



UNIVERSITA' DEGLI STUDI DI PADOVA
FACOLTA' DI SCIENZE STATISTICHE

CORSO DI LAUREA IN STATISTICA E
GESTIONE DELLE IMPRESE

TESI DI LAUREA TRIENNALE

LA METODOLOGIA STATISTICA NEL MARKETING

“ Tecniche di analisi statistica per la segmentazione e il posizionamento ”

Relatore: Prof. Martina Pertile

Laureando: Luciano Baù

A.A. 2003 – 2004

Ai Miei Genitori, ...

*Che costantemente mi aiutano a superare
e a vincere le mie paure e difficoltà, ...*

*Che condividono i miei momenti
di tristezza e di gioia, ...*

*Che ad ogni momento del bisogno, sono
presenti e pronti ad offrire il loro amore ...*

Un Grazie Di Tutto, Amorevolmente ...

Il vostro figlio, Luciano.

INDICE

Prefazione	pag.	5
------------	------	---

PARTE PRIMA

IL MARKETING E IL SUO RUOLO NELL' IMPRESA: "STORIA, FILOSOFIA, CONCETTI, STRATEGIE E PROCESSI OPERATIVI"

1. Marketing Management	>>	7
1.1. La filosofia gestionale	>>	7
1.2. Un po' di storia: "Dall' orientamento alla produzione all' orientamento al mercato"	>>	11
1.3. Il Marketing e il mercato	>>	13
2. Il consumatore, la segmentazione e il posizionamento del prodotto	>>	19
2.1. Il consumatore e il prodotto nei concetti del marketing	>>	19
2.2. La segmentazione	>>	22
2.3. Il posizionamento	>>	25
Bibliografia	>>	28

PARTE SECONDA

LA STATISTICA NEL MARKETING: "METODI DI ANALISI STATISTICA PER LA SEGMENTAZIONE E IL POSIZIONAMENTO DI PRODOTTO"

3. Analisi Multivariata nella segmentazione e posizionamento di prodotto	>>	30
3.1. Tecniche descrittive per la segmentazione	>>	31
3.1.1. probabilità composte o congiunte	>>	31
3.1.2. Regressione multipla	>>	32
3.1.3. Correlazione canonica	>>	36
3.2. Tecniche complesse per segmentazione e posizionamento	>>	39
3.2.1. Analisi discriminante	>>	39
3.2.2. Analisi fattoriale	>>	41
3.2.3. Analisi delle componenti principali	>>	43
3.2.4. Analisi delle corrispondenze	>>	45
3.2.5. Conjoint Analysis	>>	49
3.2.6. Cluster Analysis	>>	57
3.3. Tecniche specifiche per il posizionamento del prodotto	>>	63
3.3.1. Multidimensional Scaling – MDS	>>	63
3.3.2. Unfolding e mappe di preferenza	>>	71

4. Il metodo D.E.A. applicato al Marketing	>>	76
4.1. Un uso alternativo della Data Envelopment Analysis (D.E.A.)	>>	76
Bibliografia	>>	84

PARTE TERZA

L' INFORMATICA AL SUPPORTO DELLA STATISTICA : "IL PACKAGE STATISTICO"

5. Il software statistica SYSTAT	>>	86
5.1. Introduzione	>>	86
5.2. Una panoramica generale sul tool	>>	88
5.3. La segmentazione in Systat	>>	92
5.3.1. Cluster Analysis	>>	92
5.3.2. Conjoint Analysis	>>	94
5.4. Il posizionamento di prodotto in Systat	>>	96
5.4.1. MDS	>>	96
5.4.2. Unfolding	>>	98
Bibliografia	>>	101

PREFAZIONE

Nella seguente relazione verranno affrontati gli argomenti della segmentazione di mercato e del posizionamento del prodotto (o marca) con particolare riguardo ai metodi statistici di analisi associati ad essi.

Tale relazione è suddivisa in tre aree principali. La prima parte rappresenta una sorta di introduzione, ove vengono spiegati i vari concetti che sono alla base della segmentazione e del posizionamento, e più in generale del Marketing Management.

Nella seconda parte, invece si descrivono le varie metodologie statistiche per fare analisi relative agli argomenti precedentemente citati. In particolare verranno viste le tecniche di analisi multivariata, descrivendo con piccoli cenni alcune tecniche che tale insieme mette a disposizione, e cercando di approfondire le tecniche di Cluster (CLA) e Conjoint Analysis (COA) per la segmentazione, di Multidimensional Scaling (MDS) e Unfolding (UNF) per il posizionamento, e il metodo D.E.A. (Data Envelopment Analysis).

In merito al capitolo dedicato al metodo D.E.A. verrà presentato anche un software statistico specifico per tale analisi. Il software usato, alla quale verrà sviluppata una simulazione di funzionamento in merito all'esempio contenuto nel capitolo, è FRONTIER ANALYST 3 Professional.

La terza parte, invece è dedicata all'area informatica. In particolar modo in tale parte si descrive il software statistico SYSTAT 10.2 e verranno presentate alcune sue applicazioni di analisi attraverso dati fittizi. Si dimostrerà com'è possibile sviluppare informaticamente un'analisi di Clustering e utilizzare le tecniche COA, MDS, e UNF.

1. IL MARKETING E IL SUO RUOLO

NELL' IMPRESA:

**“STORIA, FILOSOFIA, CONCETTI,
STRATEGIE E PROCESSI OPERATIVI”**

1. Marketing Management

1.1. La filosofia gestionale

Il Marketing visto come il complesso delle attività volte alla migliore conoscenza del mercato potenziale di un prodotto per organizzarne la vendita nel modo più efficace, trova fondamentale importanza all'interno dell'impresa sia dal punto di vista organizzativo che strategico.

Ma perché parlare d'impresa?

Innanzitutto ogni impresa ha un obiettivo fondamentale (o primario), ossia quello di trasformare economicamente un insieme di risorse.

Ottenere quindi attraverso un processo produttivo, prodotti con valore economico superiore al costo dei mezzi usati per produrli.

Fondamentale per un progressivo sviluppo tecnologico, atto al fine di migliorarne la stessa produzione, e quello di reinvestire l'utile (o ritorno economico) ottenuto dalle vendite.

Quindi possiamo sintetizzare tutto ciò attraverso una semplice formula imprenditoriale che vede come elementi fondamentali il mercato, la concorrenza, la struttura aziendale, e il sistema del prodotto.

In particolare, di vitale importanza per il Marketing, sono la concorrenza, il sistema prodotto e il Mercato. (fig.1)

Essi, combinati tra loro, rappresentano le linee guida per lo studio di mercato.

Oltre a ciò è interessante notare come un'impresa interagisce con il suo ambiente esterno e come attraverso una sua struttura interna, crea attività imprenditoriali che permettono di raggiungere risultati economici, sociali e competitivi o concorrenziali.



fig. 1 - Linee guida per lo studio di Marketing

Possiamo quindi vedere, sotto un certo aspetto, l'impresa come una black-box o scatola nera. (fig.2 - Schema INPUT – OUTPUT di V.Coda)

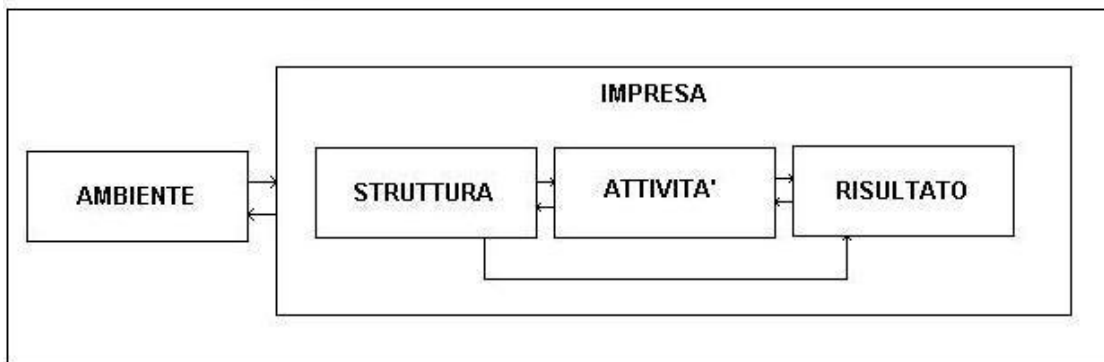


fig. 2 - Black-box dell'impresa

Il legame tra ambiente e impresa è rappresentato dagli scambi che vedono in entrata all'impresa, le risorse umane, fisiche, e tecnologiche, ed in uscita il prodotto. All'interno dell'impresa si trova la sua struttura aziendale (es. S.p.a, S.r.l., ecc...), l'attività e il risultato che si raggiunge.

L'attività può essere caratterizzata in due tipologie: Strategica o Tattica. La Strategica rappresenta la capacità di pianificare a lungo periodo, la tattica rappresenta la capacità di svolgere qualcosa al fine di raggiungere un obiettivo prefissato nel breve (o al più medio) periodo.

Importante rilievo trovano invece i risultati che rappresentano gli obiettivi primari dell'imprenditore.

I risultati possono essere classificati in due macro aree: risultati a breve periodo e risultati nel lungo periodo.

Nel breve periodo possiamo evidenziare risultati di tipo economico/finanziario, competitivi (es. misure della forza di mercato, quota di mercato, margini di contribuzione ecc...), sociali (peso dell'impresa sul consumatore, turn-over, tasso di assenteismo, ecc...).

Nel lungo periodo trova di vitale importanza lo sviluppo, ossia valutare quale risultato si deve premiare per sviluppare l'impresa, cioè la capacità di valutare, analizzare e pianificare i risultati alla ricerca di un miglioramento continuativo.

Cosa fondamentale è avere indicatori (o risultati) che siano oggettivi perché **si può migliorare solo ciò che si può misurare.**

Ricordando che un'impresa è condizionata dall'ambiente in cui è inserita, possiamo individuare una serie di ambienti che possono interagire con l'impresa.

Essi possono essere classificati in ambiente politico, socio-culturale, demografico, tecnologico, fisico, ed economico.

Con particolare riguardo l'ambiente economico rappresenta l'insieme di regole che gestiscono le relazioni economiche con l'impresa (parte interna) e i legami esterni (come altre Nazioni).

Oltre all'ambiente, l'impresa è particolarmente correlata con i suoi fornitori, che possono rappresentare a sua volta dei possibili avversari, e i concorrenti stessi.

Molte volte l'impresa non trova un diretto contatto con il suo cliente finale, ma tra i due intervengono degli intermediari commerciali (ad esempio altre imprese che trasportano a conto dell'impresa madre, il prodotto finale al cliente).

Concludendo possiamo quindi evidenziare come un'impresa proceda tramite un suo circolo virtuoso.

Esso trova fondamenti sul soddisfacimento delle attese sociali, sull'attrazione delle risorse con un obiettivo di avere ricompense in seguito, sulla produzione di un sistema di prodotto competitivo, sulla soddisfazione dei bisogni di mercato e sulla produzione di adeguati flussi di reddito.

Grazie proprio ad un suo circolo virtuoso possiamo considerare il Marketing, oltre ad una specifica funzione aziendale e una specifica area di studio, una filosofia che ispira la direzione e monitorizza il controllo e la gestione delle attività. Possiamo quindi riassumere il tutto attraverso i processi decisionali che spingono l'impresa ad orientarsi al Marketing.

Essi sono sintetizzabili in quattro elementi fondamentali:

1. Individuazione dei bisogni dei consumatori e del mercato di cui si intende valutare le opportunità e le minacce, il funzionamento e le potenzialità dell'impresa.
2. Implicazione della scelta contestuale dei segmenti di mercato a cui rivolgere l'offerta, dei benefici da offrire all'utilizzatore finale, e i vantaggi competitivi difendibili su cui basare il rapporto con i concorrenti
3. Gestire in modo ottimale (Tra i metodi innovativi ricordiamo **D.E.A. Data Envelopment Analysis**) le risorse e le capacità disponibili dell'impresa
4. Controllo sistematico dell'efficacia dell'attività dell'impresa in funzione del grado di soddisfazione dei clienti. (Ricordando che il cliente non è solo esterno ma anche interno, concetto VISON 2000)

1.2. Un po' di storia: "Dall' orientamento alla produzione all' orientamento al mercato"

L' evoluzione del management aziendale ha visto nella sua storia tre grandi orientamenti: "L' orientamento alla produzione, alle vendite, al mercato."

Il primo orientamento, quello alla produzione, è stato caratterizzato dall' avvenuta della rivoluzione industriale (fine '700 primo '800). In quel periodo nasce la necessità di produrre a grande quantità, dovuta al fatto di richiesta di prodotto. Si inizia a produrre a basso prezzo con tecnologie produttive meccaniche. Ciò porta ad uno sviluppo dell' industria e alla nascita di nuovi concorrenti spinti al desiderio di soddisfare la richiesta.

Tale richiesta viene così soddisfatta a punto tale però da avere i magazzini pieni, e tale dilemma comporta perdite economiche.

Nasce quindi un nuovo orientamento, l' orientamento alle vendite, che aveva come focalizzazione una pressione di vendita.

Il dato di partenza sulla quale ci si focalizzava era sempre il prodotto, ma l' obiettivo era quello di svuotare i magazzini, dovuto dal fatto che c' era un' offerta superiore alla domanda.

Non si usò più quindi una tecnologia produttiva meccanica ma si cercò di creare comunicazione e tecniche di vendita, ossia si andò a pubblicizzare il prodotto. Nascono così le reti di vendita, ossia uomini messi nel territorio per piazzare al meglio il prodotto medesimo (Agenti di vendita).

La cosa tuttavia sembrò non funzionare al meglio, poiché non ci si era posto il problema di chiedersi cosa volesse il consumatore e quali fossero le sue disponibilità economiche, così i magazzini rimasero sempre pieni in quanto ciò che si produceva non sempre era utile al mercato. Tuttavia proprio in quest' epoca, il Marketing diventa un' autonoma disciplina nell' ambito degli studi aziendali e inizia ad assumere un' identità specifica all' interno

dell'impresa, tanto che i pionieri e i sistematizzatori del nuovo approccio lo descrivono come una “Rivoluzione copernicana”.

Inizia quindi una nuova era ricca di eventi e fatti che spinge il concetto di Marketing a divenire ciò che ora è: una scienza che sta spingendo e gestendo i mercati d'impresa in un raggiungimento di miglioramento sistematico e continuativo nelle ricerche & sviluppo.

La nuova era, vista come un orientamento al mercato si focalizza quindi al bisogno del consumatore, studiando i bisogni e i desideri e creando così il concetto di predisporre un'offerta adeguata alle esigenze del mercato: “Il Marketing Mix”.

Philip Kotler è riuscito forse a stilizzare nella forma più sintetica tale trasformazione, marcando in modo netto una differenza tra l'impresa orientata alle vendite e l'impresa orientata al Marketing nella quale l'obiettivo è quello di non vendere tutto ciò che si produce, ma produrre ciò che si è in grado di vendere.

Kotler in tale modo arriva a definire il Marketing come l'analisi, l'organizzazione, la pianificazione e il controllo delle risorse, delle politiche e delle attività aziendali rivolte al consumatore con l'obiettivo di soddisfare i bisogni e i desideri di un gruppo selezionato di consumatori, ottenendo un profitto. In tale definizione data da Kotler si esprime un concetto di Marketing che si basa su 4 elementi fondamentali:

1. La focalizzazione del mercato: ossia capire che tipo di cliente si deve servire
2. Orientamento al cliente: capire il cliente per soddisfare le sue aspettative (o attese).
3. Coordinamento delle attività di Marketing: ogni decisione deve essere posta al corrente a tutte le altre funzioni aziendali e il Marketing è una di esse e che aiuta al coordinamento delle restanti.

4. Redditività: ottenere ciò che ci si è prefissati, ossia raggiungere gli obiettivi. Il Marketing aiuta a raggiungere tali obiettivi portando nuove informazioni sul funzionamento del prodotto.

In sintesi:

Orientamento	Produzione	Vendita	Mercato
Focalizzazione	Prodotto a basso costo	Pressione di vendita	Bisogno del cliente
Dato di partenza	Prodotto	Prodotto	Bisogni e Desideri
Strumento	Tecnologie produttive meccaniche	Comunicazione e tecniche di vendita	Marketing Mix
Condizioni	Domanda superiore all'offerta	Offerta superiore alla domanda	Disponibilità di reddito

tab. 1 - tipi di orientamento del mercato

1.3. Il Marketing e il mercato

Questo paragrafo del primo capitolo ha lo scopo di presentare brevemente ciò che è racchiuso all'interno del Marketing e come esso è strutturato e le sue interazioni esistenti con il mercato.

Base fondamentale per il Marketing e quindi per l'impresa è quello di conoscere il mercato nel quale si vuole operare.

Trova importanza a questo livello lo studio della funzione di domanda dell'impresa sottostante agli studi di Marketing.

In sintesi essa può essere formulata nel seguente modo:

$$q_{ij} = f(p_{ij}, p_k, m_{ij}, m_k, y, t)$$

Ove :

- q_{ij} = quantità domandata del bene (o servizio) i offerto dall'impresa j
- p_{ij} = prezzo del bene i offerto dall'impresa j
- p_k = prezzi dei beni k sostitutivi il bene i
- m_{ij} = altre politiche di mercato diverse dal prezzo che l'impresa j pone in atto per accreditare il bene i
- m_k = altre politiche di mercato, sempre diverse dal prezzo dell'impresa j , ma relative al bene sostitutivo
- y = variabili esogene specifiche che influiscono sullo sviluppo della domanda delle classi di beni a cui appartiene i
- t = fattori sottostanti alla formazione delle aspettative degli utilizzatori cui si rivolge l'impresa j e loro variazioni temporali.

Si possono individuare tre gruppi di variabili che influiscono sulla domanda d'impresa: quelle (p_{ij} e m_{ij}) direttamente controllabili dalla stessa impresa, quelle (p_k e m_k) riguardanti la concorrenza che nell'ambito oligopolistico l'impresa è legata da relazioni d'interdipendenza, e quelle (y e t) che sono del tutto esogene dall'impresa e dal mercato.

La relazione che si stabilisce tra l'impresa e il suo mercato di sbocco, non è del tutto adattiva perché l'impresa deve costantemente valutare le politiche di mercato dei concorrenti e in secondo luogo perché la dinamica delle variabili esogene, o in particolare quelle sulle aspettative dei consumatori, influiscono solo indirettamente sulla domanda.

A tale proposito vediamo cosa si intende per mercato nel concetto di Marketing. Per mercato si intende l'insieme dei potenziali clienti (coloro che hanno disponibilità economica) che possono essere interessati ad un prodotto e che hanno adeguate risorse per l'acquisto dello stesso.

In uno studio di mercato, bisogna quindi tenere in considerazione la domanda, le tipologie di mercato Business to Business o Business to

Consumer e la segmentazione fonte fondamentale (o stadio iniziale) per conseguire un ottimale posizionamento di prodotto/marca.

Di vitale importanza trova anche lo studio dell'ambiente esterno: dall'analisi della concorrenza, all'analisi del settore industriale (ossia l'insieme di produttori di una certa tipologia di prodotti che agiscono nel mercato nel quale svolgono le loro transazioni) e all'analisi dello scenario ambientale (ossia aspetto economico, andamento tecnologico, componente sociale e ideologie politiche).

Nel mercato il dato essenziale è il prodotto con il quale si vuole posizionarsi. Il prodotto deve essere percepito dal consumatore, e particolare importanza hanno gli attributi che si possono individuare nel prodotto medesimo. Infatti, il prodotto può essere classificato come un insieme di attributi tangibili e intangibili ed essi devono essere in grado di soddisfare le aspettative richieste dal cliente.

Un'impresa quindi si trova a dover svolgere delle ricerche di mercato più o meno specifiche con il fine di analizzare il mercato d'interesse, il cliente che ne fa parte e tante altre sfumature nei confronti degli stessi concorrenti. In quest'ambito trova di vitale importanza la statistica, argomento principale di questo contesto alla quale è dedicata in modo specifico una parte riportata in seguito.

La statistica nel nostro contesto sarà approfondita per lo studio della segmentazione e del posizionamento.

Parlando appunto di segmentazione, ossia quel processo mediante il quale si suddivide il mercato in un numero limitato di clienti sufficientemente omogeneo, si evidenzia come essa è esperienza dell'azienda, e come l'azienda stessa debba attraverso varie tipologie di Marketing:

“indifferenziato, differenziato o concentrato”, trovare un giusto dosaggio e una corretta composizione dei fattori di Marketing, in un insieme integrato

e coordinato, sulla base di un'opportuna strategia ragionata per un particolare periodo di tempo.

Tale processo rappresenta il Marketing Mix composto di quattro fondamentali elementi: “il prodotto, il prezzo, la distribuzione e la comunicazione”.

Dalla segmentazione si arriva al completamento del lungo percorso di gestione del prodotto secondo una sintesi che permette di verificare la distanza tra desiderio e la realtà.

Ciò è rappresentato dal posizionamento, ossia sintetizzando attraverso una frase: “ Trasformare un piccolo sogno in realtà”.

La gestione del prodotto non è tuttavia così semplice, è in questo paragrafo non ci soffermeremo di più di quanto citato, ma tuttavia è importante ricordare come il prodotto ha una sua vita che può essere suddivisa in quattro momenti temporali di base: “la nascita, lo sviluppo, la maturità e il declinio”

Nel periodo della nascita il prodotto è visto come un pioniere e in questa fase interviene processo innovativo, l'innovazione sul prodotto. Nello sviluppo si inizia a distinguere i prodotti a seconda della loro quota di mercato: alta si hanno le stelle, bassa i bambini difficili (o wild-cats). In questa fase interviene un'innovazione di processo. La maturità è caratterizzata dal tentativo del miglioramento dei prezzi e i prodotti vengono classificati nell'alta quota come vacche da denaro (cash cows) e nella bassa come cani (dog). Il declinio viene caratterizzato da una concorrenza distruttiva dei prezzi e i prodotti classificati nell'alta quota come veterani e nella bassa come vecchi fossili.

Questa è una visione molto sintetica delle caratteristiche di un ciclo di vita del prodotto, ma è tuttavia sufficiente per raggiungere il concetto del portafoglio prodotti, visto anche come ASA (area strategica d'affari), strutturato dalla BCG (Boston Consulting Group) che classifica il binomio

prodotto-mercato in quattro settori essenziali: Wild-CATS o Question Marks, STARS, CASH-COWS e DOGS. Questi quattro settori vengono rappresentati in una matrice ove gli assi della matrice stessa sono il tasso di sviluppo del mercato e la quota di mercato relativa. La quota di mercato è il rapporto tra il fatturato dell'impresa e il fatturato realizzato complessivamente dalle imprese che rientrano nell'aggregato di riferimento (fig.3).Rapportando questo indicatore alla quota di mercato del concorrente principale si ottiene la quota di mercato relativa dell'impresa.

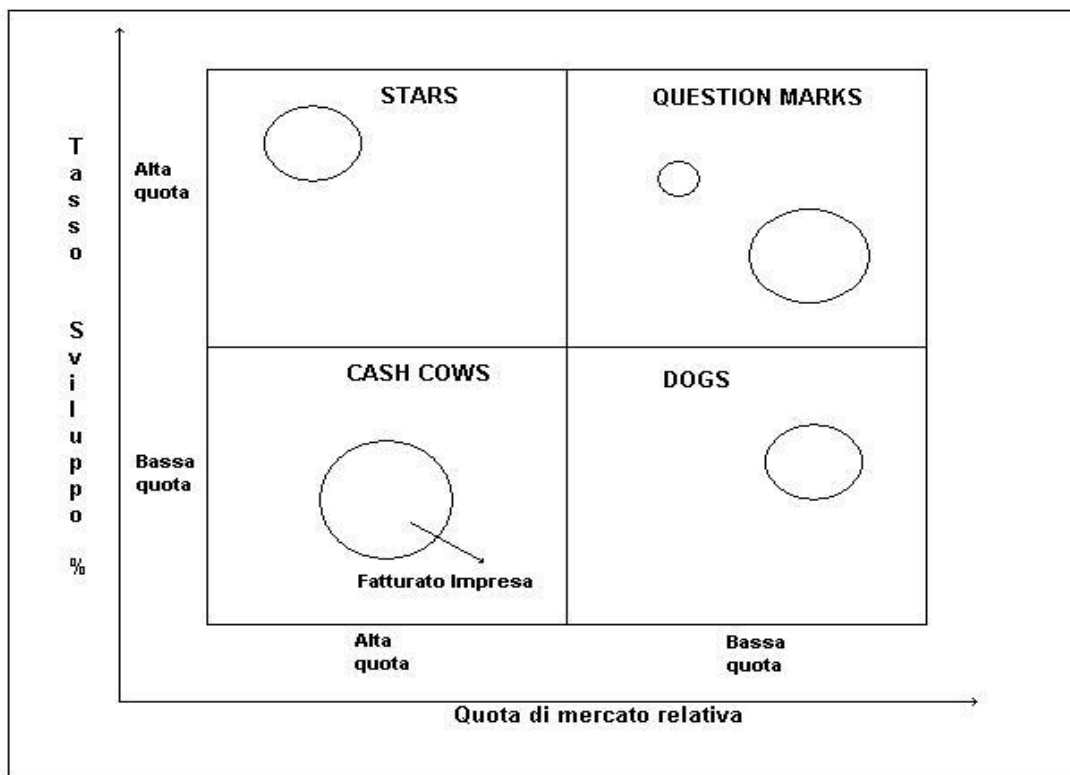


fig. 3 – Matrice BCG

Il prodotto, non è creato e sviluppato solo da un'azienda e spesso la stessa tipologia di prodotto viene venduta nel mercato da imprese diverse. Nasce quindi il concetto di differenziazione del prodotto e nel suo insieme di prezzo.

Ogni acquirente definisce un valore al bene che decide di acquistare e sceglie il fornitore che offre il maggiore valore aggiunto. A questo

proposito l'impresa deve essere in grado di creare un vantaggio competitivo, ossia a creare attività che danno valore all'impresa. Ogni impresa è costituita da una serie di attività che hanno come oggetto la progettazione, la commercializzazione, la consegna del prodotto, il servizio offerto dei prodotti nel mercato. Tali attività vanno a formare un catena definita del valore. L'azienda quindi deve essere in grado di esaminare i costi e i risultati relativi ad ogni attività creatrice di valore. Deve confrontarsi per poter capire come lei opera con gli altri (fare Benchmarks). A questo punto i costi e i risultati dei concorrenti dovrebbero essere valutati e assunti come standard di riferimento, e nella misura in cui l'impresa può operare meglio dei propri concorrenti, valutare su come conseguire il vantaggio competitivo, ossia capire dove l'azienda può divenire la migliore nei confronti degli avversari stessi.

Raggiunto tale vantaggio competitivo l'azienda è in grado di posizionarsi con una certa forza nel mercato di sbocco prescelto (o segmentato) con il suo prodotto. Tuttavia questo deve richiedere inizialmente anche un'accurata analisi, attraverso ricerche di mercato, su quali sono i bisogni e le relative aspettative che il consumatore cerca nel prodotto.

2. Il consumatore, la segmentazione e il posizionamento del prodotto-marca

2.1. Il consumatore e il prodotto nei concetti del marketing

Ogni individuo cerca di soddisfare i propri bisogni in varie forme e con vari mezzi. Il Marketing si interessa di una sola forma (il consumo) e di un solo mezzo (il prodotto). I prodotti sono mezzi nel duplice significato del termine: da un lato sono strumentali alla soddisfazione dei bisogni, dall'altro collegano la produzione al consumo.

