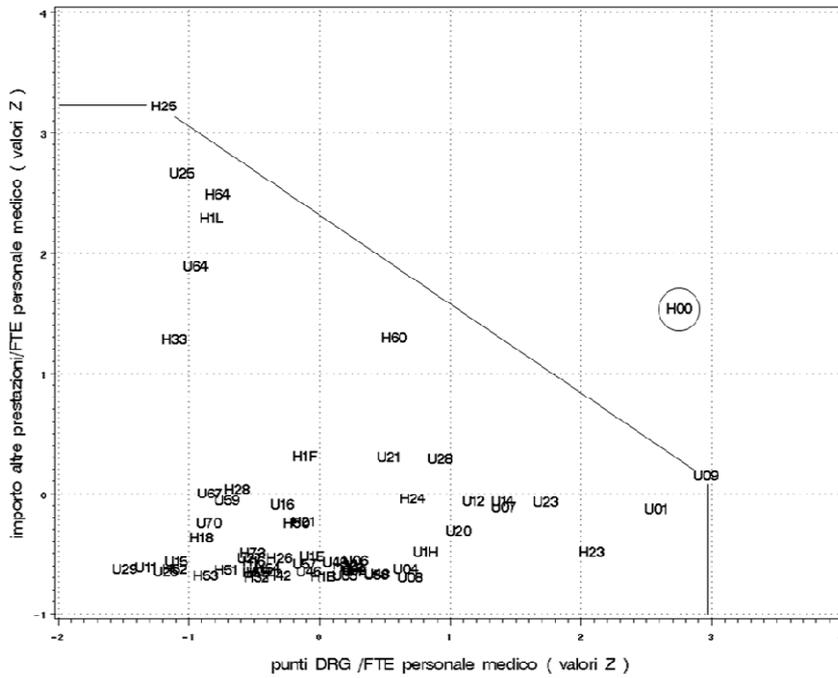


Fig.5: Individuazione progressiva delle strutture esterne alla retta di regressione

- c) costruzione delle rette di regressione tra il nuovo punto e i due best performers che lo circoscrivono nella frontiera (figura 6);

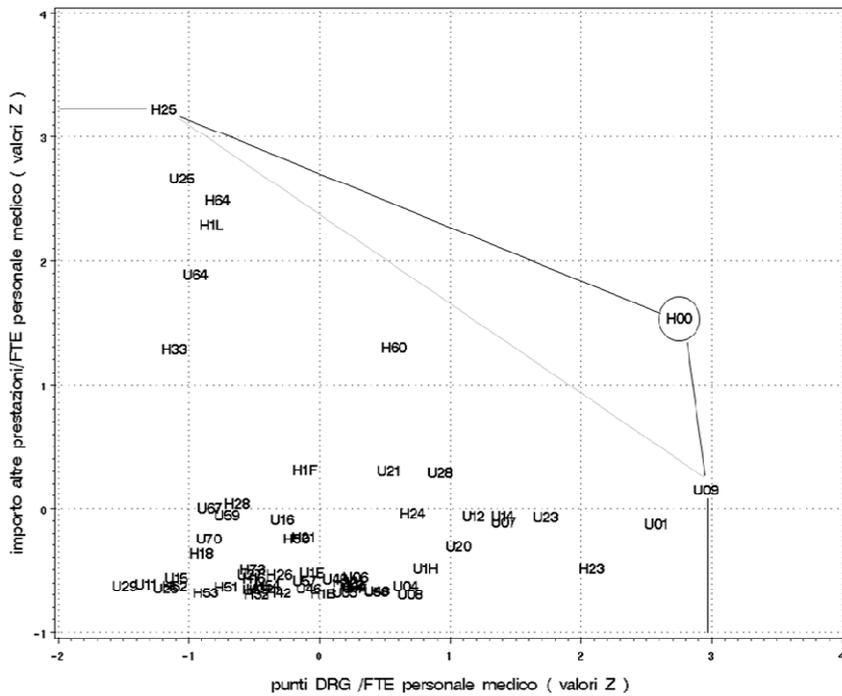


Fig.6: Costruzione delle nuove rette di regressione sulla struttura individuata all'esterno delle precedenti

d) ripetizione dei punti b) e c) fino ad esaurimento dei CDR esterni alle rette di regressione (figura 7).

