

Analisi multivariata: Analisi cluster e Analisi fattoriale

Erika Pignatti
Dipartimento di Economia e Ingegneria Agrarie
A.A. 2007-2008

La segmentazione

- Il concetto di macro-segmentazione
- Il concetto di micro-segmentazione
 - suddivisione del mercato
 - scelta
 - posizionamento

Contesto strategico

- Elaborare una strategia di presenza su determinati mercati
 - scomposizione del mercato (segmentazione)
 - scelta della strategia
 - scelta del mercato di riferimento

Posizionamento

- Percezione del prodotto o della marca da parte dei clienti, relativamente alla marca o ai prodotti concorrenti

Posizionamento

- identificazione delle **dimensioni** rilevanti per la percezione
- identificazione dei termini di riferimento (es. concorrenti)
- identificazione e comprensione di motivazioni e aspettative

Posizionamento

- dipende dalla segmentazione
- è sempre in termini relativi
- è influenzato dalle decisioni
 - dell'impresa
 - dei concorrenti
- implica una differenziazione

Dimensioni di posizionamento

- attributi del prodotto
 - benefici (valori) ricercati
 - occasioni e modalità di uso
 - caratteristiche della concorrenza (sulla base della similarità/differenza)
- BENCHMARKING

Dimensioni di posizionamento

- Nella fase analitica possono essere impiegate opportune tecniche di marketing, finalizzate a costruire le cosiddette

MAPPE DELLE PERCEZIONI

(PERCEPTUAL MAPPING)

Segmentazione

- Segmentazione: processo mediante il quale le imprese suddividono la domanda in insiemi di clienti potenziali (similarità interna, diversità esterna: funzione di domanda ben definita)
- L'azienda sceglie i gruppi di clienti su cui personalizzare l'offerta, differenziando le sue strategie di marketing (frammentazione e sofisticazione della clientela)

Segmentazione

■ Vantaggi di una corretta segmentazione:

- Aiuto all'azienda nella definizione del mercato, in termini di bisogni del cliente e obiettivi aziendali
- Rafforzamento della capacità di percezione e interpretazione dei mutamenti
- Analisi dei punti di forza e debolezza della propria offerta nei confronti dei concorrenti
- Razionalizzazione del portafoglio prodotti (NO cannibalismo)
- Se opportunamente impiegata, crea barriere all'ingresso di ulteriori concorrenti
- Maggiore precisione nella misura degli effetti delle azioni di marketing su vendite e profitti

Segmentazione

■ Prerequisiti per l'attuazione della segmentazione:

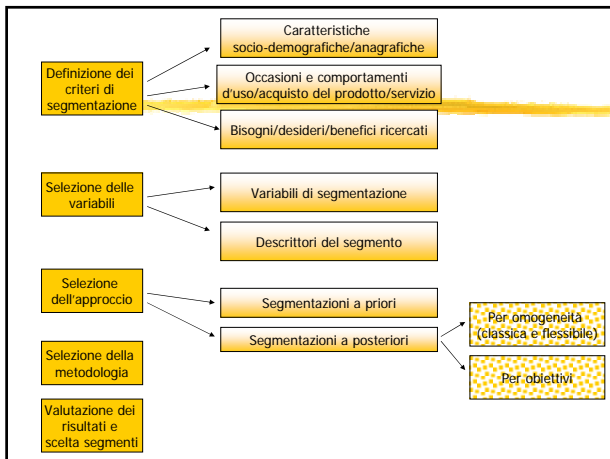
- Omogeneità/Disomogeneità segmenti
- Dimensione e redditività segmento (potenziale del segmento, ritorno sugli investimenti)
- Accessibilità al segmento (raggiungibilità commerciale)
- Durata del segmento (arco di tempo in cui le caratteristiche del segmento restano stabili)