Lo stato psicologico e sociale che si va ad instaurare con il consumatore trova fondamentale e primaria importanza nella valutazione e nell'analisi di segmentazione e del posizionamento del prodotto.

Ogni individuo si ritrova a vivere in stati d'insoddisfazione ed essi vanno a rappresentare il bisogno che l'individuo cerca, in quanto tale insoddisfazione pone il soggetto in una situazione negativa (che diviene in seguito stimolo).

Per compensare a tale insoddisfazione, l'acquisto di beni rappresenta la risposta orientata alla soddisfazione.

Kotler distingue tra bisogni generici e non influenzabili e desideri specifici e influenzabili: “ i desideri umani sono costituiti dall'individuazione di qualcosa di specifico in grado di soddisfare i bisogni più profondi”. Questo qualcosa viene rappresentato quindi dal prodotto.

I desideri tuttavia sono condizionati dalle variabili dell'ambiente in cui si è inseriti e quindi mutevoli.

Altri studiosi hanno designato con più precisione lo spazio che si va a creare tra bisogni e preferenze, definendo così il bisogno derivato come la

risposta tecnologica particolare (il bene) data al bisogno generico e allo stesso tempo l'oggetto del desiderio. I bisogni derivati costituiscono la base su cui i consumatori sviluppano i desideri ed elaborano un proprio fittizio prodotto. (fig.4)

Nell'ambito della psicologia della personalità umana i bisogni sono stati classificati in altri modi. Lo schema più noto è quello di Maslow (1943 e 1970) (fig.5), che comprende, partendo dal livello più basso:

1. bisogni fisiologici;
2. bisogno di sicurezza;
3. bisogni d'amore e d'appartenenza;
4. bisogno di stima;
5. bisogno di autorealizzazione.

Maslow, afferma che l'individuo attraversa nella sua evoluzione i diversi stadi della gerarchia. La soddisfazione dei bisogni ad un certo livello spinge alla creazione d'altri bisogni più specifici.

A prescindere dal tipo di classificazione, il dato da considerare è la varietà dei bisogni umani e tale dato costituisce il riferimento primario per il Marketing.

Quindi sulla base di queste considerazioni è utile sottolineare come il prodotto possa essere caratterizzato da un insieme di attributi che possono essere tangibili e/o intangibili (come servizi di consulenza, garanzie, installazioni, consegne, ecc..). In particolare luogo (fig.4) è importante notare come il prodotto venga visto come un prodotto atteso in cui si iscrive gli attributi e i benefici che di norma l'acquirente si aspetta di ritrovare (Levitt). Passando ad un livello astratto di prodotto all'offerta concreta dell'impresa questo concetto mette in relazione il bene venduto con le aspettative e preferenze del consumatore.

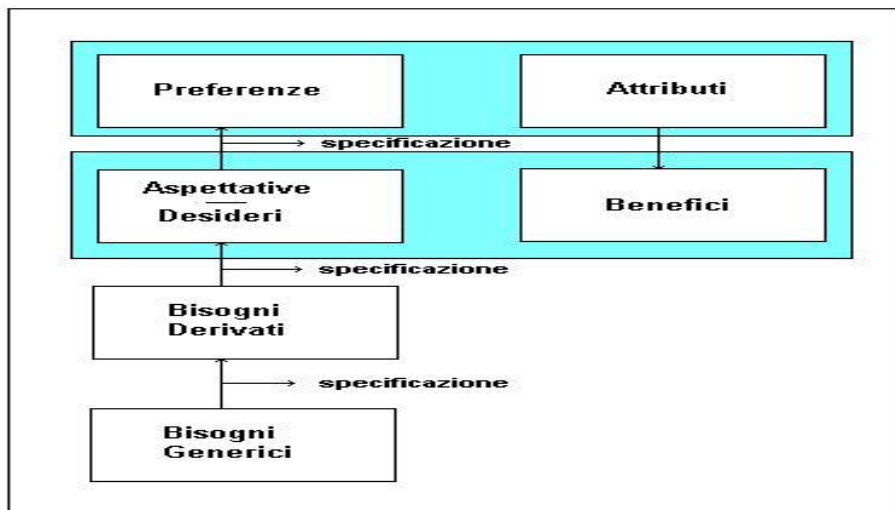


fig. 4 – Relazione tra la domanda e l’offerta

L’ acquisto di un bene che presenta o promette determinati benefici sulla base delle sue caratteristiche dovrebbe in qualche modo esaudire le aspettative e corrispondere alle preferenze ricercate a partire dal bisogno derivato, nel confronto con le alternative di offerta e sotto l’influenza di molteplici variabili. E’ già a tali livelli che la statistica ci permette di sottoporre a studi e analisi tali concetti con metodologie più o meno complesse di analisi multivariata nello studio degli attributi sul prodotto che in seguito verranno analizzate nello studio della segmentazione e posizionamento.

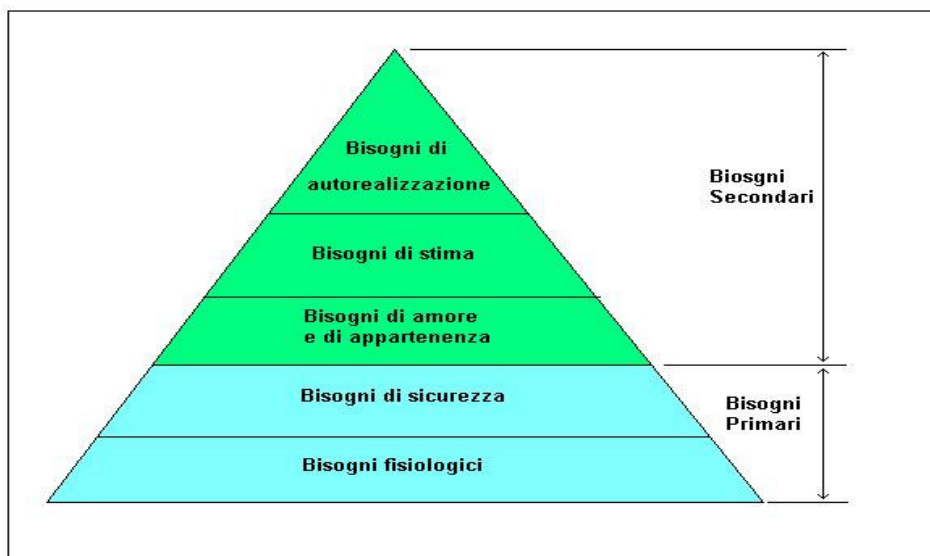


fig. 5 – Piramide di Maslow

2.2. La segmentazione

Alla varietà dal lato di un'offerta si contrappone la varietà dei consumatori e il problema di come andare a rappresentare la varietà dei consumatori che compongono un mirato mercato è il primo che l'impresa deve affrontare nelle sue strategie di mercato.

Per definire in modo rigoroso la segmentazione è necessario partire ad esprimere in modo opportuno la domanda per un determinato prodotto, nell'ipotesi che essa sia funzione del prezzo e degli attributi del prodotto.

$$Q = F(p, x_1, x_2, \dots, x_n)$$

La forma assunta dalla funzione di domanda relativa al prodotto di una singola impresa dipende dalle preferenze dei consumatori, dal reddito pro-capite che si possiede e dalla concorrenza presente nel mercato. Tale funzione di domanda differisce da quella classica dal fatto che in quest'ultima si assumono costanti gli attributi. Le variabili x_1, x_2, \dots, x_n , vengono interpretate come percezione degli attributi del prodotto. Infatti ricordiamo come il prodotto può essere costituito da un insieme di attributi di tipo tangibile e/o intangibile e su tale concetto il consumatore pone un particolare atteggiamento nei confronti del prodotto (o marca).

Un modello da prendere in considerazione per svolgere accurate analisi e studi sull'atteggiamento del consumatore è quello proposto da Fishbein che si basa sull'uso di una funzione matematica applicata all'atteggiamento del consumatore nei confronti di un determinato prodotto:

$$A_j = \sum_n v_{ij} = \sum_n x_{ij} w_i \quad \text{con} \quad \sum_n w_i = 1$$

dove A_j rappresenta l'atteggiamento nei confronti della marca-j, v_{ij} la valutazione dell'attributo i nella marca j , x_{ij} la presenza percepita dell'attributo nella marca e w_i l'importanza assegnata all'attributo. La funzione ha il fine di ipotizzare che l'atteggiamento del consumatore nei

confronti di un'alternativa d'offerta deriva dalla somma delle valutazioni relative agli n attributi presi in esame.

Quindi la funzione di domanda e l'analisi posta sullo studio degli attributi definisce la base teorica per definire i segmenti di mercato. Un concetto relativo alla segmentazione "perfetta" è quello di considerare un'unica funzione di domanda per rappresentare tutti i consumatori di un segmento.

Esistono diverse tipologie di segmentazione, una in particolare è quella proposta da Russel Haley nella quale i consumatori di un segmento si distinguono dal fatto di assegnare la massima importanza relativa ad un determinato beneficio ricercato nel prodotto. Tale segmentazione prende il nome di benefit segmentation (segmentazione per benefici attesi).

A differenza dei classici metodi di segmentazione come quelli per variabili socio-demografici o per variabili geografiche essa presenta caratteristiche molto particolari. Nei tradizionali metodi si cerca di trovare un approccio alla segmentazione di tipo descrittivo o a priori: "si cerca di assumere come ipotesi che le modalità delle variabili prescelte identifichino segmenti di consumatori diversi senza un'indagine sui comportamenti d'acquisto. Esempio nel caso di variabili geografiche che nella segmentazione di una Nazione si assume come base di segmentazione la regione e di conseguenza (esempio Italia) si ottengono subito 20 segmenti ai quali si dovrà solamente determinare l'ampiezza relativa".

La benefit segmentation invece è comportamentale e a posteriori: "si cerca di costruire segmenti di persone che condividono i benefici ricercati nel prodotto e l'approccio quindi diviene selettivo e applicato a mercati di prodotto specifici". Inoltre tale metodo stabilisce collegamenti utili con il Marketing Mix in modo particolare con il prodotto inteso come un insieme di attributi. Altri approcci utili per segmentare il mercato possono essere quello che si basa su tipologie di segmentazione flessibile come analisi psicografica per rappresentare la varietà dei consumatori.

Esso centra l'attenzione sulle caratteristiche personali dei consumatori e riconosce nel modo di vivere l'aggregazione per segmentare il mercato.

In questo caso i consumatori adottano modi di comportarsi simili tra loro, condividendo gli stessi valori ed esprimono opinioni e atteggiamenti omogenei. Tale approccio prende il nome di segmentazione per "stili di vita".

Tuttavia non va dimenticato l'approccio su variabili socio-demografiche, ove l'età rimane ancora un criterio fondamentale per segmentare il mercato dell'abbigliamento e più in generali per tutti quei mercati che sono influenzati dalla moda.

La segmentazione è quindi un importante processo strategico per l'azienda. Infatti una volta svolta l'analisi di segmentazione, ossia come rappresentare sotto certi aspetti la varietà di consumatori, è necessario trovare un orientamento strategico di fondo nei confronti delle condizioni che il mercato presenta. Un'impresa avrà sicuramente un suo adattamento all'interno del mercato che viene caratterizzato a seconda delle strategie prese dall'azienda stessa. Tra le varie tipologie di adattamento trova importanza la strategia di modifica che si distingue per il tipo di cambiamento che l'impresa cerca di indurre nella mappa dei segmenti presente: "si possono sviluppare nuovi segmenti in aggiunta a quelli già presenti, svilupparne uno di nuovo come traslazione di un altro, dividere un segmento in più segmenti, trovare una fusione tra più segmenti".

Non si deve però dimenticare che identificati i segmenti di un mercato-prodotto e definito l'orientamento strategico l'impresa deve qualificare tale strategia scegliendo un accurato livello di copertura, ossia decidere a quanti e a quali segmenti rivolgersi e in seguito come ultimo step si deve posizionarsi in modo adeguato ed efficace nei segmenti prescelti, ricordando che i requisiti base della segmentazione si basano sulla tendenziale uniformità delle risposte degli acquirenti, sulle dimensioni del

segmento, le quali devono permettere il raggiungimento di un certo profitto, e sulla possibilità dell'azienda nell'accedere al segmento tramite gli strumenti di Marketing che ha a disposizione. (fig.6)



fig. 6 – Relazione tra segmentazione e posizionamento

In seguito verranno presentate le tecniche statistiche più note per analizzare e fare segmentazione attraverso tecniche di analisi multivariata, soffermando l'attenzione alle tipologie di tipo flessibile e a posteriori. Tuttavia si ricorda, che nell'ambito della segmentazione a priori esistono tecniche binarie, quali AID e CHAID, per lo studio dei profili dei suoi segmenti.

2.3. Il posizionamento

Individuati i segmenti in cui entrare o mantenere la propria presenza l'impresa deve posizionarsi all'interno di essi e di conseguenza deve sviluppare un appropriato Marketing Mix per ciascuno di essi. Tale concetto risulta intrecciato anche con il concetto di differenziazione del prodotto. Le variabili prodotto, prezzo, comunicazione e distribuzione utili da un lato per raggiungere un vantaggio competitivo, indispensabili per creare un'offerta adeguata al mercato, vanno declinate ai singoli segmenti. Infatti quando la strategia di copertura adottata dall'impresa coinvolge più segmenti, un'attenta gestione del Marketing Mix risulta essenziale affinché

la differenziazione per il raggiungimento del vantaggio competitivo possa ottenere un effettivo riscontro nella mappa percettiva dei consumatori.

Il posizionamento riguarda gli interventi che permettono di definire gli attributi tangibili e intangibili del prodotto (compresa la marca) e tutte quelle azioni che hanno un impatto sul sistema cognitivo del consumatore. In particolare, Kotler intende per posizionamento l'insieme di iniziative volte a definire le caratteristiche del prodotto e a impostare le altre variabili del Marketing Mix al fine di attribuire una certa posizione al prodotto nella mente del consumatore. Tuttavia il problema di posizionamento si gioca nella mappa percettiva del consumatore.

Tra le varie tipologie di posizionamento possiamo ricordare il posizionamento per benefici attesi. Il vantaggio principale di tale approccio permette all'impresa di affrontare in modo unitario e con le medesime tecniche di analisi la rappresentazione dello stato di eterogeneità della domanda e dell'offerta.

Altro tipo di posizionamento è quello basata sul rapporto prezzo/qualità. In realtà questo approccio è una variante del precedente, ma molto più diffuso. Alcune marche possono offrire maggiori prestazioni fisico-funzionali ma ad un prezzo più elevato che ha lo scopo di segnalare la qualità del prodotto al potenziale acquirente, oppure anche porre l'accento sulla qualità elevata offerta a parità di prezzo. Quest'ultima può risultare coerente sia con una leadership di costo sia sulla differenziazione di prodotto.

Altre tipologie di posizionamento si basano sull'associare il prodotto al suo utilizzo, a uno stato d'animo, o a qualche altro tratto caratteristico degli utilizzatori che compongono il segmento target.

La scelta di un criterio dipende comunque da come è stata concepita la strategia di posizionamento. Tuttavia si deve sottolineare che oltre alla presenza di un'identificazione di un'immagine da far percepire all'acquirente è da tenere in considerazione il caso quando il posizionamento

deve sostenere una strategia di differenziazione fortemente centrata sul valore simbolico del prodotto. In questa situazione le associazioni mirate alla sfera emotiva prevalgono sulla comunicazione degli attributi fisici dei prodotti e delle loro prestazioni.

Concludendo, si evidenzia come il posizionamento chiude il processo strategico di segmentazione e a sua volta rappresenta un processo proiettato nel tempo. Ogni impresa deve costantemente effettuare controlli sul suo posizionamento di prodotto nella mappa percettiva dei consumatori presa in considerazione. Sul piano strategico, il problema da risolvere è quello di trovare una posizione soddisfacente per il proprio prodotto (o marca) e di individuare gli strumenti (anche politiche di mercato) in modo che essi permettano di raggiungerla e di difenderla da possibili attacchi dei concorrenti. (fig.7)

Esempio di posizionamento competitivo di marche in base ai benefici percepiti e di segmentazione di mercato fondata sulla valutazione degli stessi benefici.

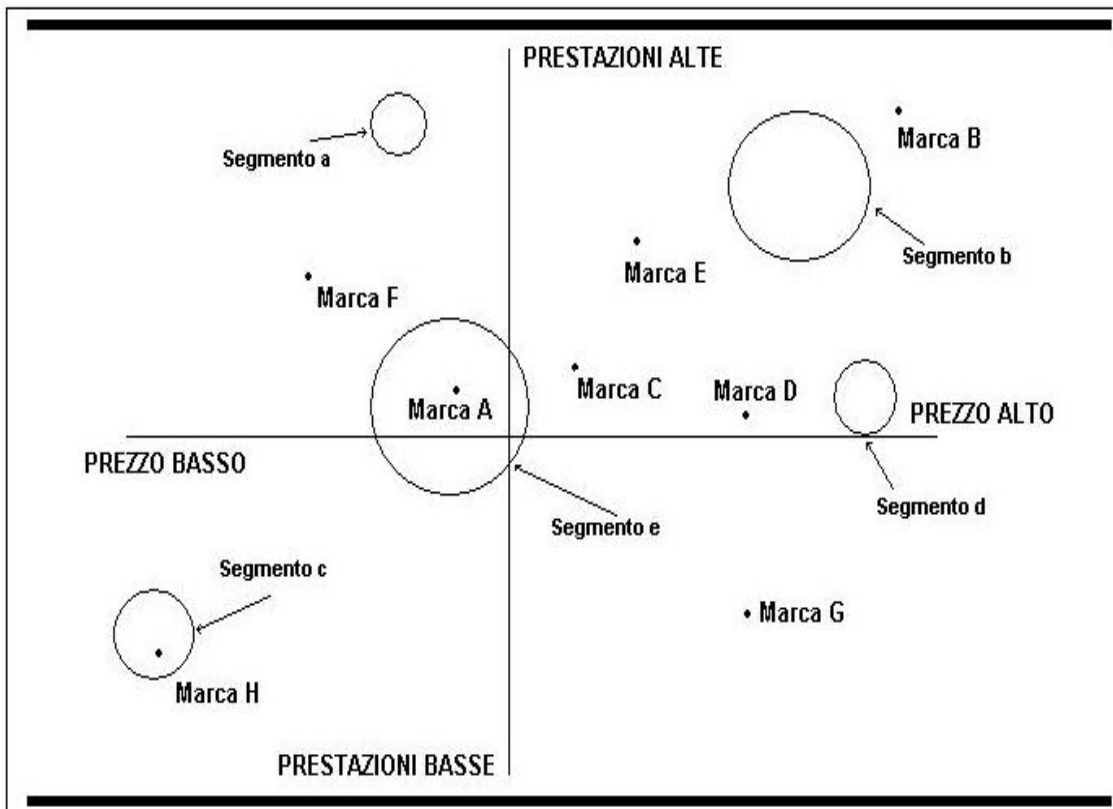


fig. 7 – Mappa di percezione

Per comprendere meglio l'attività competitiva è importante quindi avere elementi informativi sul dinamismo del mercato, sulle caratteristiche dei consumatori e sulla posizione dei prodotti nella mente di essi. Le mappe di percezione hanno quest'ultimo fine e forniscono un contributo per capire la struttura competitiva del mercato e per creare un'appropriata offerta (Marketing Mix). Tali mappe contengono le marche (o prodotti) posizionate in un certo modo all'interno della mappa stessa a seconda di come i consumatori ritengono simili (o non) le alternative dei prodotti in base anche a preferenze sulle stesse o sui attributi che caratterizzano il prodotto medesimo. In tal modo si identificano gli spazi vuoti, o aree disponibili al lancio di nuovi prodotti, oppure si possono tracciare i confini della propria area competitiva e riformulare riposizionamenti, o in fine verificare se il lancio del nuovo prodotto soddisfa le attese del consumatore nel segmento a cui è destinato.

Bibliografia

- Russell S. Winer, "Marketing Management", Apogeo Education, 2000.
- Grandinetti R., "Concetti e Strumenti di Marketing", Etas, 2002.
- Cozzi G., Ferrero G., "Marketing: principi, metodi, tendenze evolutive", G. Giappichelli Editore Torino, 2000.
- Brasini S., Tassinari F., Tassinari G., "Marketing e pubblicità: Metodi di analisi statistica", Il Mulino, 1999
- Cozzi G., Ferrero G., "Le applicazioni del marketing nelle imprese italiane", G. Giappichelli Editore Torino, 2000.
- Canarini F., "La segmentazione del mercato dei beni di consumo", Cedam Padova, 1987
- Caldani E., "Definizione e segmentazione del mercato per i beni industriali e di largo consumo", Giuffrè Milano, 1984
- Kotler P., "Marketing Management, 7ª Ed.", Isedi Torino, 1993
- Dispense della prof. M.Pertile relative al corso di Marketing e del prof. R.Grandinetti del corso di Economia e Gestione delle Imprese II (rif: www.statistica.unipd.it).

2. LA STATISTICA NEL MARKETING:

**“METODI DI ANALISI STATISTICA PER
LA SEGMENTAZIONE E IL
POSIZIONAMENTO DI PRODOTTO”**

3. Analisi Multivariata nella segmentazione e posizionamento di prodotto-marca

Il seguente capitolo è dedicato alla presentazione dei metodi statistici più usati e conosciuti relativi alla macro-area della statistica multivariata.

In particolare il capitolo sarà suddiviso in tre paragrafi, nei quali verrà fatto cenno ad alcune tecniche descrittive per la segmentazione, a tecniche complesse sia per essa che per il posizionamento e a tecniche specifiche per il solo posizionamento di prodotto (o marca). Saranno approfondite invece le tecniche di Multidimensional Scaling e Unfolding per il posizionamento e la creazione di mappe percettive, la Conjoint Analysis e la Cluster Analysis per la segmentazione di mercato. (fig.8)

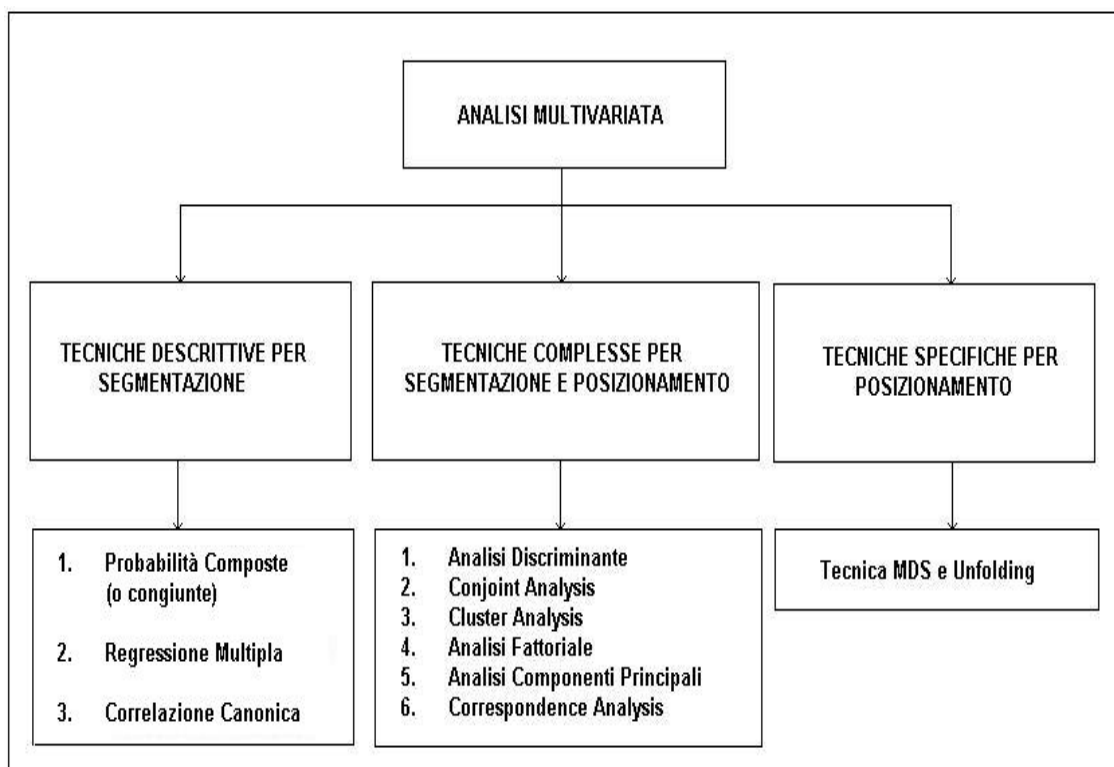


fig. 8 – Suddivisione delle tecniche di analisi multivariata

3.1. Tecniche descrittive per la segmentazione

Tra le tecniche di tipo descrittivo specifiche per la segmentazione ricordiamo la tecnica della probabilità congiunta, nonché le tecniche di correlazione canonica e regressione multipla. In questo paragrafo si descriverà con brevi cenni tali tecniche.

3.1.1. Probabilità composte o congiunte

Tale metodo viene impiegato principalmente per determinare la rilevanza di vari segmenti socio-demografico nei confronti del consumo di uno specifico prodotto e per calcolare l'ampiezza numerica del target group.

Partendo dai dati della struttura della popolazione riguardanti una serie di caratteristiche descrittive si va a considerare un certo universo di riferimento, ad esempio la popolazione femminile, e si va ad analizzare la composizione secondo una serie di parametri oggettivi identificabili per esempio in età, reddito, ripartizione territoriale e altro.

In sintesi disponendo di una fonte di dati che riportano i valori di composizione di una popolazione campione è possibile calcolare la probabilità che una donna italiana si trovi al Nord (per esempio 0.40) e la probabilità che essa abbia un'età compresa tra i 18-24 anni (ad esempio 0.30), in seguito risulta facile determinare la probabilità che una donna di 18-24 anni si trovi nel Nord ($0.40 \cdot 0.30 = 0.12$). La probabilità finale è una probabilità congiunta determinata dalla probabilità che si verifichi sia l'evento x che l'evento y: $P(X=x, Y=y) = P(X=x) \cdot P(Y=y|X=x)$ (nell'ipotesi di indipendenza tra le v.c. X e Y si ha $P(X=x, Y=y) = P(X=x) \cdot P(Y=y)$)

Va inoltre sottolineato che le sottoclassi in cui è suddiviso ciascun parametro possono essere considerate come elementi disgiunti in cui è possibile calcolare la percentuale di composizione sull'universo di riferimento e la probabilità corrispondente.

3.1.2. Regressione Multipla

Essa viene impiegata per prevedere i valori di una variabile dipendente attraverso i valori di più variabili indipendenti, ad esempio le vendite di un prodotto o le leve del Marketing Mix.

Nel campo della segmentazione della domanda può essere utile per stabilire quale variabile indipendente sia più utile per determinare la variabile dipendente.

Pur essendo considerata una tecnica funzionale, quindi legata alla previsione di un fenomeno, essa può essere impiegata nelle fasi preliminari della segmentazione di mercato quando non sono chiari i legami tra un fenomeno e le variabili indipendenti usate per spiegarlo.