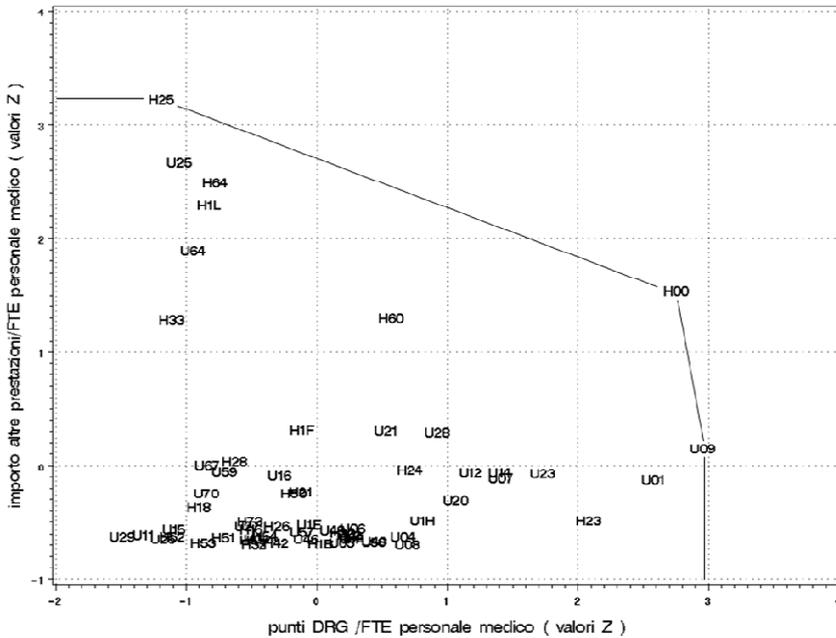


Fig.7: Costruzione della frontiera di produzione definitiva

Al termine del ciclo di processo si ottiene una rappresentazione grafica dove i best performers tracciano una linea (la frontiera di produzione) in corrispondenza della quale si trova la migliore performance possibile teoricamente realizzabile per ognuno degli infiniti microsettori angolari del quadrante. In altre parole ciascuna struttura che non contribuisce alla costruzione della frontiera può progredire in senso centrifugo sulla retta che la unisce all'origine degli assi fino a trovare il punto di migliore performance possibile nel contesto in cui opera e per il mix di attività che la caratterizza, elaborata sulla base della distribuzione delle performance di tutta l'azienda.

Simulando tale progressione per ciascuna unità operativa si ottiene la matrice di efficienza relativa. Il procedimento utilizzato è descritto nel sottocapitolo seguente.

3.4. La matrice di efficienza relativa

Anche per l'elaborazione della matrice dell'efficienza relativa di ciascuna unità operativa è stato scelto un approccio di tipo geometrico.

Risulta evidente dalla lettura del grafico in **figura 8** che l'efficienza della struttura presa ad esempio (H60) è rappresentata dalla percentuale che la sua distanza dall'origine degli assi (segmento A) rappresenta rispetto alla lunghezza totale del segmento di retta ottenuto prolungando tale distanza fino alla linea di frontiera (segmento A+B).

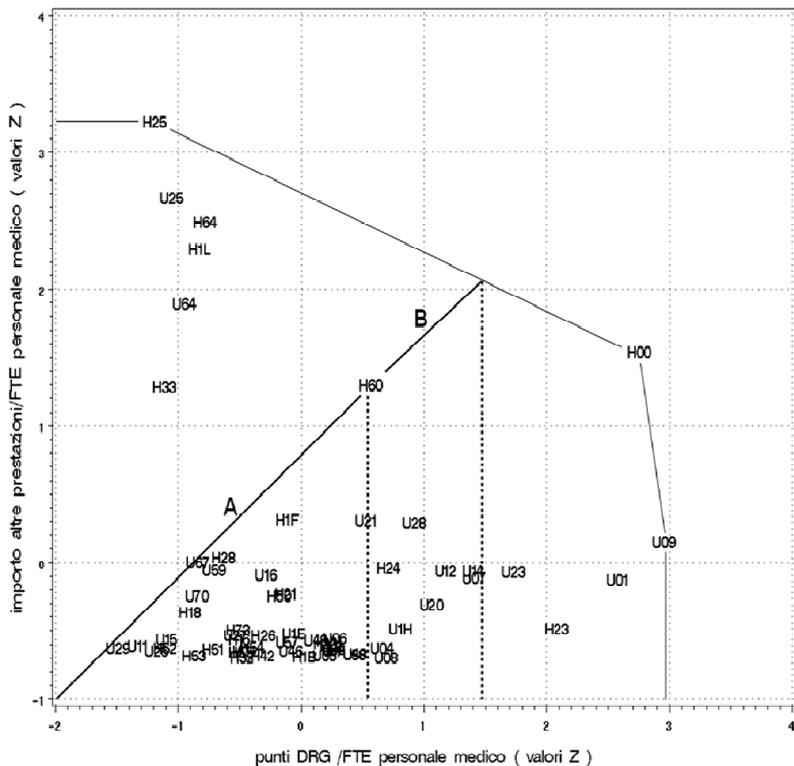


Fig.8: Calcolo dell'efficienza relativa di ciascuna struttura come rapporto dei segmenti A e A+B tramite l'applicazione del teorema di Pitagora

Risulta altrettanto evidente che la distanza tra l'origine degli assi e H60 (segmento A) insieme alle coordinate cartesiane dell'unità operativa forma un triangolo rettangolo inscritto in un altro triangolo di maggiore superficie i cui lati sono il segmento A+B sopra citato e le coordinate cartesiane del suo incrocio con la linea di frontiera. Queste ultime coordinate sono calcolabili risolvendo l'appropriata equazione di regressione tra quelle che hanno contribuito a costruire la frontiera (sottocapitolo 3.3).

