

Università degli Studi di Padova  
Facoltà di Scienze Statistiche ed Economiche  
Anno Accademico 2002-2003

LA SEGMENTAZIONE DELLA DOMANDA E IL  
POSIZIONAMENTO DEL PRODOTTO:  
ANALISI STATISTICHE PER TRE  
UNIVERSITA' A CONFRONTO.

Laureanda: Nucibella Anna

Relatori: prof.ssa Pertile Martina  
prof.ssa Bassi Francesca



*The essence of mathematics  
is not to make simple things complicated,  
but to make complicated things simple.*



# INDICE :

	<i>pag.</i>
<i>Introduzione</i>	5

## *Capitolo primo*

### **LA SEGMENTAZIONE DELLA DOMANDA**

1.1	La segmentazione della domanda	7
1.2	Requisiti che devono possedere i segmenti di mercato	9
1.3	I criteri su cui si può basare la segmentazione del mercato	11
1.4	Le principali variabili utilizzate nella segmentazione del mercato	12
1.5	Utilizzo del criterio della strumentalità del prodotto	13
1.6	L'analisi multivariata nelle ricerche di marketing	16

## *Capitolo secondo*

### **L'ANALISI FATTORIALE E LA CLUSTER ANALYSIS**

2.1	L'Analisi Fattoriale	25
2.2	Il modello di Analisi Fattoriale	26
2.3	Il modello di analisi delle componenti principali	28
2.4	Comunanza e unicità dei fattori	30
2.5	Il procedimento dell'Analisi Fattoriale	34
2.6	Criteri per determinare il numero dei fattori	37
2.7	Rotazione dei fattori	38
2.8	La Cluster Analysis	40

	<i>pag.</i>
2.9 Il percorso di analisi	42
2.10 Selezione della misura di prossimità tra le variabili	43
2.11 Selezione di un algoritmo di classificazione	44
2.11.1 Tecniche gerarchiche aggregative	47
2.11.2 Metodi gerarchici scissori o divisivi	48
2.11.3 Criteri che generano partizioni non gerarchiche	49
2.11.4 Tecniche non gerarchiche con sovrapposizione	50
2.12 Scelta tra metodi di analisi	53

*Capitolo terzo*

**IL POSIZIONAMENTO COMPETITIVO DEL PRODOTTO  
IN UNO SPAZIO MULTIDIMENSIONALE**

3.1 Il mapping multidimensionale	53
3.2 Gli obiettivi del mapping multidimensionale	54
3.3 Il procedimento logico del mapping	55
3.4 Le tecniche statistiche per il mapping multidimensionale	58

*Capitolo quarto*

**L'ANALISI DISCRIMINANTE E IL MULTIDIMENSIONAL  
SCALING**

4.1 L'analisi discriminante	63
4.2 Aspetti metodologici	66
4.3 I test per verificare le differenze tra i gruppi	68

	<i>pag.</i>
4.4 Multidimensional scaling	70
4.5 La raccolta dei dati	71
4.6 Multidimensional Scaling metrico	74
4.7 Multidimensional Scaling non metrico	77
4.8 Una misura per la valutazione del modello	80

### *Capitolo quinto*

## **UN CASO APPLICATIVO: SEGMENTAZIONE DELLA DOMANDA E POSIZIONAMENTO DELLA FACOLTA' DI SCIENZE STATISTICHE DI PADOVA, BOLOGNA E MILANO.**

5.1 Il caso di studio	81
5.2 Le Facoltà analizzate	82
5.3 Definizione del campione	86
5.4 Il questionario	91
5.5 Descrizione del campione	96

### *Capitolo sesto*

## **L'ANALISI DEI DATI**

6.1 La segmentazione degli studenti frequentanti la Facoltà di Scienze Statistiche ed Economiche	107
6.2 L'analisi fattoriale	107
6.3 Cluster analysis	116

## Indice

	<i>pag.</i>
6.4 Un confronto tra l'importanza data ad ogni variabile ed il livello di soddisfazione ottenuta	124
6.5 Il posizionamento della Facoltà di Scienza Statistiche ed Economiche dell'Università di Padova, Bologna e Milano Bicocca in una mappa delle percezioni	130
6.6 L'analisi discriminante	131
<i>Conclusioni</i>	139
<i>Appendice</i>	145
<i>Riferimenti bibliografici</i>	149

# INTRODUZIONE

Attualmente le aziende operano in un mercato complesso ed eterogeneo per cui risulta strategicamente vincente la scelta di focalizzare le proprie energie in uno specifico segmento di clientela, al quale proporre prodotti specializzati ed in grado di soddisfare appieno le loro esigenze.

Per creare un ottimo prodotto o per migliorare quello già esistente, le imprese devono in primo luogo conoscere le caratteristiche del mercato di riferimento e suddividere la clientela in segmenti omogenei al loro interno e tra loro eterogenei (*segmentazione del mercato*) (cap. 1) e quindi capire quale sia il prodotto che meglio gli si addice. In un secondo momento devono confrontare il prodotto ideale per il segmento target con quello che attualmente viene loro proposto, in modo da poter operare le opportune modifiche e migliorarlo (*posizionamento del prodotto in una mappa delle percezioni*) (cap.3).

Le tecniche statistiche che possono aiutare a definire la segmentazione della clientela sono principalmente l'analisi fattoriale per ridurre il numero delle variabili considerate e la cluster analysis per delineare i vari gruppi (cap. 2), mentre, per quanto riguarda il posizionamento del prodotto, le tecniche maggiormente utilizzate sono l'analisi discriminante, l'analisi delle corrispondenze e il multidimensional scaling, ognuna delle quali considera una differente tipologia di dati in input (cap.4).

Si è voluto in seguito svolgere questo tipo di analisi relativamente al prodotto "Facoltà di Scienze Statistiche ed Economiche" ed in particolare per le lauree triennali ad indirizzo economico proposte dalla Facoltà di Scienze Statistiche dell'Università degli Studi di Padova (Laurea in Statistica e Gestione delle imprese ed in Statistica, Economia e Finanza), dell'Università di Bologna (Laurea in Statistica, Impresa e Mercati) e dell'Università Bicocca di Milano (Laurea in Scienze Statistiche ed Economiche).

E' stato somministrato un questionario (Appendice) a centotrentacinque studenti frequentanti il secondo e terzo anno, iscritti ad uno dei corsi di laurea sopra citati, in modo da ottenere una valutazione generale della Facoltà. Si è cercato di sondare tutti gli elementi ritenuti rilevanti dagli studenti, per cui le variabili considerate ricoprono vari aspetti del mondo universitario e non solo quello didattico (cap.5).

Gli studenti appartenenti al nostro campione sono stati suddivisi in cluster omogenei al loro interno e tra loro eterogenei in base all'importanza data alle variabili analizzate. In seguito le Facoltà sono state posizionate in una mappa delle percezioni in base al livello di soddisfazione dichiarato dagli studenti e confrontate con la Facoltà per loro ideale (cap.6).

## Capitolo 1

# LA SEGMENTAZIONE DELLA DOMANDA

### *1.1 La segmentazione della domanda*

L'orientamento del marketing consiste nel focalizzare l'attività dell'impresa sul consumatore, nel realizzare quindi gli obiettivi aziendali attraverso la soddisfazione del cliente e dei suoi bisogni. Ovviamente però il bisogno è avvertito dai consumatori in modo differente ed il mercato si presenta quindi eterogeneo.

A tale situazione l'impresa può rispondere con due soluzioni estreme: fornire una risposta specifica - marketing mix – per ogni cliente, costruire cioè un prodotto/servizio ad hoc, oppure fare una proposta indifferenziata.

Fra i due estremi citati esiste una situazione intermedia di mercato eterogeneo, composto da gruppi di consumatori che presentano bisogni e comportamenti d'acquisto differenti. E' necessario quindi effettuare una segmentazione del mercato. Con questo termine si intende la scomposizione del mercato di sbocco in parti – segmenti – ciascuna delle quali presenti al suo interno un sufficiente grado di omogeneità quanto ad aspettative ed a percezione dei benefici offerti dal prodotto di un'impresa e dalla relativa marca e, al tempo stesso, sia sufficientemente diversa dalle altre.

Il comportamento dell'azienda può ispirarsi a quattro diverse strategie di segmentazione del mercato:

1. *Strategie differenziate*: consistono nel delineare specifiche politiche di mercato appropriate a ciascun segmento obiettivo. Queste considerano una pluralità di politiche di prodotto, di prezzo, di comunicazione e di distribuzione e richiedono ingenti risorse umane, tecniche e finanziarie

2. *Strategie concentrate*: consistono nell'individuare vari segmenti obiettivo, ma nel porre in atto solo le politiche di mercato maggiormente appropriate al più importante di questi, che possono avere effetti vantaggiosi anche sugli altri
3. *Strategie focalizzate*: consistono nello scegliere un solo segmento, il più vantaggioso in termini di redditività, e nell'orientare tutte le politiche di marketing su questo, senza preoccuparsi degli effetti che si possono ripercuotere sugli altri segmenti
4. *Strategie indifferenziate o di contro-segmentazione*: consistono nell'offrire prodotti standardizzati e con buone funzionalità d'uso, destinati indifferentemente a tutto il mercato, a prezzi competitivi.

Mentre è evidente che strategie differenziate, concentrate e focalizzate richiedano precedenti analisi di segmentazione del mercato, non bisogna ritenere che nel caso di strategie indifferenziate queste siano superflue. E' necessario infatti conoscere comunque le funzionalità del prodotto apprezzate dal maggior numero di consumatori, anche in segmenti diversi, per elaborare un'offerta efficace, anche se indifferenziata. Inoltre, spesso, attraverso tali analisi, le imprese che hanno adottato strategie indifferenziate si accorgono che il loro sistema prodotto risponde comunque alle esigenze di uno specifico segmento e non di altri e su questa base possono migliorare la loro proposta.

Nell'elaborare la strategia di marketing, le imprese devono valutare attentamente l'attrattività dei segmenti e non lasciarsi ingannare dalla cosiddetta miopia di marketing. I segmenti di maggiore consistenza, infatti, a motivo delle maggiori vendite potenziali che sembrano offrire, sono quelli che registrano una elevata concentrazione di concorrenti. Risulta perciò difficile ottenere una adeguata quota di mercato o per lo meno oneroso il suo mantenimento. Al contrario, la scarsa presenza di concorrenti rende più congeniale alle piccole imprese la scelta di segmenti minori e l'adozione di una strategia di nicchia.

## ***1.2 Requisiti che devono possedere i segmenti di mercato***

Sotto il profilo analitico, i segmenti che vengono individuati, per essere utili dal punto di vista del marketing, devono possedere alcuni requisiti fondamentali:

- *misurabilità*, non soltanto in termini di numero di individui che compongono il segmento, ma anche di dimensione del mercato potenziale e della sua incidenza sul potenziale complessivo
- *omogeneità interna di ogni segmento* per aspettative e comportamenti degli individui che lo compongono. Con questo intendiamo quindi che vi sia similarità nella reazione agli stimoli di marketing, nei benefici attesi e nei processi di acquisto e di consumo.  
Il livello di omogeneità interna ed il giudizio di significatività che ne consegue deve essere ponderato e messo in relazione con il numero e le dimensioni dei segmenti e le variabili utilizzate per la segmentazione
- *eterogeneità tra i segmenti* individuati, che devono essere significativamente diversi l'uno dall'altro
- *sostanzialità del segmento*, ossia la sua idoneità a produrre reddito e ad essere proficuamente sfruttabile sotto il profilo commerciale. Questo requisito non può essere definito con tecniche standardizzabili in quanto bisogna tener conto delle dimensioni aziendali rispetto a quelle dei segmenti, dei potenziali di crescita di questi ultimi e degli specifici programmi aziendali; lo stesso segmento può cioè presentare gradi diversi di sostanzialità al variare delle imprese e delle loro politiche di mercato

- *accessibilità del segmento*, ossia la sua idoneità ad essere agevolmente raggiunto con gli strumenti operativi a disposizione dell'impresa. In tal caso spesso le aziende titolari di prodotti di marca e che utilizzano strumenti di comunicazione pubblicitaria, individuano i cosiddetti *media habits* che caratterizzano i vari segmenti, ossia la loro specifica esposizione ai diversi media; anche questo requisito deve essere valutato in relazione alle capacità e alle politiche di mercato di ogni singola impresa
  
- *aggregabilità*, grado di affollamento del segmento; la presenza di imprese di grandi dimensioni che controllano il segmento può creare delle barriere all'entrata
  
- *grado di variabilità* dei segmenti, dovuta ad una crescente dinamicità della domanda, che deve essere individuata e tenuta sotto controllo. Questa variabilità comporta una riduzione dei tempi di accessibilità e di sfruttamento di un segmento ed implica la necessità di limitare l'arco temporale coperto da un piano di marketing e di sottoporlo a continue verifiche; per tenere sotto controllo la crescente variabilità, è necessario confrontare i risultati di processi di segmentazione eseguiti ad intervalli di tempo limitati e regolari, in modo che emergano anche i lievi mutamenti delle aspettative del consumatore e queste informazioni possano poi essere utilizzate nella definizione di strategie di marketing appropriate.

### ***1.3 I criteri su cui si può basare la segmentazione del mercato***

La scelta dei criteri e delle variabili con cui procedere alla segmentazione del mercato deve essere coerente con gli obiettivi che l'impresa si propone di raggiungere attraverso le strategie funzionali di marketing e con i problemi che tenta di risolvere. Nonostante la soggettività della scelta delle tecniche di analisi da utilizzare e delle variabili da considerare, vi sono due principali criteri che solitamente vengono utilizzati congiuntamente:

- *il criterio descrittivo*
- *il criterio della strumentalità del prodotto.*

Seguendo il primo criterio, si parte dai bisogni dei consumatori e si suddividono questi ultimi in segmenti di mercato sulla base di più variabili (di carattere demografico, economico, sociale, psicologico, culturale) cui si connettono significative differenze dei comportamenti d'acquisto dei prodotti oggetto di analisi. Tale via può risultare particolarmente utile nel definire prodotti fortemente innovativi in quanto non si basa su comportamenti d'acquisto già consolidati ma sui bisogni percepiti. Unico limite di tale tecnica è la limitata funzione discriminante delle variabili considerate.

Seguendo il secondo criterio invece si parte dalla classe dei prodotti oggetto di analisi e si distribuiscono i suoi acquirenti sulla base di variabili come le quantità acquistate in un dato arco temporale, le diverse modalità e occasioni d'uso, il grado di fedeltà alle marche e i diversi benefici ricercati nelle varie categorie di prodotti (*benefit segmentation*) e le diverse reazioni agli stimoli provocati da specifiche leve di marketing. In questo caso non vengono approfonditi i fattori, specie socio-culturali, sottostanti ai comportamenti d'acquisto e neanche i mezzi più efficaci per raggiungere i vari segmenti, ma viene dato un ottimo contributo per le analisi del posizionamento competitivo dei prodotti e delle marche.

Dato il progresso delle tecniche multivariate e la disponibilità di software specifici, attualmente vengono utilizzati contemporaneamente entrambi i criteri.

#### ***1.4 Le principali variabili utilizzate nella segmentazione del mercato***

Le variabili che possono essere impiegate nelle analisi finalizzate alla segmentazione del mercato, per essere efficaci devono possedere:

- a. una capacità *identificatrice*, ossia devono consentire di fornire l'identikit del consumatore medio che caratterizza tale segmento
- b. una capacità *discriminante* in quanto devono differenziare in maniera significativa gli appartenenti ai vari segmenti per aspettative, atteggiamenti e comportamenti d'acquisto
- c. una capacità *esplicativa*, ossia fornire fondate ipotesi sui fattori sottostanti alla varietà dei fenomeni osservati.

Le tecniche di segmentazione che fanno capo ad un criterio descrittivo si basano solitamente su variabili di tipo socio-economico, demografico e geografico (sesso, reddito, livello di istruzione, nazionalità...). I principali vantaggi che hanno portato alla diffusione di tali variabili sono la loro facilità di misurazione, che avviene su scale definite e di semplice comprensibilità, e la veloce reperibilità dei dati. Inoltre questi sono altamente attendibili ed oggettivi e possono essere elaborati con semplici tecniche.

Per quanto riguarda invece il criterio basato sulla strumentalità del prodotto, le variabili maggiormente utilizzate sono le modalità e la frequenza d'uso del prodotto e i benefici associati (benefit segmentation) ed in fine la fedeltà alle marche.

Spesso tali analisi vengono supportate da una preliminare fase di ricerca qualitativa, volta a registrare tutte le possibili motivazioni alla base dell'utilizzo del prodotto, seguita da uno studio pilota per eliminare le motivazioni ridondanti o scarsamente influenti. In seguito viene valutata la capacità del consumatore di associare le proprie esigenze alle modalità d'uso del prodotto. Ciascuna occasione d'uso viene poi registrata e descritta in termini di quantità consumata, marca, orario e luogo. La fase quantitativa di raccolta dei dati termina con la richiesta fatta ad ogni individuo del campione, di indicare nella lista delle esigenze potenziali quali abbia avvertito e con quale intensità per ogni occasione d'uso del prodotto. Infine tali dati vengono elaborati mediante l'incrocio con altre variabili relative ai consumi di marca, in modo da ottenere interessanti informazioni sulla capacità delle varie marche di soddisfare specifiche esigenze e che sia possibile passare da un'analisi di segmentazione ad una sul posizionamento competitivo del prodotto.

### ***1.5 Utilizzo del criterio della strumentalità del prodotto***

Da una ricerca basata sul criterio della strumentalità del prodotto possono emergere diverse informazioni in base alle variabili di segmentazione utilizzate e all'obiettivo della nostra analisi.

Alcune variabili consentono di analizzare il livello di disponibilità all'acquisto (consapevole, non consapevole, interessato, primo acquisto, acquisto regolare) e le motivazioni d'acquisto (economicità, prestigio, affidabilità) e l'uso finale (proprio, regalo, uso saltuario, regolare).

La segmentazione del mercato che si basa sulla fedeltà alla marca mira invece principalmente ad analizzare le sequenze oggettive degli acquisti effettuati dai consumatori rispetto alle marche. I clienti vengono suddivisi per comportamenti omogenei e il ricercatore deve cercare di estrapolare le

motivazioni che differenziano i consumatori fedeli da quelli infedeli alla marca, con l'obiettivo di porre in atto politiche di marketing finalizzate alla fidelizzazione dei consumatori infedeli. Queste analisi però, per risultare realmente efficaci, devono essere integrate a quelle di tipo descrittivo in modo da ottenere un profilo sufficientemente preciso del consumatore fedele e di quello infedele.

Le forme più avanzate di segmentazione comportamentale, come già accennato, sono quelle che raggruppano i consumatori di una determinata classe di prodotti in base alla similarità dei benefici, primari o secondari, ricercati ed ottenuti.

Le difficoltà che incontra la benefit segmentation sono duplici: da una parte i segmenti di consumatori non si differenziano per la ricerca di un singolo particolare beneficio, ma per un insieme di benefici che si connettono ad una pluralità di attributi offerti con gradi diversi di intensità (*benefit bundle analysis*), e questo aspetto rende le analisi assai complesse. D'altra parte, affinché l'impresa sia in grado di sviluppare politiche efficaci per la penetrazione nei segmenti di interesse, è necessario che queste analisi siano supportate ad altre forme di segmentazione descrittiva. Questo perché il ricercatore possiede numerose informazioni utili per definire politiche di prodotto e messaggi pubblicitari efficaci, ma gli mancano i riferimenti per scegliere i media comunicazionali preferiti dal segmento e i canali distributivi di cui si avvale maggiormente.

**Il vantaggio maggiore della benefit segmentation è quello di consentire all'impresa di affrontare in modo unitario e con l'uso delle medesime tecniche di analisi sia i problemi della segmentazione del mercato che quelli di posizionamento competitivo del prodotto.** Tale metodologia di analisi consente infatti di definire, in caso di prodotti già esistenti nel mercato, il profilo comparato degli attributi percepiti per il proprio prodotto e per quelli concorrenti, in relazione ai benefici ricercati da ogni segmento di mercato.

Da quanto detto risulta evidente che, data la vastità e la varietà delle decisioni di marketing, l'individuazione e l'uso di un'unica base per la segmentazione può condurre a decisioni errate ed a uno spreco di risorse. Rimane pur sempre vero però che alcune variabili siano più idonee di altre per certi tipi di segmentazione. (Tabella 1.1)

Tabella 1.1 *Criteri utili come base per la segmentazione*

*Per una comprensione generale del mercato:*

- benefici ricercati
- modalità d'uso e d'acquisto del prodotto
- bisogni
- fedeltà alla marca

*Per studi di posizionamento:*

- uso del prodotto
- preferenze verso i prodotti
- benefici ricercati

*Per l'introduzione di nuovi prodotti:*

- reazione a nuove proposte (intenzione ad acquistare, preferenza rispetto alle marche esistenti...)
- benefici ricercati

*Per decisioni di prezzo:*

- sensibilità al prezzo
- propensione alla trattativa
- sensibilità al prezzo in rapporto a specifiche d'acquisto e d'uso

*Per decisioni relative alle politiche pubblicitarie:*

- benefici ricercati
- atteggiamenti nei confronti dei diversi media
- stili di vita

## ***1.6 L'analisi multivariata nelle ricerche di marketing***

Comunemente l'analisi multivariata viene divisa in due aree: *l'analisi della dipendenza* e *l'analisi dell'interdipendenza*.

Nello studio della **dipendenza** viene individuata una variabile od un gruppo di variabili che costituisce l'obiettivo dell'analisi; queste variabili sono appunto definite dipendenti e vengono spiegate dalle rimanenti, che vengono definite indipendenti od esplicative.

Le tecniche di analisi dell'**interdipendenza** studiano invece l'interrelazione tra un insieme di variabili. Lo scopo è quello di comprendere quale sia la struttura sottostante alle variabili e di creare nuovi fattori che aiutino nel processo interpretativo della realtà multidimensionale.

Cerchiamo ora di schematizzare le tecniche di analisi multivariata in Figura 1.1, 1.2 e 1.3. I metodi più frequentemente usati per la segmentazione del mercato e il posizionamento del prodotto verranno in seguito approfonditi.