Definiamo un modello lineare una scrittura matematica del seguente tipo:

$$y_i = f(x_i; \beta) + \varepsilon_i \quad [3.1]$$

dove $f(x_i; \beta)$ è una funzione matematica di n variabili esplicative $x_i = (x_{i1}, x_{i2}, x_{i3}, \dots, x_{in})'$ e di m parametri ignoti $\beta = (\beta_1, \beta_2, \dots, \beta_m)'$, mentre ε_i rappresenta gli errori rispetto alla risposta sistematica e si assume che abbiano distribuzione marginale $N(0, \sigma^2)$ e siano indipendenti. In termini matriciali la funzione può essere scritta come $Y = X\beta + \varepsilon$.

Dalle proprietà del modello [3.1] si può definire, dato un i fissato, che y_i è una variabile casuale con media e varianza rispettivamente uguali a:

$$E[y_i] = f(x_i; \beta); \quad \text{VAR}[y_i] = \sigma_\varepsilon^2$$

Una volta definite tali osservazioni il problema che si presenta è quello di stimare il vettore β dei parametri in modo che la $f(x_i; \beta)$ sia il più vicino al valore della y_i . Il criterio maggiormente adottato è quello dei minimi quadrati. Ossia si tratta di trovare la stima del vettore dei parametri β in modo da minimizzare la funzione:

$$S(\beta) = \sum_i [y_i - f(x_i; \beta)]^2 \text{ con } i = 1 \dots N \quad [3.2]$$

Considerando il modello di regressione $y_i = \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_p x_{ip} + \varepsilon_i$ in termini matriciali $Y = X\beta + \varepsilon$ e assumendo il fatto che X è una matrice non stocastica $n \times p$ e il suo rango è pari a p con $p < n$ e i termini di disturbo hanno media zero e varianza costante e finita la funzione [3.2] diventa:

$$S(\beta) = (y - X\beta)'(y - X\beta) = y'y - 2y'X\beta + \beta'X'X\beta$$

E derivando $S(\beta)$ rispetto a β e uguagliandola a zero, nell'ipotesi che tale funzione sia differenziabile si ottiene la seguente equazione: $X'X\beta = X'y$.

Lo stimatore dei minimi quadrati risulta ad essere:

$$\hat{\beta} = (X'X)^{-1}X'y$$

Importante è la verifica delle condizioni precedentemente citate in merito al fattore di disturbo. Una sua stima può essere data dalla relazione $e_i = y_i - \hat{y}_i$. Essi vengono definiti anche residui e possono essere espressi in funzione degli ε_i :

$$\begin{aligned} e &= y - \hat{y} = y - X\hat{\beta} = y - X(X'X)^{-1}X'y = [I_n - X(X'X)^{-1}X']y = \\ &= [I_n - X(X'X)^{-1}X'](X\beta + \varepsilon) = [I_n - X(X'X)^{-1}X']\varepsilon \end{aligned}$$

Tali residui risultano ortogonali rispetto alle variabili indipendenti:

$$X'e = X'y - X'X(X'X)^{-1}X'y = 0$$

e usando tale equazione si ottiene che:

$$y'y = (\hat{y} + e)'(\hat{y} + e) \text{ da cui si ricava che } e'e = y'y - (\hat{\beta})'X'\hat{y}$$

L' espressione corrisponde ad una scrittura del seguente tipo:

$$\sum_i (y_i - \bar{y})^2 = \sum_i (\hat{y}_i - \bar{y})^2 + \sum_i e_i^2 \quad \text{con } i = 1 \dots n \text{ e } \bar{y} \text{ è l'intercetta del modello.}$$

Il termine a primo membro rappresenta la devianza totale (TSS) il primo termine a secondo membro rappresenta la devianza spiegata (ESS) e l'ultimo termine la devianza residua (RSS).

Questo ci permette di determinare una misura della bontà della regressione definita dal coefficiente di determinazione multipla:

$$R^2 = ESS / TSS = 1 - (RSS / TSS)$$

Si può considerare una versione corretta che tiene conto del numero di variabili indipendenti

$$\bar{R}^2 = 1 - [(RSS/(n - p)) / (TSS/(n - 1))]$$

Infine lo stimatore di σ_ε^2 risulta ad essere $s^2 = e'e / (n - p)$.

Nei casi più semplici la stima dei parametri non risulta difficoltosa, per esempio in un modello di regressione lineare del tipo $y_i = \beta_1 + \beta_2 x_i$ attraverso la metodologia precedentemente descritta si perviene ai seguenti valori dei parametri:

$$\hat{\beta}_1 = (\sum_i y_i) / n - (\sum_i x_i \hat{\beta}_2) / n$$

$$\hat{\beta}_2 = [(\sum_i x_i y_i - (\sum_i x_i \sum_i y_i) / n) / (\sum_i x_i^2 - (\sum_i x_i)^2 / n)]$$

Esempio:

Tale esempio è preso dal testo “Introduzione alla statistica”, di L.Pace e A. Salvan, CEDAM, 2001.

E' stato sottoposto ad abrasione costante e per una certa durata fissata un campione di 30 elementi di un certo tipo di gomma. L'obiettivo è quello di valutare quanto la resistenza di abrasione della gomma è influenzata dalla durezza della gomma stessa e dalla sua resistenza alla trazione. E'un esempio che può essere visto nel concetto di individuare gli attributi che meglio spiegano un certo particolare funzionale del mio prodotto (gomma). Si ricorda come il prodotto può essere inteso come un insieme di attributi e come i consumatori ricerchino in esso certe qualità (preferenze) nell'obiettivo di soddisfare i bisogni.

In questo esempio si individua un modello di regressione che permetta di descrivere l'abrasione (Y_i) in funzione delle caratteristiche strutturali del prodotto gomma.

Ipotizzando che $x_{i2} = D_i$ e $x_{i3} = R_i$ con D_i l'attributo durezza e R_i l'attributo resistenza alla trazione si ottiene un modello del tipo:

$$Y_i = \beta_1 + \beta_2 x_{i2} + \beta_3 x_{i3} + \varepsilon_i$$

Il modello ipotizzato è un modello lineare normale e usando le formule sopra descritte con i dati riportati nella seguente tabella (tab.2):

y_i	372	206	175	154	136	112	55	45	221	166
D_i	45	55	61	66	71	71	81	86	53	60
R_i	162	233	232	231	231	237	224	219	203	189

y_i	164	113	82	32	228	196	128	97	64	249
D_i	64	68	79	81	56	68	75	83	88	59
R_i	210	210	196	180	200	173	188	161	119	161

y_i	219	186	155	114	341	340	283	267	215	148
D_i	71	80	82	89	51	59	65	74	81	86
R_i	151	165	151	128	161	146	148	144	134	127

tab. 2 - y_i = perdita di peso dovuta all' abrasione in grammi/ora , D_i = durezza in gradi Shore, R_i = resistenza in Kg/cm²

si ottengono i seguenti valori:

$$\begin{array}{c}
 (X^T X)^{-1} \\
 \left| \begin{array}{ccc}
 30 & 2108 & 5414 \\
 2108 & 152422 & 376562 \\
 5414 & 376562 & 1015780
 \end{array} \right| \\
 \\
 X^T y \\
 \left| \begin{array}{c}
 5263 \\
 346867 \\
 921939
 \end{array} \right|
 \end{array}$$

da cui la stima dei parametri è $\hat{\beta}^T = (885.16, -6.571, -1.374)$ e il modello di regressione che ne deriva è: $\hat{y}_i = 885.16 - 6.571x_{i2} - 1.374x_{i3}$ con una stima della varianza pari a: $\hat{\sigma}^2 = \sum(y_i - \hat{y}_i)^2 / 30 = 1198.325$

3.1.3. Correlazione canonica

La correlazione canonica presenta forti affinità con la regressione multipla. Essa infatti cerca di studiare il legame che esiste tra più variabili dipendenti a quelle indipendenti.

Diviene utili nei casi in cui si voglia collegare la fedeltà alle diverse marche di prodotto presenti in un mercato ai benefici ricercati dai consumatori nei prodotti (o marche) medesimi.

Il processo di determinazione delle componenti canoniche non è banale e in questo paragrafo ne faremo solo un cenno di come esse vengono determinate. In modo particolare imposteremo i sistemi matematici necessari per determinare il primo stadio delle componenti canoniche.

Si supponga che su un collettivo di unità statistiche si siano osservati due gruppi di k e m variabili corrispondenti ad altrettanti aspetti di un fenomeno (o prodotto) in esame. L'obiettivo dell'analisi della correlazione canonica è quello di studiare le relazioni di interdipendenza tra i due gruppi di variabili. Si considerino le due variabili X e Y rispettivamente di dimensione k e m che osservate su un insieme di n elementi danno origine alle matrici di dati

$$\mathbf{X} = \begin{pmatrix} x_{11} & \cdots & x_{1k} \\ \vdots & & \\ x_{n1} & \cdots & x_{nk} \end{pmatrix}, \quad \mathbf{Y} = \begin{pmatrix} y_{11} & \cdots & y_{1k} \\ \vdots & & \\ y_{n1} & \cdots & y_{nk} \end{pmatrix}$$

e caratterizzate da matrici di varianza e covarianza campionarie del tipo:

$$S_x = [(\mathbf{X} - \mu_n \mathbf{X}')'(\mathbf{X} - \mu_n \mathbf{X}')] / n$$

$$S_y = [(\mathbf{Y} - \mu_n \mathbf{Y}')'(\mathbf{Y} - \mu_n \mathbf{Y}')] / n$$

$$S_{xy} = [(\mathbf{X} - \mu_n \mathbf{X}')'(\mathbf{Y} - \mu_n \mathbf{Y}')] / n$$

Si definisce inoltre i vettori \mathbf{V} e \mathbf{U} di n -dimensioni come combinazione lineare delle colonne delle matrici \mathbf{X} e \mathbf{Y} : $\mathbf{U} = \mathbf{X}\mathbf{a}$ e $\mathbf{V} = \mathbf{Y}\mathbf{b}$ con \mathbf{a} e \mathbf{b} vettori di k e m costanti per la quale vale la scrittura $\mathbf{U} = \mathbf{a}'\mathbf{X}$ e $\mathbf{V} = \mathbf{b}'\mathbf{Y}$.

Il coefficiente di correlazione esistente tra \mathbf{U} e \mathbf{V} viene definito dalla seguente formula:

$$\rho_{uv} = \frac{\mathbf{a}' S_{xy} \mathbf{b}}{\sqrt{\mathbf{a}' S_x \mathbf{a} \mathbf{b}' S_y \mathbf{b}}}$$

I vettori \mathbf{a} e \mathbf{b} sono definiti in modo da massimizzare il coefficiente di correlazione e poiché esso è invariante per trasformazioni lineari delle variabili, ciò implica l'introduzione di alcuni vincoli per l'unitarietà delle varianze associate alle variabili \mathbf{U} e \mathbf{V} , ossia si deve risolvere un problema di massimo vincolato come riportato nel seguente sistema:

$$\begin{cases} \max_{\mathbf{a}} \mathbf{a}' S_{xy} \mathbf{b} \\ \max_{\mathbf{b}} \mathbf{a}' S_{xy} \mathbf{b} \\ \mathbf{a}' S_x \mathbf{a} = 1 \\ \mathbf{b}' S_y \mathbf{b} = 1 \end{cases}$$

A questo punto si introduce la funzione di Lagrange con $v/2$ e $\eta/2$ i

moltiplicatori associati: $L(a,b,v,\eta) = a'S_{XY}b - \frac{v}{2}(a'S_Xa - 1) - \frac{\eta}{2}(b'S_Yb - 1)$.

Impostando il sistema classico associato alla funzione di Lagrange, ossia impostando il sistema con le derivate prime fatte rispetto ad a,b,v,η ed eguagliandole a zero si perviene ad un sistema (qui già ridotto) del seguente tipo:

$$\begin{cases} S_X^{-1}S_{XY}S_Y^{-1}S_{XY}a = v^2a \\ S_Y^{-1}S_{YX}S_X^{-1}S_{XY}b = v^2b \\ a'S_Xa = 1 \\ b'S_Yb = 1 \end{cases}$$

Osservando le prime due equazioni si ricava che $v^2 = \lambda$ è un autovalore di $E1 = S_X^{-1}S_{XY}S_Y^{-1}S_{XY}$ e di $E2 = S_Y^{-1}S_{YX}S_X^{-1}S_{XY}$ e che a e b sono gli autovettori corrispondenti in modo che siano vere le altre due equazioni del sistema.

Data la covarianza di U e V : $COV(U,V) = a'S_{XY}b = v = \sqrt{\lambda}$, gli autovettori a_1 e b_1 che corrispondono al più grande degli autovalori di $E1$ (o $E2$) indicato con λ_1 individuano le combinazioni delle colonne di X e Y che risultano in tal modo massimizzate nella loro correlazione.

Si definisce così prima correlazione canonica la quantità:

$$\rho_{U_1V_1} = \sqrt{\lambda_1}$$

A tale punto il meccanismo di determinazione delle altre correlazioni canoniche diventa ricorsivo fino al raggiungimento dell'ultima correlazione canonica ammessa.

Quindi le coppie delle componenti canoniche successive alle prime saranno $U_2 = Xa_2$ e $V_2 = Yb_2$ da cui si riprende il ciclo sopra citato tramite la funzione di Lagrange, e così via fino alla $U_n = Xa_n$ e $V_n = Yb_n$.

3.2. Tecniche complesse per segmentazione e posizionamento

3.2.1. Analisi discriminante

Tale metodo consente di risolvere due problemi:

a1) l'esame delle differenze che esistono tra gruppi di individui (nelle ricerche di segmentazione) al fine di trovare quali variabili riescono a discriminarli maggiormente.

a2) la ricerca di funzioni di classificazione che consentono di inserire nuovi individui all'interno dei gruppi che sono già definiti a priori.

Quindi l'analisi discriminante viene condotta per definire varie modalità di assegnazione di casi a differenti gruppi in funzione di una serie di variabili tra loro correlate, e l'interesse primario è rivolto a definire un modello che consente di assegnare un nuovo individuo ad un gruppo predefinito in funzione di un certo n di variabili.

Il modello matematico che descrive tutto ciò è rappresentato da una legge matematica che combina in modo lineare una serie di variabili ipotizzate tra loro indipendenti e tale combinazione permette di calcolare il coefficiente di discriminazione D:

$$D = b_0 + b_1 x_1 + \dots + b_n x_n$$

Dove D = coefficiente di discriminazione, b_0, \dots, b_n i coefficiente della funzione di discriminazione e le x_1, \dots, x_n le n variabili indipendenti che compongono la funzione sopra definita.

Il metodo di calcolo che si adotta è quello già visto per la regressione multipla, quello dei minimi quadrati. Per semplicità in questo paragrafo si considera il caso della discriminazione fra due gruppi, ma le considerazioni sono estendibili a k gruppi che daranno origine a k-1 funzioni discriminanti. Dati due gruppi si può quindi ottenere solo una funzione discriminante i cui

coefficienti si ottengono dalla seguente equazione: $b = \Delta x W$ con

$$b = (b_0, \dots, b_n)' ,$$

$$\Delta X = ((\bar{X}_{1A} - \bar{X}_{1B}), \dots, (\bar{X}_{nA} - \bar{X}_{nB}))' ,$$

W = matrice di dispersione comune,

\bar{X}_{nA} = media della n-ma variabile del gruppo A,

\bar{X}_{nB} = media della n-ma variabile del gruppo B.

I valori medi dei coefficienti di discriminazione per i due gruppi sono uguali a: $\bar{D}_A = b_0 + b_1 \bar{X}_{1A} + \dots + b_n \bar{X}_{nA}$ e $\bar{D}_B = b_0 + b_1 \bar{X}_{1B} + \dots + b_n \bar{X}_{nB}$.

Le varianze del coefficiente di discriminazione per i due gruppi invece sono pari a: $S_{DA}^2 = b X_A' X_A b$ e $S_{DB}^2 = b X_B' X_B b$.

La combinazione $X_A' X_A$ rappresenta la matrice di dispersione del relativo gruppo A, analogamente per la combinazione $X_B' X_B$ al gruppo B.

La soglia discriminante deve essere tanto più vicina possibile ad una delle due medie tanto minore è la deviazione standard del gruppo corrispondente, e tale soglia viene calcolata dalla seguente formula:

$$D_0 = \frac{\bar{D}_A S_{DB} + \bar{D}_B S_{DA}}{S_{DA} + S_{DB}}$$

Pertanto un nuovo elemento (individuo nelle ricerche di segmentazione) viene assegnato al gruppo in funzione della soglia discriminante. Se il valore di D è superiore alla soglia viene assegnato al gruppo con la media di D più alta, viceversa se il valore è più piccolo. Si può assegnare una probabilità all'assegnamento di un elemento ad un gruppo.

Tale probabilità può essere determinata in funzione dello scarto standardizzato tra la media di un gruppo e il valore della soglia discriminante.

Quindi, per il gruppo A la probabilità è:

$$P(Z_A) = \text{Tavole } f(Z) \text{ con } Z_A = \frac{D_0 - \bar{D}_A}{S_{DA}}$$

Analogamente per il gruppo B, la probabilità è:

$$P(Z_B) = \text{Tavole } f(Z) \text{ con } Z_B = \frac{D_0 - \bar{D}_B}{S_{DB}}$$

3.2.2. Analisi fattoriale

L'obiettivo principale è quello di sintetizzare le relazioni esistenti tra un collettivo di variabili che sono sottoposte ad analisi, le quali possono essere espresse attraverso una serie di dimensione che risultano essere a sua volta combinazioni lineari delle variabili di partenza. Ulteriore obiettivo è quello di individuare una o più strutture che possano in qualche modo combinare o aggregare fenomeni diversi sulla base di caratteristiche comuni.

In questo paragrafo si farà cenno solamente al modello fattoriale che costituisce la base della seguente analisi, tralasciando le metodologie di stima e rappresentazioni grafiche. Il modello di tale analisi è definito, data una variabile X aleatoria di k -dimensioni, nel seguente modo:

$$X = \mu + \Lambda F + U$$

dove Λ viene definita la matrice di $k \times m$ dimensioni di costanti dette pesi fattoriali, F è il vettore dei fattori comuni di m -dimensioni e U quello di k -dimensioni dei fattori specifici.

Si assume inoltre che:

$$E(F) = 0 ,$$

$$\text{COV}(F) = E(FF') = I_m ,$$

$$E(U) = 0 ,$$

$$\text{COV}(U) = E(UU') = \text{diag}(\Psi_1, \dots, \Psi_k),$$

$$E(FU') = 0 ;$$

Ossia si assume che i fattori comuni abbiano media zero e varianza unitaria e che essi siano incorrelati tra loro, mentre per quanto riguarda i fattori specifici usati si assume lo stesso che media sia pari a zero e varianza unitaria ma che essi siano incorrelati tra loro e anche con i fattori comuni.

Supponiamo per semplicità che la $\mu = 0$ ottenendo così un modello ridotto del tipo: $X = \Lambda F + U$ e indicando con λ_j la j -esima riga della matrice Λ si può scrivere $X_j = \lambda_j F + U_j$.

Da quest'ultima osservazione si può definire la varianza di ciascuna componente attraverso la formula:

$$\text{VAR}(X_j) = \sum_{h=1}^m \lambda_{jh}^2 + \psi_j$$

dove la quantità $\sum_{h=1}^m \lambda_{jh}^2$ detta comunalità, rappresenta la quota della varianza di X_j spiegata dai fattori comuni, mentre la quantità ψ_j è denominata specificità ed è la parte residua della varianza di X_j non spiegata dai fattori comuni.

Inoltre, nell'ipotesi precedentemente assunta, quella che la media sia zero, la covarianza tra due componenti X_j e X_i è pari a:

$$\begin{aligned} E(X_j X_i) &= E[(\lambda_j F + U_j)(\lambda_i F + U_i)'] \\ &= \lambda_j E(F F') \lambda_i' + \lambda_j E(F U_i') + E(U_j F') \lambda_i' + E(U_j U_i') \\ &= \lambda_j \lambda_i' \end{aligned}$$

Concludendo dato il modello fattoriale precedentemente descritto la covarianza tra le due componenti X_j e X_i è spiegata dai fattori comuni e nel complessivo del modello essa risulta ad essere pari a:

$$\begin{aligned}
E(\mathbf{X}\mathbf{X}') &= E[(\Lambda\mathbf{F} + \mathbf{U})(\Lambda\mathbf{F} + \mathbf{U})'] \\
&= \Lambda E(\mathbf{F}\mathbf{F}')\Lambda + \Lambda E(\mathbf{F}\mathbf{U}') + E(\mathbf{U}\mathbf{F}')\Lambda + E(\mathbf{U}\mathbf{U}') \\
&= \Lambda\Lambda' + \Psi
\end{aligned}$$

dove $E(\mathbf{F}\mathbf{F}') = \mathbf{I}_m$ e $E(\mathbf{F}\mathbf{U}') = E(\mathbf{U}\mathbf{F}') = 0$ e $E(\mathbf{U}\mathbf{U}') = \Psi$. Inoltre i pesi che necessita il modello coincidono con le covarianze tra i fattori e le variabili del modello:

$$\text{COV}(\mathbf{X}, \mathbf{F}) = E(\mathbf{X}\mathbf{F}') = E[(\Lambda\mathbf{F} + \mathbf{U})\mathbf{F}'] = \Lambda.$$

3.2.3. Analisi delle componenti principali (ordinali)

L'obiettivo di tale metodo, consiste nell'individuare un insieme ridotto di componenti, rispetto un particolare collettivo di variabili d'origine.

Esso è considerato un metodo di analisi multivariata che ha il fine di ridurre il più possibile le dimensioni dei dati multivariati presi in considerazione e di estrapolare una sintesi delle informazioni in modo da rendere più agevole lo studio del fenomeno in questione.

Un esempio di applicazione di tale metodo può essere riferito allo studio delle valutazioni da parte di consumatori relativamente a un particolare prodotto (o il servizio che dovrebbe offrire) in cui ci si può riferire alle seguenti classificazioni:

1. = per niente soddisfatto
2. = poco soddisfatto
3. = mediamente soddisfatto
4. = soddisfatto
5. = molto soddisfatto

Se a tali punteggi associati al tipo di soddisfazione, viene attribuito anche un concetto di equidistanza tra due categorie contigue, la soddisfazione dei consumatori verrà valutata come variabile quantitativa.

Poiché l'assegnazione del punteggio è piuttosto arbitraria ciò comporta che l'operazione di quantificazione non sempre è generalizzabile. In questo paragrafo, quindi, si cercherà di impostare il metodo delle componenti principali mantenendo la natura ordinale delle variabili prese in questione.

Si definisce un vettore casuale X di m componenti: $X = (x_1, x_2, \dots, x_m)'$ avente una sua distribuzione con una sua opportuna media: $\mu = (\mu_1, \dots, \mu_m)'$ e matrice di varianza e covarianza:

$$\Sigma = \begin{pmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1m} \\ \sigma_{21} & \sigma_{22} & \cdots & \sigma_{2m} \\ \vdots & & & \\ \sigma_{m1} & \sigma_{m2} & \cdots & \sigma_{mm} \end{pmatrix}$$

A questo punto si standardizza il vettore X nel vettore Z attraverso la trasformazione $z_1 = (x_1 - \mu_1)/\sigma_1$ (lo stesso per tutte le altre $m-1$ variabili).

La nuova matrice di varianza e covarianza sarà:

$$P = \begin{pmatrix} 1 & \rho_{12} & \cdots & \rho_{1m} \\ \rho_{21} & 1 & \cdots & \rho_{2m} \\ \vdots & & & \\ \rho_{m1} & \rho_{m2} & \cdots & 1 \end{pmatrix}$$

$$\text{Dove } \rho_{ij} = \frac{\text{COV}(x_i, x_j)}{\sigma_i \sigma_j}.$$

Considerando ora la variabile casuale, combinazione lineare delle variabili standardizzate $Y_1 = a_1'Z$, si deve massimizzare la sua varianza $V(Y_1) = (a_1')Pa_1$ sotto il vincolo $a_1'a_1 = 1$.

Si risolve tale problema massimizzando un sistema Lagrangiano e poiché il sistema ammette come soluzione il vettore nullo, la condizione per cui si ottengono soluzioni diverse dal vettore nullo è: $|P - \lambda I_m| = 0_m$.

$$L(a, \lambda) = a_1' P a_1 - \lambda (a_1' a_1 - 1)$$

$$\frac{\partial L(a, \lambda)}{\partial a} = 2(P a_1 - \lambda a) = 2(P - \lambda I_m) a_1 = 0_m$$

$$(P - \lambda I_m) a_1 = 0_m$$

Per stabilire quale autovalore fornisce il massimo della funzione, si moltiplica per a_1' i membri della equazione $P a_1 = \lambda a_1$: $a_1' P a_1 = a_1' \lambda a_1$.

Essendo valido il vincolo di normalizzazione $a_1' a_1 = 1$ la $a_1' P a_1 = a_1' \lambda a_1$ può essere scritto come $V(Y_1) = \lambda$, e l'obiettivo è quello di massimizzare quest'ultima scrittura.

Tale soluzione si ricava in corrispondenza del massimo autovalore λ_1 della matrice di varianza e covarianza di Z .

In conclusione la prima componente principale delle variabili standardizzate è la combinazione lineare $Y_1 = a_{11} Z_1 + a_{21} Z_2 + \dots + a_{m1} Z_m$.

La seconda componente principale Y_2 si ottiene sempre come combinazione lineare delle variabili standardizzate in cui viene massimizzata la $V(Y_2) = (a_2') P a_2$, sotto il vincolo di normalizzazione $a_2' a_2 = 1$ e d'incorrelazione con la precedente componente principale $(a_2') P a_1 = 0$, e impostando il seguente problema di massimo vincolato:

$$L(a, \lambda, \delta) = a_2' P a_2 - 2\lambda (a_2' a_2 - 1) + \lambda - \delta a_2' P a_1$$

$$\frac{\partial L(a, \lambda, \delta)}{\partial a} = 2P a_2 - 2\lambda a_2 - \delta P a_1 = 0$$

La soluzione che risulterà è fornita dall'autovettore di P corrispondente al secondo autovalore. Questo procedimento diviene così ricorsivo fino alla determinazione della Y_m componente principale.

3.2.4. Analisi delle corrispondenze (Correspondence Analysis)

Tale metodo viene impiegato per descrivere, rappresentare graficamente

e sintetizzare le righe e le rispettive colonne di una matrice delle informazioni di tipo categorico in un unico spazio bi-dimensionale o al più tri-dimensionale permettendo così una migliore lettura e interpretazione delle informazioni raccolte in merito ad un particolare fenomeno oggetto di studio e di ricerca di mercato.

Prendiamo in esame due variabili qualitative X e Y rispettivamente con k e h modalità. Tali modalità possono essere rappresentate in una tabella di contingenza, definendo: n_{ij} il numero di unità con il risultato i-esimo per la prima rilevazione rappresentata dalla variabile X e j-esimo risultato della seconda rilevazione (Y), $f_{ij} = n_{ij} / n$ le frequenze relative e f_{i*} e f_{*j} le frequenze cumulative marginali della variabile X e Y. Dividendo le f_{ij} per la frequenza cumulative corrispondente come mostra la tab. 3, si ottengono i profili riga o le ripartizioni percentuali all'interno della riga. (Esso rappresenta un peso relativo)

	1	2	3	...	j	...	h	marg.x
1								1
2								1
3								1
...								1
i					f_{ij} / f_{i*}			1
...								1
k								1
marg.y					f_{*j}			1

tab. 3 – Matrice delle frequenze

Analogamente, tali considerazioni si riflettono anche per le colonne, ove le ripartizioni percentuali all'interno delle colonne prendono il nome di profili colonna.