Le misure dei segmenti A e A+B (ipotenuse) sono calcolabili applicando il teorema di Pitagora dove i cateti sono rappresentati dalle due coppie di coordinate cartesiane. Il rapporto tra le due ipotenuse esprime l'efficienza relativa della struttura in esame.

La formula applicata per il calcolo dell'efficienza relativa è stata pertanto la seguente:

$$ER_{H60} = \frac{\sqrt{(X_A^2 + Y_A^2)}}{\sqrt{(X_{A+B}^2 + Y_{A+B}^2)}}$$

A questo punto sono indispensabili due considerazioni.

La prima relativa al fatto che l'elaborazione del modello è stata effettuata separatamente per le strutture con soli ricoveri diurni rispetto a quelle dotate anche di posti letto di regime ordinario e pertanto due valori uguali di efficienza nei due diversi modelli non sono perfettamente confrontabili.

La seconda, di valenza generale nei sistemi di valutazione delle performance, è invece relativa alla quota residua di variabilità tra le strutture non spiegata dai sistemi di classificazione utilizzati (DRG, nomenclatori tariffari, analisi dei costi). Ciò rende inappropriato utilizzare in termini rigorosamente parametrici i risultati dell'analisi per il confronto tra le unità operative; è

pertanto consigliabile utilizzare un approccio non parametrico e ragionare per quantili di distribuzione del risultato.

Per rispondere alle due considerazioni si è deciso di calcolare per le due matrici di efficienza la variabile Z standardizzata a media=0 e deviazione standard=1 unificando così le matrici derivanti dai due modelli in un'unica base dati e stratificando le strutture in tre macroaree i cui cut-off sono rappresentati dai valori $Z=-1$, $Z=1$.

In questo caso un valore Z pari a 1.96 individua comunque una o più strutture posizionate al 99mo percentile della loro reciproca distribuzione (modello con soli ricoveri diurni o modello con anche ricoveri ordinari) indipendentemente dal valore non standardizzato della loro misura di efficienza e quindi consente l'unificazione delle due matrici.

Inoltre la suddivisione delle strutture in un gruppo che insiste intorno al valore medio delle due distribuzioni e due gruppi rispettivamente a destra e a sinistra di una deviazione standard rispetto alla media ($p=0,33$) riduce il bias conseguente alla disomogeneità intrinseca dei sistemi di classificazione. In un processo di miglioramento continuo della qualità è infatti più corretto e più produttivo tenere conto delle strutture con performance nella media, superiore alla media (oltre 1 deviazione standard a destra) o inferiore alla media (oltre 1 deviazione standard a sinistra) e costruire un obiettivo su queste ultime, piuttosto che agire sull'interpretazione aritmetica di minime differenze di performance che, come si è detto, potrebbero dipendere dalla disomogeneità dei sistemi di classificazione utilizzati.

4. L'incrocio con l'efficienza allocativa della diagnostica

Il processo produttivo nelle aziende sanitarie risente fondamentalmente di due principali fattori produttivi: le risorse umane, che come si è detto in premessa nel capitolo 3

rappresentano una quota di costi variabile dal 65% al 70%, ed i beni & servizi che aggiungono a tale quota un altro 15%-20%. Il controllo di questi due fattori rappresenta pertanto una leva in grado di mobilitare circa l'85% dei costi di produzione.

In un precedente articolo (Copello-Pellicanò, 2009) è stato presentato un modello di analisi multidimensionale dell'efficienza allocativa mirato principalmente a beni & servizi, diagnostica di laboratorio e diagnostica per immagini e strumentale. Tale modello è stato applicato in Azienda in questi ultimi anni soprattutto per migliorare l'efficienza allocativa del materiale di consumo (beni & servizi). A partire dall'anno in corso l'attenzione si è spostata sulle altre due voci che in realtà sono più complesse in quanto racchiudono al loro interno sia la gestione delle risorse umane che quella dei beni di consumo.

La diagnostica costituisce infatti la quota prevalente delle cosiddette prestazioni intermedie; di quelle prestazioni cioè richieste dalle unità operative degenziali ai servizi diagnostici (laboratori, radiologie, altri servizi specialistici) ed effettuate nei confronti dei pazienti ricoverati, necessarie per la produzione dell'oggetto di costo finale rappresentato dai DRG. La restante quota delle prestazioni intermedie, minoritaria, è costituita dalle consulenze specialistiche. I fattori produttivi principali utilizzati dai servizi diagnostici per la produzione delle prestazioni intermedie sono naturalmente rappresentati dalle risorse umane e dai beni di consumo.