■ Approccio di segmentazione: indifferenziato (mass market target)/macro/micro

Segmentazione

■ Fasi della segmentazione

- Definizione dei criteri di segmentazione
- Selezione delle variabili utili per costruire e descrivere i segmenti
- Scelta dell'approccio di segmentazione
- Scelta della metodologia quantitativa più adatta
- Valutazione dei risultati e **scelta dei segmenti su cui concentrare le risorse dell'azienda**



Segmentazione: criteri

- 1) Caratteri socio-demografici/anagrafici:
ad essi sono correlate le percezioni di bisogno, i benefici ricercati, i comportamenti d'uso e le preferenze di prodotto differenziate (analogia del profilo socio-demografico=analogia nei comportamenti d'uso). Spesso sono trattati come variabili complementari
- 2) e 3) Occasioni e comportamenti/Bisogni, benefici ricercati:
richiedono un'indagine sulla strumentalità del prodotto/servizio, alla quale si aggiunge in seguito l'analisi dei caratteri socio-demografici (per verificarne la rilevanza in sede di segmentazione della domanda).

Segmentazione: variabili

La scelta delle variabili per la segmentazione dipende dalla peculiarità del prodotto/servizio indagato e dagli obiettivi decisionali a cui la ricerca deve fungere da supporto:

	CONSUMO	INDUSTRIALE
Situazioni d'acquisto, uso e consumo	Comportamento d'acquisto Modalità d'uso e acquisto Reazione a nuove concezioni Rilevanza dei criteri di scelta	Volume d'acquisto Fedeltà alla marca Motivazioni d'acquisto
Caratteristiche socio-demografiche/anagrafiche	Età, sesso, titolo di studio Localizzazione geografica Ampiezza del nucleo familiare	Fase del ciclo di vita Stili di vita Professione e livello di reddito
Bisogni/desideri/benefici ricercati	Funzionali Simbolici Esperienziali	Funzionali Simbolici

Segmentazione: approccio

■ Segmentazione a priori:

le caratteristiche del segmento sono definite sulla base di informazioni già in possesso (si basano generalmente sulle caratteristiche socio-demografiche/anagrafiche)

■ Segmentazione a posteriori:

le caratteristiche del segmento emergono dall'applicazione di tecniche di analisi quantitativa e non sono note. Si dividono in tecniche di segmentazione *per omogeneità* e *per obiettivi*.

Segmentazione: approccio

■ Tecniche di segmentazione *per omogeneità*

Gli elementi della popolazione target sono raggruppati in base alla loro similarità, riferita ad un insieme di variabili (bisogni, attitudini, benefici ricercati, motivazioni d'uso, stili di vita...), in gruppi caratterizzati da ELEVATA VARIABILITA' ESTERNA E ALTA OMOGENEITA' INTERNA.

■ Tecniche di segmentazione *per obiettivi*

Il target è diviso in sub-popolazioni usando una variabile dipendente nota (es. frequenza d'acquisto, sensibilità a campagne pubblicitarie..) e si individuano variabili esplicative che influiscono su essa in modo rilevante

Segmentazione: approccio

■ Tecniche di segmentazione per OMOGENEITA':

> Modalità classica (Analisi fattoriale+cluster analysis)

Richiesta di valutazioni di importanza su caratteristiche del prodotto/servizio indagato, sintesi in macro-elementi di scelta di tali caratteristiche mediante l'analisi fattoriale e costruzione di gruppi omogenei del target rilevante attraverso l'analisi cluster, applicata sui macro-elementi individuati (incrocio finale dei cluster con le caratteristiche demografiche del target, per verificarne la raggiungibilità)

Segmentazione: approccio

■ Tecniche di segmentazione per OMOGENEITA':

> Modalità flessibile (Conjoint analysis+cluster analysis)

Valutazione di profili d'offerta globali e scomposizione dei giudizi globali in valutazioni di utilità degli attributi considerati. La scomposizione avviene attraverso la Conjoint Analysis, alla quale si aggiunge l'analisi cluster per ottenere gruppi omogenei in relazione alle valutazioni di utilità fornite implicitamente.