Scopo di tale analisi è quello di studiare i risultati delle due rilevazioni e di evidenziare le eventuali dipendenze esistenti tra loro.

Per la verifica di indipendenza tra X e Y, ossia se i profili riga (o colonna) sono tutti simili tra loro, si può usare il test χ^2

$$n \sum_{i=1}^K \sum_{j=1}^h \frac{(f_{ij} - f_{i*} f_{*j})^2}{f_{i*} f_{*j}}$$

con $m = (k-1)(h-1)$ gradi di libertà.

Per effettuare l'analisi delle corrispondenze si opera nel seguente modo:

- I profili riga sono considerati come punti di una certa combinazione \mathbb{R}^h con pesi f_i .
- Si opera una trasformazione in modo da rappresentare i dati in modo grafico in uno spazio più ridotto.
- Si interpreta la rappresentazione grafica
- Si ripete il tutto per i profili colonna
- Si interpreta i legami fra le due rappresentazioni ottenute.

La matrice R dei profili riga è definita come $R = P_r^{-1}N$ ove N la matrice di dimensione (k,h) corrispondente alla tabella di contingenza originale, il cui elemento generico è f_{ij} , P_r^{-1} è la matrice diagonale di dimensione (k,k) contenente l'inverso dei pesi delle righe, il cui elemento (i,i) della diagonale è $1/f_{i*}$. Il baricentro che rappresenta il profilo medio è il totale margine ed è rappresentato dal punto $(f_{*1}, f_{*2}, \dots, f_{*h})$.

Tuttavia per meglio confrontare in seguito su pari scala le rappresentazioni grafiche, è necessario a procedere ad una trasformazione normalizzante della matrice R.

Se indichiamo con:

- P_c la matrice della diagonale contenente sulla diagonale i valori $(f_{*1}, f_{*2}, \dots, f_{*h})$.
- $P_c^{-1/2}$ la matrice con i valori nella diagonale $(1/\sqrt{f_{*1}}, \dots, 1/\sqrt{f_{*h}})$

la matrice dei profili riga trasformati è $R^* = RP_c^{-1/2}$.

L'elemento generico sarà: $f_{ij}/(f_{i^*}\sqrt{f_{*j}})$ dove il nuovo baricentro sarà pari alla seguente combinazione $(\sqrt{f_{*1}}, \dots, \sqrt{f_{*h}})$ e indicando con $\alpha=1,2,\dots,rg$ un asse principale dei profili riga, il vettore contenete le proiezioni della nuvola dei dati su tale asse sarà: $C_\alpha^r = R^*u_\alpha$ con u_α autovetture normalizzato della matrice Γ^r di varianza e covarianza di R^* il cui termine generico è:

$$\lambda_{j,j'} = \sum_{i=1}^K f_{i^*} \left(\frac{f_{ij} - f_{i^*}f_{*j}}{f_{i^*}\sqrt{f_{*j}}} \right) \left(\frac{f_{ij'} - f_{i^*}f_{*j'}}{f_{i^*}\sqrt{f_{*j'}}} \right)$$

Lo stesso vale per i profili colonna con $C = P_c^{-1}N'$ e $C^* = CP_r^{-1/2}$ il cui elemento generico è $f_{ij}/(f_{*j}\sqrt{f_{i^*}})$ e il baricentro è $(\sqrt{f_{1^*}}, \dots, \sqrt{f_{k^*}})$. Il vettore contenente le proiezioni della nuvola è: $C_\alpha^c = C^*v_\alpha$ con v_α autovetture normalizzato della matrice di covarianza e varianza di C^* .

Data la varianza della proiezione di un profilo su un asse principale λ_α la

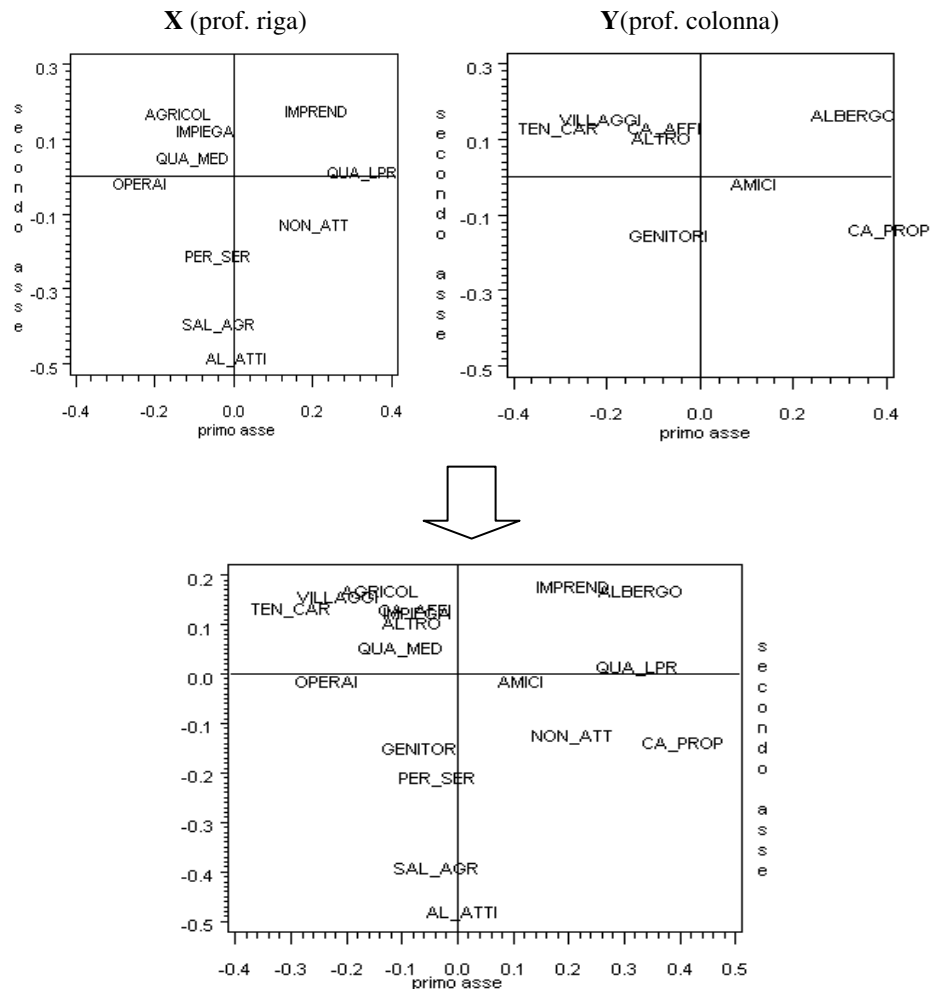
$\sum_{\alpha}^{rg} \lambda_\alpha = I = \text{Inerzia}$ e date le seguenti relazioni:

$$c_\alpha^r = \frac{1}{\sqrt{\lambda_\alpha}} R c_\alpha^c$$

$$c_\alpha^c = \frac{1}{\sqrt{\lambda_\alpha}} C c_\alpha^r$$

si ottiene che le proiezioni dei profili riga su un asse sono esprimibili come combinazioni lineari delle coordinate delle proiezioni dei profili colonna sullo stesso asse. In tale modo e attraverso la rappresentazione grafica simultanea dei due profili si può risalire e quindi mettere in evidenza le variabili responsabili di certe vicinanze.

Esempio: Rappresentazione grafica dati qualitativi



Unione dei profili riga e colonna.

Nei precedenti paragrafi è stato fatto cenno ad una serie di tecniche tendenzialmente complesse maggiormente usate sia nel campo della segmentazione che in quello del posizionamento. Ora e nei paragrafi successivi, si analizzerà in modo più approfondito delle tecniche specifiche di segmentazione (Conjoint e Cluster Analysis) e di posizionamento (MDS e Unfolding). Inoltre nel successivo capitolo (CAP. 4) verrà descritto il metodo DEA, nella sua struttura e funzione nativa, evidenziando delle possibili applicazioni in un'ottica orientata al Marketing.

3.2.5. Conjoint Analysis

Tale tecnica viene spesso usata per l'analisi della segmentazione flessibile.

La segmentazione flessibile, o meglio il suo schema si basa sull'integrazione dei risultati ottenuti dalla Conjoint Analysis (COA) con quelli relativi al comportamento dei consumatori. I concetti fondamentali su cui si basa la COA sono quelli sviluppati dallo studioso Lancaster relativamente alla teoria del consumatore ove l'utilità d'uso del bene deriva dalle singole caratteristiche che lo descrivono.

Si può in tal modo scomporre l'utilità che un consumatore ottiene dal prodotto in varie utilità separate in merito alle diverse caratteristiche del prodotto medesimo. Le caratteristiche del bene non vengono tuttavia giudicate una alla volta ma, congiuntamente nelle varie modalità che si manifestano.

La COA è nata per analizzare i diversi compromessi (trade-off) che specifiche tipologie di consumatori svolgono nella valutazione e scelta delle alternative possibili. Grazie a tale tecnica si possono valutare l'importanza che il consumatore attribuisce agli attributi, o caratteristiche del prodotto studio, e il grado di utilità corrispondente ad ogni caratteristica del prodotto.

Il processo operativo di tale tecnica si basa sulla selezione di un campione di consumatori potenziali ai quali si sottopone una lista di alternative di un certo prodotto (o servizio). Tali alternative vengono denominate stimoli e descritte attraverso attributi considerati rilevanti al fine dello studio in questione. In seguito si contattano i potenziali consumatori ai quali si chiede di dare un giudizio, o meglio un punteggio di gradimento (su scala già prefissata) alle alternative offerte. Si ottiene così una relazione biunivoca tra preferenze e utilità.

In seguito si tratta di stabilire la regola che permette di studiare tali preferenze, e tra i modelli più usati si è soliti a considerare un modello additivo in cui le utilità parziali di ogni attributo vengono sommate tra loro per ottenere l'utilità complessiva.

Di solito vengono usati tre schemi principali:

- il modello vettore
- il modello punto ideale
- il modello part-worth

Il modello vettore ipotizza la seguente relazione: $y_j = \sum_{k=1}^n w_k f_{jk}$ ove w_k rappresentano i pesi attribuiti all'importanza del singolo attributo e f_{jk} è il livello di intensità che l'attributo quantitativo k si manifesta nello stimolo j . Il modello punto ideale si basa sul definire una misura di utilità che cresce al diminuire della distanza esistente tra un profilo ideale e quello considerato: maggiore è la somiglianza del prodotto considerato con quello ideale tanto più sarà maggiore l'utilità del consumatore di quel bene.

La sua relazione è: $d_j^2 = \sum_{k=1}^n w_k (f_{jk} - i_k)^2$ con i_k il livello che costituisce il profilo ideale relativo ad ogni attributo.

Il modello part-worth invece esprime la preferenza (utilità) y_j per lo stimolo relativo j -esimo attraverso funzioni di discontinuità s definita da un insieme di attributi o per le modalità degli stessi: $y_j = \sum_{k=1}^n s_k(f_{jk})$ e tale modello risulta, paragonato con i precedenti, il più flessibile in quanto permette alla funzione di utilità (o preferenze) di assumere forme diverse a seconda della struttura della funzione s . Tale modello costituisce un semplice schema additivo che considera gli effetti delle singole modalità e non della loro interazione.

Tra le tante considerazioni da tener presente nella tecnica COA, si ricorda la scala di misura degli attributi. Gli attributi da considerare sono quelli che l'azienda ritiene corretti ai fini dell'analisi che lei stessa è in grado di monitorare e che influenzano le scelte del consumatore. Le modalità degli

attributi devono tuttavia essere contenute per non gravare sul modello e devono tuttavia cogliere i punti chiave delle scelte del consumatore.

Trova importanza tra gli attributi, il prezzo il quale può influire in modo considerevole sulle preferenze di scelta del consumatore. Buona cosa è quello di inserire un intervallo di prezzo che sia accessibile ai potenziali consumatori. Inoltre, spesse volte il prezzo è indice di qualità percepita dal consumatore e la presenza di tale attributo con l'attributo qualità potrebbe causare ai fini della stima una minore rilevanza su uno dei due attributi.

Sintetizzando il tutto, lo studio con l'analisi COA rappresenta una successione di azioni che si fondano:

1. Individuare gli attributi del prodotto
2. definire i profili del prodotto da sottoporre a giudizio
3. selezionare un campione di potenziali acquirenti
4. stima delle utilità parziali associate ad ogni modalità dell'attributo
5. determinazione dell'importanza di ogni attributo
6. valutazione dell'utilità totale relativa al profilo del prodotto.

Il punto di forza che la COA rappresenta è la capacità di analizzare la disponibilità dei consumatori a combinare tra loro diverse modalità relative ad un particolare prodotto, attraverso stime semplici. Tuttavia essa presenta rischi e tra questi, si ricorda il fatto che prodotti a forte contenuto di immagine non valutati in modo razionale da parte del consumatore. Su questo la COA può dare solo un'approssimazione di stima. Inoltre le analisi svolte con la COA si basano sulla scelta degli attributi e una scelta di essi errata può pregiudicare l'analisi sull'effettivo comportamento dei consumatori nel mondo reale. Quindi non solo bisogna cercare di trovare gruppi da sottoporre ad analisi il più omogenei possibile al loro interno, ma è necessario anche una buona e attenta analisi strategia a priori da parte dell'analista nella scelta del paniere più idoneo di attributi.

Un'altra analisi permessa tramite il metodo COA è quella di valutare le preferenze per ogni alternativa non sottoposte a test diretto, ed è sufficiente sommare le utilità parziali ottenute con la part-worth per ottenere la preferenza (o utilità) totale di un certo profilo di prodotto.

Tuttavia ci sono varie simulazioni per valutare la preferenza totale. Una di queste si basa sul valutare l'alternativa alla quale corrisponde il valore più alto della preferenza totale da cui si può determinare la quota di preferenza di un certo profilo di prodotto, ottenuta come rapporto tra gli intervistati che hanno preferito quel profilo e il numero complessivo di intervistati.

Un'altra procedura invece, si basa su assegnare una certa probabilità alle

alternative possibili definita da: $p_{ij} = \frac{u_{ij}}{\sum_{j=1}^J u_{ij}}$ con u_{ij} la preferenza totale che il

consumatore i attribuisce all'alternativa j . In tale caso la quota di

preferenza sarà: $QP_j = \frac{1}{n} \sum_{i=1}^n p_{ij}$.

Le quote di preferenza, sono delle simulazioni e come tali possono rappresentare una possibile interpretazione di quote di mercato da porre a confronto con il mondo reale.

Il seguente esempio ha come fine la stima del mercato potenziale nel lancio di una marca di alimenti per animali domestici (Tale esempio è ripreso dal libro "Marketing e Pubblicità: metodi di analisi statistica", si veda Bibliografia). Lo studio ha lo scopo di inserire nella sua linea di produzione una nuova marca di cibo per gatti appena nati. Si considera un campione di 150 consumatori prelevati dall'archivio della società e a partire dalle valutazioni ottenute dall'intervista si cerca di capire quali attributi siano più rilevanti su quelli scelti e capire quale combinazione delle modalità possa determinare il più alto tasso di acquisto.

In questo esempio il disegno ottenuto relativo all'associazione dati, stimoli, giudizi, ecc.. è riportato nella tabella seguente:

Stimoli	Nome	Confezione	Prezzo	Approv. Ordine	Garanzia	Preferenza	Gradimento
1	Mamacat	Scatoletta	2.000	N	S	12	50
2	Formula 9	Scatoletta	1.000	S	N	7	78
3	Plus	Scatoletta	1.500	N	N	9	66
4	Mamacat	Vasetto	2.000	S	N	18	15
5	Formula 9	Vasetto	1.000	N	S	8	72
6	Plus	Vasetto	1.500	N	N	15	32
7	Mamacat	Tetrapack	1.000	N	N	4	90
8	Formula 9	Tetrapack	1.500	S	N	6	80
9	Plus	Tetrapack	2.000	N	S	5	84
10	Mamacat	Scatoletta	1.500	N	N	10	52
11	Formula 9	Scatoletta	2.000	N	N	16	28
12*	Plus	Scatoletta	1.000	S	S	1	95
13	Mamacat	Vasetto	1.000	N	N	13	42
14	Formula 9	Vasetto	1.500	N	S	11	50
15	Plus	Vasetto	2.000	S	N	17	20
16	Mamacat	Tetrapack	1.500	S	S	2	92
17	Formula 9	Tetrapack	2.000	N	N	14	38
18	Plus	Tetrapak	1.000	N	N	3	90

tab. 4 - *= è la maggiormente preferita

Esso è stato creato attraverso software specifici in quanto tali disegni e relative compilazioni dei dati non risultano facili.

Ottenuti quindi i dati di base, lo step successivo è quello di determinare la stima delle utilità parziali associate alle modalità di ogni attributo e in seguito di ottenere stime in cui la somma delle utilità part-worth riferite alle modalità dia luogo a valori di utilità totale più vicini alle vere preferenze del consumatore.

I modelli di analisi sono vari, e per quanto detto precedentemente si cerca di trovare un modello di tipo additivo.

In tale situazione avviene utile richiamare un modello di regressione lineare multipla descritto nei precedenti paragrafi con variabili dummy. Tale modello è definito in tale modo:

$$Y = a + b_1X_1 + b_2X_2 + b_3X_3 + b_4X_4 + b_5X_5 + b_6X_6 + b_7X_7 + b_8X_8$$

dove Y rappresenta la valutazione dell'intervistato le variabili X_i rappresentano variabili dummy. Tali variabili sono la codifica degli attributi, o meglio l'espressione di ogni modalità sotto forma di variabile dicotomica. Ciò è reso possibile in quanto le valutazioni fornite sono misurate su scala ad intervallo.

La X_1 è la modalità Mamacat dell' attributo nome, X_2 quella di Plus dell' attributo nome, X_3 modalità tetrapak dell'attributo tipo confezione, X_4 quella della scatoletta sempre per tipo di confezione, X_5 la modalità 1000 relativa al prezzo, X_6 la 1500 sempre del prezzo, X_7 modalità si dell' attributo approvazione dei medici, X_8 è la modalità no relativa ai medici. Altre modalità sono state omesse per l'esistenza di collinearità, che anche in questo caso si è lasciato al software a svolgere le opportune semplificazioni ottenendo in sintesi i seguenti risultati:

Stimoli	X1	X2	X3	X4	X5	X6	X7	X8
1	1	0	0	1	-1	-1	-1	1
2	-1	-1	0	1	1	0	1	-1
3	0	1	0	1	0	1	-1	-1
4	1	0	-1	-1	-1	-1	1	-1
5	-1	-1	-1	-1	1	0	-1	1
6	0	1	-1	-1	0	1	-1	-1
7	1	0	1	0	1	0	-1	-1
8	-1	-1	1	0	0	1	1	-1
9	0	1	1	0	-1	-1	-1	1
10	1	0	0	1	0	1	-1	-1
11	-1	-1	0	1	-1	-1	-1	-1
12	0	1	0	1	1	0	1	1
13	1	0	-1	-1	1	0	-1	-1
14	-1	-1	-1	-1	0	1	-1	1
15	0	1	-1	-1	-1	-1	1	-1
16	1	0	1	0	0	1	1	1
17	-1	-1	1	0	-1	-1	-1	-1
18	0	1	1	0	1	0	-1	-1

tab. 5 – codifica relativa a variabili Dummy (-1 non corrisponde nessuna Dummy allo stimolo)

Omettendo i passaggi di stima di tale modello, (si usa il metodo dei minimi quadrati descritto nei paragrafi precedenti) si perviene ad una stima del modello $Y = 64.13 - 2.83X_1 + 4.83X_2 + 19.33X_3 + 1.83X_4 + 18.17X_5 + 2.33X_6 + 2.75X_7 + 10.63X_8$. Tuttavia per i calcoli sempre i software più evoluti, come vedremo nel CAP 5, permettono di ottenere tali stime e relativi passaggi di calcolo tutto in modo automatico.

I parametri di tale modello rappresentano le utilità ricercate (part-worth). Per il discorso della collinearità, manca la stima dell'utilità Formula 9, ma essa si ricava per le ipotesi poste in precedenza nel seguente modo:

$U(\text{Formula9}) = - [U(\text{Mamacat}) + U(\text{Plus})] = - [-2.83 + 4.83] = -2.00$ (si veda parametri del modello). Di conseguenza tale relazione vale anche per tutte le altre modalità non stimate direttamente nel modello.

Al fine di facilitare una comparazione tra le utilità è buona norma

trasformare i valori attraverso la formula: $U_{ij} = \frac{V_{ij} - P}{G - P}$ ove G è la più

grande utilità parziale, P la più piccola e V_{ij} l'utilità da trasformare.

Inoltre a tale punto si può procedere a determinare l'importanza relativa

dell' attributi A_i : $IRA_i = \frac{[Utilità\ più\ grande\ A_i - Utilità\ più\ piccola\ A_i]}{\sum_{i=1}^n [utilità\ più\ grande\ A_i - utilità\ più\ piccola\ A_i]}$.

Nella seguente tabella riportiamo una sintesi dell'importanza assegnata ad ogni attributo.

Attributo	U più grande	U più piccola	Differenza	Somma diff.	Importanza relativa
Nome	4.83	-2.83	7.66	113.58	0.07
Tipo confezione	19.33	-21.16	40.49	113.58	0.35
Prezzo	18.17	-20.50	38.67	113.58	0.34
Approvazione dell' ordine	2.75	-2.75	5.50	113.58	0.05
Garanzia del rimborso	10.63	-10.63	21.26	113.58	0.19

tab. 6 – Importanza degli attributi

Mentre nella seguente tabella riportiamo una sintesi relativa alle utilità parziali originarie e trasformate.

Attributo	Modalità	Utilità Originaria	Utilità Trasformata
Nome	Mamacat	-2.83	0.45
	Formula 9	-2.00	0.47
	Plus	4.83	0.64
Tipo	Scatoletta	1.83	0.57
Confezione	Vasetto	-21.16	0.00
	Tetrapak	19.33	1.00
Prezzo	1.000	18.17	0.97
	1.500	2.33	0.58
	2.000	-20.50	0.02
Approvazione	Si	2.75	0.59
Ordine	No	-2.75	0.45
Garanzia	Si	10.63	0.79
	No	-10.63	0.26

tab. 7 – valore delle utilità parziali

Concludendo, oltre al fatto come già accennato in precedenza che la combinazione migliore è relativa al 12° stimolo (tab.4), si possono formulare vari quesiti in base ai risultati ottenuti come ad esempio il fatto se un consumatore sarebbe disposto a spendere 2.000 al posto di 1.500 in cambio dell'approvazione da parte dell'ordine dei medici, e così via con altre possibili simulazioni di ipotesi. Tuttavia va ricordato che i passi di analisi si devono replicare anche per gli altri consumatori appartenenti al campione e i risultati ottenuti si possono tradurre in termini di strategie di Marketing con il fine di centrare il miglior modo per aggiungere la marca all'interno della gamma già esistente e di stimare la quota di mercato potenziale.

3.2.6. Cluster Analysis

La Cluster Analysis (CLA) è una tecnica che permette di scomporre una particolare realtà, anche complessa, in sottoinsiemi specifici. Ossia, dato

un certo collettivo di individui, con tale tecnica si può dividere tale insieme in sottoinsiemi omogenei al loro interno.

Essa trova larga applicazione nelle fasi di studio del comportamento di acquisto del consumatore e nelle ricerche di sviluppo di nuovi prodotti. Oltre ciò, torna utile nei studi di concorrenza da parte delle aziende, permettendo di valutare la sua gamma di prodotti a confronto con quella degli avversari e identificare di conseguenza i punti critici (o punti di forza e di debolezza) del suo prodotto con quello avversario.

Il procedimento base della CLA è quello di costruire un campione di n elementi ciascuno rappresentato da p variabili caratteristiche. Per esempio se l'elemento è il cliente, le variabili possono essere il comune di residenza, il reddito, la composizione del nucleo familiare, il tipo di lavoro, il grado di istruzione, ecc...

I dati raccolti si organizzano in una matrice $n \times p$, ove ogni riga rappresenta il profilo individuale dell'elemento attraverso le p variabili caratteristiche. A questo punto si procede con una serie di elaborazioni (ove anche in questo caso l'informatica ha dato un grande appoggio all'analista semplificando notevolmente i passaggi di calcolo risparmiando in tempo e quindi anche in denaro) per ottenere G gruppi dagli n elementi presi in considerazione.

In sintesi le fasi specifiche della CLA possono essere riassunte nel seguente modo:

1. Selezionare il campione
2. Scegliere le variabili di segmentazione o caratteristiche
3. Selezionare un criterio per la dissomiglianza esistente tra gli elementi
4. Scegliere un accurato algoritmo di raggruppamento
5. determinare il numero di gruppi ottimale
6. Verificare e Interpretazione dei dati ottenuti.

Certe volte, si può verificare che alcune variabili scelte possano presentare valori simili. In questo caso è buona cosa procedere a delle semplificazioni attraverso trasformazioni, per esempio usando il metodo delle componenti principale (descritto nei paragrafi precedenti), con il fine di ridurre la numerosità delle variabili ma lasciando invariato il risultato finale.

Fissate quindi le variabili e eseguite le opportune correzioni, il passo successivo è quello di scegliere un criterio di dissomiglianza rispetto all'insieme delle stesse variabili caratteristiche. Tuttavia bisogna tener presente che per determinare una misura di dissomiglianza è necessario che le variabili oggetto di studio siano tra loro confrontabili su stesa scala, e molto spesso si deve procedere ad una standardizzazione classica che porta ad ottenere un insieme di variabili standardizzate a partire da quelle iniziali con media pari a zero e ad una varianza unitaria.

Di solito per la misura della dissomiglianza si fa uso di particolari coefficienti di associazione e di distanze appartenenti alla metrica di Minkowski.

I coefficienti di associazione misurano la somiglianza esistente tra le unità quando i caratteri possono essere espressi su scala binaria nominale. I dati vengono rilevati a coppie e su ciascuna coppia i dati vengono riportati su una matrice del seguente tipo:

		Individuo j	
		1	0
Individuo i	1	a	b
	0	c	d

Tra i coefficienti più noti ricordiamo quello di Jaccard che assume valori in

un intervallo $[0,1]$: $J_{ij} = \frac{a}{a + b + c}$.

Un altro coefficiente che offre una maggiore importanza alla presenze simultanee (su frequenza a) è quello di Dice : $D_{ij} = \frac{2a}{2a + b + c}$.