Allocare correttamente la diagnostica significa pertanto poter allocare correttamente sia le risorse umane che il materiale di consumo dei servizi. In altre parole una organizzazione teoricamente perfetta dell'attività dei servizi diagnostici viene vanificata a livello Aziendale da un uso inappropriato dei loro oggetti di costo da parte delle altre strutture.

Per questo motivo, avendo costruito la DEA sulle sole strutture degenziali, si è ritenuto opportuno interpretare ed utilizzare congiuntamente per la valutazione ed i processi di

miglioramento il dato dell'efficienza relativa rispetto alla frontiera di produzione con il dato del costo della diagnostica per punto DRG elaborato come descritto nell'articolo citato in bibliografia.

4.1. Costruzione degli indicatori

Per ogni unità operativa è stato calcolato il costo della diagnostica di laboratorio e della diagnostica per immagini o strumentale ricavato dalla base dati delle prestazioni intermedie. Al fine di limitare la variabilità distributiva si è deciso di non considerare le prestazioni di radiologia interventistica per i seguenti motivi: l'alto costo di erogazione; la loro concentrazione in determinate discipline (chirurgia vascolare, neurochirurgia, ecc.); la loro verosimile maggiore appropriatezza rispetto alle tecniche esclusivamente diagnostiche.

Sono stati parallelamente elaborati i punti DRG prodotti da ciascuna struttura utilizzando il metodo standard:

- sommatoria dei pesi DRG per i ricoveri ordinari;
- calcolo dei punti equivalenti ottenuto dal rapporto tra valorizzazione economica del ricovero e valore medio del punto DRG (vedi formula al sottocapitolo 3.1) per le altre tipologie di ricovero.

Dal rapporto tra i due dati si è ottenuto il costo medio delle due tipologie di diagnostica per punto DRG per ciascuna unità operativa, la cui distribuzione è illustrata nelle figure 9 e 10. E' evidente una elevata disomogeneità della spesa, maggiore per la diagnostica di laboratorio ma in ambedue i casi superiore alla presunta variabilità intrinseca del sistema DRG (75%), misurata tramite il coefficiente di variazione.

Infine i due costi sono stati sommati per ottenere un unico vettore che è stato successivamente standardizzato alla variabile Z con media 0 e deviazione standard 1 in analogia a quanto già effettuato sui dati della efficienza relativa.

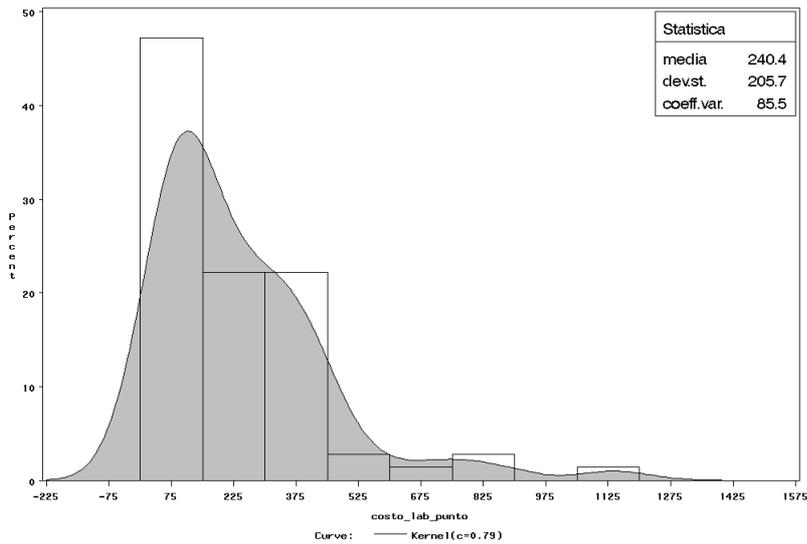


Fig.9: Distribuzione per unità operativa del costo per punto DRG in diagnostica di laboratorio

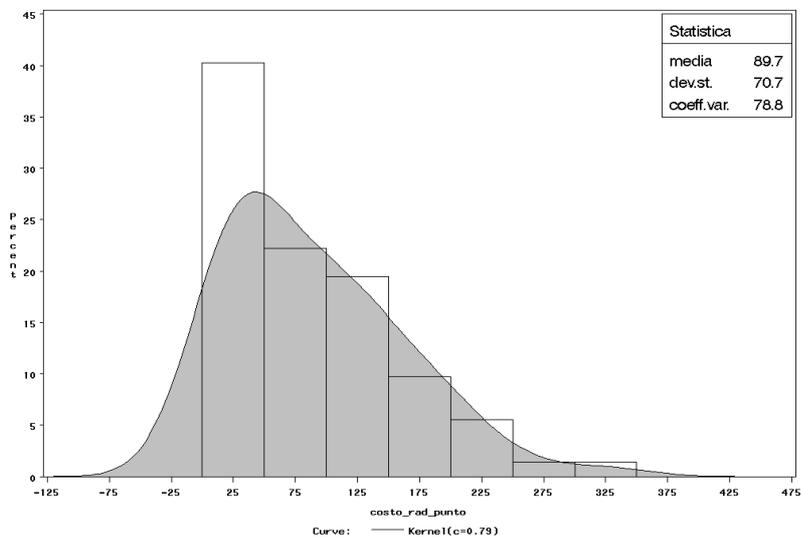


Fig.10: Distribuzione per unità operativa del costo per punto DRG in diagnostica per immagini