Metodi di segmentazione

■ Tecniche basate sugli attributi

- Analisi di raggruppamento (Cluster analysis)
- Analisi fattoriale + cluster

■ Tecniche basate sulla valutazione comparativa degli oggetti

- Analisi congiunta (Conjoint analysis) + cluster

Metodi di posizionamento

■ Tecniche basate sugli attributi

- Analisi discriminante
- Analisi delle corrispondenze

■ Tecniche basate sulla valutazione comparativa degli oggetti

- Scalatura multidimensionale (Multidimensional scaling)

Tecniche di analisi multivariata

- Metodi statistici che analizzano simultaneamente più attributi/caratteristiche di ogni "oggetto" analizzato
- Multivariata: molteplicità di variabili considerate nello studio degli oggetti (multivariable)/casualità delle variabili, interrelate in modo che i loro effetti non possano essere interpretati significativamente se separate.

Tecniche di analisi multivariata

- Elementi base dell'analisi multivariata:
 - Variata: combinazione di variabili con pesi determinati empiricamente (risultato: valore che rappresenta la combinazione di tutte le variabili che meglio conseguono l'obiettivo dell'analisi)

Tecniche di analisi multivariata

- Elementi base dell'analisi multivariata:
 - Scale di misura: non metriche o qualitative (descrivono gli oggetti sulla base della presenza/assenza dell'attributo); metriche o quantitative (descrivono gli oggetti sulla base del "quantitativo" o grado di attributo posseduto)

Tecniche di analisi multivariata

- Scale di misura non metriche: **nominali** (attribuzione di numeri agli oggetti per identificarli, senza alcuna connotazione); **ordinali** (classificazione degli oggetti sulla base della quantità di attributo posseduto: più di-meno di)
- Scale di misura metriche: interval scales; ratio scales

Tecniche di analisi multivariata

- Classificazione delle variabili: tecniche di dipendenza (variabili dipendenti-variabile indipendente); tecniche di **interdipendenza** (nessuna relazione tra le variabili, che sono analizzate simultaneamente)
- Numero di variabili dipendenti
- Modalità di misurazione delle variabili dipendenti
- Base della struttura dei rapporti

Tecniche di analisi multivariata

- **Approccio alla tecniche di Analisi multivariata:**
Definizione del problema di ricerca, degli obiettivi dell'analisi e della tecnica da usare (sulla base degli elementi a disposizione)
Definizione di un piano di analisi (basi di partenza: dimensione del campione; metodi estimativi..)
Valutazione del modello di analisi (accettabilità statistica)
Applicazione del modello
Validazione del modello (affinchè i risultati forniti si possano considerare generalizzabili e altamente descrittivi dei dati in possesso)

Analisi cluster

Analisi di raggruppamento (Cluster Analysis)

- Analisi utilizzata nel momento in cui si vuole individuare una serie di sottogruppi di oggetti significativamente diversi all'interno di un campione
- I gruppi non sono predefiniti. Il risultato dell'analisi è una classificazione sulla base delle similarità esistenti tra gli oggetti esaminati.
- Passaggi dell'analisi cluster (almeno 3):
 - Misura di similarità o associazioni esistenti tra gli oggetti in studio, al fine di comprendere il numero di gruppi presente nel campione
 - Processo di clustering (ripartizione degli oggetti in gruppi)
 - Definizione del profilo degli appartenenti ai gruppi

Analisi di raggruppamento (Cluster Analysis)

- **Scopo** dell'analisi:
Consente di ottenere l'identificazione di gruppi omogenei al fine di supportare decisioni manageriali, strategiche e/o operative
- ❖ Analisi di segmentazione (ricerca di gruppi di entità che condividano determinate caratteristiche: attitudini, benefici ricercati..)
- ❖ Identificazione di nicchie competitive per la ricerca di opportunità per nuovi prodotti (attraverso la classificazione di marchi e prodotti, e con l'ausilio delle tecniche di posizionamento)
- ❖ Individuazione di aree test per la verifica di politiche promozionali o di vendita (creazione di cluster costituiti da aree-test omogenee, al fine di poter estendere i risultati conseguiti in un'area su tutte quelle appartenenti al medesimo cluster).