Figura 1.1 Una classificazione dei metodi di analisi multivariata

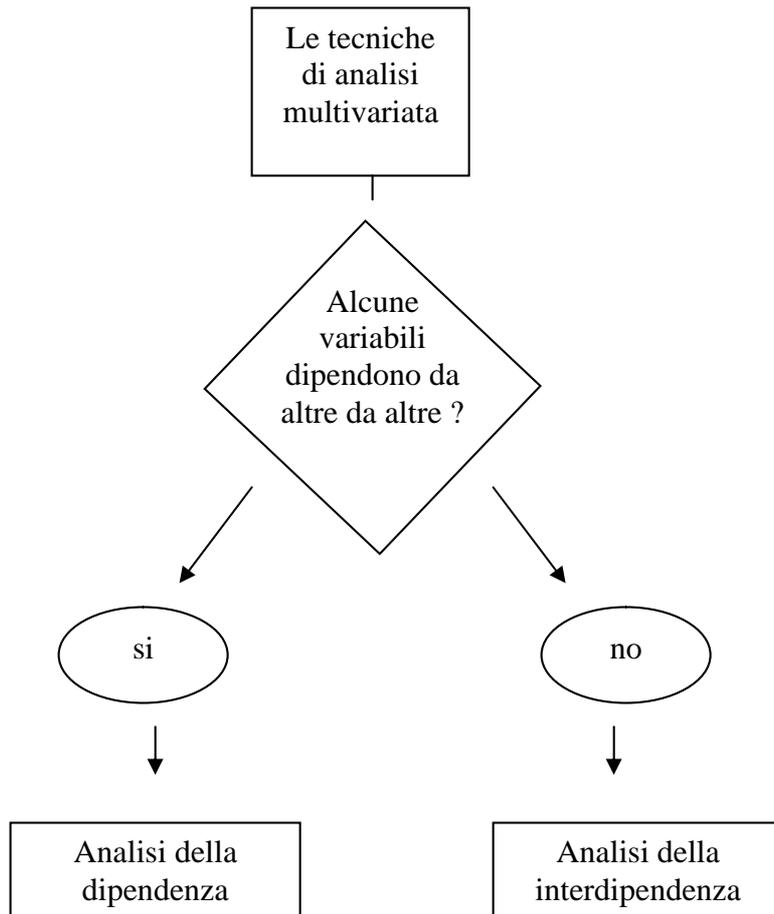
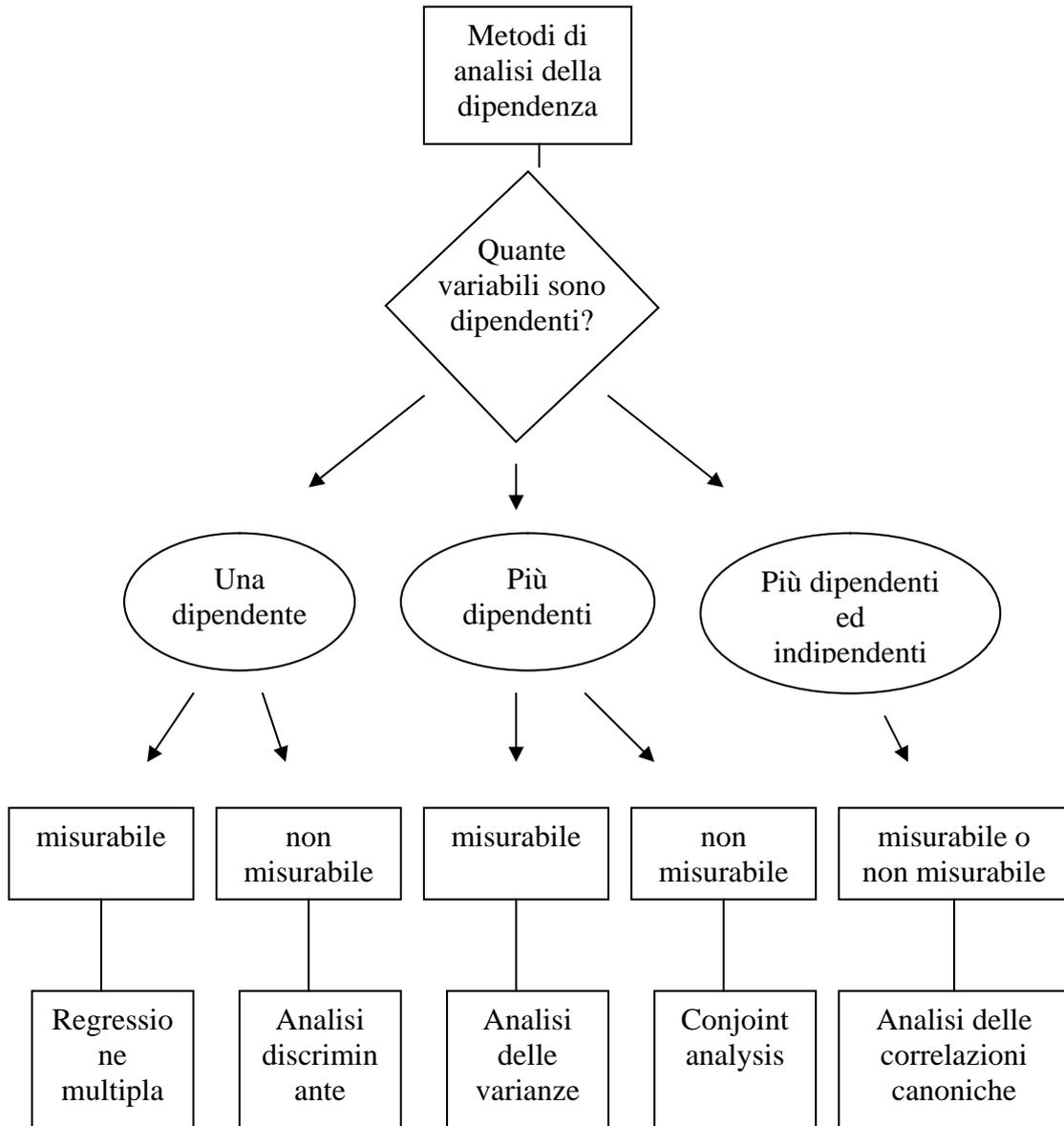


Figura 1.2 Classificazione dei metodi di analisi della dipendenza



Il *metodo della regressione multipla* ha l'obiettivo di investigare gli effetti che simultaneamente hanno due o più variabili indipendenti su quella dipendente e quindi può risultare utile nel caso in cui si voglia **prevedere le vendite di un prodotto** (variabile dipendente) **tramite le leve del marketing mix** (variabili indipendenti). Nel campo della segmentazione della domanda, questo tipo di analisi può venire utilizzata per stabilire quale variabile indipendente sia più importante nel determinare la variabile dipendente. Quindi, pur essendo una tecnica funzionale, legata alla previsione di un fenomeno, la regressione multipla può essere impiegata nelle fasi preliminari delle ricerche di segmentazione quando non sono chiari i legami, e l'intensità di questi, tra un fenomeno e alcune variabili indipendenti in grado di spiegarlo.

L'*analisi discriminante* è una tecnica statistica che consente di risolvere due problemi. In una prima fase permette di studiare ed esaminare le differenze esistenti tra due o più gruppi, noti a priori, di oggetti o individui al fine di **determinare quali variabili li discriminino maggiormente**. Ciò avviene mediante una procedura che consente di individuare una serie di funzioni discriminanti, ossia di combinazioni lineari delle variabili esplicative, la cui costruzione segue un ordine preciso: la prima funzione spiega il massimo della varianza, la seconda ne spiega una quota inferiore fino a spiegare il totale della varianza complessiva.

Grazie a questa tecnica, le informazioni sulle medesime variabili possono essere in un secondo momento usate per classificare opportunamente nuovi elementi di cui non si conosca il gruppo di appartenenza (funzione di classificazione).

L'*analisi multivariata delle varianze* (MANOVA) vuole verificare se e quanto siano **statisticamente significative le differenze sulle medie** tra i gruppi per due o più variabili dipendenti.

La *conjoint analysis* è una tecnica di analisi multivariata che consente di misurare l'importanza relativa di una serie di attributi di un prodotto/servizio, fornendo anche indicazioni sul gradimento di varie specifiche degli attributi stessi. Il punto di partenza dell'analisi è costituito da una serie di valutazioni globali fornite da consumatori/clienti riguardo un insieme di alternative. La conjoint analysis permette di decomporre tali valutazioni globali in **scale di utilità corrispondenti a ciascun attributo in modo tale che le considerazioni iniziali possano poi essere ricostruite come somma delle parziali**.

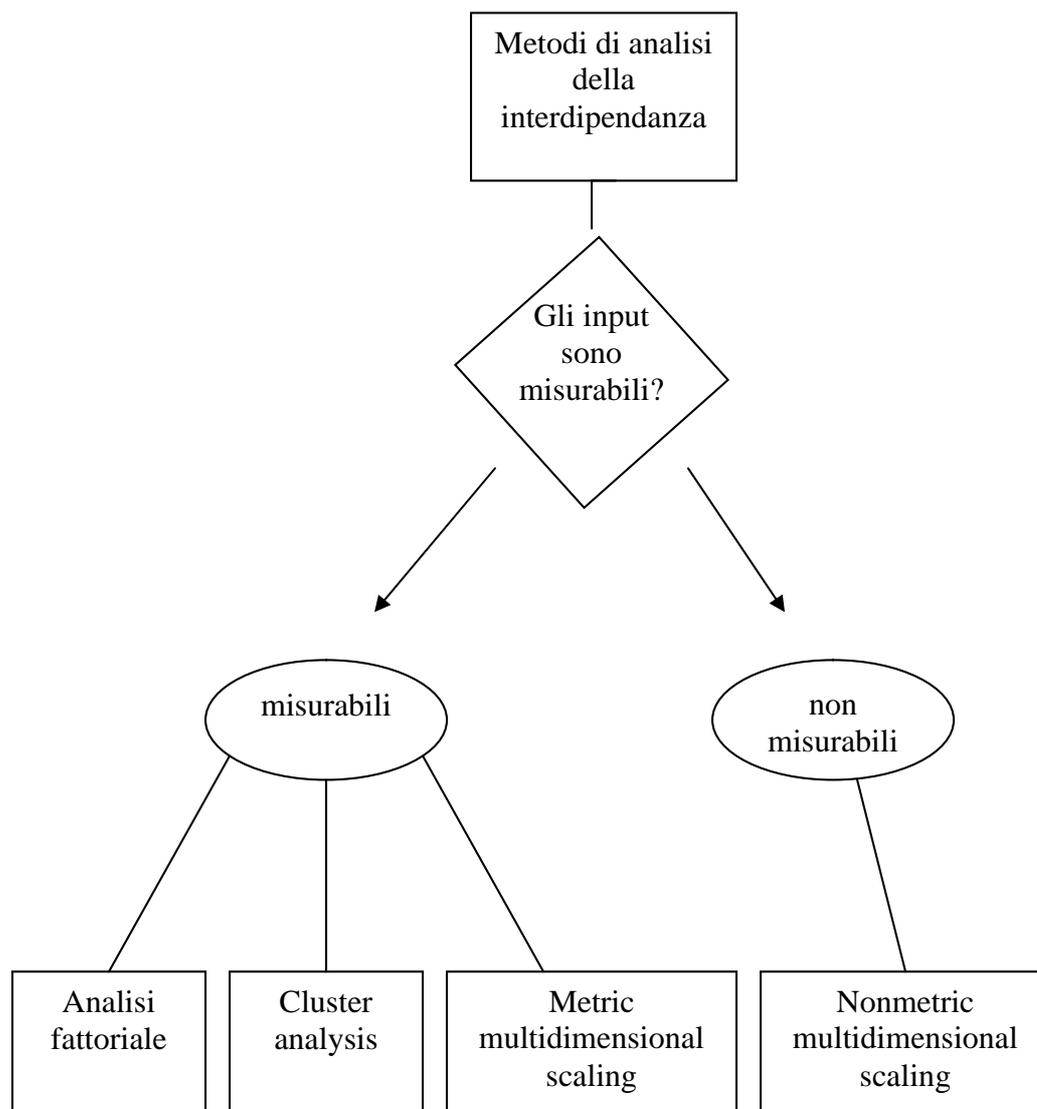
La conjoint analysis risulta uno strumento valido per la segmentazione del mercato per diversi motivi:

- il focus della tecnica è nella misurazione dei benefici che il cliente ricava dalle caratteristiche del prodotto. Questa misurazione avviene, al contrario delle procedure di segmentazione classiche, senza una valutazione diretta dell'importanza degli attributi stessi ma solo evidenziando i trade-off esistenti tra attributi e livello degli attributi
- le preferenze per gli attributi sono misurate a livello individuale e il ricercatore può quindi facilmente verificare se ci sono omogeneità o eterogeneità nelle preferenze
- vengono tipicamente considerate caratteristiche socio-demografiche dell'intervistato e i suoi comportamenti d'acquisto in modo tale che poi risulti semplice descrivere i segmenti individuati
- in genere gli studi di conjoint analysis includono una fase di simulazione in cui il ricercatore può testare profili di prodotti

nuovi o modificati e verificare se e con quale frequenza vengono preferiti rispetto ai prodotti concorrenti.

Il *metodo delle correlazioni canoniche* presenta forti affinità con la regressione multipla. Esso serve a determinare il livello della **relazione lineare esistente tra due gruppi di variabili**, uno dipendente e l'altro indipendente (in questo caso abbiamo più variabili dipendenti e non una soltanto come nella regressione multipla). Tralasciando l'aspetto previsivo, che interessa in minima parte le ricerche di segmentazione, la correlazione canonica può essere ad esempio utile nei casi in cui si voglia collegare la fedeltà alle diverse marche presenti sul mercato ai benefici ricercati dai consumatori nelle marche stesse.

Figura 1.3 Classificazione dei metodi di analisi della interdipendenza



L'*analisi fattoriale* si pone principalmente l'obiettivo di **sintetizzare le relazioni esistenti tra un insieme di variabili oggetto di analisi**, esprimibili attraverso una serie di dimensioni che risultano essere combinazioni lineari delle variabili di partenza. Ciò avviene attraverso l'analisi di alcuni indicatori del livello di correlazione esistente tra le variabili stesse, al fine di individuarne un sottoinsieme ridotto rispetto a quello originario. Le dimensioni a cui si perviene prendono il nome di fattori o variabili latenti.

Ulteriore obiettivo di questa tecnica è quello di individuare una o più strutture che permettano di aggregare fenomeni diversi sulla base di caratteristiche comuni, **condensando l'insieme di variabili originarie in un sottoinsieme di variabili maggiormente significative** ai fini dell'analisi ed in grado di spiegare buona parte della variabilità iniziale nel campione osservato.

La *cluster analysis* si propone invece di classificare gli oggetti o gli individui appartenenti all'insieme campionario di osservazioni in gruppi (cluster), i quali devono presentare due caratteristiche fondamentali. Devono **massimizzare il livello di omogeneità al loro interno**, essendo gli individui del cluster individuati in base ad un criterio di similarità o ad una misura della distanza tra di loro (massimizzando la similarità tra gli elementi del cluster o minimizzando la somma delle distanze). I gruppi devono inoltre risultare **il più possibile differenziati tra loro massimizzando l'eterogeneità**.

La tecnica del *mutidimensional scaling* è composta da un insieme di procedure che, partendo da una matrice di prossimità tra  $n$  elementi (prodotti, marche, servizi) tra le quali si assumono relazioni simmetriche, trova una configurazione, rappresentabile geometricamente, dei prodotti in un numero usualmente limitato di

dimensioni. Tale tecnica permette quindi di **misurare gli oggetti in uno spazio multidimensionale** sulla base dei giudizi dati da un campione di individui relativamente alla similarità tra i prodotti o servizi valutati.

Per risolvere i problemi relativi alla segmentazione del mercato, le tecniche di analisi multivariata più frequentemente utilizzate sono quelle di analisi fattoriale e successiva cluster analysis o di conjoint analysis. Noi analizzeremo in modo dettagliato le prime due.

## Capitolo 2

# L'ANALISI FATTORIALE E LA CLUSTER ANALYSIS

### *2.1 L'Analisi Fattoriale*

L'analisi fattoriale è un metodo statistico idoneo a ridurre un sistema complesso di correlazioni in un numero minore di dimensioni. Inizialmente si è sviluppata e ha avuto largo impiego nella psicologia come modello matematico per la formalizzazione di teorie nell'ambito degli studi sui test mentali e attitudinali e sul comportamento umano. Il pioniere in questo campo è stato Spearman (1904) seguito poi da Thurstone(1931), il primo a proporre una teoria multifattoriale.

Attualmente tale tecnica viene utilizzata in diversi campi: sociale, psicologico, economico, e gli impieghi più ricorrenti sono:

- Ridurre la complessità di una matrice di dati, riducendo il numero delle variabili
- Semplificare la lettura di un fenomeno
- Costruire modelli previsionali più stabili
- Verificare ipotesi sulla struttura delle variabili, in termini di numero di fattori significativi, sui loro legami, sulle cause comuni che agiscono sulle loro manifestazioni
- Misurare costrutti non direttamente osservabili a partire da indicatori osservabili ad essi correlati.

Nelle analisi di mercato esse trovano varie applicazioni:

- Individuare i fattori che determinano gli atteggiamenti verso un prodotto o l'immagine di un'impresa
- Effettuare analisi strutturali sulle caratteristiche di un prodotto, per individuare i fattori di preferenza
- Costruire modelli esplicativi o indicatori economici per la misura dei potenziali di mercato
- Individuare le dimensioni che caratterizzano il comportamento dei clienti.

## ***2.2 Il modello di Analisi Fattoriale***

Si supponga di aver osservato un insieme di  $p$  variabili quantitative o dicotomiche presso  $n$  unità statistiche e che  $n$  sia abbastanza elevato rispetto a  $p$ , di aver ordinato le osservazioni nella matrice  $X$  il cui elemento generico  $x_{hj}$  denota il valore della variabile  $x_j$  osservato presso l'unità  $h$ , e di aver successivamente standardizzato i dati (le variabili hanno media nulla e varianza unitaria).

Il modello di analisi fattoriale si esprime con l'equazione:

$$\begin{aligned}x_j &= a_{j1} f_1 + a_{j2} f_2 + \dots + a_{jq} f_q + u_j c_j \\ &= \sum_i^q a_{ji} f_i + u_j c_j \quad (j = 1, \dots, p)\end{aligned}\tag{2.1}$$

dove i deponenti relativi alle unità statistiche sono stati soppressi per semplificare l'esposizione;  $f_i$  ( $i = 1, \dots, q$ ) rappresenta il fattore comune  $i$ -esimo (*variabile latente*);  $a_{ji}$  è il coefficiente che lega il fattore  $f_i$  alla variabile  $x_j$ , ed

è detto *peso fattoriale* (*factor loading*);  $c_j$  è il fattore specifico di  $x_j$  e  $u_j$  è il suo coefficiente.

Nella notazione matriciale, il modello consiste nella scomposizione della matrice di dati in matrici di fattori comuni e specifici:

$$\mathbf{X} = \mathbf{F} \mathbf{A}_q^T + \mathbf{E} \quad (2.2)$$

dove  $\mathbf{F}$  è la matrice  $n \times q$  di fattori,  $\mathbf{A}_q$  è una matrice di pesi fattoriali di ordine  $p \times q$  ( $q \leq r$ ),  $\mathbf{E} = \mathbf{C} \mathbf{U}$  è una matrice  $n \times p$  di fattori specifici e  $\mathbf{U}$  è la matrice diagonale di coefficienti dei fattori specifici  $c_1, c_2, \dots, c_p$ .

Nel modello fattoriale vengono fatte le seguenti ipotesi:

$$\begin{aligned} \text{Corr}(f_i, f_j) &= 0 & \forall i \neq j \\ \text{Corr}(c_i, c_j) &= 0 & \forall i \neq j \\ \text{Corr}(c_i, f_j) &= 0 & \forall i, j \end{aligned}$$

Il fattore  $f_i$  si dice *comune* perché è presente in tutte le  $p$  possibili equazioni; se ha coefficienti non nulli con tutte le variabili, si dice *generale*;  $c_j$  si dice *specifico* perché appartiene solo alla variabile  $x_j$ . Ogni fattore comune è combinazione di tutte le variabili osservate:

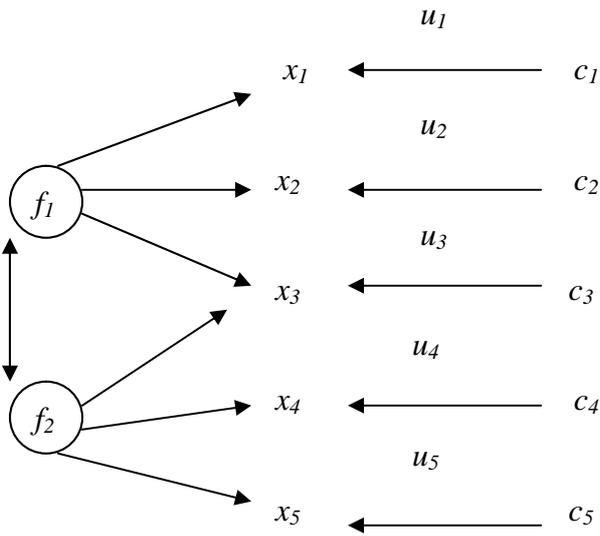
$$f_i = \sum_j^p w_{ji} x_j \quad (i = 1, \dots, q) \quad (2.3)$$

dove  $w_{ji}$  è il coefficiente fattoriale (*factor score coefficient*) della variabile  $x_j$  nella combinazione  $f_i$ . Adottando il modello di analisi fattoriale si assumono dunque relazioni lineari ed additive tra le variabili osservate.

I fattori possono essere *ortogonali*, ossia incorrelati, oppure *obliqui*, ossia correlati.

Graficamente un modello di analisi fattoriale può essere rappresentato in tal modo (Figura 2.1):

Figura 2.1 Modello di analisi fattoriale con 5 variabili e 2 fattori latenti



### 2.3 Il modello di analisi delle componenti principali

L'analisi delle componenti principali è un metodo di trasformazione matematica di un insieme di variabili in uno nuovo di variabili composite (componenti principali) ortogonali tra loro e che spiegano la totalità della variabilità del fenomeno. Si distingue dall'analisi fattoriale in quanto vengono considerate tutte le componenti principali, anche se solo alcune saranno poi utilizzate a fini interpretativi.

In tale analisi la generica variabile  $x_j$  è funzione lineare di tutte le possibili componenti principali estraibili (pari ad  $r$ , rango della matrice di correlazione):

$$\begin{aligned}
 x_j &= a_{j1} f_1 + a_{j2} f_2 + \dots + a_{jr} f_r \\
 &= \sum_i^r a_{ji} f_i \quad (j = 1, \dots, p)
 \end{aligned}
 \tag{2.4}$$

che in notazione matriciale diventa:

$$\mathbf{X} = \mathbf{F} \mathbf{A}^T
 \tag{2.5}$$

dove la matrice  $\mathbf{F}$  di ordine  $n \times r$  comprende tutte le componenti  $f_i$  e la matrice  $\mathbf{A}$  di ordine  $p \times r$  i pesi fattoriali.

Le componenti principali si ricavano identificando in sequenza la combinazione lineare delle variabili osservate che estrae la quota massima di variabilità man mano depurata della variabilità e covariabilità delle componenti principali estratte.

La prima componente sarà quella a varianza maggiore, generalmente indicata con  $\lambda_i$  e chiamata autovalore, per cui valgono le seguenti relazioni:

$$\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_r
 \tag{2.6}$$

$$\sum_i \lambda_i = \sum_i \text{var}(x_i)$$

Applicando l'analisi delle componenti principali si assiste pertanto ad una ridistribuzione della varianza totale con una forte concentrazione nelle prime componenti principali.

Se non vi è collinearità tra le variabili osservate, il numero di componenti eguaglia quello di variabili ( $r = p$ ); inoltre la varianza spiegata dalle componenti principali coincide con la variabilità osservata.

### 2.4 Comunanza e unicità dei fattori

La (2.1) ha la forma di un'equazione di regressione dove  $x_j$  è la variabile dipendente e i fattori sono le esplicative e  $c_j$  il termine residuale. Per analogia con l'analisi di regressione, se una variabile è esprimibile in funzione di fattori comuni e di un fattore specifico, anche la sua varianza è scomponibile in due parti: la varianza comune (*comunanza*) e la varianza unica (*unicità*).