Altri indici invece introducono anche la presenza della frequenza d, cosa opportuna se le variabili sono rilevanti ai fini dell'analisi dei profili in questione. Tra questi coefficienti si ricorda il coefficiente semplice di somiglianza:

$$s_{ij} = \frac{a + d}{a + b + c + d}.$$

Un altro indice è il coefficiente do Gower che può essere applicato sia

avariabili di tipo cardinale, nominale e ordinale: $G_{ij} = \frac{\sum_{k=1}^p w_k s_{kij}}{\sum_{k=1}^p w_k}$ ove s_{kij} è

l'indicatore di somiglianza semplice tra le unità i j rispetto alla variabile k e vale:

- 1 se la variabile è di tipo nominale o ordinale ed esiste concomitanza di presenza o assenza per i e j
- 0 se la variabile è di tipo nominale o ordinale e non esiste concomitanza di presenza o assenza per i e j
- $1 - [|x_{ik} - x_{jk}| / R_k]$ se la variabile è quantitativa (R_k rappresenta il campo di variazione della variabile k)

Questi coefficienti sono misure di somiglianza, per ottenere la dissomiglianza si può applicare la seguente relazione: $d_{ij} = 1 - s_{ij}$.

Se si dispone di valutazioni espresse su scala intervallare o a rapporti per ciascuna delle p variabili considerate, si può procedere a considerare una distanza esistente tra 2 individui i e j mediante misure di Minkowski:

$${}^r d_{ij} = [\sum_{k=1}^p |x_{ik} - x_{jk}|^r]^{1/r}$$

dove r assume valori interi ma minori di 1.

Nel caso si dovesse manifestare una forte interdipendenza tra alcuni caratteri come base del processo usato, si ricorre ad un'altra distanza detta di Mahalanobis, che permette di ridimensionare il peso della variabili portatrici di informazione in eccesso:

$$d_{ij} = \sqrt{\sum_{k=1}^p \sum_{h=1}^p s^{kh} (x_{ik} - x_{jk})(x_{ih} - x_{jh})}$$

ove s^{kh} indica l'elemento della matrice inversa della varianza e covarianza tra le p variabili caratteristiche.

Trovate le dissomiglianze per sintetizzare la diversità dei profili tra le copie degli elementi considerati, si ottiene una matrice del seguente tipo:

$$D = \begin{pmatrix} 0 & d_{12} & \dots & d_{1n} \\ d_{21} & 0 & \dots & d_{2n} \\ \dots & \dots & \dots & \dots \\ d_{n1} & d_{n2} & \dots & 0 \end{pmatrix}$$

Passo successivo, è quello di trovare un certo numero di gruppi omogenei, attraverso l'uso di algoritmi di raggruppamento. Tra gli algoritmi, possiamo identificare algoritmi di tipo gerarchico e non. In quelli di tipo gerarchico, si distinguono quelli di tipo aggregativi che hanno il fine di trovare una serie di partizioni concatenate ottenute aggregando due gruppi in quelle di ordine precedente, e la mappa delle aggregazioni può essere rappresentata attraverso uno schema grafico definito denogramma (fig.9):

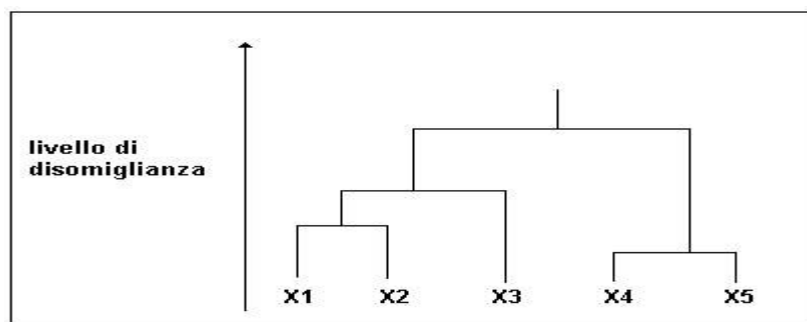


fig. 9 – Esempio di denogramma

Tra gli algoritmi gerarchici ricordiamo il metodo di Ward che riunisce ad ogni step del processo i due gruppi in esame, e che dalla fusione ne derivi il minimo incremento possibile della devianza entro i gruppi. Per esempio se date n unità si desidera ottenere G gruppi con numerosità variabile n_g rispetto a p variabili caratteristiche si ottiene che:

$$\text{Dev(Totale)} = \sum_{g=1}^G \sum_{i=1}^{n_g} \sum_{k=1}^p (x_{ikg} - \bar{x}_k)^2$$

ed è scomponibili in:

$$\text{Dev(entro)} = \sum_{g=1}^G \sum_{i=1}^{n_g} \sum_{k=1}^p (x_{ikg} - \bar{x}_{kg})^2$$

$$\text{Dev(tras)} = \sum_{g=1}^G \sum_{i=1}^{n_g} \sum_{k=1}^p (\bar{x}_{kg} - \bar{x}_k)^2$$

ove:

$$\bar{x}_k = \frac{1}{n} \sum_{g=1}^G \sum_{i=1}^{n_g} x_{ikg} \quad \text{è il valore medio della variabile } k \text{ dell'insieme.}$$

$$\bar{x}_{kg} = \frac{1}{n_g} \sum_{i=1}^{n_g} x_{ikg} \quad \text{è il valore medio della variabile } k \text{ nel } g\text{-esimo gruppo.}$$

A differenza di tali algoritmi, gli algoritmi non gerarchici mirano a classificare direttamente le n unità in G gruppi fissati a priori.

Un algoritmo di questo tipo è quello di McQueen detto anche delle K -medie. Tuttavia tali algoritmi si basano su due fondamentali fasi:

1. la determinazione di una partizione iniziale degli n individui in G gruppi.
2. spostamento successivo delle unità fra i gruppi in modo da ottenere la partizione il più omogenea possibile.

Ora, rimane solo la determinazione del numero ottimale di gruppi.

Per la determinazione ottimale del numero di gruppi, non esiste una procedura standard applicabile a tutte le occasioni, ma si può citare alcune tecniche che vengono normalmente utilizzate per tale scopo.

Una di queste è la rappresentazione grafica che pone in ordinata il numero di gruppi creato e in ascissa i corrispondenti valori di dissomiglianza. Si analizza la spezzata che si genera unendo i punti del grafico e nell'area in cui si presenta un forte appiattimento di tale spezzata si identifica il numero ottimale di gruppi.

Un'altra procedura è quella di analizzare la dissomiglianza tra i due gruppi che si forma ad ogni step successivo. Si cerca di analizzare il salto che esiste nella serie che si genera di tali valori. Il processo si ferma prima del salto maggiore identificando così il numero ottimale.

La conclusione si basa sull'analisi dei dati che si ottengono, studiando anche con l'ausilio di test F se esiste una differenza significativa tra i gruppi. In questo modo si riesce a capire come ogni gruppo si caratterizzi per dare peso ad un certo attributo al posto di un altro, o come tutti i gruppi sono collegati ad uno stesso attributo caratteristico. Ciò permette quindi di capire come il mercato è segmentato e come comportarsi di conseguenza per coprire l'area interessata. Non solo i dati che pervengono tornano utili anche per un'analisi ai fini di un miglioramento continuativo sulle performance dell'impresa, grazie al fatto che essi sono buoni indicatori per lo studio della fedeltà del cliente e di gradimento.

3.3. Tecniche specifiche per il posizionamento del prodotto

3.3.1. Multidimensional Scaling (MDS)

Il MDS è una tecnica nata inizialmente per lo studio delle scienze sociali e comportamentali, in seguito applicata nel generale mediante la quale

l'informazione presente in un insieme di dati viene rappresentata graficamente in una nube di punti.

Le tecniche dell' MDS si fondano su dati di prossimità, ossia su valutazioni di tipo percettivo e a seconda della loro scala di misurazione essi si possono presentare sotto tre tipologie:

1. Di distanza
2. Di dissomiglianza (o somiglianza)
3. Di ordinamento, qualora i dati non hanno un significato cardinale ma si limitano ad indicare posizioni relative.

Normalmente i dati di prossimità vengono rilevati secondo modalità ben definite che possono essere riassunte su quattro tipologie:

1. Attraverso giudizi diretti
2. Come valutazione assimilabile a probabilità condizionate
3. Come valutazione assimilabile a probabilità congiunte
4. Come sintesi (o distanza) dei profili degli elementi in esame rispetto a certe caratteristiche

In particolare è da tenere presente per il punto 1 che il giudizio diretto corrisponde al valore che l'intervistato attribuisce alla somiglianza percepita tra due prodotti in una specifica scala, e che il numero di intervistati sia almeno pari a $s' = 40p / (n-1)$ ove n è il numero di prodotti e p rappresenta il numero di dimensioni rispetto a cui sarà costruita la mappa dell' MDS. Per il punto 4 è utile ricordare che la distanza che normalmente viene usata per misurare l'indice relativo ai profili è quella euclidea:

$$\delta_{ij} = \sqrt{\sum_{k=1}^p (v_{ik} - v_{jk})^2}$$

ove v_{ik}, v_{jk} rappresentano il valore del k -esimo carattere sul prodotto i e j .

Tra i possibili metodi dell' MDS si ricorda il metodo di tipo metrico, non metrico e per differenze individuali.

Si è soliti parlare di MDS metrico qualora i dati pervenuti sono di tipo metrico, ossia le loro scale di misura sono di tipo intervallare o a rapporto.

L' MDS non metrico invece si basa su dati la cui scala è di tipo ordinale, mentre l' MDS per differenze individuali viene visto come una variante di quello metrico in cui i dati di prossimità sono espressi in forma individuale e non aggregata.

Nel modello MDS metrico, l'obiettivo è dunque quello di individuare un numero n di prodotti su uno spazio a p dimensioni. E' fondamentale quindi che il numero di dimensioni da scegliere sia il più limitato possibile per permettere una interpretazione grafica più leggibile ($p \leq 3$).

Di norma si è soliti determinare le distanze euclidee tra ciascuna coppia conoscendo già le coordinate di tali coppie. Nell' MDS ciò funziona al contrario, ossia si è già in possesso delle distanze euclidee che mi rappresentano le coppie e da tali informazioni si deve ricavare le coordinate delle posizioni dei singoli "prodotti" nella mappa percettiva del consumatore. Le dimensioni di tale grafico saranno le basi sulla quali si estrapoleranno informazioni sintetiche, o più semplicemente si traggono delle conclusioni in merito ai risultati ottenuti, su come il prodotto si posiziona a confronto con quelli concorrenti.

Il processo quindi per determinare le coordinate parte dal presupposto di avere come dati iniziali le distanze euclideanee, che possono essere rappresentate attraverso una matrice di $n \times n$ dimensioni contenete però i quadrati delle distanze medesime con n che rappresenta il numero di prodotti. L' elemento generico quindi di tale matrice è quello della classica

formula descritta precedentemente: $d_{ij} = \sum_{k=1}^p (x_{ik} - x_{jk})^2$ dove X_{ik} e X_{jk} sono le coordinate dei punti che dovranno in seguito essere determinate.

Infatti ora si deve procedere a determinare una matrice che mi dia le informazioni dovute sulle coordinate cercate.

Data, quindi la matrice delle distanze, si procede a calcolare la matrice delle somme dei quadrati delle coordinate dei prodotti $B = XX'$, ove X è la matrice delle coordinate il cui elemento generico della matrice B è:

$$b_{ij} = -\frac{1}{2}(d_{ij}^2 - d_{i*}^2 - d_{*j}^2 + d_{**}^2)$$

ove $d_{i*}^2, d_{*j}^2, d_{**}^2$ sono gli elementi marginali della matrice delle distanze euclidee. A questo punto si calcola di determinare la matrice X calcolando gli autovalori e gli autovettori di B .

Sapendo che il rango di B è pari a p con $p \leq n$, la matrice B ammetterà p autovalori non nulli $\lambda_1, \dots, \lambda_p$ e definendo che i corrispondenti autovettori di norma uno sono indicati come segue $e_1 \dots e_p$, una possibile matrice X di coordinate con dimensione (n, p) è rappresentata dai autovettori $X = [f_1, f_2, \dots, f_p]$ con $f_i = \sqrt{\lambda_i} e_i$.

Tuttavia anche in questa situazione l'informatica aiuta di molto l'analista, gestendo in modo automatico la determinazione di tale procedimento offrendo direttamente i risultati richiesti. Per esempio se si ha una matrice D di distanze euclidee:

	A	B	C	D	E	F	G	H
A	0.00	-	-	-	-	-	-	-
B	1.41	0.00	-	-	-	-	-	-
C	1.00	1.00	0.00	-	-	-	-	-
D	1.00	1.73	1.41	0.00	-	-	-	-
E	1.41	1.41	1.73	1.00	0.00	-	-	-
F	1.41	1.41	1.00	1.00	1.41	0.00	-	-
G	1.73	1.00	1.41	1.41	1.00	1.00	0.00	-
H	0.71	1.41	1.00	1.00	1.41	1.41	1.73	0.00

(Tale esempio è ripreso dal libro "Marketing e Pubblicità: metodi di analisi statistica", si veda Bibliografia)

attraverso la relazione precedentemente descritta si può determinare la matrice $B = XX'$:

	A	B	C	D	E	F	G	H
A	0.70	-	-	-	-	-	-	-
B	-0.21	0.89	-	-	-	-	-	-
C	0.17	0.26	0.64	-	-	-	-	-
D	0.17	-0.74	-0.37	0.64	-	-	-	-
E	-0.21	-0.12	-0.74	0.26	0.89	-	-	-
F	-0.34	-0.24	0.14	0.14	-0.24	0.64	-	-
G	-0.71	0.39	-0.24	-0.24	0.39	0.26	0.89	-
H	0.45	-0.21	0.17	0.17	-0.21	-0.34	-0.71	0.70

e da tale matrice si calcolano i corrispettivi autovalori e considerando come numero di dimensioni $p=2$ da tale matrice si ricava la matrice X delle coordinate dei otto ipotetici prodotti:

	DIM 1	DIM 2
A	-0.73	0.00
B	0.42	0.71
C	-0.28	0.71
D	-0.28	-0.71
E	0.42	-0.71
F	0.24	0.00
G	0.95	0.00
H	-0.73	0.00

Il grafico, o meglio la mappa dell' MDS associata viene così rappresentata:

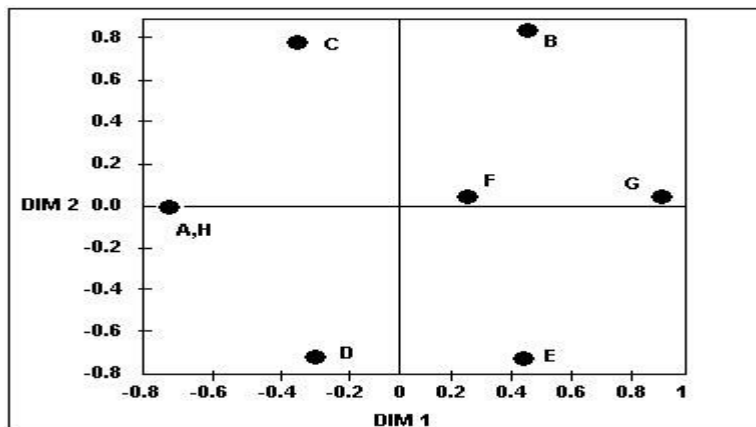


fig. 10 – Mappa di percezione

Nel caso che le misure di prossimità sono di tipo ordinale, si procede invece con il metodo MDS non metrico.

Definendo $\Delta = [\delta_{ij}]$ la matrice dei dati di prossimità, l'obiettivo è quello di determinare un p di dimensioni intero e una matrice $X = [x_{ij}]$ ($n \times p$) di coordinate. La soluzione dell' MDS deve possedere due caratteristiche fondamentali:

1. la dimensione p deve essere ridotta
2. la differenza tra le prossimità originarie δ_{ij} e le distanze d_{ij} deve essere minima.

Il processo operativo dell' MDS non metrico, consiste nel determinare una configurazione iniziale di coordinate casuali, oppure ottenute con l' MDS metrico in un numero prefissato di dimensioni.

In seguito si procede a calcolare le distanze a partire da tali coordinate con la seguente relazione:

$$d_{ij} = \sqrt{\sum_{k=1}^p (x_{ik} - x_{jk})^2}$$

ponendo il vincolo nella configurazione interessata: $\delta_{ij} < \delta_{lm} \leftrightarrow d_{ij} \leq d_{lm}$.

Si cerca quindi una trasformazione monotona, $F(\delta_{ij}) = f_{ij}$ (al fine di valutare la differenza esistente dalle distanze calcolate e le approssimazioni) crescente, se i valori sono dissimilarità, o decrescente se di similarità, in modo tale che sia soddisfatto il vincolo: $\delta_{ij} < \delta_{lm} \leftrightarrow f_{ij} \leq f_{lm}$.

Tale trasformazione F può essere lineare, esponenziale, logaritmica ecc...

In fine si spostano i punti nella mappa dell' MDS in modo da minimizzare il valore dello Stress:

$$\text{Stress} = \sqrt{\frac{\sum_{i < j} \sum_{j=2}^n (d_{ij} - f_{ij})^2}{\sum_{i < j} \sum_{j=2}^n (d_{ij})^2}}$$

Oltre a ciò si può fare una considerazione sul valore rappresentato dallo stress. Tale valore dovrebbe decrescere all'aumentare delle dimensioni p . Questo diviene un indicatore utile anche per scegliere il numero di dimensioni da usare per costruire la mappa. Si è soliti a scegliere il valore di p , in cui la spezzata che collega i punti del grafico con ascissa p e ordinata lo stress, presenta una sorta di gomito seguito da una stazionarietà nel verso dell'asse delle ascisse (fig.11).

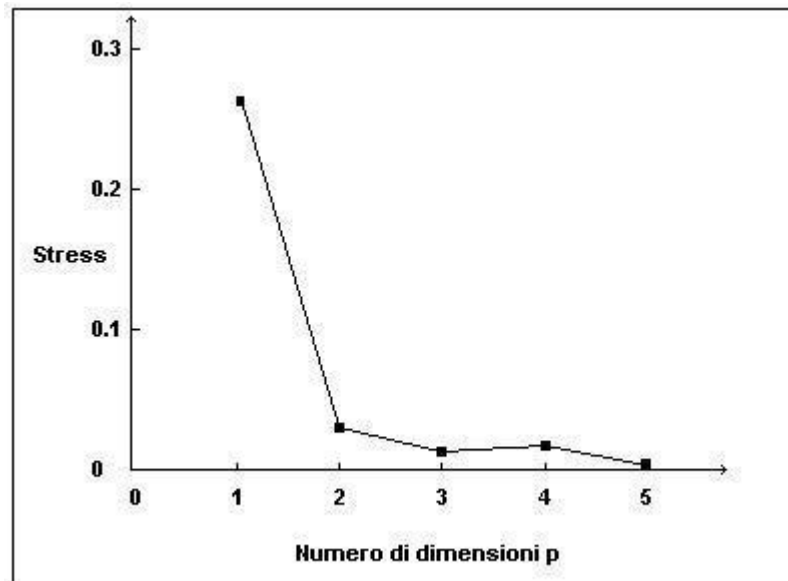


fig. 11 – Relazione dimensione - stress

Una variante del modello metrico è quella per differenze individuali. Essa viene usata qualora i dati di prossimità sono espressi in forma individuale e non aggregata.

Come in precedenza l'obiettivo è quello di determinare X , matrice delle coordinate di gruppo, e in questo caso anche X_s , matrice delle coordinate del singolo individui s e la matrice W_s rappresentante i pesi che l'individuo s assegna alla dimensione in questione.

Infatti come ipotesi di base del modello si considera che esista uno spazio geometrico condiviso da tutti gli individui, ma le differenze di valutazione

degli individui dipendano dal peso che essi attribuiscono alle dimensioni di tale spazio.

Definendo $x_{iks} = x_{ik}w_{ks}$ gli elementi della matrice X_s delle coordinate dell'individuo s e definendo la seguente relazione $X_s = XW_s$ ove W_s è la matrice diagonale di dimensione (p,p) relativa all'individuo s che ha w_{ks} come k -esimo elemento della diagonale che a sua volta rappresenta il peso assegnato alla dimensione k -esima, si procede alla stima di tali matrici.

Il processo è lo stesso di quello adottato in quello metrico. Infatti ricordando la formula della distanza, in questo caso relativa all'individuo s :

$$d_{ijs} = \sqrt{\sum_{k=1}^p w_{ks}^2 (x_{ik} - x_{jk})^2}$$

si procede a determinare la matrice $B_s = XW_s^2X'$ il cui elemento è:

$$b_{ijs} = -\frac{1}{2}(d_{ijs}^2 - d_{i*s}^2 - d_{*js}^2 + d_{**s}^2)$$

ove:

$$d_{i*s}^2 = \text{alla media } (1/n) \sum_j d_{ijs}^2$$

$$d_{*js}^2 = \text{alla media } (1/n) \sum_i d_{ijs}^2$$

$$d_{**s}^2 = \text{alla media generale } (1/n) \sum_i \sum_j d_{ijs}^2$$

Di conseguenza si va a determinare la matrice delle coordinate e dei pesi relativi ad ogni individuo.

Anche in questo metodo è possibile determinare una misura di adattamento in grado di dare il numero ottimale di dimensioni, e tale misura è definita dalla bontà di adattamento del modello:

$$r = \frac{\sum_{i=1}^n \sum_{j=1}^n \sum_{s=1}^m (b_{ijs} - b^*)(\bar{b}_{ijs} - \bar{b}^*)}{\sqrt{\sum_{i=1}^n \sum_{j=1}^n \sum_{s=1}^m (b_{ijs} - b^*)^2 \sum_{i=1}^n \sum_{j=1}^n \sum_{s=1}^m (\bar{b}_{ijs} - \bar{b}^*)^2}}$$

dove $b^* = (1/n^2m) \sum_{i=1}^n \sum_{j=1}^n \sum_{s=1}^m (b_{ijs})$ e $\bar{b}^* = (1/n^2m) \sum_{i=1}^n \sum_{j=1}^n \sum_{s=1}^m (\bar{b}_{ijs})$.

Anche in questo caso, riportando i dati relativi a tale misura in ordinata e riportando in ascissa il numero di dimensioni, si formerà un grafico ove collegando i punti tra loro e osservando la spezzata che si genera si individuerà una sorta di gomito in riferimento alla quale corrisponde il numero ottimale di dimensioni (fig.12)

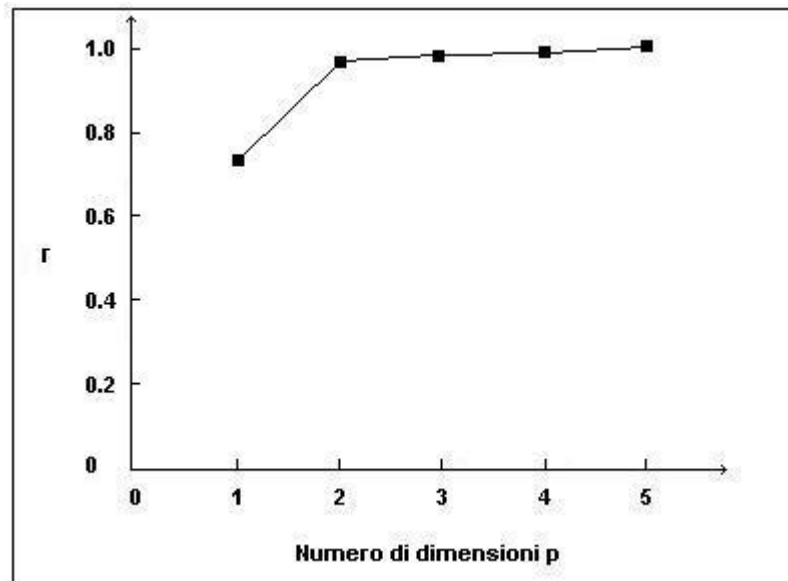


fig. 12 – Relazione dimensione - r

3.3.2. Unfolding e mappe di preferenza.

Nelle ricerche di Marketing è indispensabile studiare ed analizzare le preferenze dei consumatori su specifici prodotti per poter in seguito generare opportune strategie di posizionamento.

La tecnica scoperta da Coombs che prende il nome di Unfolding (UNF), serve per studiare in maniera sintetica le preferenze dei consumatori e

rappresentare in uno spazio geometrico sia i consumatori che le descrivono sia i prodotti che ne sono oggetto.

Obiettivo base, quindi, è quello di ottenere rappresentazioni che avvicinino tra loro soggetti in cui le preferenze per certi prodotti o marche sono analoghe.

I giudizi di preferenza possono essere espressi con diverse scale di valutazione e ai fini operativi si può individuare le seguenti forme di misura:

1. per posizione o ranghi
2. valutazione su scale
3. comparazione tra coppie

Nel primo caso, ad ogni prodotto si associa la corrispondente posizione con un ordinamento (la migliore alternativa avrà posizione 1). Nel secondo caso i prodotti vengono valutati attraverso una scala di tipo cardinale o ordinale. Nel terzo caso, invece, si ottengono informazioni sulla preferenza tra due elementi per le coppie che possono essere considerate valide ai fini dello studio. Per esempio se si analizzano quattro marche A,B,C,D, le coppie possibili sono (A,B) (A,C) (A,D) (B,C) (B,D) (C,D). I dati relativi possono essere raccolti in una matrice A a tre dimensioni (n,n,m) per cui il generico elemento a_{ijs} si associa una delle quattro seguenti relazioni:

1. $a_{ijs} = 1$ e $a_{jis} = 0$ se l'elemento i è preferito a quello j relativamente all'individuo s
2. $a_{ijs} = 0$ e $a_{jis} = 1$ se l'elemento j è preferito a quello i relativamente all'individuo s
3. $a_{ijs} = a_{jis} = 1/2$ se i e j sono indifferenti per l'individuo s
4. $a_{ijs} = a_{jis} = 0$ se i e j non sono considerati dall'individuo s

Tale modello pone per ipotesi che gli individui al momento in cui formulano un giudizio abbiano un riferimento relativo ad un prodotto ideale. Questo significa che tale modello dovrà stimare le coordinate del punto ideale dell'individuo s (x_{sk}) e di conseguenza tanto più un prodotto o una marca sono vicine alle coordinate stimate del punto ideale tanto più vengono preferite dal consumatore medesimo. Tuttavia si deve tener presente il fatto che i prodotti possono essere giudicati diversi tra loro ma essere preferiti in uguale misura oppure essere considerati simili a livello di percezione ma non di preferenza. (Rif. CAP 1)

Per valutare le coordinate del punto ideale del singolo individuo si possono adottare due approcci, uno implicito l'altro esplicito.

Nel primo caso la stima delle coordinate avviene attraverso i soli giudizi di preferenza che l'individuo esprime sui prodotti in esame. Nel secondo le stime avvengono attraverso risposte dirette dell'individuo intervistato su tale aspetto includendo di conseguenza un ipotetico prodotto ideale tra quelli da giudicare (presenti nella lista).

Oltre a questi approcci, si possono distinguere altre metodologie di analisi classificabili in analisi interne ed esterne.

Le prime forniscono stime delle coordinate dei prodotti a partire dalla sola disponibilità di dati di preferenza. Le seconde presuppongono che le coordinate siano già note ottenute con metodi come MDS visto in precedenza e utilizzano i dati per stimare particolari parametri relativi ai soli intervistati, come le coordinate del punto ideale e misure di bontà di adattamento del modello.

Per quanto riguarda l'analisi interna di dati di preferenza, il procedimento di stima delle coordinate e di eventuali parametri caratteristici è analogo al metodo MDS non metrico. Infatti tale metodo di UNF è di tipo non metrico e i risultati di un'analisi non metrica su dati di preferenza è molto simile su quelli di prossimità visti nell' MDS.