Analisi di raggruppamento (Cluster Analysis)

■ **Caratteristiche** dell'analisi:

Fondata sulla capacità dei cluster di mettere in evidenza:

- ❖ **compattezza interna** (minimizzazione della varianza tra gli elementi appartenenti allo stesso gruppo)
- ❖ **respingenza esterna** (massimizzazione della varianza tra gruppi diversi)

Sintesi delle osservazioni (n cluster $<$ n oggetti)

Intersezione tra due cluster = insieme vuoto

Unione dei cluster = insieme di partenza

Analisi di raggruppamento (Cluster Analysis)

- **INDIVIDUAZIONE DEI GRUPPI** sulla base di alcune variabili (criteri di similarità) rilevanti (misura della similarità e individuazione del numero di gruppi; clusterizzazione)

- **PROFILAZIONE DEI GRUPPI** (al fine di estrapolare elementi caratteristici in grado di descriverli)

Analisi di raggruppamento (Cluster Analysis)

- Necessità di identificare la partizione ottimale del gruppo di oggetti (ottimizzazione di una *funzione obiettivo*)
- All'aumentare del numero di oggetti, aumenta il numero di possibili partizioni
- Al fine di risolvere il problema, vengono impiegati algoritmi che consentono di giungere alla partizione maggiormente prossima all'ottimale, riducendo il numero di alternative possibili

Analisi di raggruppamento (Cluster Analysis)

Tecniche di classificazione:

- Algoritmi di **classificazione diretta**:
scelta dei cluster basata sull'ottimizzazione di una funzione obiettivo assegnata (metodo delle k-medie)
- Algoritmi di **classificazione gerarchica**:
procedura iterativa che genera una gerarchia nelle partizioni (metodo di Ward)

In entrambi i casi, deve essere definita la misura della distanza tra gli oggetti, o della loro similarità

Distanza più utilizzata: **DISTANZA EUCLIDEA**
(somma delle differenze quadratiche tra le misurazioni relative ai due oggetti considerati)

Analisi di raggruppamento (Cluster Analysis)

- **Classificazione non gerarchica (metodo delle k-medie)**
Permette l'analisi di grandi masse di dati a costi bassi.
Si basa su una procedura iterativa che ammette la riallocazione degli elementi già clusterizzati, così da migliorare le partizioni ottenute.

Premesse dell'analisi:

- Definizione del numero di centri (poli di aggregazione degli oggetti)
- Metodo per la scelta dei cluster iniziali
- Metodo di attribuzione degli elementi ai cluster iniziali
- Fissazione del criterio per fermare l'iterazione

Analisi di raggruppamento (Cluster Analysis)

- **Classificazione non gerarchica (metodo delle k-medie)**

Segmentazione su clienti

Insieme costituito da N oggetti, caratterizzati da p variabili, spazio p-dimensionale, d distanza euclidea; si intende identificare fino ad un numero massimo di k clusters

Fase 0

Determinazione dei centri: scelta di k centri provvisori come poli iniziali di aggregazione. I centri creano una prima partizione in cluster: ogni oggetto appartiene a un cluster se la sua distanza dal centro del cluster è minore della distanza rispetto al centro degli altri cluster

Analisi di raggruppamento (Cluster Analysis)

Fase 1

Rideterminazione dei centri: sono ricalcolati i centri k (centri = vettori delle medie delle variabili per ogni cluster) dei cluster iniziali, così da originare una seconda partizione (costruita secondo gli stessi criteri della prima)

.....