Se i fattori sono incorrelati tra loro e con quello specifico, per ogni  $x_j$  vale l'identità:

$$\begin{aligned} \sigma_j^2 &= \text{Var} \left[ \sum_i^q a_{ji} f_i + u_j c_j \right] = \sum_i^q a_{ji}^2 \lambda_i + u_j^2 \sigma^2(c_j) \\ &\qquad\qquad\qquad \left| \qquad\qquad\qquad \right| \\ &= \text{comunanza} + \text{unicità} \end{aligned} \tag{2.7}$$

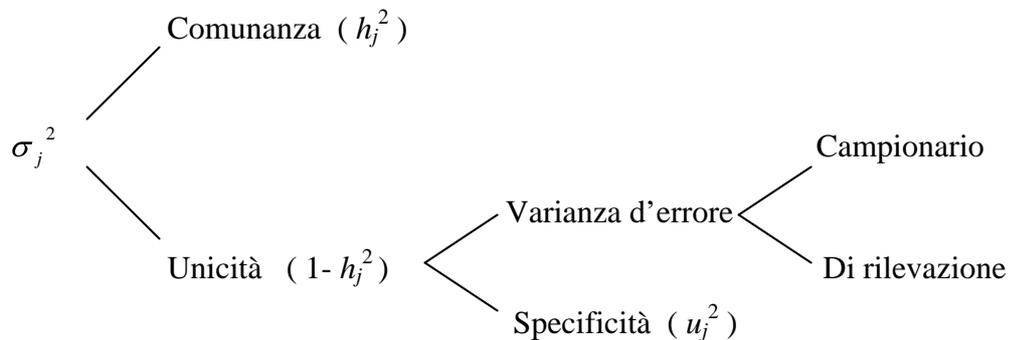
La comunanza  $h_j^2$  è la frazione di varianza di  $x_j$  spiegata dall'insieme dei fattori comuni. Essendo il coefficiente di correlazione tra la variabile  $x_j$  e il fattore  $f_j$  uguale al peso fattoriale,  $r_{ij} = a_{ij}$ , la comunanza, data dalla somma del quadrato dei coefficienti di correlazione con i singoli fattori comuni, è anche ottenibile sommando il quadrato dei pesi fattoriali:

$$h_j^2 = \sum_i^q r_{ji}^2 = \sum_i^q a_{ji}^2 \quad (j = 1, \dots, p) \tag{2.8}$$

La comunanza di una variabile è la parte di varianza che questa condivide con le rimanenti fattorizzate, mentre l'unicità della  $x_j$  è la parte complementare. Questa contiene generalmente tre componenti, che solo in certi casi possono essere misurate (Figura 2.2):

1. una di errore casuale, detta *varianza di campionamento*, addebitabile al campione
2. una di errore di rilevazione, detta *varianza di rilevazione*, dovuta alla inaccuretezza della rilevazione
3. una residuale, detta *varianza specifica*, che deriva dal fattore specifico  $c_j$ .

Figura 2.2 Decomposizione della varianza



## 2.5 Il procedimento dell'Analisi Fattoriale

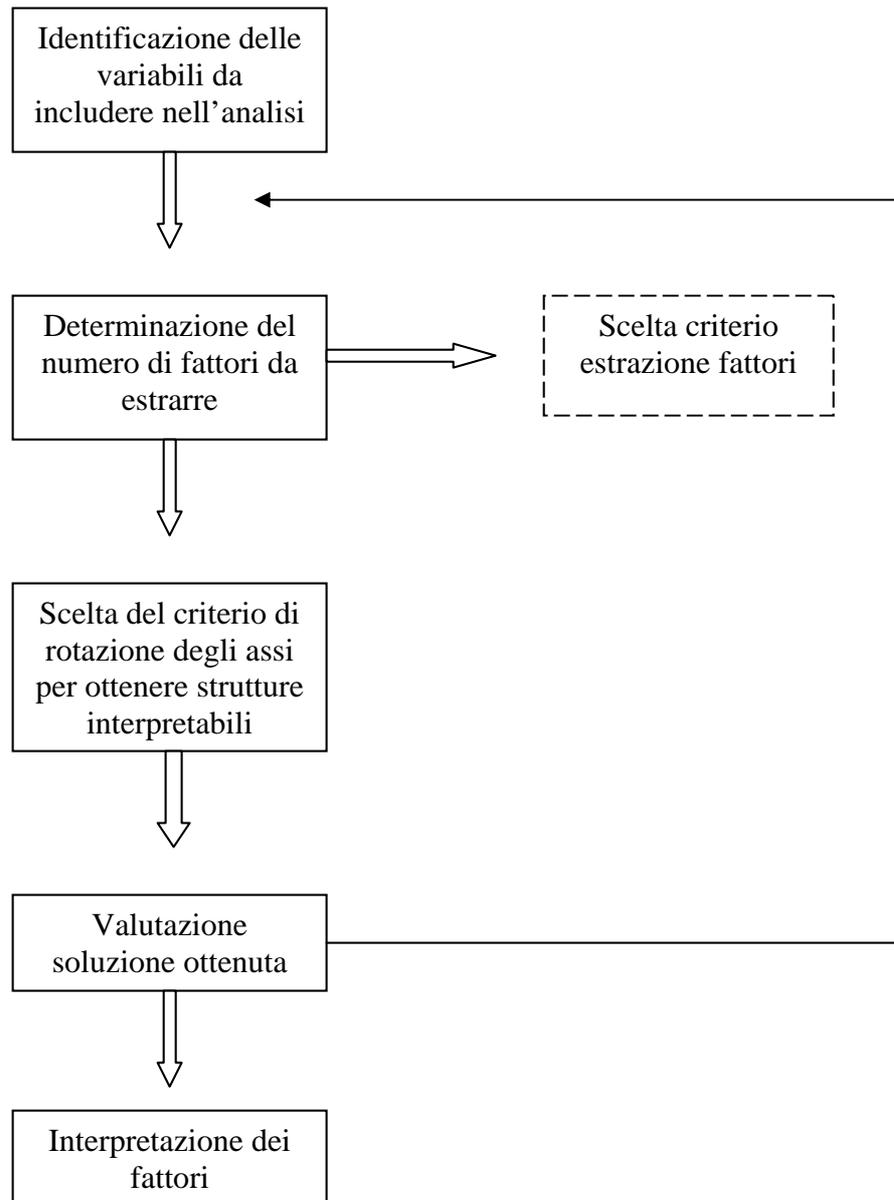
Nello svolgere un'analisi fattoriale vanno prese alcune decisioni (Figura 2.3).

Bisogna:

1. *Identificare la matrice sulla quale si svolgerà l'analisi.* Solitamente si considera la matrice di correlazione  $\mathbf{R}$  o la matrice di varianze e covarianze

2. *Stabilire il numero di fattori da estrarre.* Il numero massimo di fattori che possono essere considerati è  $r$  (rango della matrice di correlazione) anche se solitamente ne vengono utilizzati un numero inferiore. Tale scelta deve essere coerente con i principi di *parsimonia della soluzione finale*, ossia il numero di fattori deve essere inferiore a quello delle variabili; di *partecipazione di ogni fattore all'interpretazione della variabilità dei fenomeni osservati*, per cui la comunanza deve crescere significativamente all'aumentare del numero di fattori, di *semplicità e interpretabilità della soluzione finale*, per poter stabilire facilmente e chiaramente in quale misura una variabile sia attribuibile ad un dato fattore
3. *Definire il criterio di estrazione dei fattori.* Solitamente inizialmente vengono estratte tutte le componenti principali e viene trasformata la matrice delle informazioni in funzione di queste, quindi si pongono sulla diagonale della matrice  $\mathbf{R}$  le relative comunanze in modo da ottenere una soluzione fattoriale unica (tale step viene fatto in automatico dai pacchetti statistici utilizzati)
4. *Determinare il criterio di rotazione degli assi ortogonali trovati.* Le rotazioni, che possono essere ortogonali od oblique, modificano i fattori in modo da rendere più realistici e semplici i fattori e facile l'interpretazione finale dell'analisi
5. *Calcolare i punteggi fattoriali*, valore che una unità statistica ha sul fattore, *valutare e interpretare i fattori.*

Figura 2.3 Sequenza di decisioni da prendere nell'eseguire un'AF



## 2.6 Criteri per determinare il numero dei fattori

Solitamente il numero dei fattori non è noto a priori per cui l'analista inizia con quello che crede sia il numero più probabile di fattori e poi, per approssimazioni successive, trova la soluzione più congruente con gli obiettivi della ricerca. I criteri maggiormente utilizzati per la determinazione del numero dei fattori sono due: uno basato sulla *varianza spiegata dai fattori* e uno sulla *rappresentazione grafica degli autovalori*.

### *Varianza spiegata dai fattori:*

Questo criterio consiste nell'estrarre un numero di fattori tale per cui venga spiegata una certa quota di varianza.

Ricordiamo che l'autovalore  $\lambda_i$  del fattore  $i$  è la sua varianza e la somma degli autovalori è uguale alla somma delle varianze se l'analisi è condotta su una matrice di varianze – covarianze e a  $p$ , numero di variabili, se è condotta su una matrice di correlazione.

La quota di varianza estratta dal fattore  $i$  è:

$$\lambda_i / \sum_k^r \lambda_k = \lambda_i / \sum_k^p s_k^2 \quad (2.9)$$

se l'analisi è condotta su una matrice di varianze – covarianze e

$$\lambda_i / \sum_k^r \lambda_k = \lambda_i / p \quad (2.10)$$

se è condotta sulla matrice di correlazione.

Una percentuale di varianza del 75% è considerata un buon traguardo, anche se spesso si tollerano percentuali inferiori a questo valore. La frazione di varianza complessivamente estratta si valuta in funzione del numero di

variabili inserite nell'analisi e dal tipo di impiego che si farà delle nuove variabili latenti costruite.

*Rappresentazione grafica degli autovalori:*

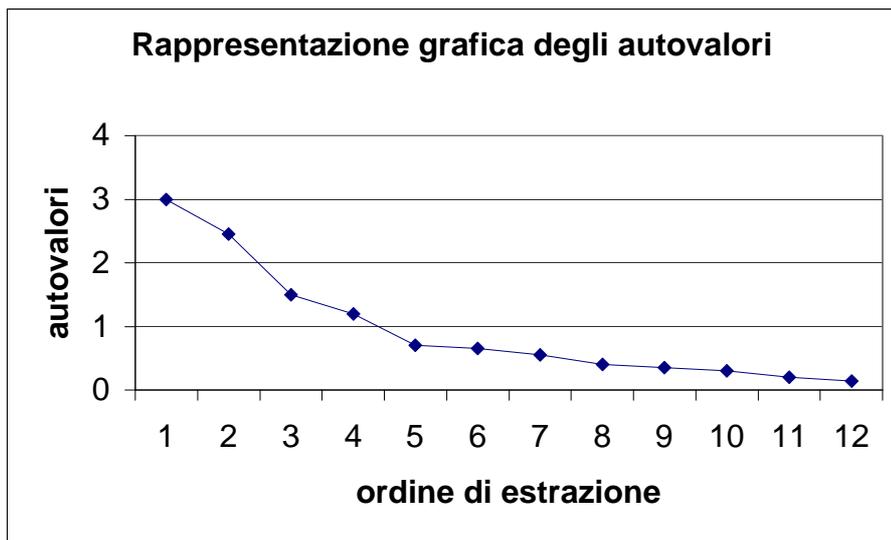
La rappresentazione grafica degli autovalori  $\lambda_i$  in relazione all'ordine di estrazione  $i$  permette di individuare gli autovalori importanti (Figura 2.4). Rappresentando i punti  $(i, \lambda_i)$  ( $i = 1, \dots, q$ ) sul piano cartesiano e collegandoli con segmenti, si ottiene una spezzata: se questa mostra due tendenze: una forte inclinazione all'altezza dei primi fattori e un successivo appiattimento che la porta ad essere quasi parallela all'asse delle ascisse, i fattori che appartengono a quest'ultima parte della spezzata possono essere ignorati. Si considerano rilevanti per l'analisi solo i fattori il cui autovalore, stando più in alto del flesso, descritto dalle due tendenze, si stacca visibilmente dagli altri.

Inoltre, per motivi che ora non analizziamo, vengono considerati di rilevante importanza solo gli autovalori superiori all'unità.

Se non ci sono fattori che prevalgono nettamente sugli altri allora significa che l'analisi fattoriale non è un metodo adatto per l'analisi di quei dati.

Figura 2.4 Esempio di rappresentazione grafica degli autovalori

<b>Ordine di estrazione</b>	<b>Autovalori</b>	<b>Percentuale cumulata di varianza spiegata</b>
<b>1</b>	<b>2,99</b>	<b>24,60%</b>
<b>2</b>	<b>2,45</b>	<b>46,70%</b>
<b>3</b>	<b>1,5</b>	<b>58,80%</b>
<b>4</b>	<b>1,2</b>	<b>72,30%</b>
5	0,7	78,30%
6	0,65	83,90%
7	0,55	87,70%
8	0,4	91,30%
9	0,35	94,30%
10	0,3	97,20%
11	0,2	99%
12	0,14	100%



Nel caso riportato ad esempio nella Figura 2.4 i fattori estratti sono quattro e complessivamente spiegano il 72,3% della variabilità totale.

## 2.7 Rotazione dei fattori

I pesi fattoriali  $a_{ji}$  (*factor loadings*) coincidono con i coefficienti di correlazione tra le variabili iniziali e i fattori ed indicano quanto la variabile sia determinante per il fattore. Dall'analisi della matrice dei pesi fattoriali è possibile riuscire a comprendere quali variabili contribuiscono maggiormente alla definizione del fattore e quindi alla sua interpretazione ed essa inizialmente viene prodotta senza essere sottoposta ad alcuna rotazione.

La rotazione dei fattori, o degli assi, è pertanto un cambiamento di posizione delle dimensioni estratte nella prima fase dell'analisi che facilita la comprensione del significato dei fattori stessi. La rotazione si sostanzia nella riduzione del valore dei pesi fattoriali marginali, ossia quelli che nella costruzione originaria dei fattori risultano essere relativamente piccoli, e nell'incremento, in valore assoluto, dei pesi più significativi.

La soluzione ideale, ai fini dell'interpretabilità dei fattori, è quella in cui tutti i pesi fattoriali siano prossimi a 0 o a 1. La rotazione comporta per tanto una ridistribuzione delle comunanze delle variabili e della varianza spiegata dai fattori.

I principali criteri di rotazione ortogonali sono *Varimax*, *Quartimax*, *Equamax*.

### *Varimax:*

La rotazione con il metodo Varimax tende a minimizzare il numero di variabili con cui ciascun fattore ha coefficienti di correlazione elevati. Tale criterio è raccomandabile se si vuole ottenere una netta separazione tra i fattori e se la rotazione è effettuata senza precisi criteri di riferimento.

### *Quartimax:*

Tale criterio semplifica le righe della matrice dei pesi fattoriali, cercando di stabilire la corrispondenza tra la variabile sulla riga e uno o pochissimi fattori. Tale criterio è adatto per identificare i fattori che governano la variabilità delle

caratteristiche osservate e dà risultati migliori del metodo precedente quando si vuole semplificare il primo fattore estratto, che tende ad essere un fattore generale.

*Equamax:*

E' un compromesso tra i due criteri precedenti in quanto tenta di realizzare la semplificazione simultanea di righe e colonne della matrice dei pesi fattoriali. Non si adatta efficacemente a strutture semplici.

## ***2.8 La Cluster Analysis***

Sotto il termine generale di analisi di raggruppamento o cluster analysis si accorpano varie tecniche operanti su dati di tipo quantitativo volte a classificare l'insieme delle unità dell'analisi in gruppi, cluster, non definiti a priori, in base alle caratteristiche possedute. I gruppi vengono formati cercando di massimizzare l'omogeneità interna e le differenze tra i vari clusters.

Per stabilire la similarità tra le varie unità campionate, vengono calcolate delle distanze: questo comporta il dover scegliere una metrica che sia in grado di esprimere al meglio la distanza tra gli elementi considerati.

In sintesi, l'input dell'analisi è costituito da una matrice di dati quantitativi che riporta, per ciascuna unità statistica, il valore delle variabili rispetto alle quali si vuole operare la classificazione, mentre l'output è una nuova variabile categoriale le cui modalità rappresentano il cluster di appartenenza a cui ciascun elemento è assegnato in modo univoco.

Fondamentalmente, esistono due differenti tipi di algoritmi di classificazione: quelli *gerarchici*, suddivisi in *scissori* e *agglomerativi*, e quelli *non gerarchici*.

- **Algoritmi gerarchici:** ogni gruppo fa parte di un gruppo più ampio, il quale è contenuto a sua volta in uno di ampiezza maggiore e così in progressione fino al gruppo che contiene l'intero insieme di unità analizzate. Gli algoritmi gerarchici si suddividono in:
  - *Scissori:* quando l'insieme delle  $n$  unità, in  $n-1$  passi, si ripartisce in gruppi che sono, ad ogni passo dell'analisi, sottoinsieme di un gruppo formato allo stadio precedente, e che termina con la situazione in cui ogni gruppo è composto da una unità
  - *Aggregativi:* se procedono a una successione di fusioni delle  $n$  unità, a partire dalla situazione di base nella quale ogni unità costituisce un gruppo a sé stante e fino allo stadio  $n-1$  nel quale si forma un gruppo che le contiene tutte (questi sono maggiormente usati in quanto richiedono un minor tempo di elaborazione)
  
- **Algoritmi non gerarchici:** in questo caso è necessario conoscere a priori il numero di cluster che si vogliono ottenere ed i centroidi iniziali di tali cluster. L'algoritmo procede in maniera iterativa cercando di ottenere la migliore classificazione degli elementi secondo il numero di classi prestabilito: ad ogni iterazione dispari vengono accorpati i due cluster più vicini mentre ad ogni iterazione pari viene separato il cluster più disomogeneo. Si procede poi al calcolo dei centroidi fino a quando lo spostamento dei centroidi da un'iterazione all'altra diventa infinitesimale.

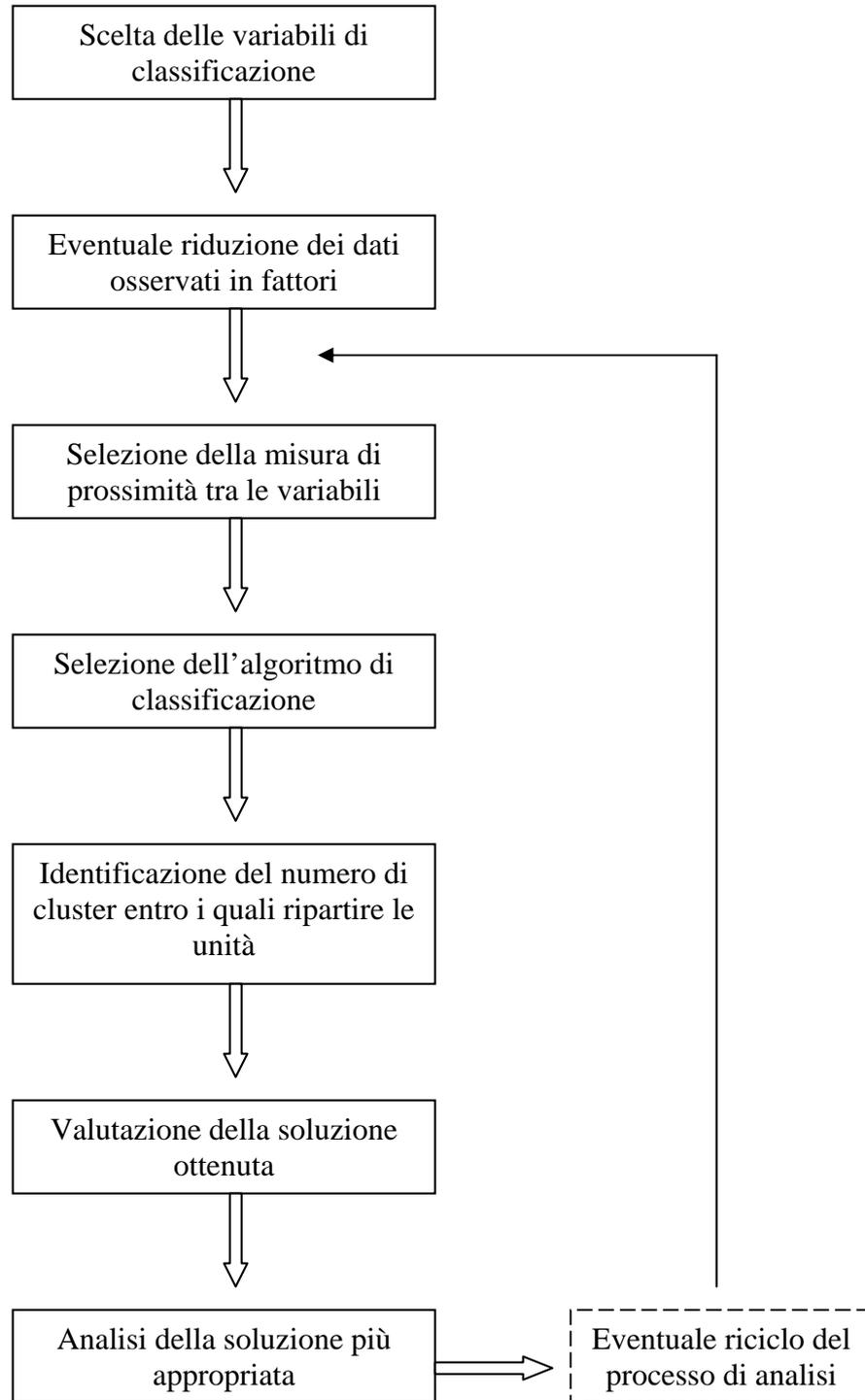
Le procedure di analisi non gerarchica si suddividono in due categorie a seconda che generino *partizioni*, ossia classi mutuamente esclusive, o *classi sovrapposte*, per le quali si ammette la possibilità che un elemento appartenga contemporaneamente a più cluster.

## 2.9 Il percorso di analisi

Per effettuare una cluster analysis si devono prendere diverse decisioni (Figura 2.5):

1. *Identificare le variabili di classificazione.* Si consideri la matrice di dati  $X = \{x_{hj}\}$  ( $h = 1, \dots, n$  ;  $j = 1, \dots, p$ ) relativa ad  $n$  osservazioni su  $p$  variabili ( nell'analisi si possono considerare o le variabili osservate o una loro opportuna trasformazione, spesso definita tramite l'analisi delle componenti principali o fattoriale). Solitamente le variabili vengono standardizzate e rese omogenee per quanto concerne l'indice di variabilità
2. *Selezione della misura di prossimità tra le unità da raggruppare.* Se l'obiettivo dell'analisi è la classificazione delle unità si userà una matrice simmetrica di ordine  $n$  (solitamente matrice di varianze e covarianze o matrice di correlazione), se invece è la classificazione delle variabili una matrice di ordine  $p$
3. *Selezione della tecnica di raggruppamento delle entità.* Le tecniche di raggruppamento proposte in letteratura sono numerose e diverse tanto che risulta difficile riuscire a capire quale si adatti meglio agli obiettivi di ogni singola analisi. Ricordiamo le gerarchiche, agglomerative e scissorie e le non gerarchiche, che generano partizioni o classi sovrapposte
4. *Identificazione del numero di gruppi entro i quali ripartire le entità.* Questo problema risulta simile a quello per la scelta del numero di fattori nell'analisi fattoriale vista precedentemente
5. *Completamento dell'analisi e interpretazione dei risultati dell'analisi.*

Figura 2.5 Diagramma delle fasi della cluster analysis



### 2.10 Selezione della misura di prossimità tra le variabili

Dopo aver deciso se operare l'analisi sulla matrice iniziale di dati  $\mathbf{X} = \{x_{hj}\}$  ( $h = 1, \dots, n$ ;  $j = 1, \dots, p$ ) o su una loro trasformazione lineare ottenuta tramite l'analisi fattoriale e aver standardizzato i dati, si determina la matrice delle distanze, ossia quella matrice quadrata il cui elemento generico  $d_{hk}$  è una misura di distanza tra le unità  $h$  e  $k$ .

Tra le misure più utilizzate per la cluster analysis vi sono:

#### *Distanza euclidea:*

La distanza calcolata tra le entità  $h$  e  $k$  basata sulla distanza euclidea viene calcolata nel seguente modo:

$$d_{hk} = \left\{ \sum_i^p (x_{hj} - x_{kj})^2 \right\}^{1/2} \quad (h, k = 1, \dots, n) \quad (2.11)$$

La distanza tra  $x_{hj}$  e  $x_{kj}$  non varia al variare dell'origine o al ruotare degli assi.

#### *Distanza media assoluta:*

La distanza media assoluta  $d_{hk}$  tra le unità statistiche  $h$  e  $k$  nello spazio  $p$ -dimensionale definito dalle  $p$  variabili osservate è data da:

$$d_{hk} = \sum_i^p |x_{hj} - x_{kj}| \quad (h, k = 1, \dots, n) \quad (2.12)$$

ed è particolarmente appropriata quando le variabili sono su scala ordinale. Rispetto la distanza euclidea, la distanza media assoluta non è invariante rispetto a traslazioni o rotazioni degli assi coordinati.