4.2. La rappresentazione grafica bidimensionale

Per una valutazione di insieme del fenomeno, propedeutica alla pianificazione degli interventi di miglioramento, si è provveduto alla costruzione di un grafico di dispersione (scatterplot) ponendo in ascissa la standardizzata Z dell'efficienza relativa, come calcolata tramite la DEA e in ordinata la standardizzata Z del costo della diagnostica per punto DRG come appena descritta.

Utilizzando la media come cut-off per quest'ultima e 1 deviazione standard attorno alla media per l'efficienza relativa secondo quanto descritto nel sottocapitolo 3.4 il grafico è suddivisibile in 6 aree (figura 11):

- due centrali, con efficienza relativa delle risorse umane attorno al valore medio suddivisa in alto e basso costo della diagnostica;
- due a destra, con alta efficienza relativa suddivisa in alto e basso costo della diagnostica;
- due a sinistra, con bassa efficienza relativa suddivisa in alto e basso costo della diagnostica.

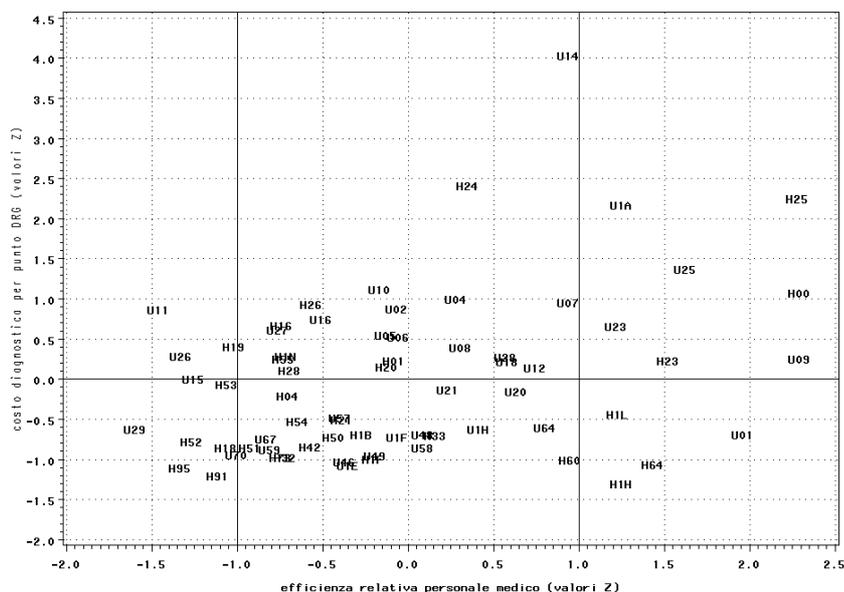


Fig.11: Distribuzione delle unità operative per efficienza relativa delle risorse umane e costo totale della diagnostica per punto DRG

Ai fini del processo di miglioramento è evidente che, a parità di efficacia clinica, l'area a tendere è quella situata nel quadrante inferiore di destra che presenta la massima efficienza delle risorse umane unitamente al minor costo per punto DRG della diagnostica mentre le unità operative dove appare più urgente un processo di riorganizzazione sono quelle localizzate nel quadrante superiore di sinistra, caratterizzato dalla minore efficienza delle risorse umane e dal maggior costo per unità di prodotto (punto DRG).

4.3. Analisi di raggruppamento

Per valutare la correttezza della stratificazione delle unità operative effettuata mediante questo modello geometrico e successiva rappresentazione grafica bidimensionale è stata altresì

effettuata una analisi di raggruppamento (cluster analysis) utilizzando come fattori tutte le variabili della matrice originaria della DEA (input e output; sottocapitolo 3.1) e le due variabili utilizzate per il costo complessivo della diagnostica (costo laboratorio e costo immagini-strumentale).

I risultati, rappresentati nel grafico in figura 12, dimostrano il raggruppamento delle unità operative in 8 cluster la cui correlazione con le sei categorie determinate dal grafico in figura 11 è pari al 61%.

Peraltro la lettura degli 8 cluster consente di interpretare in maniera statisticamente più corretta la distribuzione delle performance rispetto al semplice grafico cartesiano.