Fase m

Fine dell'iterazione: due iterazioni successive danno luogo alla stessa partizione; raggiunto il n di iterazioni stabilito; la funzione obiettivo non decresce più significativamente

Analisi di raggruppamento (Cluster Analysis)

■ Classificazione non gerarchica

Problema principale: numero dei cluster

Elementi di scelta e interpretazione delle soluzioni:

- Numero osservazioni (dimensione dei cluster): dimensioni omogenee (significatività operativa del segmento → outliers in cluster poco numerosi)
- Analisi della varianza (tutte le variabili significative al test F): analisi della significatività del segmento e della soluzione di cluster (significatività di tutti i segmenti); > cluster, > significatività (anche: variabile omogenea in tutto il campione → eliminazione dalla clusterizzazione → significatività delle variabili più rilevanti per la segmentazione nel marketing)
- Leggibilità delle caratteristiche dei centri finali (definizione delle caratteristiche dei cluster rispetto alle variabili considerate e leggibilità dei cluster in termini di marketing)

Analisi di raggruppamento (Cluster Analysis)

■ Classificazione gerarchica

Agevolata dalle migliorate capacità dei processori elettronici

Impiego di algoritmi

Strumento di interpretazione degli output dell'analisi cluster gerarchica: dendrogramma

(vedere: La cluster analysis, in Molteni L., Troilo G.
Ricerche di marketing)

Analisi fattoriale

Analisi fattoriale

Approccio usato per esaminare le interrelazioni fra un ampio numero di variabili, nello sforzo di trovarne un nuovo insieme, di dimensioni minori rispetto a quello delle variabili originarie, che esprima ciò che è in comune fra le variabili originarie stesse. Nella segmentazione della domanda, è usata per sintetizzare una serie di valutazioni fornite su alcune caratteristiche specifiche del prodotto/servizio studiato, in modo da individuare i "pilastri" attorno ai quali si forma il giudizio del consumatore.

Analisi fattoriale

Lo scopo è quello di condensare le informazioni contenute in un'ampia gamma di variabili originali, in un numero di variabili più ridotto e facilmente gestibile, con una perdita minima di informazioni.

Tipi di analisi fattoriale:

- Analisi delle componenti principali**
- Analisi dei fattori comuni**

Scopi

- analizza la struttura delle interrelazioni (correlazioni) tra variabili
- definisce un insieme di "dimensioni" comuni sottostanti a tali variabili, chiamate **fattori**
- determina quanto ogni variabile è spiegata da ciascuna dimensione. Obiettivo dell'analisi è infatti quello di individuare un numero ridotto di "combinazioni lineari" delle variabili originarie che spieghino buona parte della "varianza" (cioè del contenuto informativo) delle variabili originarie stesse.

$$c_i = a_{1i}x_1 + a_{2i}x_2 + \dots + a_{pi}x_p \quad i=1,2,\dots,p$$

c_i = i-esima combinazione lineare individuata; x_j = j-esima variabile originaria; a_{ji} = coefficiente della j-esima variabile nella i-esima combinazione lineare (detta anche fattore o componente principale)

Usi principali

- aggregazione di variabili
 - descrizione di un fenomeno usando un numero di concetti inferiore alle variabili
- riduzione dei dati
 - tramite il calcolo degli "scores" (punteggi) delle dimensioni individuate, tali dimensioni possono essere usate al posto delle variabili originali in altre elaborazioni
 - le dimensioni individuate sono (per costruzione) poco correlate fra loro

Caratteristiche

- Non è una tecnica di **dipendenza**, ma di **interdipendenza: le variabili sono analizzate simultaneamente**
- I fattori (dimensioni, o combinazioni lineari) sono impiegati per massimizzare la capacità di "riprodurre" la variabilità delle variabili di partenza, piuttosto che predire il valore di una variabile dipendente, perdendo poche informazioni

Caratteristiche

- Le combinazioni lineari o fattori (anche detti "componenti principali" laddove si adotti questo tipo di AF) sono funzioni di tutte le variabili originarie, ma si correlano in particolare ad alcune di esse; le combinazioni non sono invece correlate tra loro e quindi apportano un contributo informativo differenziato.
- Esse riassumono una percentuale decrescente di contenuto informativo dei dati.