***Distanza di Lagrange – Tchebychev:***

La distanza di Lagrange – Tchebychev tra due unità statistiche  $h$  e  $k$  è lo scostamento massimo, in valore assoluto, tra tutti gli scostamenti tra le singole variabili osservate e le unità  $h$  e  $k$ :

$$d_{hk} = \text{Max} |x_{hj} - x_{kj}| \quad (h, k = 1, \dots, n) \quad (2.13)$$

dove il valore massimo è calcolato in relazione alle  $p$  variabili osservate.

***2.11 Selezione di un algoritmo di classificazione***

Le tecniche di analisi dei gruppi possono essere divise in gerarchiche, aggregative e scissorie, e non gerarchiche.

Gli algoritmi gerarchici non necessitano della definizione a priori del numero di cluster che si vuole ottenere e risultano molto onerosi e poco efficienti dal punto di vista computazionale. Inoltre, sono fortemente influenzati dalla presenza di *outliers*.

Nel caso di dataset di elevate dimensioni, gli algoritmi non gerarchici risultano estremamente più efficienti e meno influenzati da valori anomali inoltre, essendo non monotoni, permettono che un'unità statistica, inizialmente inserita in un cluster, possa modificare il proprio gruppo di appartenenza durante il processo iterativo.

### 2.11.1 Tecniche gerarchiche aggregative

Date tre unità  $h$ ,  $k$  e  $l$  di numerosità rispettivamente  $n_h$ ,  $n_k$ ,  $n_l$ , le tecniche di analisi gerarchica aggregative prevedono di utilizzare la matrice delle distanze per trovare la coppia di elementi  $h$  e  $k$  che sono più vicine e formare così il primo cluster. Successivamente si ricalcola la matrice delle distanze sostituendo le righe e le colonne relative ai gruppi  $h$  e  $k$  con una riga e una colonna di distanze tra il gruppo  $(h, k)$  e il gruppo  $l$ . L'individuazione delle unità più prossime e il ricalcolo delle distanze si ripetono per  $n-1$  volte finché tutte le unità fanno parte di un gruppo unico.

Il calcolo della distanza  $d_{l(h,k)}$  tra l'entità  $l$  e il gruppo  $(h, k)$  può essere effettuato mediante vari criteri:

#### *Metodo della media di gruppo:*

La distanza tra l'elemento  $l$  ed il gruppo formato dalla fusione di  $h$  e  $k$  è data dalla media aritmetica delle distanze  $d_{hl}$  e  $d_{kl}$  ponderate con la numerosità degli individui appartenenti ai gruppi  $h$  e  $k$ :

$$d_{l(h,k)} = \alpha_h d_{hl} + \alpha_k d_{kl} \quad (h \neq k \neq l = 1, \dots, n) \quad (2.14)$$

dove  $\alpha_h = n_h / (n_h + n_k)$  e  $\alpha_k = n_k / (n_k + n_h)$  e  $d_{hl}$  e  $d_{kl}$  sono due misure qualsiasi di dissomiglianza, calcolate come mostrato precedentemente.

#### *Metodo del centroide:*

Operando con il metodo del centroide (vettore delle medie di una distribuzione multivariata), la distanza tra due gruppi è la distanza

euclidea tra i centroidi dei gruppi. La distanza tra l'unità  $l$  e il gruppo formato dalla fusione di  $h$  e  $k$  è data da:

$$d_{l(h,k)} = \{(\alpha_h d_{hl}^2 + \alpha_k d_{kl}^2 - \alpha_h \alpha_k d_{hk}^2)\}^{1/2} \quad (2.15)$$

$$(h \neq k \neq l=1, \dots, n)$$

dove  $d_{hk}$  indica la distanza euclidea tra due punti  $h$  e  $k$  qualsiasi e  $\alpha_l$  è il peso relativo del gruppo  $l$  ( $\alpha_l = n_l / (n_l + n_k)$ ).

*Metodo del legame singolo:*

Con la strategia del legame singolo la distanza tra l'unità  $l$  e la fusione  $(h, k)$  è la distanza minore tra  $l$  e le due unità aggregate:

$$d_{l(h,k)} = \min \{ d_{hl}, d_{kl} \} \quad (2.16)$$

$$(h \neq k \neq l=1, \dots, n)$$

*Metodo del legame completo:*

Il criterio del legame completo si contrappone, come logica e come risultati, a quello del legame singolo. Tra l'elemento  $l$  e il gruppo  $(h, k)$ , la distanza è infatti data dal valore più elevato tra  $d_{hl}$  e  $d_{kl}$ :

$$d_{l(h,k)} = \max \{ d_{hl}, d_{kl} \} \quad (2.17)$$

$$(h \neq k \neq l=1, \dots, n)$$

A differenza del metodo del legame singolo, con il metodo del legame completo, poiché si ottengono gruppi di forma circolare caratterizzati da notevole somiglianza interna, è possibile eseguire una ricerca dei gruppi omogenei.

*Metodo di Ward:*

Con il metodo di Ward, la scelta della coppia di unità da aggregare si basa sulla minimizzazione della devianza tra i centroidi dei possibili gruppi. La devianza ha un minimo pari a 0 quando tutti gli elementi sono isolati e un massimo pari alla somma delle devianze delle variabili di classificazione quando tutte le unità appartengono a un unico gruppo.

La distanza euclidea tra l'elemento  $l$  e il cluster  $(h,k)$  è data da

$$\sqrt{\frac{n_l n_{(h,k)} d_{l(h,k)}^2}{n_l + n_{(h,k)}}} \quad (2.18)$$

dove  $n_l$  è il numero di unità che compongono il gruppo  $l$  e  $n_{(h,k)} = n_h + n_k$ .

Per ogni livello gerarchico dell'algoritmo di classificazione si ottengono indicatori statistici che possono aiutarci nella scelta del numero ottimale di cluster. Tali indicatori si basano sulla scomposizione della variabilità tra e dentro i cluster: la variabilità tra i gruppi misura il livello di eterogeneità tra un cluster e l'altro (tanto più elevata è la variabilità, tanto più differenziati sono i gruppi di clienti a cui ci riferiamo); la variabilità entro i cluster misura il livello di omogeneità all'interno del gruppo (tanto più bassa è la variabilità, tanto più in ciascun cluster i comportamenti dei consumatori sono simili).

Il rapporto tra la variabilità tra i gruppi e quella totale, definita come  $R^2$ , fornisce una misura della qualità della clusterizzazione. Tra gli indicatori maggiormente usati ricordiamo:

- RSQ , valore dell'  $R^2$  per ogni livello gerarchico
- PSF (Pseudo F Statistic), misura del grado di separazione tra i cluster ad ogni livello gerarchico
- SPRSQ (Semipartial  $R^2$ ), peggioramento dell'  $R^2$  dovuto all'agglomerazione del livello gerarchico precedente
- PST2 (Pseudo  $t^2$  Statistic), misura il grado di separazione tra gli ultimi due cluster accorpati.

Relativamente ai primi due indicatori si tratta di identificare il livello gerarchico dopo il quale questi tendono a peggiorare bruscamente; per quanto riguarda invece gli ultimi due, valori relativamente elevati suggeriscono di arrestare la clusterizzazione al livello immediatamente precedente.

### ***2.11.2 Metodi gerarchici scissori o divisivi***

Il procedimento di suddivisione è concettualmente opposto a quello della aggregazione progressiva delle unità. Si parte infatti dalla situazione nella quale le  $n$  unità fanno parte di un unico gruppo e in  $n-1$  passi si perviene alla situazione nella quale ogni unità fa gruppo a sé stante.

Tra i metodi divisori, uno dei più utilizzati è il *K- Means* basato sulla distanza tra i centroidi, che prevede di effettuare una prima suddivisione in due gruppi sulla base della combinazione delle unità che minimizza la devianza interna ai gruppi. Ad ogni passo successivo,

individuato il gruppo che ha la massima devianza interna (devianza di ogni elemento dal centroide), la suddivisione dicotomica delle  $n$  unità del gruppo si effettua provando tutte le possibili combinazioni con 1 e  $n-1$  unità, 2 e  $n-2$  unità e così via, individuando quella che minimizza la funzione:

$$D = \sum_g^G \sum_h^{n_g} \sum_i^p \left( {}_g x_{hi} - \overline{{}_g x_{xi}} \right)^2 \quad (2.19)$$

dove  ${}_g x_{hi}$  ( $g = 1,2; h = 1, \dots, n_g; i = 1, \dots, p$ ) è il valore della variabile  $x_i$  osservato presso l'unità statistica  $h$  appartenente al sottogruppo  $g$  e  $\overline{{}_g x_{xi}}$  è il valore medio della variabile  $i$  nel sottogruppo  $g$ . Il metodo di analisi *K-Means* consiste nella suddivisione ad ogni passo del campione sulla base di un numero qualsiasi ma opportuno di suddivisioni.

### ***2.11.3 Criteri che generano partizioni non gerarchiche***

La maggior parte di questi criteri consiste nell'eseguire una successione, anche iterata, di tre procedure volte ad avviare il processo classificatorio, individuando una soluzione provvisoria; ad assegnare le unità ai gruppi individuati nella prima fase; ad assegnare gli elementi a gruppi diversi da quelli precedentemente individuati, ottimizzando una funzione obiettivo.

Per quanto riguarda le procedure di avvio dell'analisi, si sfruttano le informazioni sui gruppi o ottenute da altre analisi, anche gerarchiche, oppure da un'analisi *K-Means* non gerarchica vista precedentemente.

Se non sono disponibili queste informazioni, si può utilizzare la tecnica proposta da Beale che considera un numero elevato di centroidi casuali ed assegna le unità statistiche ai diversi gruppi in base alla minima distanza euclidea dai centroidi; quindi iterativamente vengono spaccati i cluster meno omogenei, fornendo i due nuovi gruppi e ricalcolando i centroidi, fino a quando gli spostamenti tra questi diventano irrilevanti. Anche per questi algoritmi è possibile ricorrere ad una serie di indicatori statistici per la valutazione della soluzione di classificazione ottenuta; in particolare:

- *Frequency*, numero di unità statistiche appartenenti a ciascun cluster, che deve essere il più possibile omogeneo
- *Max distance from seed to observation*, indica la distanza massima tra il centroide di ciascun cluster e la relativa osservazione maggiormente distante. Valori relativamente piccoli di tale indicatore segnalano una buona clusterizzazione
- *Distance between cluster centroids*, indica la distanza tra i centroidi dei cluster individuati. Valori relativamente alti segnalano una netta separazione tra i gruppi
- *R Squared*, quota di variabilità spiegata dall'analisi a livello totale e relativamente a ciascuna variabile. Il valore di questi indicatori dovrebbe essere significativamente superiore a zero.

#### ***2.11.4 Tecniche non gerarchiche con sovrapposizione***

Le tecniche di raggruppamento con sovrapposizione ammettono che, per un dato numero di gruppi, le unità appartengano a più insiemi disgiunti. Tra le varie tecniche ricordiamo brevemente:

- *Ricerca di insiemi sfuocati*: in tale tecnica i gruppi risultano compenetrati e le unità hanno un grado più o meno elevato di appartenenza ai gruppi. Appartengono al cluster gli elementi che si trovano entro un raggio fissato dal centro del gruppo, per cui un elemento può avere un livello di appartenenza non nullo su più gruppi
- *Analisi di miscugli di distribuzione*: si ipotizza una certa distribuzione delle frequenze delle  $n$  unità osservate, si identificano i gruppi e quindi si stabilisce la probabilità di appartenenza delle singole unità ai gruppi individuati
- *Analisi Fattoriale Q*: Analisi fattoriale condotta sulla trasposta della matrice dei dati, dopo una standardizzazione che rende uniforme la scala di misura delle variabili. La matrice fattorizzata è pertanto una matrice di similarità tra individui e i fattori sono combinazioni lineari di unità (non di variabili); la rappresentazione grafica degli elementi sugli assi definiti dai fattori, solitamente i primi due, è essenziale per decidere a quale gruppo assegnare le unità.

### ***2.12 Scelta tra metodi di analisi***

La qualità di una tecnica di raggruppamento può essere valutata in base a vari criteri:

- L'*oggettività* data dal fatto che se diversi ricercatori conducono la stessa analisi separatamente, questi devono giungere alla stessa conclusione
- La *stabilità* dei risultati della classificazione operando su campioni equivalenti, ossia vogliamo che i risultati dati dai

metodi di analisi non risentano significativamente di piccole variazioni del campione di riferimento

- L'*informatività* del risultato intermedio e finale
- La *semplicità* dell'algoritmo e la *rapidità* di esecuzione.

### *Tecniche gerarchiche o non gerarchiche?*

Prima di decidere quale sia la tecnica migliore da adottare può essere utile saper che:

- Le tecniche non gerarchiche sono in genere più informative delle gerarchiche perché danno anche risultati intermedi e indici relativi la qualità dei risultati
- I metodi gerarchici risentono della presenza di errori di misura o di altre fonti di variabilità presenti nelle misure di prossimità e i dati anomali creano alcuni problemi
- Se si cercano gruppi caratterizzati da forte omogeneità interna, le tecniche gerarchiche sono in genere meno efficaci di quelle non gerarchiche. I criteri che generano partizioni sono generalmente i migliori per questo tipo di situazione
- Il calcolo delle soluzioni gerarchiche, in particolare quelle agglomerative, è più rapido degli altri
- Uno svantaggio delle tecniche gerarchiche è la rigidità della soluzione: una aggregazione impropria effettuata nei primi stadi dell'analisi si trascina fino alla fine e può rendere i risultati artificiosi. D'altra parte, se un procedimento di analisi non gerarchica è avviato senza una adeguata conoscenza a priori, i risultati sono modesti. La soluzione più conveniente può essere allora quella di far precedere l'analisi non gerarchica da una gerarchica.

*Verifica statistica della bontà della soluzione*

Sulla soluzione di raggruppamento si possono applicare test statistici per verificare se questa è significativamente diversa da una ottenuta per caso. Con i test si verifica in genere se la distanza tra le medie dei gruppi è significativa.

Un test applicato su  $g$  gruppi è stato proposto da Beale:

$$F = \left[ \frac{D_{g-1}^2 - D_g^2}{D_g^2} \right] / \left[ \frac{n-g-1}{n-g} \left( \frac{g+1}{g} \right)^{p/2} - 1 \right] \quad (2.20)$$

Dove  $D_{g-1}^2 = (n-g) S_g^2$  è la deviazione dai centroidi di  $g$  gruppi e  $p$  è il numero di variabili.

Per verificare la significatività della devianza, il test va confrontato con il valore critico di una  $F$  di Snedecor con  $p$  e  $p(n-g)$  gradi di libertà. Se l' $F$  empirico supera  $F_\alpha$  critico si può affermare che il passaggio da  $g-1$  a  $g$  cluster comporta la riduzione di una quantità significativa di deviazione interna ai gruppi prima esistenti e quindi che  $g$  è il numero ottimo di gruppi.

## Capitolo 3

# IL POSIZIONAMENTO COMPETITIVO DEL PRODOTTO IN UNO SPAZIO MULTIDIMENSIONALE

### *3.1 Il mapping multidimensionale*

In mercati affollati e competitivi è fondamentale distinguersi. **Una strategia di posizionamento è la modalità per ottenere uno spazio di mercato identificabile dal cliente e difendibile in un ambiente competitivo;** è l'insieme delle analisi e delle azioni volte ad identificare e a costruire un vantaggio competitivo che differenzi un prodotto agli occhi del cliente.

Occorre individuare e valutare la posizione che il prodotto occupa nell' "insieme evocato", ossia tra le alternative di scelta percepite dal consumatore, tra le quali lo stesso consumatore, spesso inconsapevolmente, effettua valutazioni comparate, preferendo quella che più si avvicina a ciò che sta cercando (prodotto ideale) per benefici arrecati e bisogni soddisfatti.

Sul **piano analitico** il problema da risolvere è quindi quello di collocare i **benefici del prodotto**, in base a come vengono percepiti dal consumatore, **rispetto ai benefici attesi e di capire quanto questi vengano soddisfatti dal nostro prodotto e da quelli concorrenti**. In questo modo vengono collocate in uno stesso spazio percettivo (mappa multidimensionale) tutte le alternative tra cui il consumatore può effettuare la sua scelta di acquisto. Tali mappe sono in grado di condensare **in poche dimensioni effettivamente discriminanti**, sul piano dei benefici ricercati, le diverse composizioni di attributi dei prodotti oggetto di analisi.

Sul piano strategico il problema da risolvere è invece quello di trovare una posizione soddisfacente per il proprio prodotto o per la propria marca e di individuare gli strumenti (leve del marketing mix) idonei ed efficaci per occuparla e difenderla dalle azioni dei concorrenti.

### ***3.2 Gli obiettivi del mapping multidimensionale***

Gli studi concernenti il perceptual mapping hanno solitamente cinque obiettivi:

1. Evidenziare le caratteristiche rilevanti al fine di differenziare agli occhi del cliente i prodotti/servizi/marche studiati
2. Comprendere i punti di forza e di debolezza dei diversi profili di offerta, così come vengono percepiti dal consumatore
3. Rappresentare visivamente il grado di sostituibilità di prodotti concorrenti, suggerendo in via indiretta strategie di difesa e di attacco
4. Individuare le possibilità offerte dal mercato (vuoti di offerta), attraverso un'analisi di profittabilità delle aree a bassa densità competitiva
5. Comprendere la modalità ottimale per realizzare un nuovo prodotto o modificarne uno già esistente in modo da massimizzare il gradimento da parte dei consumatori target.

Quest'ultimo obiettivo può essere raggiunto **rilevando non solo le percezioni** della domanda, con riferimento alle alternative d'offerta disponibili, **ma anche le preferenze**, in modo da delineare il prodotto ideale che ciascun cliente preferirebbe a tutti gli altri, compresi i prodotti che possono essere concettualizzati ma che non esistono ancora nella realtà.

Appare quindi evidente la stretta relazione che esiste tra le analisi di segmentazione della domanda e di posizionamento dell'offerta. Una mappa di

posizionamento è tanto più significativa quanto più omogenee sono le valutazioni che hanno condotto alla costruzione della mappa stessa; è evidente quindi l'opportunità di condurre l'analisi di posizionamento sui segmenti di clientela individuati, dedicando particolare attenzione al target prescelto. L'analisi di posizionamento può tuttavia costituire a sua volta uno strumento di segmentazione, in particolare se costruita a partire da valutazioni di preferenza, consentendo di aggregare clienti che rivelino spazi percettivi assimilabili.

### ***3.3 Il procedimento logico del mapping***

La metodologia da seguire per effettuare un buon posizionamento del prodotto o della marca segue le seguenti tappe(De Luca, 1995).

In primo luogo bisogna **individuare il mercato** di riferimento ed i **potenziali acquirenti** che lo compongono verso i quali rivolgiamo il nostro interesse; quindi bisogna **identificare le caratteristiche del prodotto** che assumono maggiore importanza nel processo di valutazione delle diverse offerte e **delineare una scala per misurare i vari attributi**. Dopo aver ipotizzato quali siano gli attributi più importanti, **si rilevano i giudizi espressi dai consumatori** sul possesso di tali caratteristiche da parte dei diversi prodotti/marche e sull'importanza che il cliente attribuisce loro. Attraverso delle tecniche statistiche **vengono analizzate le correlazioni esistenti tra le variabili considerate e individuati dei fattori in grado di rappresentare gruppi di attributi originari fortemente correlati**; in questo modo si riduce consistentemente il numero di variabili da rappresentare sulla mappa. Tali mappe rappresentano in uno spazio a due o tre dimensioni, i cui assi sono costituiti dai fattori definiti in precedenza, i giudizi espressi dai consumatori con riferimento ai diversi prodotti/marche. L'addensamento di giudizi in

determinate zone della mappa consente di identificare il posizionamento delle marche analizzate.

In uno stesso studio sul posizionamento si possono costruire più mappe da analizzare prendendo a riferimento, ad esempio, a due a due, fattori diversi, ed ogni mappa può fornire molteplici e differenti informazioni utili per il ricercatore. Sarà poi compito di chi conduce lo studio capire su quale sia più opportuno focalizzare la propria attenzione.

**L'analisi del mapping multidimensionale** consente di valutare come siano percepiti dal mercato i prodotti dell'impresa e se di ognuno di essi si abbia una percezione omogenea (posizionamento appuntito) o diversificata (posizionamento disperso) e se ci sia corrispondenza tra le preferenze espresse da specifici gruppi di consumatori e il posizionamento di alcuni prodotti a loro rivolti. Inoltre è **possibile definire le posizioni occupate dai prodotti/marche concorrenti** con la conseguente possibilità di distinguere i concorrenti più diretti, con un posizionamento vicino al nostro, da quelli che difficilmente possono essere considerati una valida alternativa alla nostra offerta.

L'analisi del posizionamento permette anche di valutare i punti di forza e di debolezza dei nostri concorrenti e ci aiuta a prevedere le strategie che essi adotteranno in futuro e quindi a contrastarle in maniera efficace.

Le **strategie** che possono essere adottate sulla base delle analisi di posizionamento possono essere **di tipo conservativo**, nel caso in cui si voglia mantenere e rafforzare la posizione esistente, **o di riposizionamento**, nel caso in cui si voglia modificare la percezione che il cliente ha della nostra marca. Quando viene lanciato un nuovo prodotto, la strategia utilizzata è quella di affiancamento ai prodotti dell'azienda già esistenti anche se questo sarà diverso e tenderà di occupare lo spazio, possibilmente ideale, libero.

Nel caso in cui il nostro prodotto sia stato immesso nel mercato già da tempo e si differenzi poco dai concorrenti, il ricercatore può provare a cambiare le basi del posizionamento attraverso la valorizzazione di attributi prima trascurati. Vengono definiti dei nuovi assi in base ai quali posizionare i prodotti esistenti ed ideali, rivoluzionando le precedenti percezioni sul grado di similitudine e di

differenziazione fra le diverse marche, con l'obiettivo di far assumere al nostro prodotto un'identità fortemente differenziata dai concorrenti e cercando poi di farla percepire al consumatore attraverso delle efficaci politiche di comunicazione.

Per tradurre queste strategie in politiche operative occorre comprendere quali siano gli elementi che maggiormente influiscono sulla percezione del possesso di un determinato attributo da parte di una marca. Formulate delle ipotesi, si procede verificandone l'attendibilità attraverso dei test effettuati su un campione di consumatori e trasformando i risultati ottenuti in coerenti politiche di marketing mix.

La possibilità di conseguire in breve tempo un preciso posizionamento nella mente del consumatore dipende dalla novità delle caratteristiche sulle quali la marca vuole basare il proprio posizionamento, dalla forza con cui comunica i benefici che offre, dalla credibilità delle promesse che vengono fatte.

Per quanto riguarda l'uso che le imprese possono fare dei risultati delle analisi di posizionamento, è stato osservato che le strategie di posizionamento, per risultare efficaci, devono anzitutto basarsi su una forte e chiara identificazione dei fattori discriminanti sottostanti alle preferenze dei consumatori, devono riferirsi a segmenti di mercato ampi e dinamici, devono essere tradotte in politiche di mercato (marketing mix) coerenti rispetto agli obiettivi assunti, sinergiche al loro interno e dotate di risorse tecniche e finanziarie sufficienti per attuarle, debbono infine non trascurare un consono posizionamento del prodotto anche rispetto al *trade* ed essere sottoposte a strumenti efficaci di controllo nelle loro fasi attuative. (Collesei, 1994)

Ovviamente le sole analisi di posizionamento non sono sufficienti per la definizione di efficaci strategie per contrastare la concorrenza, in quanto devono essere integrate ad attente analisi sui punti di forza e di debolezza e sui comportamenti competitivi.