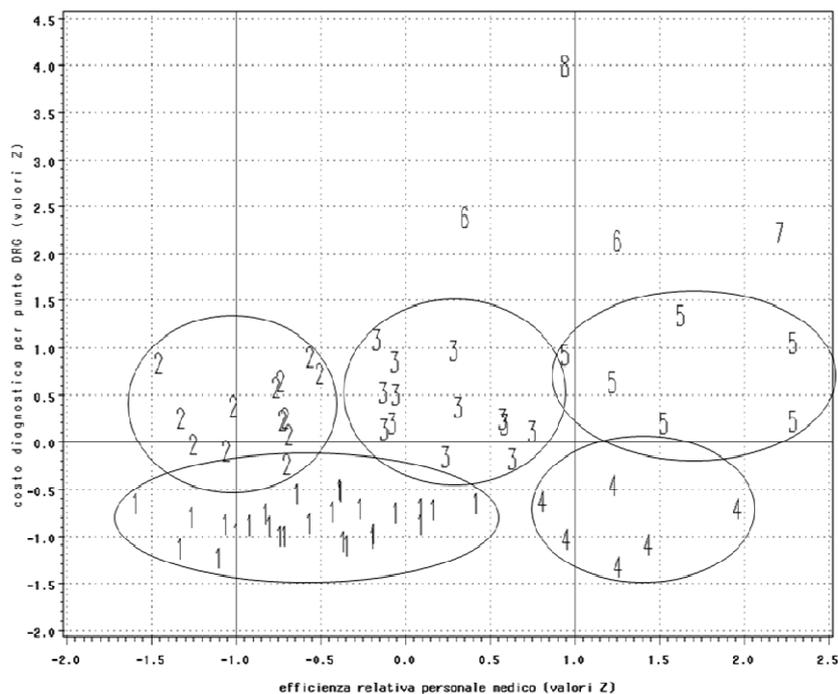


Fig.12: Risultati della cluster analysis sui dati di efficienza relativa e costo della diagnostica

Si nota un raggruppamento, classificato con il numero 3, costituito da strutture con performance attorno alla media in ambedue gli assi (efficienza delle risorse umane e costo della diagnostica per punto DRG). Due raggruppamenti, il 4 e il 5, quasi totalmente oltre 1 deviazione standard nell'asse dell'efficienza delle risorse umane (quindi con performance ottimale); il numero 4 ha anche una buona performance relativamente al costo della diagnostica mentre il 5 ha costi più elevati della media. Altri due raggruppamenti, l'1 e il 2, simili rispettivamente al 4 e al 5 per quanto riguarda il costo della diagnostica ma in posizione opposta sull'asse dell'efficienza delle risorse umane; si tratta peraltro di due cluster piuttosto dispersi su questo asse in quanto, soprattutto per quanto riguarda il numero 1, sono parzialmente compresi entro 1 deviazione standard e parzialmente oltre (quindi con performance decisamente negativa).

Infine si notano tre cluster che rappresentano in realtà outliers: il 6 composto da due strutture, il 7 e l'8 costituiti ciascuno da una sola struttura. Per questi ultimi la performance delle risorse umane è accettabile o addirittura ottima ma è l'elevato costo della diagnostica a posizionarli sulla "coda" di destra della distribuzione.

5. Costruzione del processo di budget per aree omogenee

Sulla base dei risultati finora descritti è possibile pianificare il processo di budget con obiettivi di miglioramento continuo secondo la tecnica del kaizen, già descritta nell'articolo citato in bibliografia (Copello-Pellicanò, 2009). Il kaizen prevede da un lato il mantenimento della performance per chi appartiene già alle aree di efficienza e basso costo e dall'altro la riduzione dei costi e/o l'aumento di efficienza delle risorse umane in percentuale sostenibile per le strutture localizzate nelle aree a performance

insoddisfacente. Questo processo porta ad un miglioramento della performance complessiva aziendale come risultanza della media di strutture che hanno mantenuto il loro già apprezzabile livello e di strutture che hanno invece migliorato i loro dati. Conduce altresì ad una maggiore omogeneità dei risultati e a maggiore equità nella distribuzione delle risorse in conseguenza della loro riallocazione a totale invariato.

Considerato che l'Azienda presenta più unità operative per ogni disciplina, un'ulteriore accortezza utile a rendere più sostenibile il percorso è quella di interpretare i dati per grandi aree disciplinari. Il raggruppamento delle strutture in quattro settori, medicina interna, chirurgia generale, specialità mediche, specialità chirurgiche, consente di rendere più omogenea l'analisi e di conseguenza più coerenti gli obiettivi con la realtà operativa di ciascuna struttura.

Nel grafico rappresentato in figura 13 è illustrato l'esempio delle unità operative di medicina interna. Il primo passo è stato quello di individuare i best performers nei due aspetti considerati: efficienza delle risorse umane e costo della diagnostica. Cioè delle due strutture rispettivamente con maggior valore Z sull'asse delle ascisse e con minor valore Z sull'asse delle ordinate.