Tipi di analisi

- molte variabili -> pochi fattori = tipo R
 - identificazione di variabili rappresentative
 - definizione di un nuovo insieme di fattori che sostituisce le variabili osservate
- raggruppamento di osservazioni = tipo Q
 - (si usa più spesso l'analisi di raggruppamento chiamata cluster analysis)

Operazioni

- stima dei fattori
- stima del contributo di ciascuna variabile al fattore (factor loadings)
- eventuale calcolo dei punteggi dei fattori (factor scores)

Principali problemi

■ Numero di componenti da considerare

- Rapporto n combinazioni/n variabili (riduzione della complessità)
- Percentuale di varianza spiegata (informazioni sulla capacità di sintesi delle componenti)
- Comunalità (percentuale di contenuto informativo di ogni variabile originaria spiegato dalle combinazioni lineari)
- Grafico della percentuale di varianza spiegata da ogni fattore (pendenza della spezzata)
- Interpretabilità delle combinazioni (componenti) e loro rilevanza nel marketing del prodotto

■ Interpretazione delle componenti/combinazioni

- Correlazioni tra componenti e variabili originarie
- Rotazione delle componenti

Progettazione dell'analisi

■ scelta del modo di calcolo dei dati di ingresso (matrice di correlazione)

- tra le variabili (tipo R)
- tra i casi/individui (tipo Q)

■ definizione del tipo di dati

■ definizione delle dimensioni del campione

Definizione del tipo di dati

■ Misura delle variabili

- metriche
- dicotomiche

■ Selezione delle variabili

- è opportuno avere un certo numero di variabili per ogni fattore
- si dovrebbe cercare di identificare alcune variabili-chiave, che si avvicinino a fattori ipotizzati a priori

Dimensioni del campione

- almeno 50 osservazioni
- preferibilmente oltre 100 osservazioni
- almeno 5 osservazioni per ogni variabile analizzata
- preferibile un rapporto osservazioni/variabili di almeno 10/1

Matrice dei Fattori

- Scelta del metodo di estrazione dei fattori in funzione del tipo di varianza (contenuto informativo) da analizzare
 - analisi delle componenti principali (varianza totale)
 - analisi del fattore comune (varianza comune)
- Definizione del numero di fattori da estrarre
 - autovalore o radici latenti
 - scree test

Fattori estratti

- estrazione in sequenza
- ogni fattore è una combinazione lineare di variabili
- il primo fattore è quello che estrae più varianza
- il secondo fattore è ortogonale (indipendente) al primo ed è basato sulla varianza residua, e così via

Factor loadings (o saturazioni)

- correlazioni di ogni variabile con il fattore
- ogni variabile è associata, in particolare, al fattore con il quale possiede la correlazione più elevata, e il fattore viene quindi interpretato in riferimento a ciò che accomuna le variabili a esso associate.
- è preferibile che ogni variabile sia fortemente correlata con un solo fattore. Ma questo accade raramente
- per favorire l'interpretazione, è spesso attuata la "rotazione" dei fattori
- si redistribuisce la varianza per ottenere una struttura più semplice e significativa

Rotazione dei fattori

- metodi ortogonali
 - Varimax
 - Equimax
 - Quartimax
- metodi obliqui (più flessibili)
 - Oblimin
 - Promax
 - Orthoblique

Interpretazione della matrice

- Esame della matrice di correlazione fattoriale
- Identificazione del loading maggiore per ogni variabile
- Valutazione delle comunaltà
- Assegnazione di un "nome" ai fattori in base alle variabili correlate