La differenza sostanziale sta nel fatto che nei dati di preferenza si ottiene anche una stima delle coordinate dei punti ideali dei consumatori. La numerosità ottima delle dimensioni viene calcolata in modo analogo a quella dell' MDS non metrico, e nel modello UNF la sua forma più utilizzata prevede che le preferenze si colleghino con le misure di distanza tra i prodotti dalla seguente relazione:

$$r_{is} = f_s \left(\sqrt{\sum_{k=1}^p (x_{ik} - x_{sk})^2} \right)$$

ove f_s rappresenta una funzione monotona e $\sqrt{\sum_{k=1}^p (x_{ik} - x_{sk})^2}$ è la distanza euclidea tra le coordinate di un prodotto e quelle del prodotto ideale x_{sk} del soggetto s , r_{is} è il rango assegnato all'elemento i del medesimo soggetto.

L'analisi esterna invece ha come obiettivo quello di desumere da misure di preferenza i punti ideali che i consumatori prendono in esame per formulare le loro valutazioni, e di determinare l'importanza che ogni dimensione della mappa assume. Per tale analisi si necessita quindi di giudizi di preferenza espressi in forma metrica e di un grado di valutazione della percezione del consumatore nel considerare i prodotti sottoposti a studio. Per l'analisi esterna inoltre si possono evidenziare quattro particolari modelli:

1. modello vettoriale
2. modello euclideo semplice
3. modello euclideo ponderato
4. modello euclideo generalizzato

Essi sono tutti modelli che rientrano nella categoria della regressione multipla, descritta nei paragrafi precedenti.

Ogni stima dei parametri si basa quindi sul metodo dei minimi quadrati e grazie a tale metodo e modello si possono ottenere stime relative alle coordinate dei punti ideali e dei pesi attribuiti alle diverse caratteristiche delle dimensioni della mappa di preferenza, ma non solo calcolare anche le misure di bontà di adattamento.

Oltre a ciò, dei quattro modelli solo quello vettoriale è di tipo lineare e ipotizza che le preferenze del consumatore sono in associazione in modo lineare con il livello assegnato agli attributi corrispondenti alle dimensioni della mappa. Per gli altri tre modelli la funzione che mette in relazione l'ipotesi appena descritta può essere non lineare e neppure monotona.

In seguito (CAP 5) sarà presentato un esempio di UNF e di MDS attraverso software statistico specifico per l'analisi e la costruzione di mappe di percezione, posizionamento e di preferenza.

Come nel metodo MDS anche nel metodo UNF molti passaggi matriciali richiedono processi operativi lunghi e a volte complessi, ora l'informatica a tale proposito sta offrendo un rilevante aiuto all'analista nell'ottenere, in tempi assai ridotti, risultati eccellenti offrendo di conseguenza nella gestione aziendale un'ottimizzazione di tempo e di costi.

4. Il metodo D.E.A. applicato al Marketing

4.1. Un uso alternativo della Data Envelopment Analysis (D.E.A.)

La DEA è una tecnica usata per misurare le prestazioni di unità produttive. Introdotta nel 1978 da Charnes, Cooper e Rhodes, essa permette di valutare l'efficienza di unità di produzione che vengono definite DMU (Decision Making Unit) relativamente ad un collettivo di unità DMU scelte per il confronto.

Normalmente tale tecnica trova impiego nell'analisi dell'efficienza di banche, ospedali, sistemi di trasporto, ecc...

Diversamente dalle normali analisi statistiche, che confrontano le DMU con la DMU media, la DEA analizza ogni DMU con le DMU più efficienti e ne calcola l'efficienza relativa su tali unità.

Tali unità funzionano usando certi input e producendo specifici output. Nel caso di una banca si ipotizza che gli input siano i cassieri e gli output un prodotto offerto quali ad esempio mutui e assegni.

Si dirà quindi che una data unità produttiva A è più efficiente della B se con la stessa quantità di input produce maggiori output, oppure se con input minori si riesce a produrre lo stesso output.

Tale tecnica in effetti è mirata soprattutto a valutazioni produttive dell'azienda e di una sua ottimizzazione delle risorse primarie per la produzione del prodotto (output). In questo capitolo si cercherà invece di presentare un'applicazione alternativa di tale metodo in riferimento al Marketing. Come già presentato nel capitolo 1, il Marketing concepisce il prodotto come un insieme aggregato e coerente di attributi che possono essere tangibili e/o intangibili ai quali un consumatore assegna un valore.

Infatti i beni vengono acquistati per il loro valore d'uso e per il loro valore simbolico. La seconda dimensione qui citata non è tuttavia riferita solamente ai beni da collezione ma trova ampia applicazione anche nei comuni beni di consumo, come quelli alimentari.

In merito a questa visione, in cui il prodotto è un insieme di attributi, si può assegnare ad ogni attributo un proprio valore d'uso e valore simbolico, la cui somma dei singoli valori degli n attributi andrà a rappresentare il valore d'uso totale e simbolico del prodotto esaminato.

In questo modo, nella fase di generazione di un prodotto in cui si deve decidere quali attributi dovranno costituire il prodotto medesimo si può adottare la tecnica DEA per valutare per esempio quale tra n attributi pressoché simili sia il più efficiente da inserire all'interno del prodotto.

In questo caso gli input saranno i consumatori, o meglio un campione di n potenziali consumatori che saranno selezionati dall'azienda stessa al fine di attribuire un giudizio sul valore d'uso e sul valore simbolico, degli attributi che gli saranno presentati. Gli output che si generano sono quindi tali valori (il simbolico e il valore d'uso). Prendiamo in esame un esempio di questo tipo:

Attributo	Input	Output	
	Consumatori	Valore d'uso	Valore Simbolico
A	100	1000	20
B	100	400	50
C	100	200	150

tab. 8 – Struttura tra unità/Input/Output

Nella tabella vengono riportati tre attributi considerati simili tra loro ma percepiti in modo differente dai consumatori. La domanda che viene posta è sapere quale sia il più efficiente per formare il prodotto. Come input abbiamo il nostro campione potenziale di consumatori e come output un giudizio che essi hanno espresso in merito al valore d'uso e al valore simbolico degli attributi sottoposti ad analisi.

I valori relativi al giudizio possono essere determinati con varie scale di misura, ordinali, intervallari ecc.. in questo caso si può considerare una scala intervallare a cui si assegna un valore da [1..10] considerando il valore maggiore come valore più importante.

Analizzando i dati contenuti nella tabella si nota che nessun attributo è dominato dagli altri.

Per valutare quindi l'efficienza possiamo procedere nel seguente modo:

“Preleviamo i giudizi espressi da 50 consumatori dall'attributo A in modo equidistribuito e gli assegnamo ad un nuovo attributo fittizio D, ipotizzando inoltre che i consumatori giudichino anche D in ugual modo (o con stessa efficienza equidistribuita).($50/100=0.5$ quindi $1000*0.5 = 500$ valore che si assegna al nuovo attributo relativo all'output valore d'uso).

Quindi D avrà come input 50 consumatori e come output un giudizio pari a 500 per il valore d'uso e 10 per il valore simbolico. Si fa la stessa cosa anche per C, ossia si preleva 50 consumatori e si assegna con le ipotesi formulate nel caso precedente, i valori di input e output sempre allo stesso attributo D ottenendo così un nuovo attributo con pari Input degli altri e con certi giudizi complessivi di output:

Attributo	Input	Output	
	Consumatori	Valore d'uso	Valore Simbolico
D	100	600	85

tab. 9 – Il punto virtuale

In questo caso si nota che D è dominante rispetto a B. D non esiste come attributo tuttavia esso può essere considerato come un attributo virtuale o potenziale che potrebbe in qualche modo esistere dalla combinazione di un certo numero di attributi. Infatti l'ipotesi di base della DEA si basa sul fatto che è possibile frazionare certe unità per crearne una di nuova con quantità di input e output proporzionali alle quantità delle unità coinvolte secondo la frazione attribuita.

Una rappresentazione grafica di quanto detto finora è la seguente:

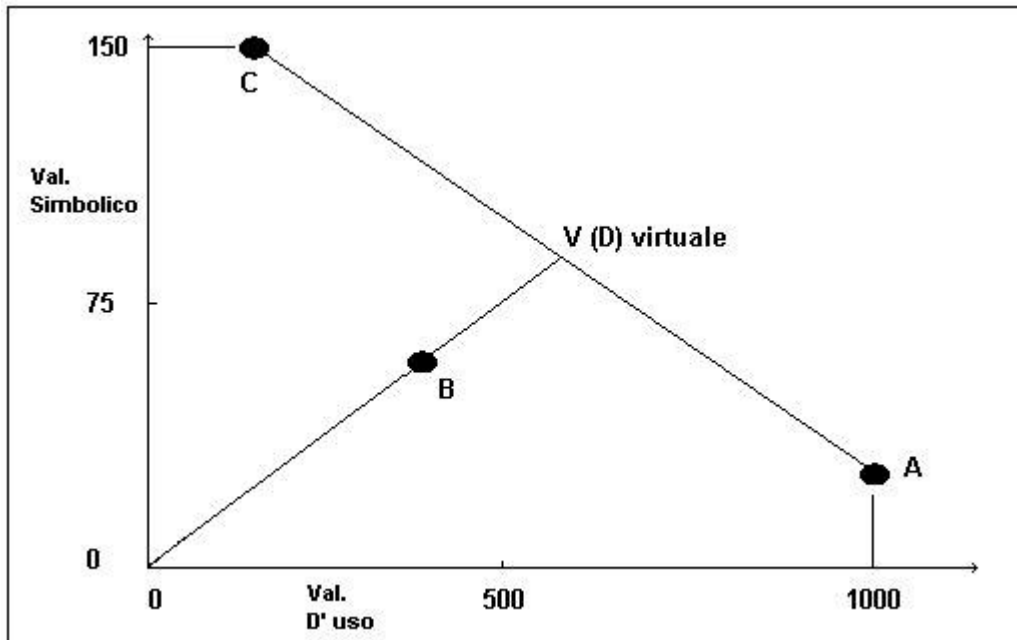


fig. 13 – La frontiera efficiente

Considerando tale figura (fig.13) in cui vengono rappresentati i tre attributi sul piano degli output in quanto gli input sono tutti uguali si può definire il segmento AC come l'insieme dei punti rappresentati gli attributi potenziali o virtuali ottenibili come combinazione di A e C. Tale segmento viene definito la frontiera efficiente. Il punto V rappresenta quindi uno dei potenziali attributi che dominano B. In particolare il punto V è il punto di incontro della retta passante per l'origine e per il punto B con il segmento AC, e tale punto è caratteristico per il fatto che offre output proporzionali a ciò che offre B. In conclusione la misura dell'efficienza di B viene misurata attraverso il rapporto di OB/OV . Infatti perché B sia efficiente bisogna moltiplicare il suo vettore di output per il fattore OV/OB .

Altre applicazioni possono essere la valutazione dell'atteggiamento del consumatori in merito sempre agli attributi da sottoporre ad analisi e in relazione con la funzione di Fishbein presentata nel capitolo 1 in cui si

afferma che l'atteggiamento del consumatore è pari alla sommatoria per tutti gli n attributi dell'importanza assegnata all'attributi i per la sua percezione da parte del consumatore: $A = \sum_{i=1}^n x_i w_i$ con x_i l'importanza attribuita e w_i la percezione. In questo caso gli input rimangono invariati e gli output saranno proprio x_i e w_i .

Il metodo DEA si fonda sui principi della programmazione lineare. Riprendendo l'esempio precedentemente fatto, si costruirà ora un modello idoneo a tale metodo.

Ricordando che un'unità è dominata se esiste una combinazione di altre che produce maggiori output con pari input oppure produce stesso output con minor input. A tale riferimento si deve scegliere una delle seguenti modalità, per esempio la seconda.

Per decidere quindi se B è dominato da A e C si associa ad A una variabile X_a che rappresenta la frazione di A usata in D e analogamente per C si associa una variabile X_c . Per esprimere il concetto che D usi meno Input si scrive la seguente relazione:

$$100X_a + 100 X_c \leq 100 - \varepsilon \text{ con } \varepsilon > 0.$$

Per esprimere che D offra la stessa quantità di output scriveremo le seguenti relazioni:

$$1000X_a + 200X_c \geq 400$$

$$20X_a + 150X_c \geq 50$$

Se il sistema definito da queste tre relazioni ammette soluzioni allora B è dominato da D e X_a , X_c rappresentano le quantità di A e C richieste per costruire D . Viceversa se non ammette soluzioni B non sarà dominato.

Tale sistema, tuttavia informa solamente se B è efficiente o pure no, ma non informa su qual è la sua efficienza.

Per risolvere tale problema, si introduce una nuova variabile y che rappresenta l'efficienza di B. In questo modo il sistema da risolvere diviene:

$$\begin{cases} \min y \\ 100X_a + 100X_c \leq 100y \\ 1000X_a + 200X_c \geq 400 \\ 20X_a + 150X_c \geq 50 \end{cases}$$

La giustificazione della funzione obiettivo si basa sul fatto che si sta cercando il migliore attributo virtuale, ossia quello che domina maggiormente B. Questo è permesso determinando il più piccolo valore di y . La soluzione ottima quindi si ottiene risolvendo il sistema precedentemente descritto. In particolare la soluzione che tale sistema restituirà sarà $X_a = 0.342466$, $X_c = 0.287671$, $y = 0.630137$. Quindi l'efficienza di B è 0.630137, ossia esiste un attributo virtuale che con circa il 63% dei consumatori offrirebbe le stesse qualità di B in termini di valore d'uso e valore simbolico. Tale processo diviene ricorsivo nella fase di valutazione degli altri attributi.

In merito a questo, l'informatica aiuta ad operare in modo sistematico tali valutazioni, generando automatismi ricorsivi per il calcolo dei dati di una programmazione lineare. Esistono infatti programmi che dai dati di Input e output determinano automaticamente le frontiere virtuali con i punti ottimi. Un programma di questi è Frontier Analyst della Banxia Software.

Riprendendo il solito esempio, verrà ora presentata alcune funzionalità di tale software nello svolgimento dell'esempio in questione.

Il software è specifico per tale metodo, e la versione che viene usata è quella professional 3.0.3.

Caricato il programma, attraverso un Wizard (fig.14), (una maschera pre-configurata che permette una semplicità di funzionamento del programma

all'utente) si introducono i parametri essenziali (label degli input,output e delle unità)

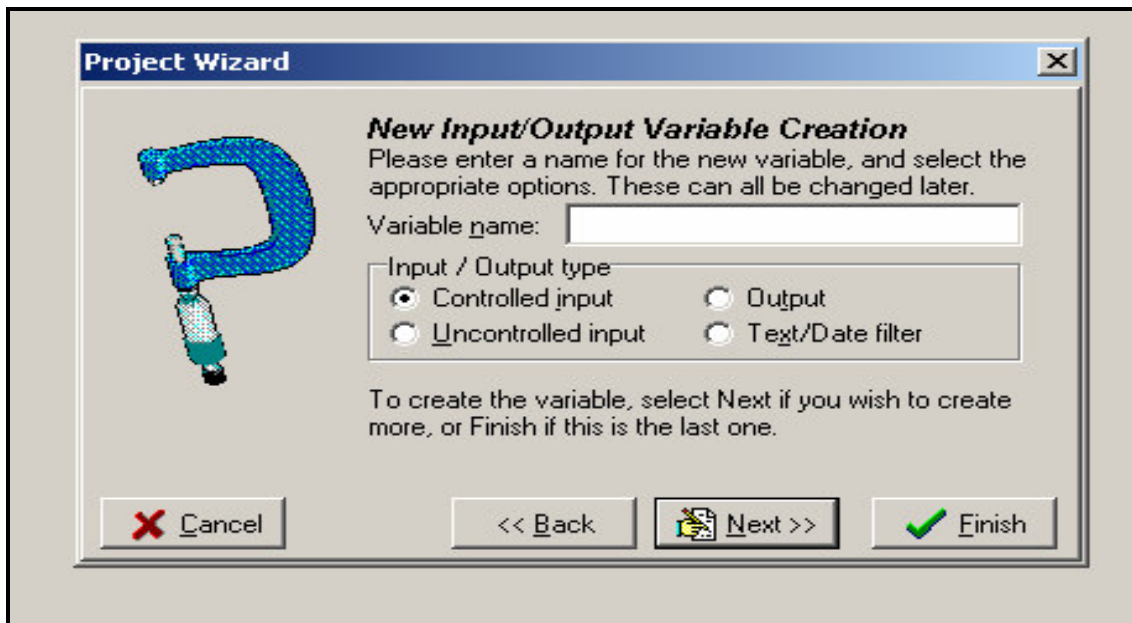


fig. 14 – Il Wizard

In seguito si introducono i dati relativi agli Input e Output originando così una tabella di dati come segue

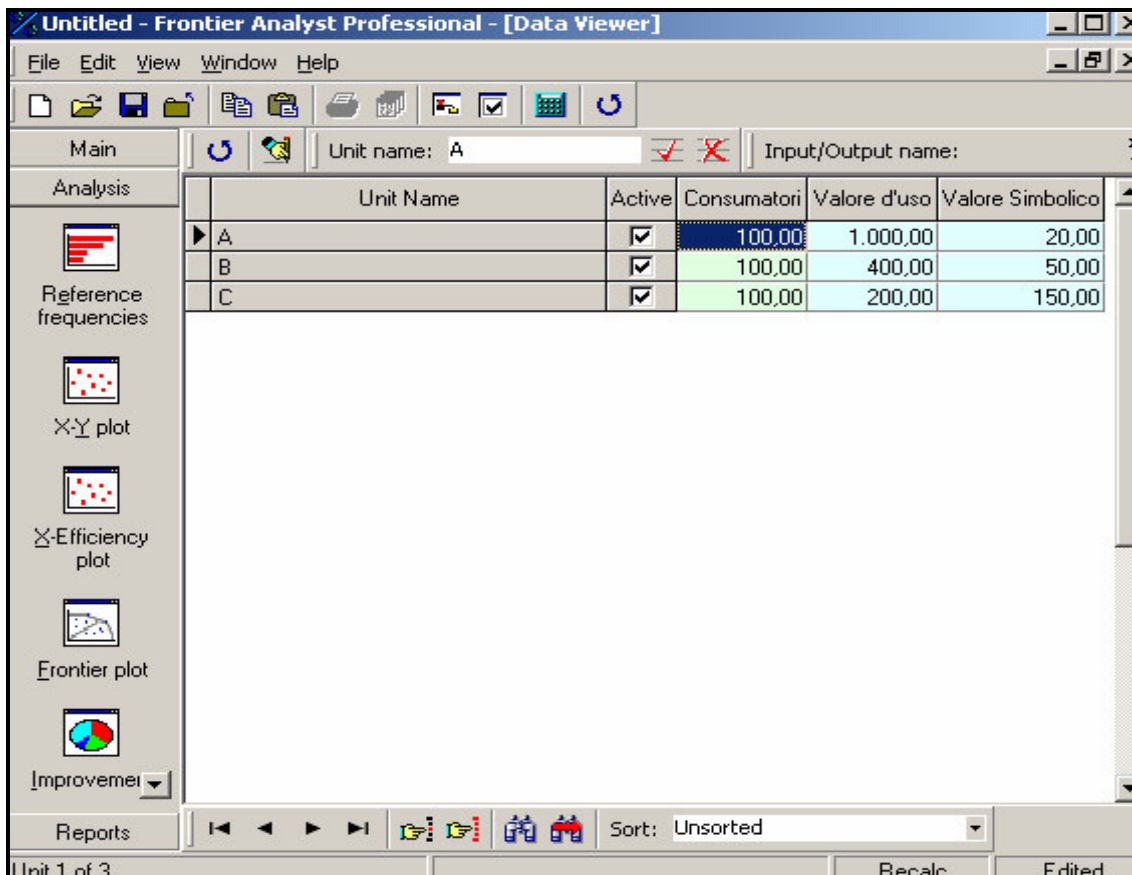


fig. 15 – Dati di base

Premendo sull'icona riportata sul lato destro con il nome Frontier Plot il programma procede con l'applicazione del metodo DEA. Si ottiene così il risultato finale che riporta graficamente la frontiera (fig.16)

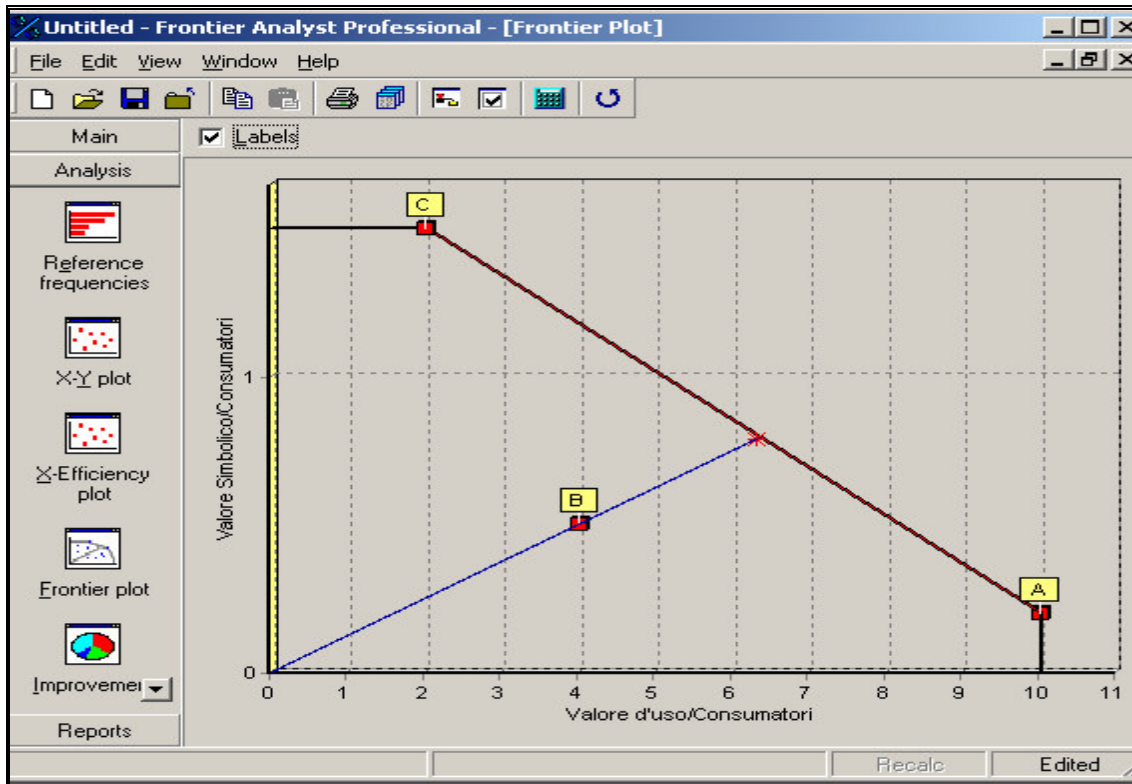


fig. 16 – Frontiera efficiente

E in una tabella le stime relative all'efficienza delle singole unità (attributi) (fig.17)

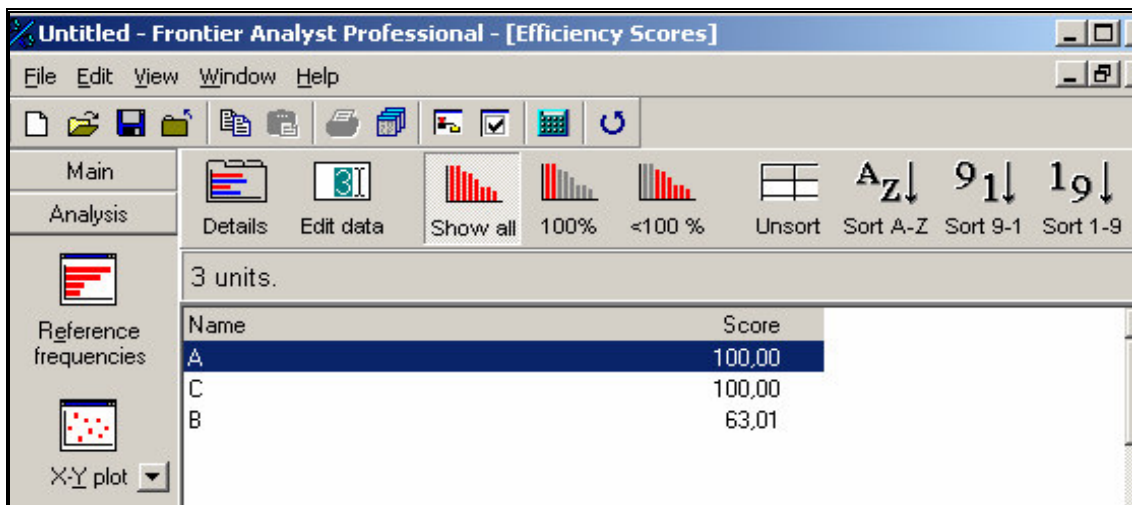


fig. 17 – efficienza delle unità (o attributi)

L'efficienza di B risulta circa al 63% come nell'esempi precedentemente presentato.

Bibliografia

Grandinetti R., “Concetti e Strumenti di Marketing”, Etas, 2002.

Cozzi G., Ferrero G., “Marketing: principi, metodi, tendenze evolutive”, G. Giappichelli Editore Torino, 2000.

Brasini S., Tassinari F., Tassinari G., “Marketing e pubblicità: Metodi di analisi statistica”, Il Mulino, 1999

Di Fonzo T., Lisi F., “Complementi di statistica economica”, Cleup, 2001.

Pace L., Salvan A., “Introduzione alla Statistica II: Inferenza, Verosimiglianza, Modelli”, CEDAM, 2001

Azzalini A., “Inferenza Statistica: Una presentazione basata sul concetto di verosimiglianza”, seconda edizione, Springer, 2001.

Dispense del corso di Statistica IV del Prof. Pollice:

- <http://www.dip-statistica.uniba.it/professori/pollice/appunti.htm> .

Dispense varie su statistica multivariata reperite dal link:

- www.sussidiario.it/ssi/href.php?keyword=http://digilander.libero.it/statistici/multivariata.htm .

Dispensa informativa sul metodo D.E.A. del dott. Bruni:

- <http://www.dis.uniroma1.it/~bruni/files/AMPLdea.pdf>

Dispense del software Frontier Analyst della Banxia Software:

- <http://www.banxia.com> .

Motore di ricerca internet usato: www.google.com

**3. L' INFORMATICA AL SUPPORTO
DELLA STATISTICA :
“IL PACKAGE STATISTICO”**

5. Il software statistico SYSTAT

In questo capitolo verrà presentato il software statistico systat nella versione 10.2 in merito ad alcune specifiche applicazioni di Marketing. Saranno presentate a grandi linee le caratteristiche e le funzionalità di base del seguente tool, evidenziando le funzionalità specifiche sulle tecniche di analisi dei cluster e della conjoint per la segmentazione, e sulle tecniche di MDS e Unfolding per il posizionamento.

5.1. Introduzione

Ai nostri giorni, l'informatica ha un ruolo rilevante nella società e nelle scienze. In merito a tale considerazione l'informatica attraverso pacchetti specifici di analisi ed elaborazione dei dati, ha permesso all'analista di poter svolgere il suo lavoro in modo assai più agevole e più ottimale. Grazie a tali software si sono utilizzare tecniche sempre più complesse e più performanti di analisi statistiche applicate casi di studio in ambiti economici, aziendali, sociali, medici, territoriali, astronomici, ecc...