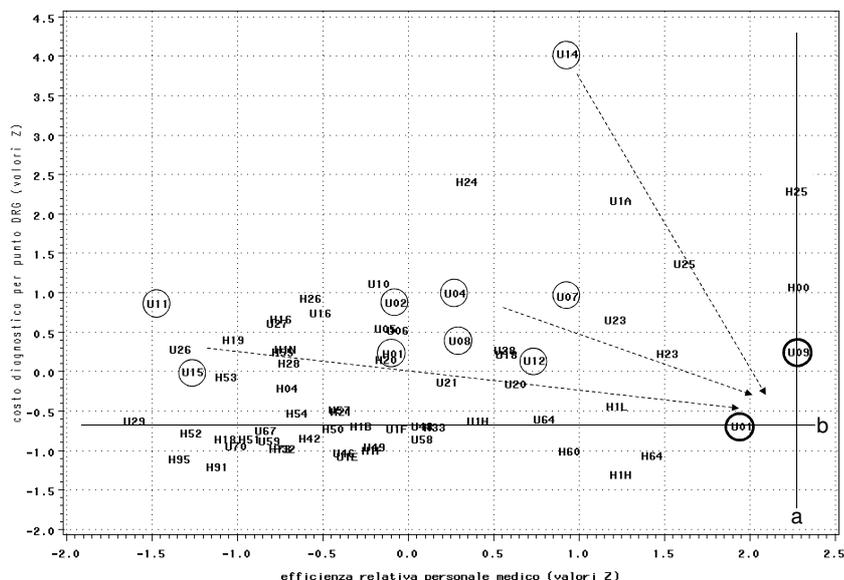


Fig.13: Strategia di assegnazione degli obiettivi di miglioramento delle performance per aree omogenee: strutture di medicina interna

I due segmenti di retta perpendicolari individuati con le lettere “a” e “b” che intersecano le due strutture, U09 e U01 rispettivamente, individuano il punto di migliore performance complessiva per l’area della medicina interna; cioè quel punto con la maggiore efficienza relativa delle risorse umane ed il minor costo della diagnostica espressi da quest’area.

Si noti che il segmento di retta “a” su cui U09 è posizionato insieme ad H00 e H25 corrisponde in questo caso alla frontiera della produzione (figura 7) in quanto il best performers delle medicine interne è uno dei tre best performers assoluti.

Individuate successivamente le posizioni delle altre strutture della stessa disciplina si è deciso di concertare un obiettivo di miglioramento in uno o ambedue gli aspetti valutati; obiettivo che,

per essere sostenibile, si ritiene non debba superare il 5%-10% in valore assoluto dei valori di input, output o costo della diagnostica che caratterizzano attualmente tali strutture. (Copello-Pellicanò, 2009).

Inoltre, facendo riferimento alle matrici organizzative citate in apertura del capitolo 3, dove vengono esplicitati i tempi e le unità di personale utilizzate per ciascuna delle attività, la posizione delle unità operative nelle 6 aree a diverso costo/efficienza costituisce un valido benchmarking per l'individuazione dei migliori assetti organizzativi da utilizzare come standard per l'assegnazione delle risorse umane.

6. Conclusioni

La DEA si è dimostrata un potente strumento per la stima dell'efficienza relativa nell'ambito delle unità operative degenziali dell'Azienda e, in associazione con il dato inerente il consumo di risorse per la diagnostica, standardizzato per il case-mix, consente una più agevole pianificazione degli obiettivi di budget in termini di risorse umane e prestazioni intermedie.

Occorre peraltro interpretare correttamente i risultati dell'analisi tenendo presente che i sistemi di classificazione delle prestazioni non sono sufficientemente omogenei per garantire una valutazione totalmente equa degli output delle singole strutture. Le diverse performance vanno pertanto aggregate in macroaree rispetto al valore medio e non confrontate sulla base della loro misurazione puntuale. A tal fine particolarmente utile si è dimostrata la standardizzazione al valore Z con media 0 e varianza unitaria.

L'approccio parametrico utilizzato nella sperimentazione ha consentito di ridurre al minimo l'investimento in risorse di calcolo basandosi su procedure software già a disposizione del controllo di gestione e su tecniche statistiche e matematiche elementari (equazione della retta di regressione ed equazione del teorema di

Pitagora). Peraltro questo tipo di elaborazione riduce di necessità a due le dimensioni di partenza e ciò si può ottenere associando due tipi di output ad un unico input o viceversa. Considerati gli obiettivi attuali dell'Azienda, mirati ad una più equa distribuzione della dirigenza medica, questo vincolo non ha costituito una criticità. In un'ottica di riconsiderazione complessiva delle risorse umane appare probabilmente più adeguato un approccio non parametrico utilizzando le tecniche della programmazione lineare che consente altresì di mantenere separata la valorizzazione dei diversi output prodotti (Gerdtham - Loethgren - Tambour - Rehnberg, 1999) (Siciliani, 2005), (Katz - Rosen - Bollino, 2009).

In ogni caso, considerata la multidimensionalità del processo produttivo e la stretta correlazione tra i diversi fattori che ne costituiscono le risorse, già dimostrata in una precedente analisi (Copello - Pellicanò, 2009), appare indispensabile l'associazione di diverse misure di efficienza allocativa così come descritto nel presente lavoro.