### ***3.4 Le tecniche statistiche per il mapping multidimensionale***

Le tecniche statistiche di analisi multivariata maggiormente usate per la costruzione di mappe di percezione sono **l'analisi discriminante, l'analisi delle corrispondenze e il multidimensional scaling.**

Le prime due sono tecniche di posizionamento *attribute based*, essendo la loro applicabilità vincolata alla iniziale scelta degli attributi e delle caratteristiche dei prodotti analizzati di cui vogliamo realizzare il posizionamento; il terzo tipo di analisi, invece, non richiede una valutazione sugli attributi.

L'input dell'*analisi discriminante* è costituito dai punteggi quantitativi dati da un campione di consumatori ai vari attributi dei prodotti oggetto di analisi. A partire da tali giudizi, l'analisi discriminante individua le dimensioni, legate agli attributi prescelti, che distinguono al meglio i prodotti da posizionare. Più precisamente, **l'analisi individua le combinazioni lineari degli attributi che spiegano al meglio le differenze nelle valutazioni dei diversi prodotti**, le quali sono tra loro ortogonali e costituiscono le dimensioni di sintesi su cui viene realizzato il posizionamento. Nella mappa di output sono rappresentati sia gli attributi di partenza, utilizzando la correlazione lineare con le dimensioni discriminanti, sia i prodotti da posizionare, utilizzando le distanze medie tra gli oggetti calcolate in base alle dimensioni discriminanti considerate.

L'applicazione dell'analisi discriminante nelle ricerche di marketing si è mossa in due direzioni distinte.

In primo luogo, è la procedura metrica più utilizzata per il mapping multidimensionale, sopra descritto. Questo è da attribuirsi sia all'ampia disponibilità di software applicativi sia alla robustezza dell'algoritmo, ossia alla possibilità di rilassare le ipotesi relative la continuità e la distribuzione statistica dei dati. L'input dell'analisi discriminante è in questo caso costituito da valutazioni fornite da ogni intervistato, concernenti una serie di caratteristiche relative agli oggetti da posizionare. Le ipotesi di base

riguardano appunto le scale di valutazione, che dovrebbero essere continue e distribuite normalmente. Tuttavia, utilizzando la tecnica con l'obiettivo del mapping, una semplice struttura ordinale delle scale si è rivelata sufficiente per ottenere risultati significativi.

Un secondo gruppo di applicazioni riguarda lo studio del comportamento della domanda, in particolare per quanto concerne la propensione all'acquisto/fruizione di determinati prodotti/servizi. In questo caso la variabile dipendente è costituita dalla scelta effettiva, attuata all'interno dell'offerta disponibile, o dall'intenzione manifestata di acquistare un nuovo prodotto. Le variabili esplicative invece possono essere caratteristiche socio-demografiche o psicografiche o le situazioni e occasioni d'uso del prodotto. Si noti che, disponendo dei risultati dell'analisi discriminante condotta su un campione opportunamente selezionato e raccogliendo informazioni sulla popolazione relative solamente le variabili esplicative, è possibile prevedere il comportamento dei consumatori/utenti non inseriti nel campione.

*L'analisi delle corrispondenze* consente di costruire mappe di percezione a partire da semplici giudizi di presenza/assenza di più caratteristiche per ogni prodotto analizzato; si tratta quindi di una tecnica statistica quantitativa che permette di gestire dati qualitativi solitamente dicotomici (presenza = 1, assenza = 0). Il dato di input è costituito da una matrice di elementi non negativi indicanti generalmente la frequenza con cui certi prodotti sono stati associati ad un attributo o ad una situazione d'uso o bisogno; mentre, come nel caso dell'analisi discriminante, le mappe di output rappresentano congiuntamente l'oggetto di analisi e gli attributi considerati.

L'analisi delle corrispondenze possiede numerose caratteristiche che contribuiscono alla sua estrema utilità nell'ambito delle ricerche di mercato, in particolare per la capacità di analizzare simultaneamente l'andamento di più variabili categoriche. La natura multivariata può consentire di rilevare l'esistenza di interrelazioni strutturali che non potrebbero essere scoperte attraverso una semplice analisi tra coppie di variabili. E' da notare inoltre la

flessibilità dei vincoli sui dati di input: l'unica richiesta in tal senso è infatti quella di utilizzare una matrice di dati rettangolare, avente elementi non negativi.

La limitata complessità dei dati in input, rispetto a quelli richiesti dall'analisi discriminante, va a scapito della significatività dei test statistici utilizzati. Inoltre, le valutazioni espresse su scale continue devono essere rese qualitative per venire utilizzate nell'analisi, con conseguente perdita di informazioni (ad esempio la scala di misurazione viene suddivisa in due parti ed una rappresenta la mancanza di una data caratteristica e l'altra la presenza).

Per quanto riguarda le analisi del posizionamento multidimensionale, l'analisi discriminante può servire sia per ripetere lo stesso studio in due momenti diversi, ad esempio prima e dopo una campagna pubblicitaria, potendo così evidenziare le modifiche nelle percezioni legate alla campagna stessa. La tecnica può rivelarsi utile anche nella fase di disegno di un nuovo prodotto o di modifica di un prodotto esistente: raccogliendo informazioni sull'apprezzamento da parte dei consumatori di una varietà di caratteristiche di una nuova offerta, è possibile ottenere indicazioni per una scelta opportuna delle strategie di marketing mix.

Il *multidimensional scaling* è una tecnica di posizionamento non *attribute based*: l'input necessario per la costruzione delle mappe è costituito da semplici valutazioni di similarità reciproca attribuite ai prodotti o marche da posizionare. Il multidimensional scaling è *metrico* nel caso in cui conosciamo già o le coordinate degli elementi in uno spazio multidimensionale o la distanza tra questi; è *non metrico* nel caso in cui non conosciamo la distanza ma ci viene fornita una valutazione di dissimilarità tra le coppie di oggetti analizzati (ad esempio ad un campione di consumatori viene chiesto di indicare in una scala da 1 a 10 quanto reputino simili due prodotti e la stessa domanda viene ripetuta per tutte le possibili combinazioni di oggetti da analizzare).

Un vantaggio di questo tipo di analisi, rispetto ai due sopra descritti, è che non si corre il rischio di inficiare l'analisi trascurando attributi rilevanti o comunque imponendo spazi di posizionamento già prestrutturati al campione da intervistare. Rimane però il problema della complessità delle procedure di raccolta dei dati di similarità ed inoltre i risultati dell'analisi sono particolarmente sensibili al numero e alle caratteristiche dei prodotti messi a confronto.



## Capitolo 4

# L'ANALISI DISCRIMINANTE E IL MULTIDIMENSIONAL SCALING

### *4.1 L'analisi discriminante*

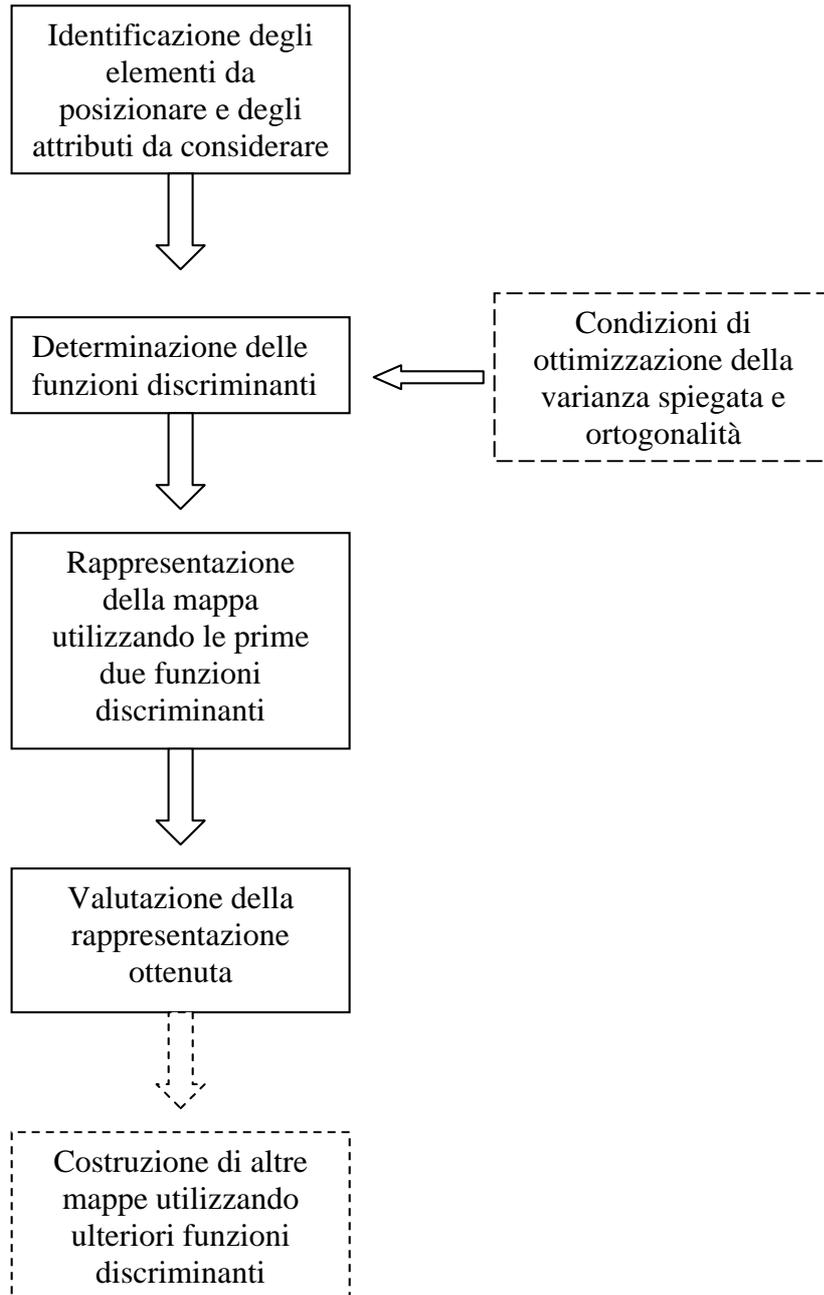
L'analisi discriminante è molto simile ad un'altra tecnica statistica multivariata: l'analisi di regressione. In entrambi i casi si utilizza il metodo dei minimi quadrati per interpolare i dati con un modello lineare, con la differenza però che, nel caso dell'analisi discriminante, la variabile dipendente è qualitativa, misurata a livello nominale, e non quantitativa. In particolare, nel caso del mapping multidimensionale, la variabile dipendente è l'oggetto da posizionare, per cui ogni oggetto valutato da ciascun intervistato costituisce un record di input e, nel caso in cui un rispondente valuti  $n$  oggetti, vengono generati  $n$  record di input.

In generale, l'analisi discriminante consente di stimare i coefficienti di un insieme di combinazioni lineari standardizzate, dette funzioni discriminanti, che spiegano al meglio le differenze nelle valutazioni dei diversi oggetti, ossia la varianza tra i punteggi ottenuti dai vari prodotti. La costruzione delle funzioni discriminanti segue un ordine logico: la prima funzione, ottenuta come combinazione lineare degli attributi, spiega la quota maggiore di varianza, la seconda spiega il massimo possibile tenendo conto della percentuale di varianza spiegata dalla prima, e così via fino a raggiungere la totalità della varianza con il numero massimo di funzioni discriminanti

ottenibili, pari al minimo tra il numero meno uno di elementi da posizionare ( $g-1$ ) e il numero di attributi ( $p$ ). Un ulteriore vincolo nella costruzione delle funzioni discriminanti consiste nel richiedere che non siano tra loro correlate, il che implica che le combinazioni individuate siano ortogonali. Le due proprietà appena descritte, ottimizzazione della varianza spiegata ed ortogonalità, costituiscono la base del mapping, le cui dimensioni sono costituite dalle funzioni discriminanti. Ovviamente la mappa con il maggior contenuto esplicativo è quella costruita utilizzando come dimensione orizzontale la prima componente principale e come dimensione verticale la seconda. Nel caso in cui la percentuale di varianza spiegata dalla mappa sia limitata, è possibile costruire una serie di altre mappe utilizzando le funzioni discriminanti successive.

Il processo di analisi viene riassunto in Figura 4.1.

Figura 4.1 Il procedimento logico dell'analisi discriminante



## 4.2 Aspetti metodologici

Si ipotizzi di avere a disposizione  $n$  osservazioni relative a  $p$  variabili  $x = (x_1, x_2, \dots, x_p)$  e di ordinare tali dati in una matrice  $\mathbf{X}$ . Si supponga inoltre di poter raggruppare a priori le osservazioni, sulla base di una variabile qualitativa opportunamente definita, ad esempio il prodotto che vogliamo posizionare.

Sia  $\bar{x}_i$  un vettore colonna di ordine  $p$ , contenente le medie riscontrate nelle variabili per quanto riguarda l' $i$ -esimo gruppo di osservazioni e si supponga di avere standardizzato i dati, in modo che la media globale per ogni variabile sia nulla. Si considerino le matrici quadrate di ordine  $p$

$$\begin{aligned} \mathbf{T} &= \mathbf{X}^T \mathbf{X} \\ \mathbf{B} &= \sum_i^g n_i \bar{x}_i \bar{x}_i' \\ \mathbf{W} &= \mathbf{T} - \mathbf{B} \end{aligned} \quad (4.1)$$

dove  $g$  è il numero complessivo di gruppi di osservazioni individuati a priori. La matrice  $\mathbf{T}$  è la matrice della somma totale dei quadrati e dei prodotti incrociati relativi alle  $p$  variabili considerate, avente sulla diagonale principale le somma totali dei quadrati per ciascuna variabile ed al di fuori della diagonale la somma dei prodotti incrociati. La matrice  $\mathbf{B}$  è la matrice della somma dei quadrati e dei prodotti incrociati tra i gruppi, mentre  $\mathbf{W}$  è la matrice della somma dei quadrati e dei prodotti incrociati nei gruppi e  $n_i$  la numerosità dell' $i$ -esimo gruppo.

Uno dei modi possibili di affrontare il problema della discriminazione è quello di cercare la funzione lineare  $a'x$  delle variabili  $x$ , che massimizzi il rapporto della varianza totale tra i gruppi con la varianza totale nei gruppi. Da cui, essendo  $\mathbf{T}$  e  $\mathbf{B}$  definite come sopra, la somma dei quadrati tra i gruppi, per la

combinazione lineare  $y = a'x$  è pari ad  $a' \mathbf{B} a$ , mentre la somma dei quadrati nei gruppi per la stessa funzione è pari a  $a' \mathbf{W} a$ ; il rapporto R tra queste due quantità è dato pertanto da

$$R = \frac{a' \mathbf{B} a}{a' \mathbf{W} a} \quad (4.2)$$

Tale massimizzazione deve avvenire subordinatamente ad opportuni vincoli sulla matrice  $\mathbf{A}$ :

- $\mathbf{A}^T \mathbf{B} \mathbf{A} = \mathbf{\Lambda}$  ossia la covarianza delle medie dei gruppi è nulla per ogni combinazione lineare.
- $\mathbf{A}^T \mathbf{W} \mathbf{A} = \mathbf{I}$  ossia le correlazioni delle medie tra i gruppi sono nulle

Se  $a_1$  è il vettore dei coefficienti che massimizza la (4.2), possiamo definire la funzione lineare  $z_1 = a_1' x$  *funzione discriminante lineare di Fisher o prima variata canonica*. Si noti come il vettore  $a$  possa essere riscalato senza che il rapporto venga modificato. Analogamente a quanto avveniva nell'analisi fattoriale, il vettore  $a_1$  nella funzione discriminante di Fisher corrisponde all'autovettore di  $\mathbf{W}^{-1} \mathbf{B}$  associato all'autovalore più grande.

La seconda funzione discriminante  $z_2$  è invece quella combinazione lineare che, di tutte le possibili combinazioni lineari indipendenti dalle prima funzione discriminante  $z_1$ , spieghi al massimo le differenze di gruppo rimaste. La variata canonica  $z_2 = a_2' x$  è in relazione con l'autovettore  $a_2$  associato al secondo autovalore più grande della matrice  $\mathbf{W}^{-1} \mathbf{B}$ .

In generale  $\mathbf{W}^{-1} \mathbf{B}$  ammette un numero di autovalori pari al minimo tra il numero di variabili  $p$  e il numero di gruppi meno uno  $g-1$ ; gli autovettori corrispondenti rappresentano le diverse variate canoniche e solitamente le prime  $k$ , con  $k < \min(p, g-1)$ , riescono a spiegare sufficientemente le differenze tra i gruppi. Tali funzioni discriminanti, considerate a coppie, rappresentano poi gli assi principali delle mappe costruite. Nel caso in cui le prime due variate spieghino la maggior parte di variabilità, non è necessario considerare

altre funzioni discriminanti, altrimenti si costruiscono più mappe in rapporto a tutte le possibili combinazioni di variate canoniche.

A volte per determinare le funzioni discriminanti non vengono calcolati direttamente gli autovalori di  $\mathbf{W}^{-1} \mathbf{B}$  in quanto o la matrice  $\mathbf{W}^{-1} \mathbf{B}$  non è simmetrica e risulta difficoltoso determinare gli autovalori ed autovettori o la matrice  $\mathbf{W}$  non ammette inversa per cui non esiste la matrice  $\mathbf{W}^{-1} \mathbf{B}$ .

Per ovviare tale problema Johnson ha proposto di considerare la matrice  $\mathbf{T}$  (vedi 4.1) al posto di  $\mathbf{W}$  e di calcolarne le componenti principali: in tal modo si giunge ad una soluzione simile alla precedente.

### 4.3 I test per verificare le differenze tra i gruppi

Dopo aver definito le funzioni discriminanti di riferimento, è necessario eseguire alcuni test statistici per verificare se vi siano differenze delle medie e delle varianze tra i gruppi analizzati.

Per verificare l'ipotesi di uguaglianza delle *medie* di ciascuna variabile nei gruppi si possono utilizzare due test: il test F e il test Lambda  $\Lambda$  di Wilks.

Il *test F* è dato dal rapporto della varianza tra i gruppi con la varianza nei gruppi per ciascuna variabile:

$$F_i = \frac{(b_{ii} - w_{ii})(n - g)}{w_{ii}(g - 1)} \quad (i = 1, \dots, p) \quad (4.3)$$

Dove  $b_{ii}$  è l'elemento  $i$ -esimo sulla diagonale della matrice  $\mathbf{B}$  e  $w_{ii}$  della matrice  $\mathbf{W}$  e tale rapporto si distribuisce come una F di Snedecor con  $g-1$  e  $n-g$

gradi di libertà, da cui si deduce che maggiore è tale rapporto tanto meno è verosimile l'ipotesi che le medie dei gruppi siano uguali.

Il test *Lambda*  $\Lambda$  di Wilks si basa invece sul rapporto tra la somma dei quadrati nei gruppi e la somma dei quadrati totali:

$$\Lambda_i = \frac{w_{ii}}{b_{ii}} \quad (i = 1, \dots, p) \quad (4.4)$$

Nel caso univariato ( $p=1$ ) tale rapporto si distribuisce come una *Lambda*  $\Lambda$  di Wilks con 1,  $n-g$  e  $g-1$  gradi di libertà, mentre nel caso multivariato abbiamo:

$$|I + W^{-1}B|^{-1} \quad (4.5)$$

dove  $I$  è la matrice identità e con la scrittura  $|\dots|$  intendiamo che dobbiamo calcolare il determinante della matrice di ordine  $p$ . Tale rapporto segue una distribuzione  $\Lambda$  di Wilks con  $p$ ,  $n-g$  e  $g-1$  gradi di libertà, ma purtroppo solo in casi particolari, e cioè solo per alcuni valori dei parametri, si può sfruttare la distribuzione al finito di  $\Lambda$ . Si può comunque dire che tale rapporto tende ad avvicinarsi ad 1 quando le medie dei gruppi sono molto simili e tende a 0 quando la variabilità nei gruppi è bassa se comparata con la variabilità totale, ovvero quando la maggior parte della variabilità è attribuibile alla differenza tra le medie dei gruppi.

Per verificare invece l'uguaglianza tra le matrici di *varianze e covarianze* tra i vari gruppi si utilizza *il test M di Box*.

Definendo con  $C = \frac{W}{n-g}$  la matrice di varianza e covarianze nei gruppi e con

$C_i = \frac{W_i}{n_i-1}$  la matrice di varianze e covarianze del gruppo  $i$  ed essendo  $n_i$  la

numerosità del gruppo  $i$  stesso, si ha che

$$M = (n-g) \log |C| - \sum_i^g (n_i - 1) \log |C_i| \quad (4.6)$$

ed anche in questo caso con la scrittura  $|\dots|$  intendiamo che dobbiamo calcolare il determinante della matrice. Nel caso in cui le matrici di varianze e covarianze dei gruppi siano uguali tra loro si ha  $C_i = C$  per cui il valore di  $M$  tende ad essere molto piccolo e asintoticamente si distribuisce come una  $X^2$  con  $\frac{p(p+1)(g-1)}{2}$  gradi di libertà. L'approssimazione di Box è soddisfacente se ciascun  $n_i$  è maggiore di 20 e se  $g$  e  $p$  sono inferiori a 5.

#### ***4.4 Multidimensional scaling***

Le percezioni e le preferenze costituiscono due fenomeni fondamentali del comportamento umano. Muovendosi all'interno della gerarchia del processo decisionale, ovvero dal riconoscimento di uno specifico problema al conseguente comportamento, che si traduce nel processo di coscienza del problema – ricerca di informazioni sul prodotto – valutazione – acquisto – valutazione successiva all'acquisto, il consumatore viene continuamente stimolato a esprimere giudizi di similarità e preferenza relativamente alle varie offerte proposte (Figura 4.2).

La tecnica del Multidimensional Scaling ha l'obiettivo di ricostruire lo spazio percettivo dei consumatori target, basando l'analisi su semplici valutazioni di similarità/dissimilarità tra gli oggetti da posizionare. Sulla mappa così costruita, per agevolare il processo interpretativo, è possibile sovrapporre eventuali valutazioni di preferenza e valutazioni di tipo quantitativo su alcuni attributi ritenuti rilevanti per l'ottenimento di un'analisi più completa, senza tuttavia che tali valutazioni contribuiscano a determinare la configurazione

spaziale degli oggetti. Tale tecnica risulta quindi particolarmente indicata nel caso in cui il processo valutativo dell'intervistato si riveli fortemente condizionato da fattori emozionali o comunque motivazioni difficilmente razionalizzabili attraverso parametri espliciti.

Il prezzo da pagare per questa riproduzione abbastanza fedele dello spazio percettivo è il fatto di non avere a disposizione una serie di giudizi quantitativi che aiutino a valutare la soluzione ottenuta e il fatto che le procedure del multidimensional scaling si complichino notevolmente all'aumentare del numero di elementi da posizionare.

#### ***4.5 La raccolta dei dati***

Uno dei problemi fondamentali nell'ambito delle tecniche del multidimensional scaling è la scelta di un metodo appropriato di raccolta dei dati. Le procedure più comuni hanno come obiettivo il completamento di una matrice quadrata simmetrica  $n \times n$ , in modo che tutte le  $[n(n-1)]/2$  coppie di offerte vengano valutate e che i valori risultino confrontabili tra righe e colonne. Tali dati solitamente derivano da misure di similarità relative le  $n$  offerte da posizionare.

Il multidimensional scaling può essere *metrico* nel caso in cui conosciamo già la distanza tra i vari elementi da posizionare; è *non metrico* nel caso in cui non conosciamo la distanza ma ci viene fornita una valutazione di dissimilarità tra le coppie di oggetti analizzati, la quale ci permette di ricavare una misura di distanza.

Il grado di similarità tra gli  $n$  elementi in genere viene misurato direttamente e la misura può essere nominale, ordinale o a livello di intervallo.