Il Systat è un package statistico evoluto e completo in tutte le sue parti, offrendo possibilità di svolgere analisi in qualsiasi ambito. Sviluppato inizialmente dalla società SPSS è ora gestito dalla Systat Intellingent software. Come gli altri pacchetti (vedi S-PLUS, SPSS, SAS, STATA, STATISTICA, ecc...), esso ha una struttura orientata ad oggetti, in stile windows. Questo permette quindi all'utente utilizzatore una facilità di uso dei comandi di analisi e una maggiore assimilazione di tale programma. Systat 10.2 è dotato di modelli estremamente avanzati di diagrammi e grafici. Una volta creato il grafico, utilizzando uno dei numerosi object

offerti dal tool, il software consente una completa gestione sui vari attributi e parametri che compongono il grafico stesso come per esempio scale di misura, simboli, colorazioni, assi, coordinate, ecc....

Esiste un command log che tiene traccia delle procedure usate, permettendo quindi un richiamo automatico di esse senza dover richiamarle. Inoltre il software permette la creazione di programmi batch che eseguiti permettono di creare un risultato atteso a seconda delle esigenze richieste in modo automatico e ricorsivo. (fig.18)

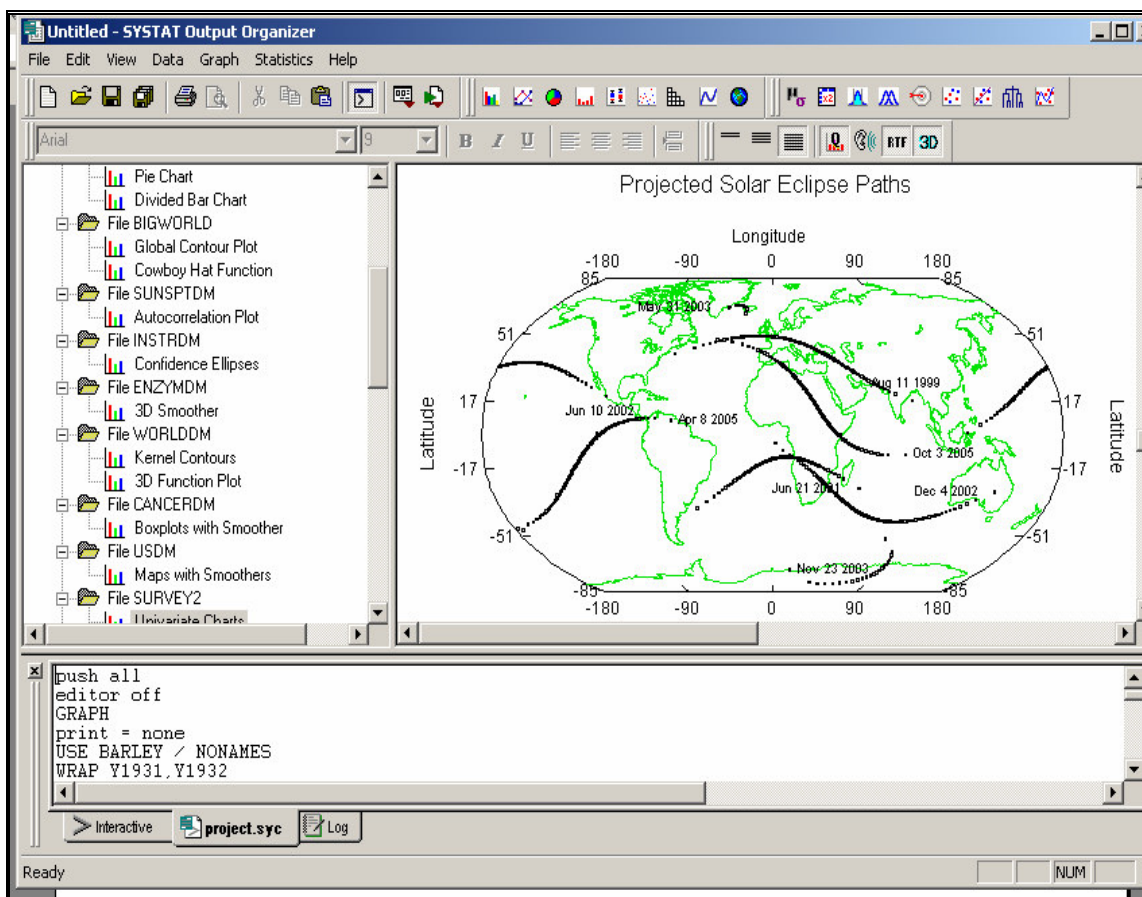


fig. 18 – interfaccia output di Systat

Tale form rappresenta l'output del software in cui viene suddiviso in tre particolari aree, una rappresentante i risultati dell'analisi (es. planisfero), una riportante un edit di scrittura per la digitazione di istruzioni batch, e un'altra che può essere definita Object Treeview in cui vengono elencati tutti gli oggetti creati e usati ai fini dell'analisi svolta.

Per quanto riguarda lo spreadsheet, esso si presenta come un normale foglio elettronico (Es. Excel) alla quale è possibile inserire per ogni variabile commenti caratteristiche particolare e raggruppamenti specifici tra esse su cui andare ad applicare le opportune tecniche statistiche offerte dal software. (fig.19)

euronew.SYD - SYSTAT Data

File Edit View Data Graph Statistics Help

Row: 11, Variable: MINLAT 47.3672

	MAPNUM	COUNTRY\$	MINLON	MAXLON	MINLAT	MAXLAT	LABLON	LABLAT
1	41.00000	Albania	19.27780	21.04190	39.69340	42.66220	20.15985	41.17780
2	42.00000	Andorra	1.41680	1.79160	42.41310	42.59230	1.60420	42.50270
3	43.00000	Austria	9.44320	17.09480	46.47710	49.00870	16.33300	48.23000
4	44.00000	Belgium	2.80600	6.22380	49.58470	51.46120	4.37000	50.87000
5	45.00000	Bulgaria	22.36490	28.66040	41.29860	44.19470	25.51265	42.74665
6	46.00000	Czechoslovak	11.91290	22.38700	47.77690	51.14090	14.43000	50.08000
7	47.00000	Denmark	8.02090	14.97790	54.59010	57.72130	12.56000	55.67000
8	48.00000	Faeroe Isl	-7.45860	-6.42230	61.39590	62.30090	-6.94045	61.84840
9	49.00000	Finland	19.58650	31.39470	59.85800	70.05830	25.00000	60.16000
10	50.00000	France	-4.63620	9.55350	41.38990	51.09520	2.33000	48.80000
11	51.00000	Germany	5.91510	14.85670	47.36720	54.93360	13.41000	52.50000
12	52.00000	Greece	19.65270	28.31860	34.89380	41.81140	23.72000	37.97000
13	53.00000	Hungary	16.04740	22.77290	45.80360	48.57220	19.00000	47.50000
14	55.00000	Ireland	-10.25910	-5.99230	51.46120	55.22670	-6.25000	53.33000
15	56.00000	Italy	6.70890	18.47300	36.70540	47.08960	12.45000	41.90000
16	57.00000	Luxembourg	5.62840	6.31200	49.54890	50.17530	5.90000	49.86000
17	58.00000	Malta	14.37160	14.58100	35.81380	35.99290	.	.
18	59.00000	Netherlands	3.53360	7.08380	50.75840	53.43220	4.88000	52.36700
19	60.00000	Norway	4.75740	31.01600	58.01460	71.11080	10.70000	59.95000
20	61.00000	Poland	14.08490	24.01870	49.12440	54.87020	21.00000	52.23000
21	62.00000	Portugal	-9.28890	-6.05840	37.06280	42.07770	-9.15000	38.73300

fig. 19 – Spreadsheet di Systat

5.2. Una panoramica generale sul tool

Possiamo suddividere l'applicazione creata con systat in quattro finestre principali (Rif. Fig.18-19):

1. Main Windows, in cui l'utilizzatore inserisce i propri comandi e ottiene l'output delle analisi statistiche
2. Spreadshett Windows, utilizzata dall'utente per la visualizzazione e l'inserimento dei dati

3. Graphic Window in cui vengono visualizzati i risultati delle analisi sui dati
4. Spin-Plot Window che permettono il cambiamento del punto di vista in maniera interattiva delle analisi di tipo spin plot.

Il software presenta tre grandi macro moduli: Data, Graph, Statistics.

Il primo modulo come mostra la fig. 20 mette a disposizione dell'utente una serie di comandi che permettono la manipolazione dei dati di origine permettendo trasformazioni, standardizzazioni, raggruppamenti, selezioni, interrogazioni, e altro ancora.

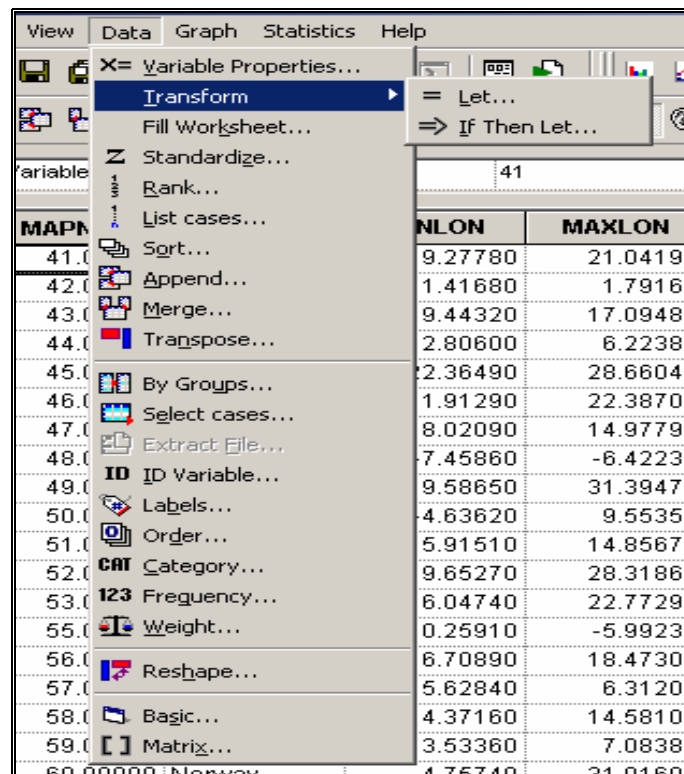


fig. 20 – Modulo Data – Manipolazione dei dati di base

In tale versione si possono infatti importare dati da database attraverso collegamenti ODBC, non solo, in questa versione sono state aggiunte funzionalità che permettono compatibilità con altri software di rilievo. Nella 10.2, si può importare anche i file di SAS, SPSS, BMDP, Excel, dBase, ASCII, ArcView, e molti ancora appartenenti ai formati standard.

Il modulo statistics, offre all'utente una serie di tecniche di analisi statistica da poter applicare ai fini del proprio studio sui dati presenti nel spreadsheet. La lista di tecniche disponibili offerta da systat 10.2 è estremamente vasta, e copre tutte le possibili aree applicative della statistica negli ambiti sociali, economici, geografici, territoriali, astronomici, aziendali, di mercato, di ottimizzazione della qualità, e altri ancora. Sintetizzando genericamente la lista possiamo individuare sottomoduli (Fig.21) che permettono analisi di:

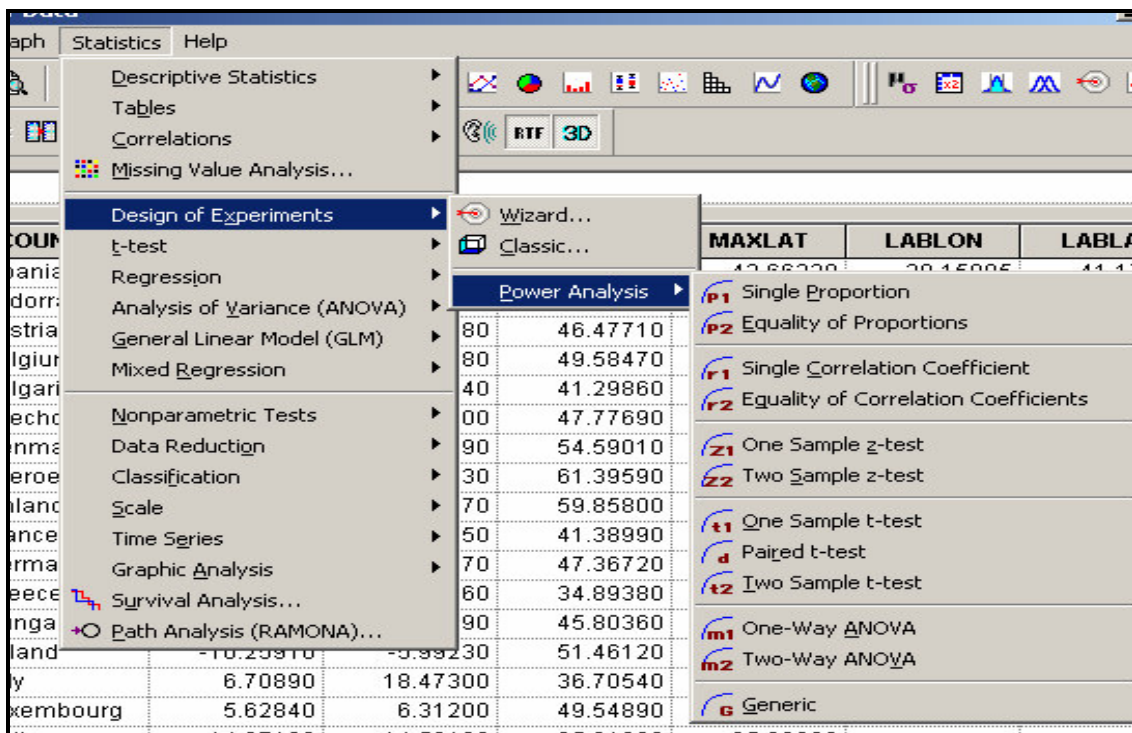


fig. 21 – Modulo Statistics

“Statistiche di tipo descrittivo, piano di un esperimento, clustering e di classificazione, di sopravvivenza, verifiche di ipotesi, analisi della varianza, test non parametrici, modelli di regressione, correlazione, studio di serie, analisi territoriali e geografici, analisi di posizionamento e studi di mappe percettive e di preferenza e molti ancora.”

Il modulo graph (fig.22) offre all'utente un'ampia gamma (che pochi altri software sono in grado di offrire) di grafici applicabili alle analisi svolte.

In particolare oltre a questa grande disponibilità grafica, il software presenta un'opzione che spesso negli altri software è assente. Tale opzione è la funzione Spin-Plot. Essa permette di rappresentare mediante marker in forma 3D i punti dove le coordinate cartesiane vengono individuate da tre colonne della tavola. In questo modo lo Spin-Plot permette la rotazione del grafico. Tale opzione trova utilità nelle analisi di fisica sperimentale in quanto è possibile osservare i dati sotto punti di vista diversi nello spazio considerato (fig.23).

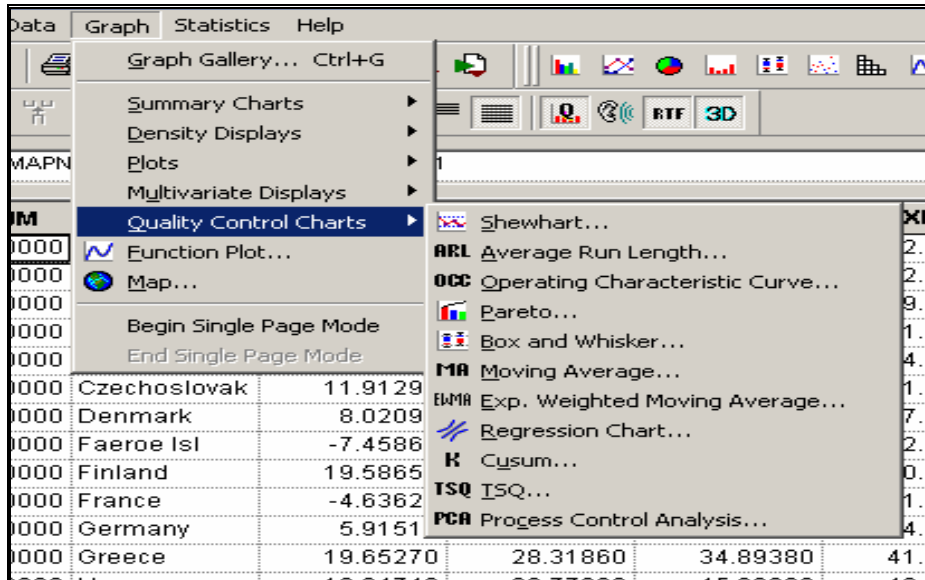


fig. 22 – Modulo Graph

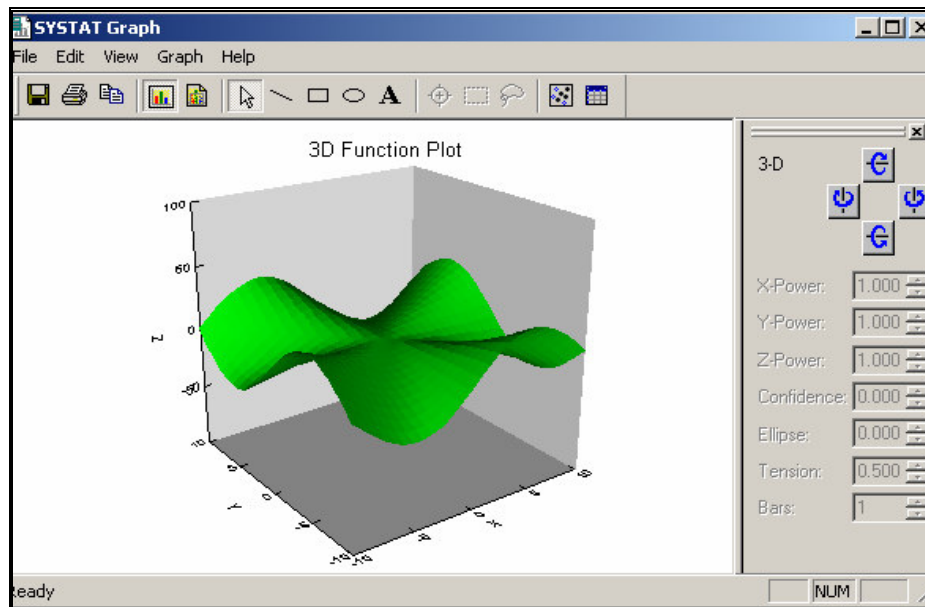


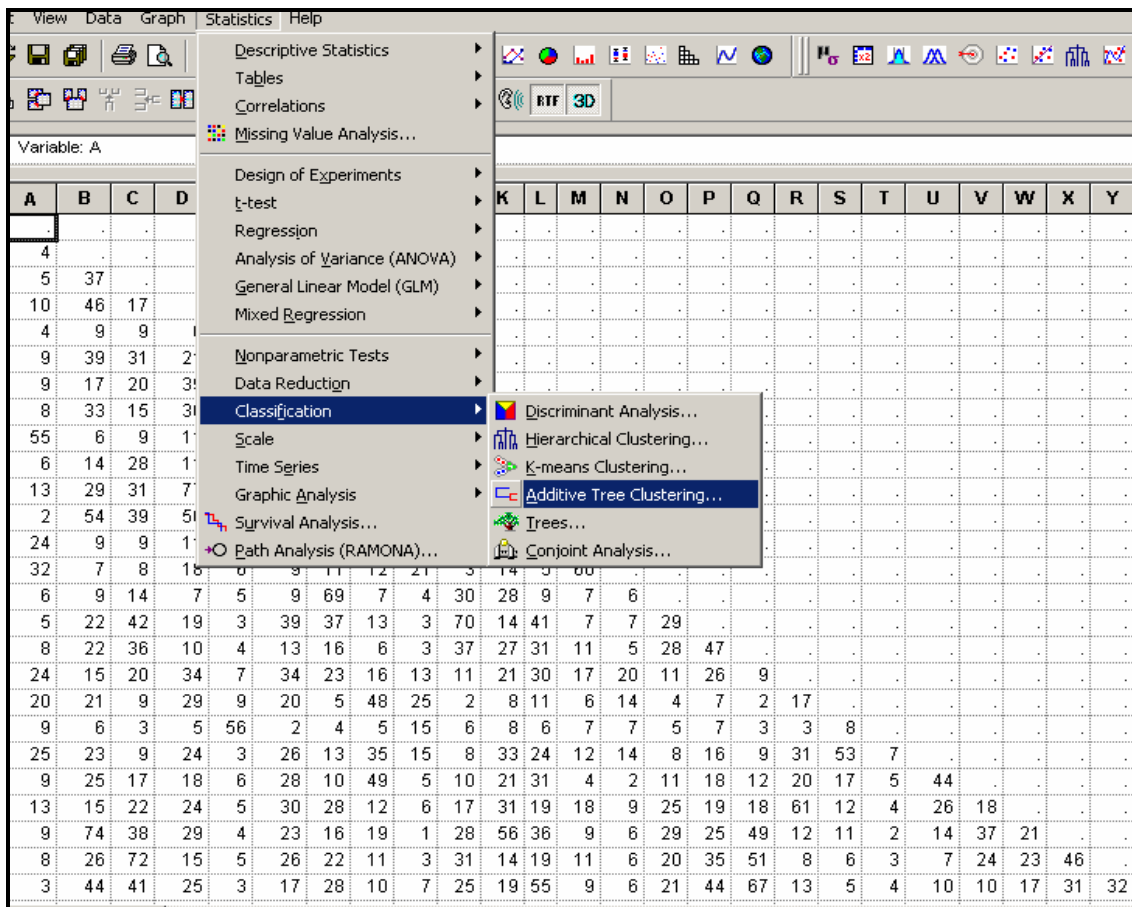
Fig. 23 SPIN-PLOT

5.3. La segmentazione in Systat

Systat offre un'ampia gamma di funzionalità per analizzare in modo statistico la segmentazione di mercato. In particolare in questo paragrafo verranno presentati degli esempi di segmentazione relativi alle tecniche di Cluster e Conjoint Analysis. I dati che saranno usati per gli esempi di segmentazione sono dati fittizi, al solo fine di creare accettabili simulazioni di tali metodi.

5.3.1. Cluster Analysis

L' esempio relativo a tale tecnica si basa sulla creazione del denogramma attraverso un modello additivo. Presi in considerazione i dati di partenza riportati nel spreadsheet in forma matriciale (fig.24):



	A	B	C	D	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y					
.					
4					
5	37					
10	46	17					
4	9	9					
9	39	31	2					
9	17	20	3					
8	33	15	31					
55	6	9	1					
6	14	28	1					
13	29	31	7					
2	54	39	51					
24	9	9	1					
32	7	8	18					
6	9	14	7	5	9	69	7	4	30	28	9	7	6						
5	22	42	19	3	39	37	13	3	70	14	41	7	7	29						
8	22	36	10	4	13	16	6	3	37	27	31	11	5	28	47	.	.	.						
24	15	20	34	7	34	23	16	13	11	21	30	17	20	11	26	9	.	.						
20	21	9	29	9	20	5	48	25	2	8	11	6	14	4	7	2	17	.						
9	6	3	5	56	2	4	5	15	6	8	6	7	7	5	7	3	3	8						
25	23	9	24	3	26	13	35	15	8	33	24	12	14	8	16	9	31	53	7					
9	25	17	18	6	28	10	49	5	10	21	31	4	2	11	18	12	20	17	5	44				
13	15	22	24	5	30	28	12	6	17	31	19	18	9	25	19	18	61	12	4	26	18			
9	74	38	29	4	23	16	19	1	28	56	36	9	6	29	25	49	12	11	2	14	37	21		
8	26	72	15	5	26	22	11	3	31	14	19	11	6	20	35	51	8	6	3	7	24	23	46	
3	44	41	25	3	17	28	10	7	25	19	55	9	6	21	44	67	13	5	4	10	10	17	31	32

fig. 24 – Dati in forma matriciale

si procede a determinare il denogramma selezionando da Statistics -> Clasification -> Additive Tree Clustering.

Comparirà una form (fig.25) che chiede di selezionare le variabili che si desidera considerare ai fini dell'analisi:

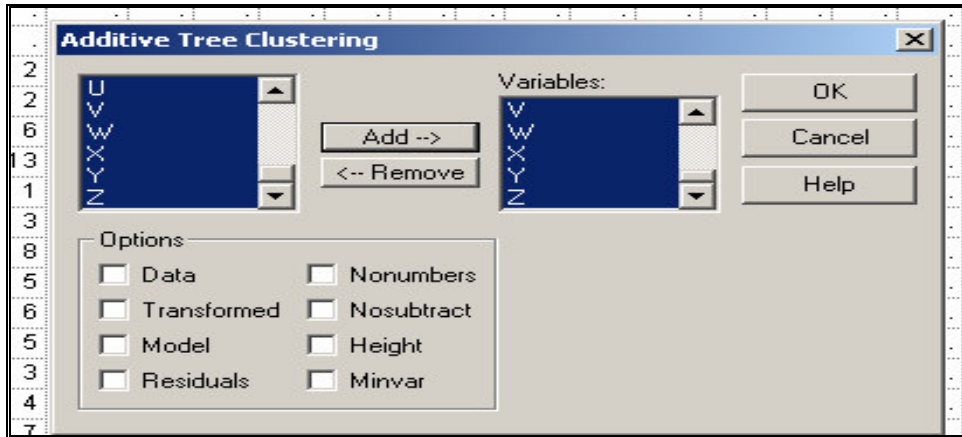


fig. 25 – Form di impostazione processo

Decise le variabili e premuto il tasto “OK”, nel tempo di qualche secondo l'analisi viene completata originando il denogramma relativo. Nella form dell'output vengono rappresentati i dati relativi alle varie trasformazioni relative al processo della CLA e il denogramma associato (fig.26-27).