L'analisi appena esposta dei punti di forza e di debolezza, o dei possibili limiti di tale tipo di tecnica, deve per definizione essere completata con l'osservazione delle opportunità che la stessa può rappresentare per altre elaborazioni o scenari futuri di applicazione (analisi SWOT, Strengths, Weaknesses, Opportunities, Threats).

Una nuova strada di applicazione del metodo potrebbe essere nel benchmarking con altre aziende sanitarie, a livello regionale ed extraregionale, al fine di permettere una migliore condivisione di modalità operative efficienti e generare a livello di settore un miglioramento globale delle prestazioni sanitarie effettuate. In particolare la Regione, nell'ambito delle sue scelte, potrebbe individuare dove concentrare i propri sforzi e, sulla base dei target di riferimento nel dataset analizzato, attuare strategie di miglioramento necessarie al sistema organizzativo sanitario locale. Infatti - come già descritto in precedenza - l'efficienza così misurata è un'efficienza di unità produttive ognuna valutata

globalmente sulla base degli input impiegati e degli output prodotti realmente dalla stessa, che è possibile migliorare sulla base non di uno standard virtuale ma del best performer che emerge a livello di sistema. A livello globale ciò è ancor più importante visti i problemi di finanziamento e di contenimento della spesa sanitaria pubblica che affliggono ormai da decenni il sistema sanitario e il dualismo italiano che emerge dalle relazioni della spesa pro-capite effettiva degli ultimi dieci anni (Cerm, 2009). Tutto ciò tenendo ovviamente sempre presente efficacia e qualità delle prestazioni offerte, escludendo dunque un uso meccanicistico del modello descritto che, in conclusione dell'analisi SWOT, potrebbe rappresentare il principale pericolo (Threats) dell'utilizzo di questa tecnica.

Bibliografia

BATTESE G., COELLI T., RAO D., (1997), *An Introduction to Efficiency and Productivity Analysis*, Kluwer Academic Publishers, Boston.

BEASLEY J.L. (2009), "Data envelopment analysis", OR Course, Department of Mathematical Sciences, Brunel University, West London

COOPER W.W., SEIDORF L.M., TONE K. (2002) *Data Envelopment Analysis*, Kluwer Academic Publishers, Boston

COPELLO F., PELLICANÒ C. (2009), "Tecniche di miglioramento dell'appropriatezza allocativa delle risorse in una grande azienda ospedaliera." *Mecosan Management ed economia sanitaria*, 18 (70), pp.57-73

GERDTHAM, U.G., LOETHGREN M., TAMBOUR M., REHNBERG C., (1999), "Internal markets and health care efficiency: a multiple-output stochastic frontier analysis", *Health Economics*, 8, pp.151-64.

KATZ M., ROSEN H., BOLLINO C., (2009), *Microeconomia*, Mc Graw Hill

MASAAKI I. (1986), "Lo spirito giapponese del miglioramento". Il Sole 24 Ore, Milano.

PAMMOLLI F., PAPA G., SALERNO N.C., (2009), "La spesa sanitaria pubblica in italia: dentro la "scatole nera" delle differenze regionali. Il modello Saniregio.", Report CERM, Quaderno 2-2009, Finanza Pubblica

REBBA V., RIZZI D., (2000) Analisi dell'efficienza relativa delle strutture di ricovero con il metodo DEA. Il caso degli ospedali del Veneto. Convegno Associazione Italiana di Economia Sanitaria "La sanità tra Stato e mercato", Padova

SEIFORD L. M., THRALL R.M., (1990) “Recent developments in DEA, the mathematical programming approach to frontier analysis”, *Journal of Econometrics*, 46, pp.7-38.

SICILIANI L., (2005) “Analisi dell'efficienza del settore ospedaliero nella letteratura economica: l'approccio delle frontiere” in *Finanziamento e valutazione dei servizi ospedalieri. Aspetti metodologici e applicativi di una sperimentazione interregionale*, Franco Angeli Editore, Milano

Working Papers recently published

(The complete list of working papers can be found at
<http://www.diseфин.unige.it>)

- n.2/2011 Stefano Capri, Rosella Levaggi, *"Shifting the risk in pricing and reimbursement schemes? A model of risk-sharing agreements for innovative drugs "*
- n.1/2011 Cinzia Di Novi, *"The Indirect Effect of Fine Particulate Matter on Health through Individuals' Life-style"*
- n.4/2010 Angelo Baglioni, Andrea Monticini, *"Why does the Interest Rate Decline Over the Day? Evidence from the Liquidity Crisis"*
- n.3/2010 Amedeo Fossati: *"The double taxation of savings: the Italian debate revisited"*
- n.2/2010 Andrea Monticini, David Peel, Giacomo Vaciago: *"The impact of ECB and FED announcements on the Euro Interest Rates"*
- n.1/2010 Amedeo Fossati: *"Vilfredo Pareto and the methodology of the Italian tradition in public finance"*