Nel caso della misura *nominale*, una delle procedure più utilizzate è quella dei raggruppamenti soggettivi, ossia si chiede agli intervistati stessi di raggruppare gli elementi da analizzare in un numero prefissato di gruppi. Un indice di similarità si ottiene calcolando il numero di volte che gli oggetti  $i$  e  $j$  compaiono nello stesso gruppo:

$$s_{ij} = \sum_k^s x_{ij} \quad (4.7)$$

dove  $x_{ij} = 1$  se l'individuo  $k$  ha posto gli oggetti  $i$  e  $j$  nel medesimo gruppo, altrimenti vale 0.

La misura del grado di similarità a livello *ordinale* avviene in genere utilizzando due procedure alternative.

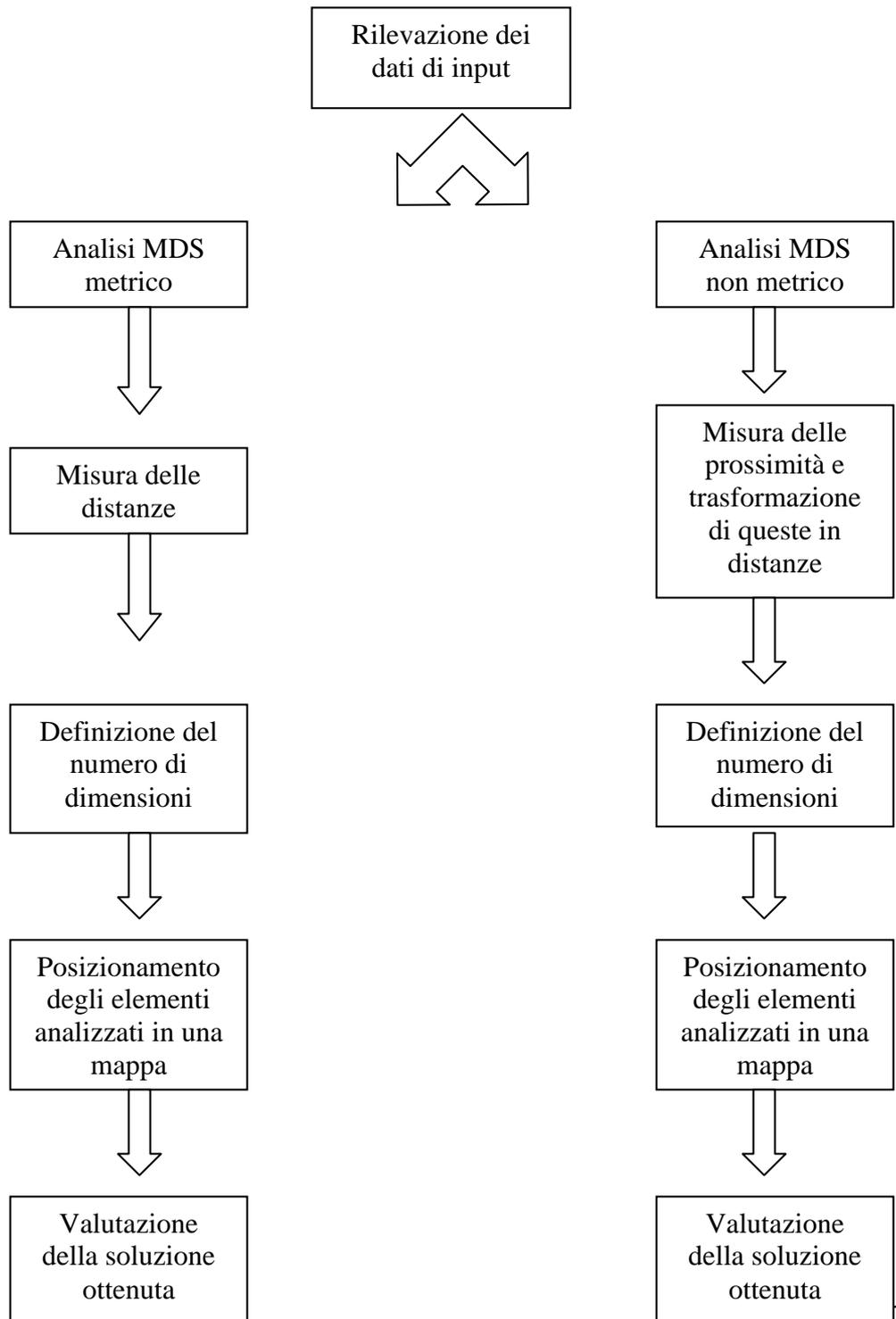
La prima è costituita dai paragoni a coppie con punti ancora mobili. Ossia si richiede all'intervistato di paragonare un determinato oggetto, punto ancora, ai rimanenti e di fornire un grado di similarità; ognuno degli  $n$  oggetti a turno costituisce il punto ancora. La matrice di similarità può essere ottenuta considerando per ciascuna coppia il grado di similarità medio.

Una seconda procedura per ottenere misure ordinali prevede di sottoporre agli intervistati l'elenco di tutte le coppie possibili e di richiedere l'ordinamento delle coppie stesse dalla più simile alla più dissimile. Al crescere del numero di elementi da posizionare, la gestione dei dati risulta però complicata per cui si preferisce richiedere un punteggio di similarità all'interno di una scala, ottenendo in questo modo una misura a livello di intervallo.

Le misure di similarità necessarie come input delle tecniche di multidimensional scaling possono anche essere ricavate *indirettamente*, per esempio a partire da valutazioni degli oggetti su un insieme di attributi giudicati rilevanti, con il conseguente calcolo di opportune misure di associazione. In questo caso si perde però la caratteristica peculiare di tale

tecnica, che consiste proprio nella possibilità di ottenere un posizionamento multidimensionale non basato sugli attributi.

Figura 4.2 Procedimento di analisi per il Multidimensional Scaling



### 4.6 Multidimensional Scaling metrico

Un modello spaziale per la matrice di similarità osservata consiste in un insieme di punti  $x_1 \dots x_n$  in  $d$  dimensioni, con ciascun punto che rappresenta uno degli elementi da analizzare, e una misura di distanza tra coppie di punti. Obiettivo di tecniche come il Multidimensional Scaling è determinare sia la dimensionalità del modello, ossia  $d$ , sia la posizione dei punti nello spazio  $d$ -dimensionale risultante, in modo che vi sia la massima corrispondenza tra le similarità osservate e le distanze tra i punti. In generale, all'aumentare della non somiglianza, deve crescere anche la distanza tra i punti nello spazio geometrico.

Si possono utilizzare diverse misure di distanza tra i punti, ma quella maggiormente usata è la *Distanza Euclidea*:

$$d_{ij} = \left[ \sum_k^d (x_{ik} - x_{jk})^2 \right]^{1/2} \quad (4.8)$$

dove  $x_{ik}$  e  $x_{jk}$  sono rispettivamente elementi dei vettori  $x_i$  e  $x_j$ .

Tale distanza può essere calcolata anche considerando la matrice  $\mathbf{B} = \mathbf{X} \mathbf{X}^T$  da cui si ottiene

$$b_{ij} = \sum_k^d x_{ik} x_{jk} \quad (4.9)$$

$$d_{ij}^2 = b_{ii} + b_{jj} - 2 b_{ij} \quad (4.10)$$

Si consideri ora il problema opposto, ovvero si supponga di conoscere le distanze e di voler determinare le coordinate. In primo luogo bisogna osservare che non vi è un'unica rappresentazione che dà origine alle medesime distanze, in quanto queste ultime rimangono immutate spostando la posizione

dell'intero insieme dei punti o ruotando la configurazione. Non è quindi possibile determinare univocamente né la posizione né l'orientamento della configurazione. Il primo problema viene risolto ponendo il vettore medio della configurazione all'origine degli assi, mentre la possibilità di sottoporre la configurazione ottenuta ad una trasformazione ortogonale arbitraria viene utilizzata per facilitare l'interpretazione della soluzione.

La procedura utilizzata per ottenere le coordinate richieste consiste in due fasi: la prima per l'ottenimento della matrice  $\mathbf{B}$  e la seconda per la fattorizzazione di  $\mathbf{B}$  in  $\mathbf{X X}^T$ .

L'ottenimento di  $b_{ij}$  in termini di  $d_{ij}$  comporta l'inversione dell'equazione (4.10) per la quale non esiste un'unica soluzione se non imponendo il vincolo

$$\sum_i^n x_{ij} = 0 \text{ per tutti gli } i \text{ e } j \quad (4.11)$$

La considerazione di tali vincoli implica che la somma per riga e per colonna di  $\mathbf{B}$  sia nulla. Sommando pertanto l'equazione (4.11) per  $i$  e per  $j$  e sia per  $i$  che per  $j$  si giunge alle tre seguenti equazioni :

$$\begin{aligned} \sum_i^n d_{ij}^2 &= D + n b_{jj} \\ \sum_j^n d_{ij}^2 &= n b_{jj} + D \\ \sum_i^n \sum_j^n d_{ij}^2 &= 2 n D \end{aligned} \quad (4.12)$$

dove  $D = \sum_i^n b_{ii}$  è la traccia della matrice  $\mathbf{B}$ .

Risolviendo le equazioni si ottiene che

$$b_{ij} = -\frac{1}{2} [d_{ij}^2 - d_{i.}^2 - d_{.j}^2 + d_{..}^2] \quad (4.13)$$

dove

$$d_{i.}^2 = \frac{1}{n} \sum_j^n d_{ij}^2$$

$$d_{.j}^2 = \frac{1}{n} \sum_i^n d_{ij}^2$$

$$d_{..}^2 = \frac{1}{n^2} \sum_i^n \sum_j^n d_{ij}^2$$

L'equazione (4.12) fornisce gli elementi della matrice  $\mathbf{B}$  in funzione delle distanze Euclidea al quadrato.

Per fattorizzare  $\mathbf{B}$  nella forma  $\mathbf{X} \mathbf{X}^T$  occorre determinare gli autovettori di  $\mathbf{B}$  e scalarli opportunamente. La matrice  $\mathbf{X}$  è quindi data da:

$$\mathbf{X} = [s_1 \ s_2 \ \dots \ s_n] \quad (4.14)$$

Dove con  $s_i$  si indicano gli autovettori della matrice  $\mathbf{B}$ .

Nel caso in cui si cerchi una configurazione in un determinato numero di dimensioni  $d$ , è possibile semplicemente considerare gli autovettori associati ai  $d$  autovalori più grandi.

Nel caso in cui non sia chiaro il numero di dimensioni da utilizzare, è possibile far ricorso al criterio  $Q$ , dato da:

$$Q = \frac{\sum_i^d \lambda_i}{\sum_i^n \lambda_i} \quad (4.15)$$

Dove  $\lambda_1, \dots, \lambda_n$  sono gli autovalori di  $\mathbf{B}$ .

Quando  $\mathbf{B}$  viene calcolata a partire da una matrice di distanze Euclidee, questa risulta essere semidefinita positiva per cui tutti gli autovalori sono positivi o nulli e quando viene fattorizzata come  $\mathbf{X} \mathbf{X}^T$  permette di ricavare valori reali e non immaginari delle coordinate. Non si può arrivare alla stessa conclusione se  $\mathbf{B}$  viene ricavata da matrici di dissimilarità non Euclidee. In questo caso la matrice può avere alcuni autovalori negativi e la fattorizzazione  $\mathbf{X} \mathbf{X}^T$  può condurre a valori immaginari per alcune delle coordinate. Se il numero di autovalori negativi è limitato e la loro grandezza non è elevata in valore assoluto, una configurazione dei punti che rispecchi sufficientemente la matrice di similarità può essere comunque ottenuta dagli autovettori associati ai primi autovalori positivi e l'adeguatezza della rappresentazione può essere misurata con la quantità:

$$Q_1 = \frac{\sum_i^d \lambda_i}{\sum_i^n |\lambda_i|} \quad (4.16)$$

#### ***4.7 Multidimensional Scaling non metrico***

Si supponga che i dati raccolti consistano in valutazioni di dissimilarità  $\delta_{ij}$  tra ciascuna coppia di oggetti  $(i, j)$  e che tali valutazioni siano inserite in una matrice simmetrica, con la diagonale principale composta da elementi nulli. Obiettivo del Multidimensional Scaling è individuare un insieme di coordinate in  $d$  dimensioni, solitamente due o tre, con associate le distanze  $d_{ij}$  che rappresentano opportunamente le dissimilarità osservate.

Per verificare il grado di adeguatezza tra distanze e dissimilarità, occorre definire una funzione che assuma valore nullo qualora le distanze rappresentino perfettamente le non somiglianze osservate e che cresca in valore al peggiorare della qualità della rappresentazione. Un esempio di tale funzione è dato dalla somma dei quadrati

$$S = \sum_i^{n-1} \sum_{i=i+1}^n (\delta_{ij} - d_{ij})^2 \quad (4.17)$$

Si noti che la quantità  $S$  è funzione delle  $n$  coordinate  $d$ -dimensionali attraverso le distanze  $d_{ij}$ . La somma dei quadrati è invariante a trasformazioni rigide come rotazioni ortogonali o traslazioni degli assi, ma è influenzata da trasformazioni di scala. Per esempio, se al posto di considerare la configurazione  $x_1, x_2, \dots, x_n$  si considera la configurazione  $kx_1, kx_2, \dots, kx_n$ , il valore di  $S$  cambia, anche se la relazione tra le distanze e quindi tra queste e le dissimilarità osservate, non muta.

Per questo motivo Kruskal ha proposto la misura STRESS, invariante a cambiamenti di scala:

$$\text{STRESS} = \left[ \frac{\sum_{i=1}^{n-1} \sum_{i=i+1}^n (\delta_{ij} - d_{ij})^2}{\sum_{i=1}^{n-1} \sum_{i=i+1}^n d_{ij}^2} \right]^{1/2} \quad (4.18)$$

Takane, Young e De Leeuw hanno proposto una seconda misura di adeguatezza della soluzione, detta S-STRESS:

$$\text{S-STRESS} = \left[ \frac{\sum_{i=1}^{n-1} \sum_{i=i+1}^n (\delta_{ij}^2 - d_{ij}^2)^2}{\sum_{i=1}^{n-1} \sum_{i=i+1}^n d_{ij}^4} \right]^{1/2} \quad (4.19)$$

Dove le distanze e dissimilarità osservate sono elevate al quadrato.

La relazione tra le distanze e le dissimilarità osservate, se si ricava la soluzione minimizzando le due equazioni precedenti (4.18, 4.19), è data da:

$$d_{ij} = \delta_{ij} + \varepsilon_{ij} \quad (4.20)$$

dove con  $\varepsilon_{ij}$  si indica una combinazione di errori di misura e distorsioni, legate al fatto che le dissimilarità osservate possono non corrispondere esattamente ad una configurazione in  $d$ -dimensioni.

In generale è tuttavia possibile ipotizzare che la relazione sia del tipo:

$$d_{ij} = f(\delta_{ij}) + \varepsilon_{ij} \quad (4.21)$$

dove con  $f(\delta_{ij})$  si indica una opportuna funzione delle dissimilarità osservate.

Il numeratore della misura STRESS diventa in questo caso:

$$S = \sum_{i=1}^{n-1} \sum_{j=i+1}^n [f(\delta_{ij}) - d_{ij}]^2 \quad (4.22)$$

Una possibile scelta per  $f$  è quella del modello lineare

$$d_{ij} = a + b\delta_{ij} + \varepsilon_{ij} \quad (4.23)$$

per cui si avrebbe

$$S = \sum_{i=1}^{n-1} \sum_{j=i+1}^n [a + b\delta_{ij} - d_{ij}]^2 \quad (4.24)$$

La procedura di minimizzazione delle quantità STRESS o S-STRESS avviene in due fasi distinte:

1. Per una data configurazione  $x_1, x_2, \dots, x_n$ , attraverso la regressione lineare semplice di  $d_{ij}$  su  $\delta_{ij}$ , si determinano i valori di  $a$  e  $b$  che minimizzano l'espressione (4.24)
2. Per questi valori di  $a$  e  $b$  si trovano le nuove quantità  $x_1, x_2, \dots, x_n$ , che minimizzano il criterio STRESS o S-STRESS.

Le due fasi vengono iterate fino a che non risulti soddisfatto un opportuno criterio di convergenza.

#### 4.8 Una misura per la valutazione del modello

Una misura idonea a valutare i risultati ottenuti sia tramite l'analisi metrica che quella non metrica è l'indice di interpretazione  $\alpha$ :

$$\alpha = \frac{1}{2} + \frac{1}{2} \frac{\sum_{i=1}^{n-1} \sum_{j=i+1}^n d_{ij}^* d_{ij}}{\left[ \left( \sum_{i=1}^{n-1} \sum_{j=i+1}^n d_{ij}^{*2} \right) \left( \sum_{i=1}^{n-1} \sum_{j=i+1}^n d_{ij}^2 \right) \right]^{1/2}} \quad (4.25)$$

dove  $d_{ij}$  è la distanza tra i punti  $i$  e  $j$  nelle configurazione ottenuta e  $d_{ij}^*$  è la disparità tra  $i$  e  $j$  ( la disparità è una trasformazione metrica delle prossimità non metriche iniziali), la cui sommatoria è estesa a tutte le  $n(n-1)/2$  possibili coppie di punti.

L'indice  $\alpha$  varia tra 0, corrispondente all'ipotesi di massima discrepanza fra i due tipi di dati, e 1, quando la prevedibilità dei dati di partenza sulla base della configurazione è perfetta.

## Capitolo 5

# UN CASO APPLICATIVO: SEGMENTAZIONE DELLA DOMANDA E POSIZIONAMENTO DELLA FACOLTA' DI SCIENZE STATISTICHE DI PADOVA, BOLOGNA E MILANO

### *5.1 Il caso di studio*

Il nostro interesse ora si focalizza sullo studio della segmentazione della domanda e del posizionamento delle Facoltà di Scienze Statistiche ed Economiche dell'Università degli Studi di Padova, dell'Università degli Studi di Bologna e dell'Università Bicocca di Milano. Per la realizzazione di tale analisi, è stato somministrato un questionario a quarantacinque studenti frequentanti una delle sopra citate Facoltà, per un numerosità campionaria complessiva di centotrentacinque individui. I dati rilevati sono quindi stati rielaborati con l'ausilio del pacchetto statistico SPSS 12.0 for Windows, grazie al quale si è potuto effettuare l'analisi fattoriale e la cluster analysis per ottenere una segmentazione della domanda. In seguito è stata svolta l'analisi discriminante per il posizionamento delle tre Facoltà in mappe di percezione, grazie all'ausilio del software statistico R 1.6.1 .

## ***5.2 Le Facoltà analizzate***

Le Facoltà di Scienze Statistiche presenti in Italia sono cinque, dislocate nelle seguenti città: Padova, Bologna, Milano, Roma e Messina. Il nostro studio si focalizza sulle prime tre e questo, oltre che per motivi logistici, per il fatto che reputiamo che uno studente residente in una città del Nord Italia, nel decidere quale Facoltà di Scienze Statistiche frequentare, abbia a disposizione una rosa di tre candidati: l'Università di Padova, Bologna e Milano Bicocca si trovano così ad essere dirette concorrenti.

Presso altre Università troviamo il dipartimento di Scienze Statistiche, ma noi abbiamo deciso di concentrare la nostra attenzione solo sulle Facoltà, le quali offrono le opportunità sotto descritte:

### *Facoltà di Scienze Statistiche ed Economiche di Padova*

La Facoltà di Scienze Statistiche ed Economiche di Padova propone due corsi di laurea triennale: in Statistica, Economia e Finanza e in Statistica e Gestione delle Imprese.

Il corso di laurea triennale in ***Statistica, Economia e Finanza*** offre una formazione interdisciplinare in statistica ed economia, fornendo le conoscenze di base di natura statistico-matematica ed economico-finanziaria utili per le analisi sia dei comportamenti individuali che di sistema.

In particolare, ogni laureato acquisisce:

- un'adeguata conoscenza dei metodi e delle procedure statistiche, con una particolare attenzione alle applicazioni alle analisi dei dati economici e finanziari
- un'adeguata conoscenza delle discipline di base nell'area delle scienze sociali, in particolare a carattere economico-finanziario, con una particolare attenzione agli aspetti quantitativi
- una buona padronanza del metodo della ricerca e della metodica statistica e di parte almeno delle tecniche statistico-economiche

- competenze pratiche ed operative, relative alla misura, al rilevamento ed al trattamento dei dati economici di sistema ed individuali, nonché alla formulazione di scenari previsivi a breve e medio-lungo termine
- gli strumenti logico-concettuali e metodologici per la progettazione ed esecuzione di indagini statistiche riguardanti fenomeni economici e per il trattamento informatico di basi di dati
- un'adeguata conoscenza della cultura dei contesti lavorativi
- adeguate competenze e strumenti per la comunicazione e la gestione dell'informazione.

Sono inoltre proposti, a scelta dello studente, due percorsi formativi formati da insegnamenti che permettono approfondimenti, rispettivamente, su *Finanza ed Economia*. Il primo prevede un approfondimento di tematiche riguardanti l'economia finanziaria e le analisi quantitative tipiche di questo ambito. Il secondo prevede un approfondimento di tematiche riguardanti la teoria e la politica economica e le analisi quantitative tipiche di questo ambito applicativo.

Il corso di laurea triennale in *Statistica e Gestione delle Imprese* offre invece una formazione interdisciplinare in statistica ed economia, con particolare attenzione alle problematiche aziendali. In particolare, ogni laureato acquisisce:

- un'adeguata conoscenza dei metodi e delle procedure statistiche, con una particolare attenzione alle applicazioni per la gestione operativa e strategica delle aziende
- un'adeguata conoscenza delle discipline di base nell'area delle scienze sociali, in particolare a carattere economico-aziendale, con una particolare attenzione agli aspetti quantitativi
- una buona padronanza del metodo della ricerca e della metodica statistica e di parte almeno delle tecniche statistiche economico-aziendali

- competenze pratiche ed operative, relative alla misura, al rilevamento ed al trattamento dei dati economici d'impresa, nonché alla formulazione di scenari previsivi a breve e medio-lungo termine
- gli strumenti logico-concettuali e metodologici per la progettazione ed esecuzione di indagini statistiche riguardanti fenomeni economici e per il trattamento informatico di basi di dati
- un'adeguata conoscenza della cultura dei contesti lavorativi
- adeguate competenze e strumenti per la comunicazione e la gestione dell'informazione.

Sono inoltre proposti, a scelta dello studente, due percorsi formativi formati da insegnamenti che permettono approfondimenti, rispettivamente, su *Analisi di Mercato e Gestione delle Imprese*. Il primo prevede lo sviluppo di tematiche come il marketing strategico, l'analisi dei mercati obiettivo e la previsione delle vendite, mentre il secondo la pianificazione e controllo e la certificazione della qualità.

### *Facoltà di Scienze Statistiche ed Economiche di Bologna*

La Facoltà di Scienze Statistiche ed Economiche di Bologna propone un corso di laurea triennale in *Statistica, Impresa e Mercati* .

Tale corso è caratterizzato dall'insegnamento dei metodi statistici e delle tecniche informatiche affiancati da discipline di approfondimento di natura economica, finanziaria ed aziendale, orientate a profili professionali di livello dirigenziale e manageriale.

La padronanza dei metodi statistici, matematici e informatici assicura una preparazione rigorosa utilizzabile professionalmente per raggiungere competenze specifiche in svariati campi.

In particolare, il corso di laurea prepara a professioni che richiedono competenze nel campo del marketing e delle ricerche di mercato, delle analisi di settore e della concorrenza, nella programmazione e nelle analisi economiche e finanziarie a livello nazionale e internazionale.