	F	G	H	I	J
F	0.00000				
G	92.00000	0.00000			
H	74.00000	97.00000	0.00000		
I	98.00000	99.00000	93.00000	0.00000	
J	76.00000	84.00000	97.00000	101.00000	0.00000
K	88.00000	75.00000	94.00000	99.00000	80.00000
L	67.00000	90.00000	77.00000	92.00000	74.00000
M	92.00000	83.00000	98.00000	94.00000	95.00000
N	95.00000	93.00000	92.00000	93.00000	101.00000
O	95.00000	35.00000	97.00000	100.00000	74.00000
P	65.00000	67.00000	91.00000	101.00000	34.00000
Q	91.00000	88.00000	98.00000	101.00000	67.00000
R	70.00000	81.00000	88.00000	91.00000	93.00000
S	84.00000	99.00000	56.00000	79.00000	102.00000
T	102.00000	100.00000	99.00000	89.00000	98.00000
U	78.00000	91.00000	69.00000	89.00000	96.00000
V	76.00000	94.00000	55.00000	98.00000	94.00000
W	74.00000	76.00000	92.00000	98.00000	87.00000
X	81.00000	88.00000	85.00000	103.00000	76.00000
Y	78.00000	82.00000	93.00000	101.00000	73.00000
Z	87.00000	76.00000	94.00000	97.00000	79.00000

	K	L	M	N	O
K	0.00000				
L	74.00000	0.00000			
M	88.00000	97.00000	0.00000		
N	90.00000	99.00000	44.00000	0.00000	
O	76.00000	95.00000	97.00000	98.00000	0.00000

fig. 26 – Output dell' analisi

Additive Tree

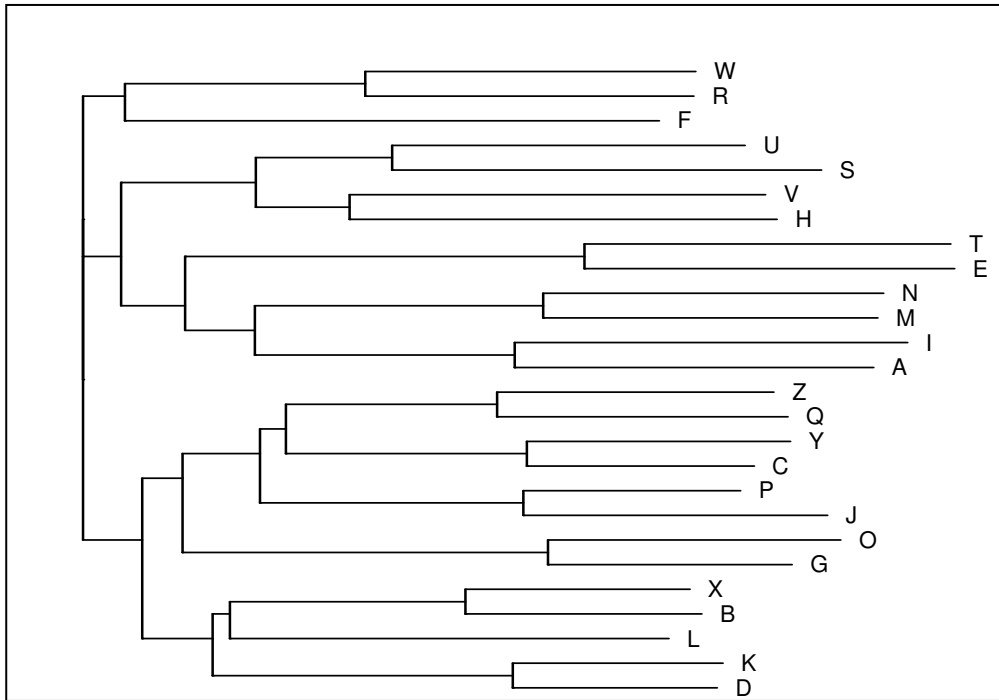


fig. 27 - Denogramma

5.3.2. Conjoint Analysis

L' esempio relativo a tale tecnica si basa su una sua applicazione classica nella scelta di prodotto attraverso un modello di tipo regressivo. Presi in considerazione i dati di partenza riportati nel spreadsheet (fig.28):

CARD	DESIGN\$	BRAND\$	PRICE	SEAL\$	GUARANT\$	RESPONSE	
1	1.00000	A	K2R	1.19000	NO	NO	13.00000
2	2.00000	A	Glory	1.39000	NO	YES	11.00000
3	3.00000	A	Bissell	1.59000	YES	NO	17.00000
4	4.00000	B	K2R	1.39000	YES	YES	2.00000
5	5.00000	B	Glory	1.59000	NO	NO	14.00000
6	6.00000	B	Bissell	1.19000	NO	NO	3.00000
7	7.00000	C	K2R	1.59000	NO	YES	12.00000
8	8.00000	C	Glory	1.19000	YES	NO	7.00000
9	9.00000	C	Bissell	1.39000	NO	NO	9.00000
10	10.00000	A	K2R	1.59000	YES	NO	18.00000
11	11.00000	A	Glory	1.19000	NO	YES	8.00000
12	12.00000	A	Bissell	1.39000	NO	NO	15.00000
13	13.00000	B	K2R	1.19000	NO	NO	4.00000
14	14.00000	B	Glory	1.39000	YES	NO	6.00000
15	15.00000	B	Bissell	1.59000	NO	YES	5.00000
16	16.00000	C	K2R	1.39000	NO	NO	10.00000
17	17.00000	C	Glory	1.59000	NO	NO	16.00000
18	18.00000	C	Bissell	1.19000	YES	YES	1.00000

fig. 28 – Dati fittizi di base per la simulazione

Applichiamo la tecnica COA selezionando dal menù Statistics -> classification -> Conjoint Analysis. Alla comparsa della form si seleziona la variabile dipendente e quelle indipendenti del nostro modello (fig.29).

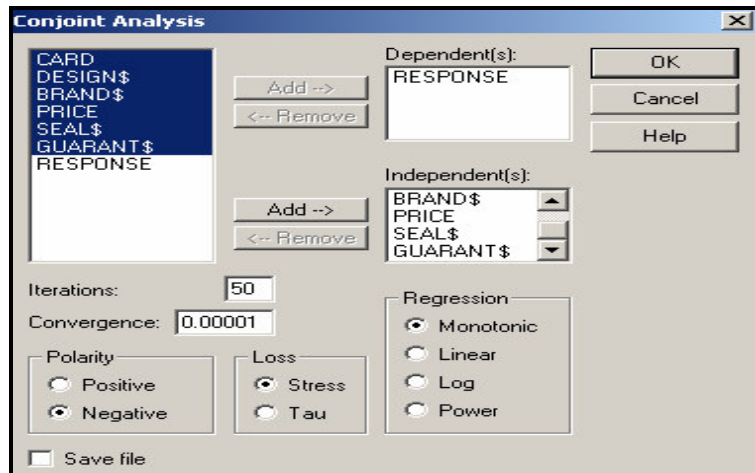


fig. 29 – Form di impostazione processo

Premendo il tasto “OK” il software procederà a determinare le stime dei parametri del modello e restituirà i risultati relativi all’analisi e i grafici associati. Di seguito riportiamo solo alcuni dei grafici relativi alla stima delle part-worths degli attributi (fig.30)

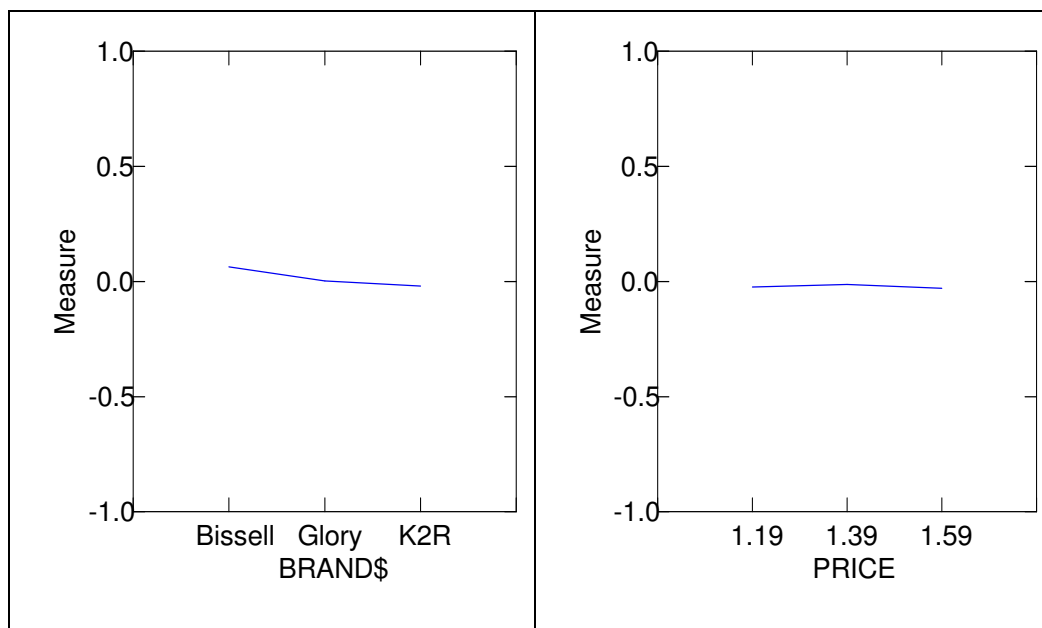


fig. 30 – Part-worths degli attributi Price e Brand

5.4. Il posizionamento di prodotto in Systat

Systat offre anche in questo caso un'ampia gamma di funzionalità per analizzare in modo statistico il posizionamento di prodotto. In particolare in questo paragrafo verranno presentati degli esempi di posizionamento relativi alle tecniche di MDS e Unfolding. I dati che saranno usati per gli esempi di posizionamento sono dati fittizi, al solo fine di creare accettabili simulazioni di tali metodi.

5.4.1. MDS

L' esempio relativo a tale tecnica si basa sulla creazione della mappa percettiva attraverso l' MDS per differenze individuali. Presi in esame i dati di partenza riportati in uno spreadsheet in forma matriciale (fig.31) si procede a determinare la mappa attraverso la selezione statistics -> Scale -> Multidimensional Scaling.

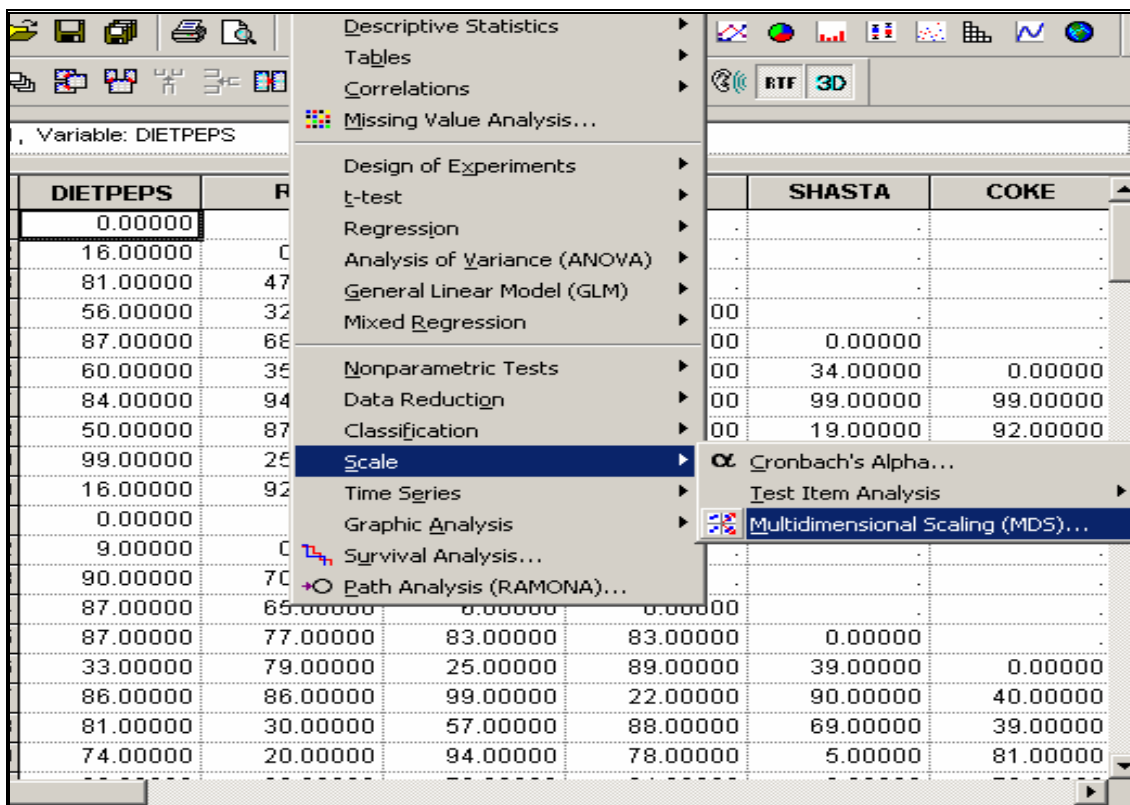


fig. 31 – Menù MDS

Successivamente comparirà una form nella quale si deve selezionare le variabili da considerare ai fini dell'analisi e impostare le dimensioni di rappresentazione della mappa. Gli altri parametri sono già settati di default per il seguente caso (fig.32).

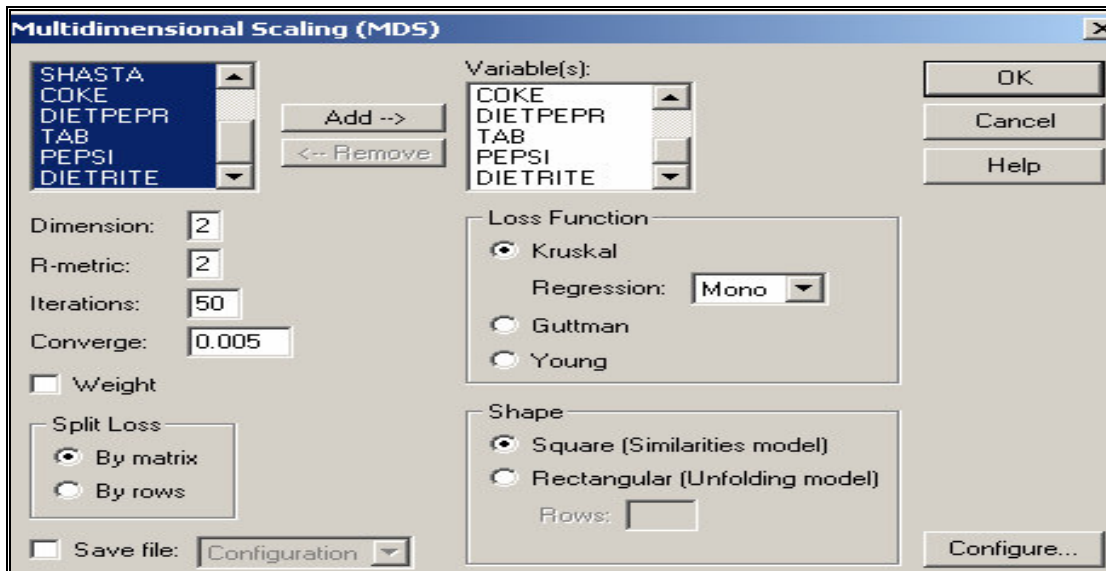


fig. 32 – Form di impostazione processo

Premendo il tasto “OK” il software elaborerà la procedura MDS e come output genera le stime dei valori relativi a tale metodo, quali per esempio lo STRESS e le coordinate della mappa, compreso il suo grafico (fig.33-34).

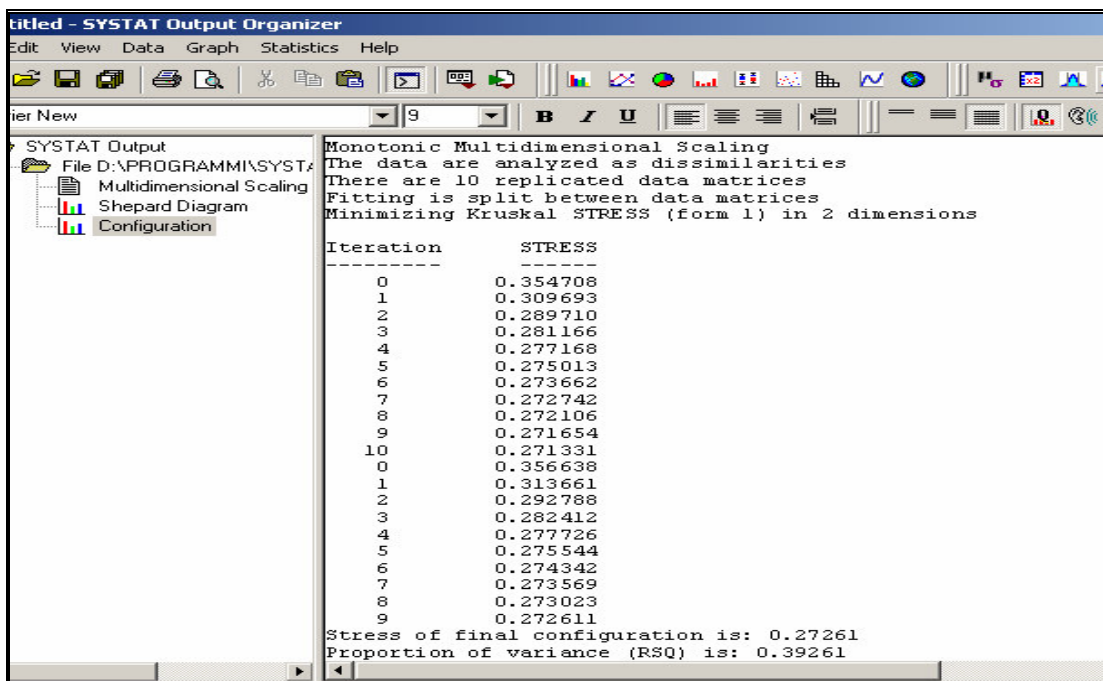


fig. 33 - Output

Coordinate della mappa: Coordinates in 2 dimensions

Variable	Dimension	
-----	-----	-----
	1	2
DIETPEPS	.14	-.72
RC	.15	.72
YUKON	.55	.25
PEPPER	1.33	.35
SHASTA	-.23	.99
COKE	-1.02	.24
DIETPEPR	.92	-.79
TAB	-.92	-.74
PEPSI	-.62	.61
DIETRITE	-.31	-.91

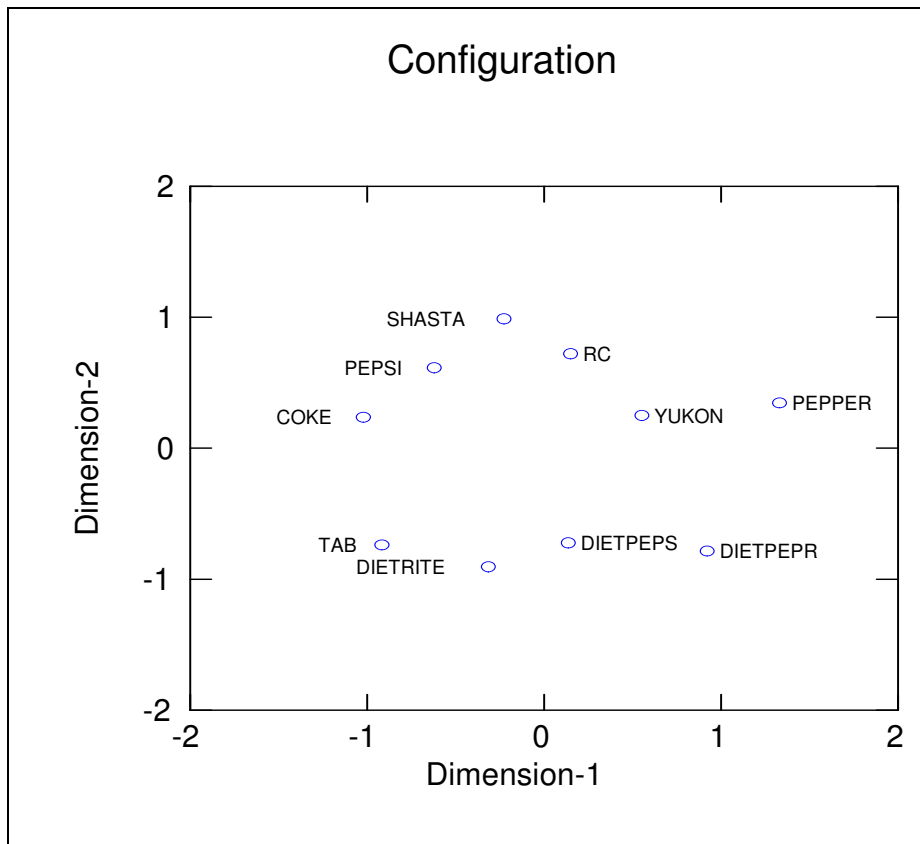


fig. 34 – Output Grafico

5.4.2. Unfolding

L' esempio relativo a tale tecnica si basa sulla creazione di mappa attraverso il metodo UNF nell'analisi esterna. Presi in esame i dati di partenza riportati in uno spreadsheet (fig.35) si procede a determinare la mappa attraverso la selezione statistics -> Data Reduction -> Perceptual Mapping.

np.synd - SYSTAT Data

dit View Data Graph Statistics Help

13, Variable: DIM2

SYMPTOM\$	DIM1	DIM2	LYME	MALARIA	YELLOW	RABIES	FLU
FEVER	0.01000	0.25000	1.00000	1.00000	1.00000	1.00000	1.00000
DIARRHEA	1.10000	-0.45000	0.00000	0.00000	0.00000	0.00000	1.00000
CONSTIP	1.18000	0.63000	0.00000	0.00000	1.00000	0.00000	0.00000
NAUSEA	0.38000	-0.56000	0.00000	1.00000	1.00000	0.00000	1.00000
STUFFY	-1.37000	0.27000	0.00000	0.00000	1.00000	0.00000	1.00000
CHEST	-0.37000	0.72000	0.00000	0.00000	1.00000	0.00000	1.00000
HEADACHE	0.13000	0.17000	1.00000	1.00000	1.00000	1.00000	1.00000

fig. 35 – Dati di base

Si presenterà come per gli altri casi una form che chiede di selezionare le variabili dipendenti e quelle indipendenti del nostro modello di analisi, in particolare le variabili indipendenti sono rappresentate dalle dimensioni della mappa. (fig.36)

Method: Vector

Options: Star

Dimens: MDPREF

Polarity: Positive, Negative

Dependent(s): YELLOW, RABIES, FLU

Independent(s): DIM1, DIM2

Buttons: Add -->, <-- Remove, OK, Cancel, Help

fig. 36 – Form di impostazione processo

Il metodo che useremo sarà quello vettoriale. Premuto “OK” il software inizierà ad elaborare il processo selezionato. L’ output corrispondente è la stima dei relativi parametri del modello e la generazione della mappa relativa (fig.37-38).

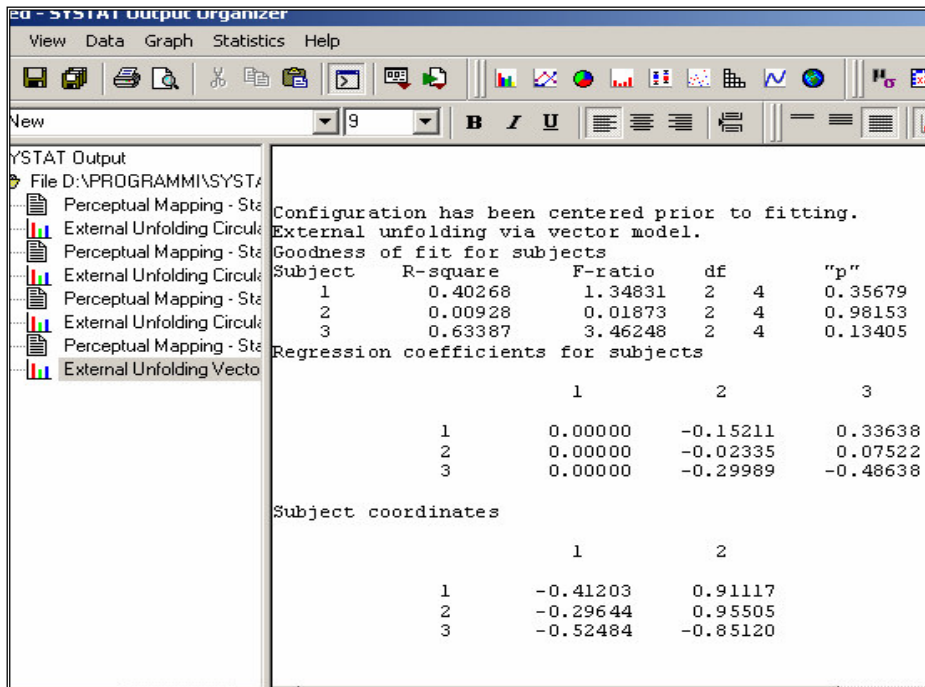


fig. 37 - Output

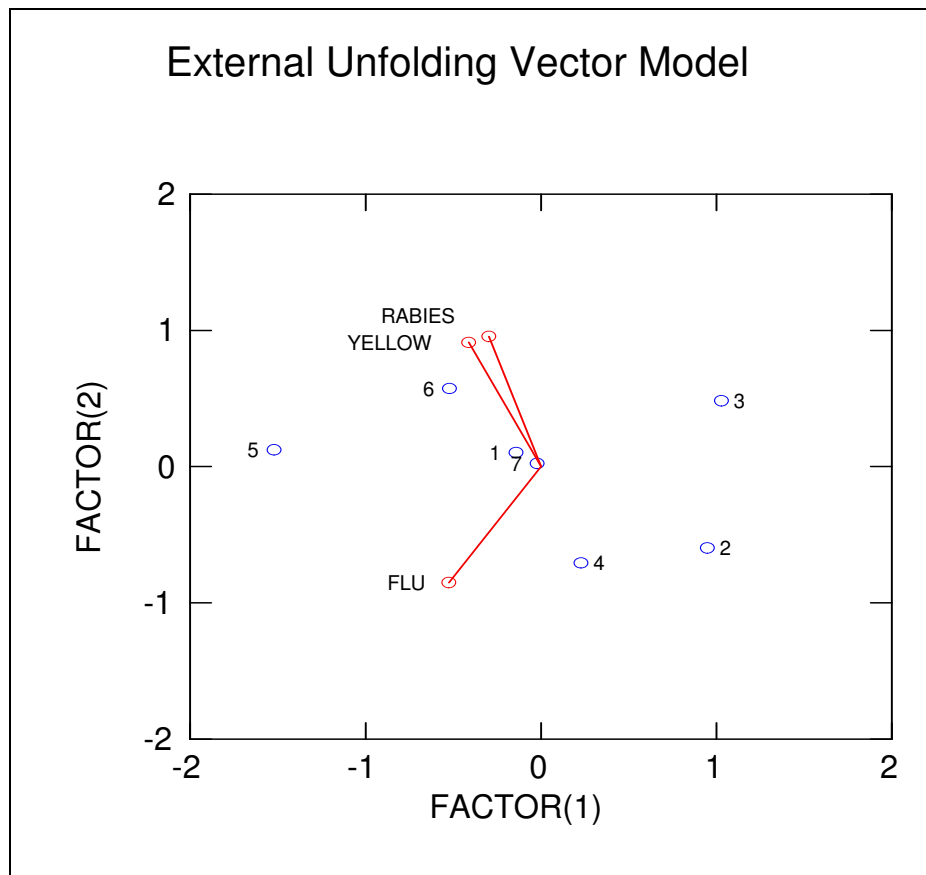


fig. 38 - Output Grafico

Systat offre molti altri metodi di stima di mappe di preferenza e/o di percezione. Per esempio in riferimento a quanto appena trattato, nel caso si

volesse determinare un'analisi con metodi UNF non metrici, è sufficiente ripercorrere i passaggi descritti precedentemente per il metodo MDS e nella form appropriata selezionare la voce Rectangular (Unfolding Model). Il processo di analisi che Systat eseguirà è quello di UNF non metrico (fig. 32).

Bibliografia

Brasini S., Tassinari F., Tassinari G., "Marketing e pubblicità: Metodi di analisi statistica", Il Mulino, 1999

Manuale del software Systat della Systat Intelligent Software: <http://www.systat.com> .

Materiale informativo reperito dai link:

- <http://webmaker.web.cern.ch/WebMaker/examples/ma-thesis/>
- http://www.lab2000.com/Lab_01_05/analisi_e_computer.htm
- <http://www.net-eng.it/eng/products/systat/fsystat.html>
- <http://www.systat.com/products/Systat/?sec=1051>
- <http://www.sxst.it/snodo.htm>

Motore di ricerca internet usato: www.google.com

Luciano Baù