*Facoltà di Scienze Statistiche ed Economiche di Milano*

La Facoltà di Scienze Statistiche ed Economiche dell'Università di Milano Bicocca propone un corso di laurea triennale in *Scienze Statistiche ed Economiche*, che intende fornire le conoscenze teoriche, le competenze operative e le abilità pratiche indispensabili per:

- procedere alla misura dei fenomeni economici, tramite la rilevazione e il trattamento dei dati ad essi relativi, sia a livello microeconomico, sia a livello macroeconomico
- analizzare, processare, sintetizzare e trasmettere le informazioni economiche rilevanti nei diversi contesti lavorativi
- descrivere, interpretare e spiegare gli eventi e i processi statistici ed economici propri della realtà locale, nazionale ed internazionale e formulare previsioni accurate e affidabili sull'evoluzione delle variabili economiche rilevanti
- governare i processi decisionali caratteristici dei sistemi organizzativi complessi e contribuire, con le competenze specifiche fornite dalla teoria e dalla metodologia statistica ed economica, alla soluzione dei problemi decisionali propri delle diverse realtà aziendali e professionali.

Il Corso di laurea in Scienze Statistiche ed Economiche è strutturato in maniera tale da consentire ai propri laureati di possedere:

- una conoscenza ampia e accurata dei vari campi della statistica, nonché dei metodi ad essa propri
- una buona conoscenza delle discipline matematiche di base e un'adeguata padronanza delle tecniche e degli strumenti per la comunicazione e la gestione dell'informazione
- una competenza specifica approfondita nell'area delle discipline statistico-economiche, economico-politiche ed economico-aziendali
- una sicura padronanza delle teorie e delle tecniche di analisi dei dati qualitativi e quantitativi, di previsione economica e di stima econometrica.

Il Corso di laurea in Scienze Statistiche ed Economiche prevede infine, nelle fasi più avanzate del percorso formativo, un'articolazione in curricula differenziati, comprendenti insegnamenti e altre attività formative (quali laboratori, tirocini, ecc.) orientate all'apprendimento di competenze e capacità operative in specifici settori applicativi.

### ***5.3 Definizione del campione***

La popolazione di riferimento è composta da tutti gli studenti frequentanti uno dei corsi di laurea del nuovo ordinamento sopra descritti (Statistica e Gestione delle Imprese e Statistica, Economia e Finanza per la Facoltà di Padova; Statistica, Impresa e Mercati per la Facoltà di Bologna; Scienze Statistiche ed Economiche per la Facoltà di Milano) iscritti al secondo o al terzo anno, in modo che avessero già acquisito un'esperienza universitaria tale da potergli permettere di valutare in maniera idonea la Facoltà frequentata.

Ad ogni studente è stato somministrato e poi ritirato da me medesima un questionario, in aula durante l'orario di lezione (questo per la rilevazione di 130/135 questionari) oppure in aula studio (questo per la rilevazione di 5/135 questionari) ed è stato selezionato un campione di quarantacinque studenti per ognuna delle tre Facoltà, ottenendo una numerosità campionaria complessiva di centotrentacinque individui.

La popolazione di riferimento è composta nel seguente modo (vedi tabelle 5.1, 5.2, 5.3, 5.4):

Per quanto riguarda l'Università degli Studi di Padova, la popolazione di riferimento, ossia gli studenti iscritti al secondo o terzo anno, è composta da 281 individui (vedi Tabella 5.1 e 5.2) dei quali 103 iscritti al secondo anno e 178 al terzo ed il 38% di questa risulta fuori corso. Circa il 65% della popolazione di riferimento è iscritta alla laurea in Statistica e Gestione delle Imprese e i rimanenti a Statistica, Economia e Finanza.

Il nostro campione, formato da 45 studenti, rappresenta il 25% della popolazione.

*Tabella 5.1 Iscritti al corso di laurea triennale in Statistica e Gestione delle Imprese presso la Facoltà di Scienze Statistiche dell'Università degli Studi di Padova suddivisi per anno di iscrizione*

<b>Anno di corso</b>	<b>Tipo iscrizione</b>	<b>Numero studenti</b>
1	In corso	87
2	In corso	70
3	In corso	47
3	Fuori corso	67
<b>Totale</b>		271

*Tabella 5.2 Iscritti al corso di laurea triennale in Statistica, Economia e Finanza presso la Facoltà di Scienze Statistiche dell'Università degli Studi di Padova suddivisi per anno di iscrizione*

<b>Anno di corso</b>	<b>Tipo iscrizione</b>	<b>Numero studenti</b>
1	In corso	39
2	In corso	43
3	In corso	26
3	Fuori corso	28
<b>Totale</b>		136

Il campione selezionato è composto da 13 studenti fuori corso, il 29%, per cui rappresenta quasi il 17% degli studenti regolari iscritti al secondo e terzo anno di corso ed il 14% degli studenti fuori corso.

Per quanto riguarda l'Università degli Studi di Bologna, la popolazione di riferimento è composta da 161 individui (vedi Tabella 5.3) dei quali 55 iscritti al secondo anno e 54 al terzo e solo il 12% di questa risulta fuori corso.

Il nostro campione rappresenta oltre il 40% della popolazione.

*Tabella 5.3 Iscritti al corso di laurea triennale in Statistica, Imprese e Mercati presso la Facoltà di Scienze Statistiche dell'Università degli Studi di Bologna suddivisi per anno di iscrizione*

<b>Anno di corso</b>	<b>Tipo iscrizione</b>	<b>Numero studenti</b>
1	In corso	52
2	In corso	55
3	In corso	41
3	Fuori corso	13
<b>Totale</b>		161

Il campione selezionato è composto da soli due studenti fuori corso, il 4%, per cui rappresenta quasi il 45% degli studenti regolari iscritti al secondo e terzo anno di corso e il 15% dei fuori corso.

Per quanto riguarda l'Università degli Studi Bicocca di Milano, la popolazione di riferimento è composta da 140 individui (vedi Tabella 5.4): 48 iscritti al secondo anno e 92 al terzo e di questi il 37% risulta essere fuori corso.

Il campione, formato da 45 studenti, rappresenta il 32% della popolazione di riferimento.

*Tabella 5.4 Iscritti al corso di laurea triennale in Scienze Statistiche ed Economiche presso la Facoltà di Scienze Statistiche dell'Università degli Studi Bicocca di Milano suddivisi per anno di iscrizione*

<b>Anno di corso</b>	<b>Tipo iscrizione</b>	<b>Numero studenti</b>
1	In corso	56
2	In corso	48
3	In corso	40
3	Fuori corso	52
<b>Totale</b>		196

Il campione selezionato è composto da soli quattro studenti fuori corso, il 9%, per cui rappresenta quasi il 47% degli studenti regolari iscritti al secondo e terzo anno di corso e solo il 7% dei fuori corso.

La non rilevazione degli studenti fuori corso probabilmente è da attribuirsi al fatto che questi sostengono gli esami da non frequentanti.

#### ***5.4 Il questionario***

Obiettivo dell'analisi è quello di fornire una valutazione generale della Facoltà e si è cercato di sondare tutti gli aspetti ritenuti rilevanti da un neodiplomato impegnato nella scelta della Facoltà da frequentare. Per questo motivo le variabili considerate ricoprono vari aspetti del mondo universitario e non solo quello didattico; ovviamente nessuno di questi è stato approfondito in maniera significativa.

Potrebbe quindi risultare utile, in un secondo momento, somministrare dei questionari che esaminino in maniera più dettagliata le variabili che da questa prima indagine sono risultate fondamentali per gli studenti.

Agli universitari che rientrano nel nostro campione è stato chiesto di compilare un questionario (vedi Appendice) e di dare una valutazione da 1 a 5 ad ogni item. Abbiamo scelto questa scala di valutazione e non una con range 1-10 per il fatto che reputiamo che, in una di dimensioni limitate, l'intervistato sia maggiormente propenso ad utilizzare tutte le varie possibilità e non scelga solo i valori intermedi. In questo modo egli è maggiormente orientato a dare valutazioni che rispecchiano un giudizio nettamente positivo o negativo e non solo intermedio.

Agli studenti è stato chiesto sia di valutare l'importanza data ad ogni variabile analizzata, sia il livello di soddisfazione raggiunto in questi anni di Università. La misurazione del livello di importanza data ad ogni caratteristica serve poi per segmentare il mercato di riferimento: vogliamo capire in questo modo quali siano le esigenze degli studenti e i benefici che si aspettano di ricevere dal frequentare la Facoltà di Statistica. Tali dati verranno poi analizzati simultaneamente tramite l'analisi fattoriale e la cluster analysis.

La misurazione del grado di soddisfazione raggiunto indica invece quanto la Facoltà sia stata capace di soddisfare le esigenze dello studente e abbia risposto alle aspettative. Tali dati verranno considerati separatamente per ogni Facoltà ed ogni gruppo servirà per la definizione del posizionamento della

corrispondente Università. In questo caso la tecnica statistica multivariata utilizzata sarà l'analisi discriminante.

Gli items considerati sono venti e sono stati suddivisi in quattro classi: Contenuti proposti dai corsi universitari (sette items), Corpo docente (cinque items), Organizzazione e logistica (cinque items) e Rapporto Università e ambiente (tre items).

### *Contenuti*

Le materie studiate in ogni corso di laurea presso la Facoltà di Scienze Statistiche ed Economiche possono essere suddivise in tre macro aree: matematico-statistica, statistica applicata ed economica; e risulta interessante cercare di capire come debba essere secondo gli studenti la distribuzione del tempo da dedicare ad ognuna di queste tre aree.

Si vuole sondare anche l'importanza data alla trattazione di argomenti attuali come l'e-commerce o il CRM o la lettura di articoli tratti da giornali come Il Sole 24 ore, che aiutano gli studenti ad essere aggiornati e li spronano ad informarsi in maniera più approfondita. Molti corsi vengono integrati da cicli di seminari tenuti o da docenti esterni, sempre relativamente un argomento di attuale importanza, o da persone provenienti dal mondo del lavoro e della ricerca che possono raccontare la loro esperienza.

Per uno studente di Statistica è fondamentale anche imparare ad utilizzare degli specifici software che lo aiutino nell'elaborazione dei dati ed anche conoscere almeno una lingua straniera.

### *Corpo docente*

Si vuole sondare quale sia la disponibilità da parte dei docenti per chiarimenti relativamente agli argomenti del corso, per spostamenti di orario e in generale per andare incontro alle esigenze degli studenti.

La loro competenza viene percepita dai ragazzi in base alla loro capacità di suscitare interesse e di rendere le lezioni particolarmente piacevoli e alla loro

chiarezza espositiva, rendendo magari anche i concetti più ostici di facile comprensione.

L'Università può acquistare prestigio non solo perché possiede un nome storico, come nel caso di Padova e Bologna, ma anche perché ha ospitato insegnanti rinomati, che hanno segnato la storia con loro scoperte e scritti. Questo elemento, che può riguardare docenti del passato ma anche dei nostri giorni, potrebbe influenzare in maniera significativa la scelta di frequentare una Università rispetto ad un'altra.

Acquista sempre più importanza anche il tessuto di relazioni che i vari docenti e la Facoltà in generale, magari con il supporto di un ufficio stage, riescono ad instaurare con il mondo del lavoro, composto da enti pubblici ed imprese private, e con il mondo della ricerca, formato da istituti pubblici o privati di ricerca ed altre Università internazionali.

### *Organizzazione e logistica*

Una idonea pianificazione delle attività universitarie, come l'organizzazione dell'orario delle lezioni, può facilitare il compito dello studente, in particolare dei ragazzi pendolari, che in questo modo riducono al minimo il tempo perso in spostamenti (basti pensare che alcune Università come Ca' Foscari di Venezia hanno, negli ultimi anni, deciso di compattare le varie lezioni in soli tre giorni in modo da agevolare i pendolari).

Altro elemento importante è la possibilità per lo studente di creare un piano di studi personalizzato così che egli possa approfondire le materie che maggiormente lo interessano e magari gli serviranno per intraprendere la carriera tanto sognata.

Una biblioteca ben fornita può essere un ottimo supporto per la preparazione di alcuni esami e per la stesura della tesi e così anche la disponibilità di spazi dove studiare come aule studio e aule computer.

### *Rapporto Università e Ambiente*

Quando un neodiplomato si trova di fronte all'ardua scelta dell'Università da frequentare e spesso alla conseguente possibilità di andare a vivere in un'altra città, entrano in gioco anche variabili di tipo "ambientale". E così magari un giovane ragazzo, oltre a chiedere certi requisiti sopra descritti, valuta anche la possibilità di vivere una vita universitaria divertente e spensierata. Due delle città analizzate, Padova e Bologna, sono proprio il simbolo di questa vita goliardica ed offrono serate dedicate solo agli universitari e mille altre occasioni per divertirsi.

Uno studente prima di scegliere valuta anche la possibilità e la facilità di fruizione di servizi correlati come la mensa o librerie specializzate in testi universitari, copisterie ed alloggi o collegi a loro rivolti ed ovviamente il costo medio della vita nella città universitaria.

### *Dati anagrafici*

All'intervistato vengono inoltre richiesti alcuni dati anagrafici come sesso ed età e poi se sia fuori corso e se abbia partecipato ad un progetto Erasmus per capire se i giudizi dati precedentemente possano essere stati influenzati dal confronto con altre Università straniere.

Vogliamo anche indagare quanto la distanza residenza-Facoltà abbia influito nella scelta universitaria e se gli studenti siano disposti a spostarsi magari prendendo in affitto un appartamento o siano più propensi a fare i pendolari. Questa valutazione risulta rilevante dato che vogliamo capire quanto un giovane sia disposto a spostarsi per motivi di studio, e quindi la scelta di frequentare l'Università di Padova o Bologna o Milano sia maggiormente determinata da altri fattori. D'altra parte se pensiamo che per tutti gli studenti del Nord Italia la rosa dei nomi delle Facoltà di Statistica è composta da soli tre elementi, essi quasi sempre sono costretti a trasferirsi; ma anche un rapido sguardo al mondo del lavoro e degli scambi interculturali ci fa prevedere già la risposta.

*Ulteriori considerazioni*

Quando lo scopo dell'indagine è il posizionamento di più prodotti o marche in un'unica mappa delle percezioni, risulta interessante chiedere agli intervistati quanto reputino simili i prodotti analizzati. Vengono allora inseriti all'interno del questionario items relativi la similarità generale di due elementi (nel nostro caso ad esempio avremmo chiesto di dare una valutazione da 1 a 5, dove il valore minimo indica poca similarità e il massimo molta, relativamente la somiglianza generale tra le Università di Padova e Bologna, Padova e Milano, Bologna e Milano, e in seguito l'affinità tra due elementi relativamente una singola variabile, per es. la somiglianza tra le Università di Padova e Bologna per quanto riguarda la loro notorietà).

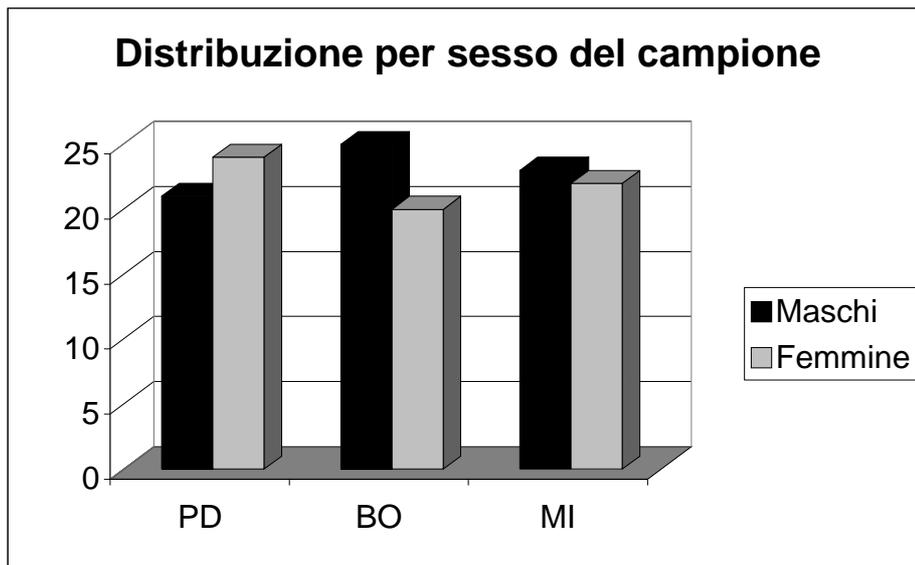
Il questionario somministrato non contiene questo tipo di item in quanto riteniamo che, dato che il singolo intervistato ha esperienza diretta di un solo prodotto e non di tutti e tre, i suoi giudizi possono risultare poco attendibili e quindi anche i risultati finali sarebbero poco significativi.

**5.5 Descrizione del campione**

Il campione si distribuisce uniformemente tra maschi e femmine per tutte e tre le Facoltà analizzate (Tabella 5.5):

*Tabella 5.5 Distribuzione per sesso del campione*

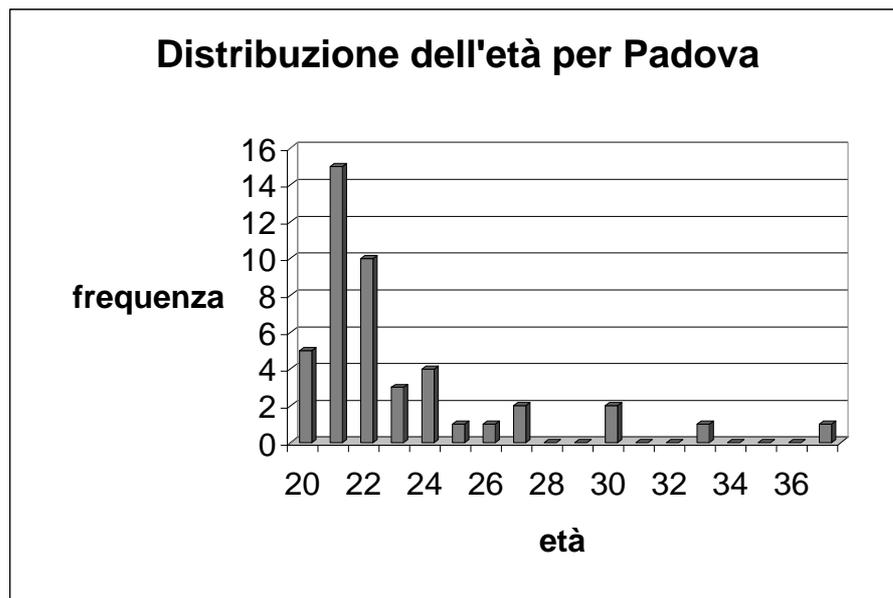
	<b>Padova</b>	<b>Bologna</b>	<b>Milano</b>
<b>Maschi</b>	21	25	23
<b>Femmine</b>	24	20	22



Per quanto riguarda l'*Università di Padova* (vedi Tabella 5.6), il campione intervistato ha un'età compresa tra i 20 e i 37 anni ed il 55% di questo è composto da ventunenni e ventiduenni. L'età media è di quasi 22 anni e mezzo, dovuta al fatto che nel campione sono presenti studenti con un'età abbastanza elevata (30, 33 e 37 anni). Questo è conseguenza sia dal fatto che quasi il 30% degli intervistati risulta fuori corso, sia che molti giovani non intraprendono la loro carriera universitaria al termine della scuola secondaria superiore ma solo dopo alcuni anni. Infatti i due studenti trentenni e il ventiseienne sono regolari negli studi.

Tabella 5.6 Distribuzione per età del campione relativo la Facoltà di Scienze Statistiche di Padova

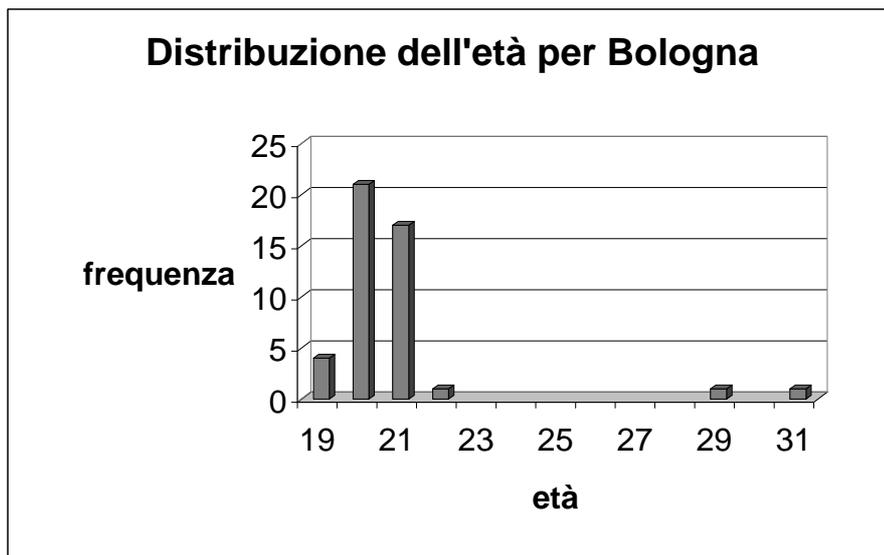
Età	n° studenti	Studenti in %	% cumulata
20	5	11%	11%
21	15	34%	45%
22	10	23%	68%
23	3	7%	75%
24	4	9%	84%
25	1	2%	86%
26	1	2%	88%
27	2	4%	92%
30	2	4%	96%
33	1	2%	98%
37	1	2%	100%
<b>Totale</b>	<b>45</b>	<b>100%</b>	<b>100%</b>



Per quanto riguarda l'Università di Bologna, il campione intervistato ha un'età compresa tra i 19 e i 31 anni e quasi l'85% di questo è composto da ventenni e ventunenni. L'età media è di quasi 21 anni e solo il 4% degli intervistati, ossia il ragazzo di 29 e 31 anni, risultano fuori corso (Tabella 5.7).

Tabella 5.7 Distribuzione per età del campione relativo la Facoltà di Scienze Statistiche di Bologna

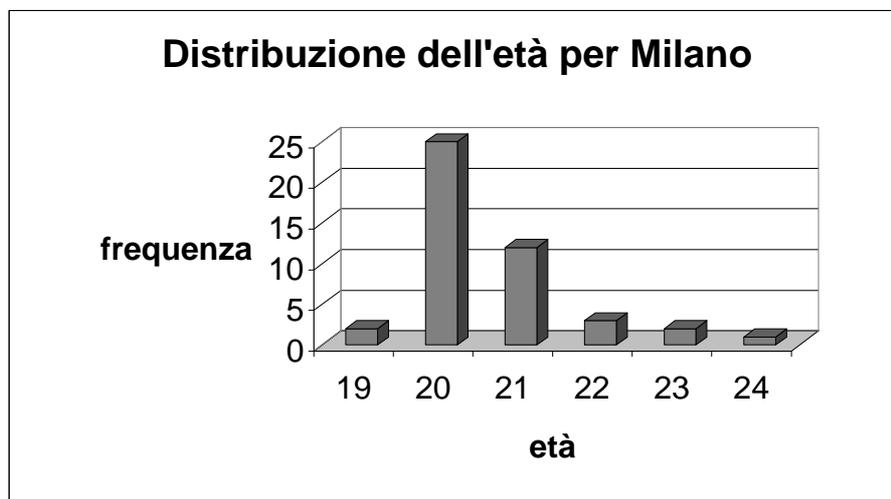
Età	n° studenti	Studenti in %	% cumulata
19	4	9%	9%
20	21	47%	56%
21	17	38%	94%
22	1	2%	96%
29	1	2%	98%
31	1	2%	100%
<b>Totale</b>	45	100%	100%



Per quanto riguarda l'Università di Milano, il campione intervistato ha un'età compresa tra i 19 e i 24 anni e oltre l'82% di questo è composto da ventenni e ventunenni. L'età media è di 20 anni e mezzo e solo il 9% degli intervistati risulta fuori corso, anche se ben il 37% degli iscritti non è regolare, questo probabilmente perché decidono di sostenere gli esami da non frequentanti (Tabella 5.8).

Tabella 5.8 Distribuzione per età del campione relativo la Facoltà di Scienze Statistiche di Milano

Età	n° studenti	Studenti in %	% cumulata
19	2	4%	4%
20	25	56%	60%
21	12	27%	87%
22	3	7%	94%
23	2	4%	98%
24	1	2%	100%
<b>Totale</b>	<b>45</b>	<b>100%</b>	<b>100%</b>

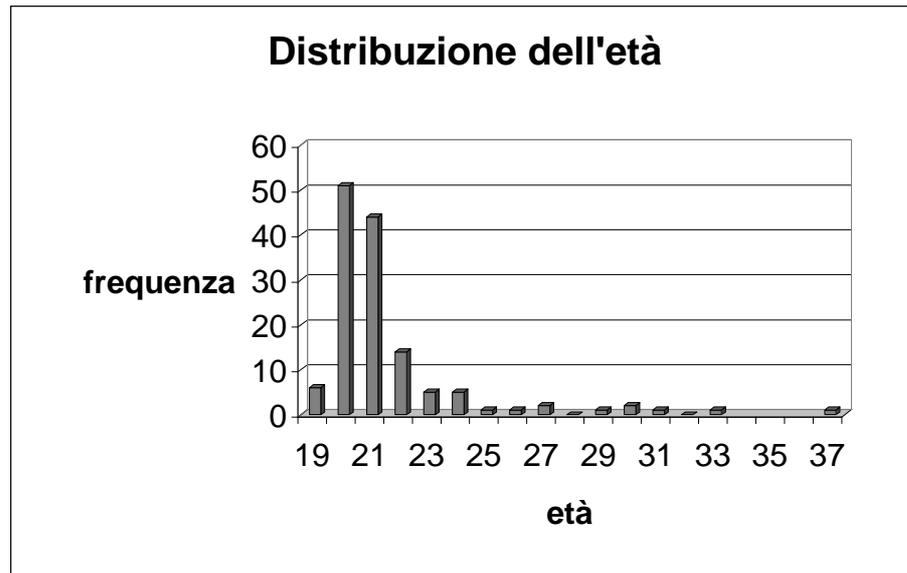


L'intero campione intervistato ha un'età compresa tra i 19 e i 37 anni e oltre l'80% di questo è composto da ventenni, ventunenni e ventiduenni. L'età media è di 21 anni e mezzo (Tabella 5.9).

Gli studenti con un'età superiore ai 25 anni frequentano quasi tutti l'Università di Padova e questo sia perché la Facoltà patavina registra un maggior numero di fuori corso rispetto all'Università di Bologna (38% contro 15%), ma pari a quello di Milano (37%), sia perché in tale città gli studenti non regolari continuano a frequentare i corsi (il 29% degli intervistati risulta fuori corso), mentre a Bologna (4%) e Milano (9%) la maggior parte di questi preferisce sostenere gli esami da non frequentante.

*Tabella 5.9 Distribuzione per età dell'intero campione*

<b>Età</b>	<b>n°studenti</b>	<b>Studenti in %</b>	<b>% cumulata</b>
19	6	4%	4%
20	51	38%	42%
21	44	33%	75%
22	14	10%	85%
23	5	4%	89%
24	5	4%	93%
25	1	0,7%	93,7%
26	1	0,7%	94,4%
27	2	1,4%	95,8%
29	1	0,7%	96,5%
30	2	1,4%	97,9%
31	1	0,7%	98,6%
33	1	0,7%	99,3%
37	1	0,7%	100%
<b>Totale</b>	135	100%	100%



*Luogo di residenza del campione*

Il campione selezionato è stato suddiviso in base alla distanza Facoltà-Luogo di residenza ed inoltre ad ogni studente è stato domandato se avesse preso in affitto un appartamento nella città in cui studiava. Tramite questi dati si vuole capire quanto i giovani siano disposti anche a lunghi spostamenti o a vivere lontani dalla famiglia pur di frequentare il corso di laurea prescelto.

Relativamente alle tre Università analizzate, il campione si suddivide nel seguente modo (Tabelle 5.10, 5.11 e 5.12):

*Tabella 5.10 Suddivisione del campione in base alla distanza Facoltà-Residenza e all'aver preso o no un appartamento in affitto nella città universitaria di Padova.*

<i>Università di Padova</i>	<b>n° studenti</b>	<b>n° studenti pendolari</b>	<b>n° studenti non pendolari</b>
<b>Meno di 10 Km</b>	13	12	1
<b>Meno di 20 Km</b>	5	5	0
<b>Meno di 30 Km</b>	5	5	0
<b>Meno di 40 Km</b>	3	1	2
<b>Meno di 50 Km</b>	3	0	3
<b>Più di 50 Km</b>	16	3	13
<b>Totale</b>	45	26	19

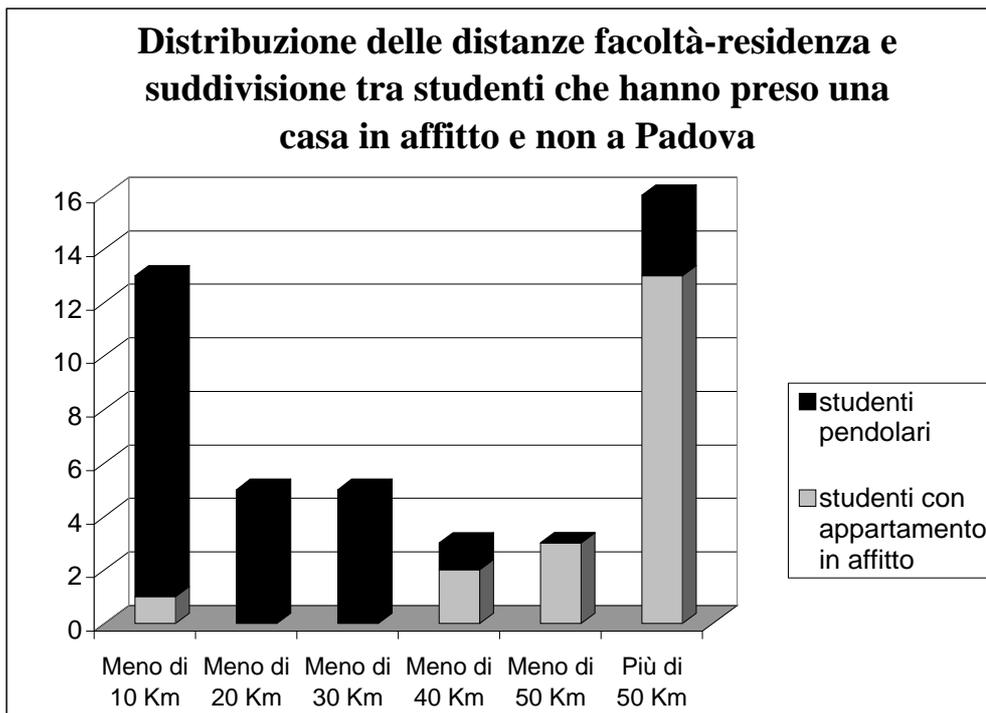
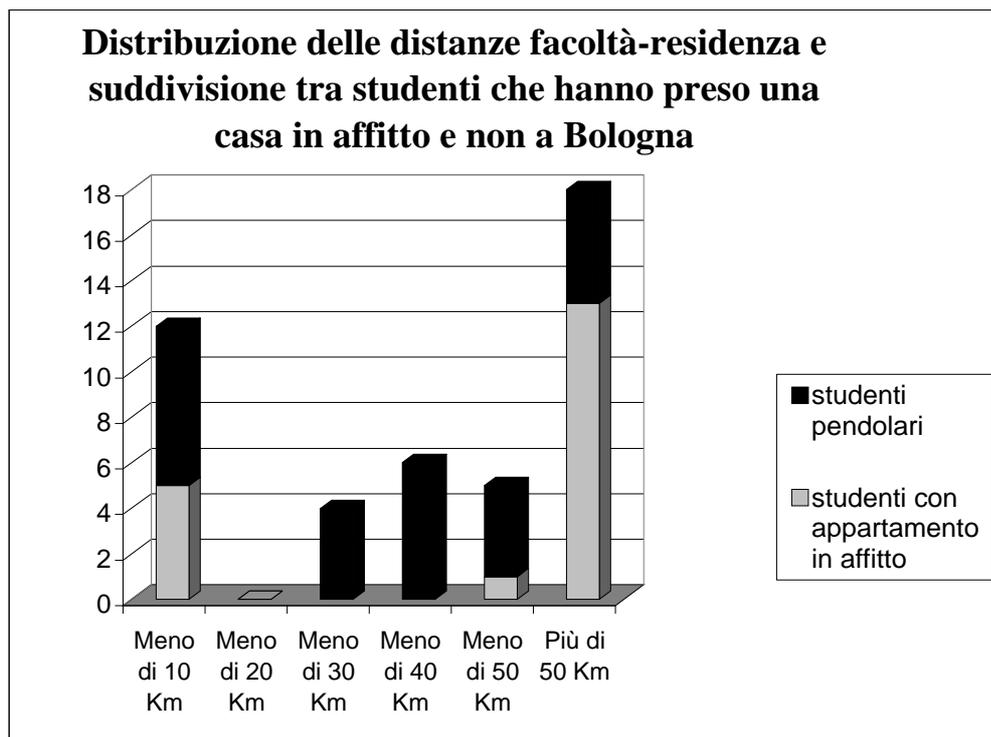


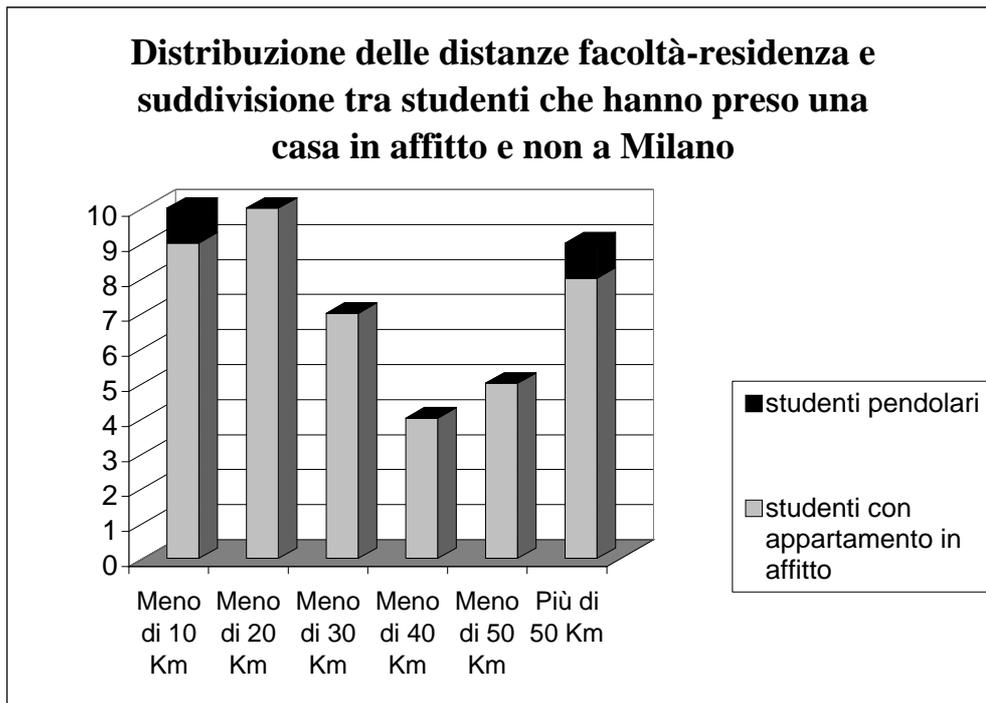
Tabella 5.11 *Suddivisione del campione in base alla distanza Facoltà-Residenza e all'aver preso o no un appartamento in affitto nella città universitaria di Bologna*

<i>Università di Bologna</i>	n° studenti	n° studenti pendolari	n° studenti non pendolari
<b>Meno di 10 Km</b>	12	7	5
<b>Meno di 20 Km</b>	0	0	0
<b>Meno di 30 Km</b>	4	4	0
<b>Meno di 40 Km</b>	6	6	0
<b>Meno di 50 Km</b>	5	4	1
<b>Più di 50 Km</b>	18	5	13
<b>Totale</b>	45	26	19



*Tabella 5.12 Suddivisione del campione in base alla distanza Facoltà-Residenza e all'aver preso o no un appartamento in affitto nella città universitaria di Milano*

<i>Università di Milano</i>	<b>n° studenti</b>	<b>n° studenti pendolari</b>	<b>n° studenti non pendolari</b>
<b>Meno di 10 Km</b>	10	9	1
<b>Meno di 20 Km</b>	10	10	0
<b>Meno di 30 Km</b>	7	7	0
<b>Meno di 40 Km</b>	4	4	0
<b>Meno di 50 Km</b>	5	5	0
<b>Più di 50 Km</b>	9	8	1
<b>Totale</b>	45	43	2



Dai tre grafici sopra riportati si può notare come l'Università di Padova e di Bologna siano tra loro molto simili: entrambe contano un elevato numero di studenti residenti nella provincia universitaria e in altre città che distano più di cinquanta chilometri dalla Facoltà. Oltre il 42% degli intervistati ha deciso di prendere un appartamento in affitto vicino alla sede universitaria in modo da rendere più agevoli gli studi.

La classe maggiormente popolata è quella relativa a una distanza superiore ai cinquanta chilometri, questo a significare che le due storiche Università attirano ancora molti studenti residenti in città lontane. E' da notare inoltre che, nel caso di Bologna, cinque studenti hanno dichiarato di aver preso un appartamento in affitto e di avere una residenza che dista meno di dieci chilometri dalla Facoltà. Questo dato fa sospettare che loro abbiano considerato come residenza l'appartamento "universitario", per cui la classe relativa la massima distanza conterebbe anche altri elementi.

La situazione di Milano si presenta invece in maniera nettamente diversa dalle due precedenti. Non vi sono sostanziali differenze tra le classi: solo quelle relative una distanza compresa tra i trenta e i cinquanta chilometri sono un po' meno popolate ed in questo caso quasi l'intero campione è composto da pendolari. Anche coloro che distano più di cinquanta chilometri dalla Facoltà non prendono un appartamento in affitto nella cittadina milanese e questo probabilmente è dovuto sia ai costi elevati degli affitti sia alla massiccia presenza di mezzi di trasporto veloci che collegano Milano con le altre città. Questo dato fa dedurre però che l'Università Bicocca accoglia principalmente studenti lombardi, mentre un ragazzo residente in un'altra regione sia più propenso ad iscriversi alla Facoltà patavina o bolognese.

Agli studenti intervistati è stato chiesto se avessero mai partecipato ad un progetto Erasmus, trascorrendo un periodo della loro vita universitaria presso una Facoltà estera. Questo dato serviva a capire se le valutazioni date fossero influenzate dal confronto con il sistema universitario estero.

Solo uno studente iscritto alla Facoltà patavina ha aderito a questo tipo di iniziativa, per cui non possiamo concludere che i dati siano influenzati da un possibile confronto.

## Capitolo 6

# L'ANALISI DEI DATI

### *6.1 La segmentazione degli studenti frequentanti la Facoltà di Scienze Statistiche ed Economiche.*

L'analisi svolta nel seguente capitolo è composta da una prima fase di suddivisione degli studenti appartenenti al nostro campione in classi omogenee al loro interno e tra loro eterogenee in base all'importanza data alle venti variabili considerate; questo utilizzando la tecnica di Cluster Analysis.

Per semplificare la clusterizzazione, si è voluto prima ridurre il numero delle variabili in un numero inferiore di fattori, che fossero combinazione lineare delle precedenti; e solo in seguito effettuare una segmentazione degli individui relativamente le nuove caratteristiche di riferimento.

### *6.2 L'analisi fattoriale*

Tramite l'ausilio del pacchetto statistico SPSS for Windows 12.0 si è potuto eseguire l'analisi fattoriale.

Il procedimento effettuato è di seguito descritto:

Partendo dalla matrice delle correlazioni relative le 20 variabili, sono state definite altrettante componenti principali ed estratte in una quantità tale che spiegasse una certa quota di variabilità. Per migliorare la soluzione ottenuta, i

fattori sono stati ruotati secondo il metodo Varimax che tende a minimizzare il numero di variabili con cui ciascun fattore ha coefficienti di correlazione elevati ed è utile nel caso in cui si voglia ottenere una netta separazione tra i fattori. Questo è apparso il criterio che meglio si addicesse alla nostra situazione.

*Tabella 6.1 Tavola delle comunanze*

**Communalities**

	<b>Initial</b>	<b>Extraction</b>
VAR 1	1	0,724863
VAR 2	1	0,708172
VAR 3	1	0,701688
VAR 4	1	0,673125
VAR 5	1	0,598046
VAR 6	1	0,661124
VAR 7	1	0,498932
VAR 8	1	0,420018
VAR 9	1	0,513409
VAR 10	1	0,668652
VAR 11	1	0,543693
VAR 12	1	0,454051
VAR 13	1	0,704264
VAR 14	1	0,650127
VAR 15	1	0,575266
VAR 16	1	0,561372
VAR 17	1	0,568713
VAR 18	1	0,468826
VAR 19	1	0,670832
VAR 20	1	0,548155

Extraction Method: Principal Component Analysis.

La comunanza  $h_j^2$  è la frazione di varianza della variabile  $x_j$  spiegata dall'insieme dei fattori comuni.

In questo caso i fattori spiegano una quota consistente di variabilità, in particolare per le prime variabili relative la didattica (Tabelle 6.2 e 6.3 e Grafico 6.1).

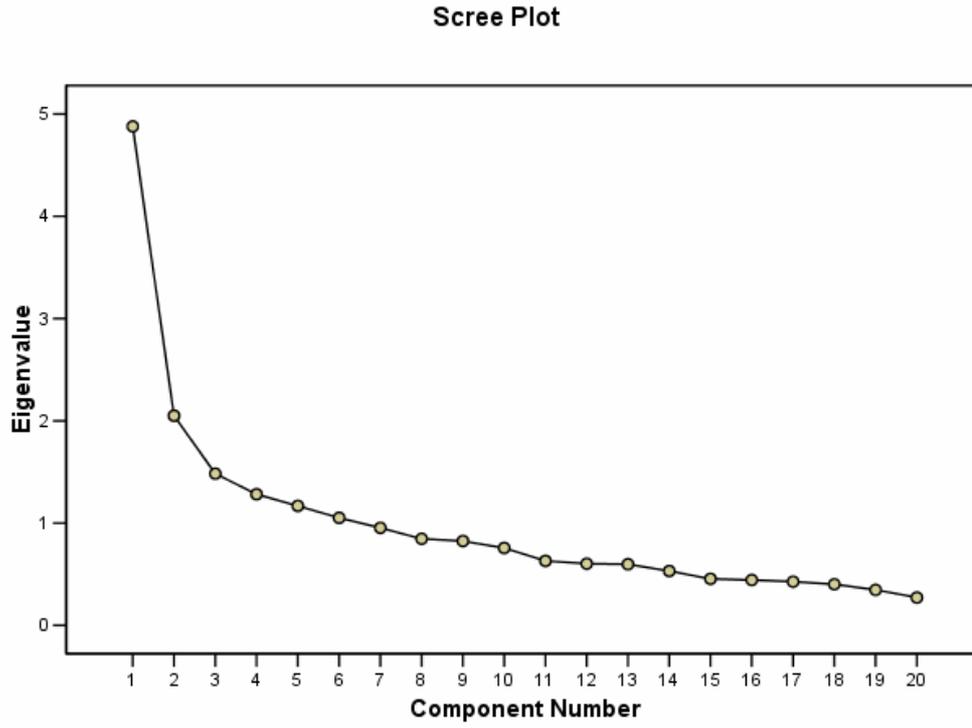
*Tabella 6.2 Quota di varianza spiegata da ogni singolo fattore*

### Total Variance Explained

Component	Initial Eigenvalues		Cumulative %
	Total	% of Variance	
1	4,879063	24,39531	24,39531
2	2,049755	10,24878	34,64409
3	1,483355	7,416777	42,06087
4	1,281879	6,409393	48,47026
5	1,168252	5,841258	54,31152
6	1,051024	5,25512	59,56664
7	0,953316	4,766579	64,33322
8	0,846082	4,230411	68,56363
9	0,824257	4,121286	72,68492
10	0,756379	3,781897	76,46681
11	0,629808	3,149039	79,61585
12	0,602767	3,013836	82,62969
13	0,597693	2,988463	85,61815
14	0,53114	2,655702	88,27385
15	0,454481	2,272406	90,54626
16	0,443231	2,216154	92,76241
17	0,427263	2,136314	94,89873
18	0,401648	2,008239	96,90696
19	0,346843	1,734216	98,64118
20	0,271764	1,358819	100

Extraction Method: Principal Component Analysis

Grafico 6.1 Rappresentazione grafica degli autovalori



Analizzando la tabella sopra riportata si osserva che i primi sei fattori spiegano quasi il 60% della variabilità totale e gli autovalori ad essi relativi sono tutti superiori all'unità: queste considerazioni inducono quindi ad estrarre sei fattori.

Il grafico relativo gli autovalori (Grafico 6.1) non ci mostra una spezzata come nell'esempio riportato al paragrafo 2.6 (Figura 2.4), in quanto in questo caso il sesto e settimo autovalore non differiscono di molto.

Tabella 6.3 Quota di varianza spiegata dai fattori estratti non ruotati

**Extraction Sums of Squared Loadings**

Component	Total	% of Variance	Cumulative %
1	4,879063	24,39531	24,39531
2	2,049755	10,24878	34,64409
3	1,483355	7,416777	42,06087
4	1,281879	6,409393	48,47026
5	1,168252	5,841258	54,31152
6	1,051024	5,25512	59,56664

Tabella 6.4 Quota di varianza spiegata dai fattori estratti ruotati secondo il criterio Varimax

**Rotation Sums of Squared Loadings**

Component	Total	% of Variance	Cumulative %
1	2,406193	12,03097	12,03097
2	2,201314	11,00657	23,03754
3	2,178431	10,89216	33,92969
4	2,036792	10,18396	44,11365
5	1,694925	8,474627	52,58828
6	1,395672	6,978359	59,56664

Tramite la rotazione dei fattori, avvenuta secondo il criterio Varimax, la percentuale di varianza da questi spiegata viene ridistribuita (Tabella 6.4).

In questo caso i primi quattro fattori spiegano una stessa quota di varianza (circa 11%), mentre gli ultimi due una parte inferiore, complessivamente pari al 60%.

Tabella 6.5 Matrice dei pesi fattoriali per fattori non ruotati

**Component Matrix(a)**

	Component					
	1	2	3	4	5	6
VAR 1	0,255	0,233	-0,331	0,477	0,517	0,028
VAR 2	0,442	0,462	-0,386	0,372	-0,028	0,104
VAR 3	0,393	0,169	0,204	-0,082	0,396	-0,560
VAR 4	0,411	0,523	0,371	-0,290	-0,067	0,072
VAR 5	0,558	0,264	0,339	0,231	0,217	0,033
VAR 6	0,435	0,609	-0,057	0,050	-0,293	0,095
VAR 7	0,359	0,379	0,397	-0,198	0,162	-0,058
VAR 8	0,605	-0,024	0,174	0,081	-0,128	0,019
VAR 9	0,577	0,116	-0,373	-0,011	-0,154	-0,067
VAR 10	0,525	0,258	-0,440	-0,358	0,060	-0,022
VAR 11	0,361	-0,143	0,027	0,091	0,318	0,532
VAR 12	0,493	0,110	0,065	-0,323	-0,277	0,114
VAR 13	0,513	-0,368	-0,304	-0,325	0,181	-0,274
VAR 14	0,614	-0,401	-0,123	-0,196	0,038	0,241
VAR 15	0,590	-0,289	-0,162	-0,240	0,194	0,146
VAR 16	0,578	-0,153	-0,032	0,158	-0,421	0,025
VAR 17	0,512	-0,260	0,406	0,272	-0,027	-0,005
VAR 18	0,386	-0,361	0,315	0,009	0,127	0,272
VAR 19	0,482	-0,326	-0,007	0,419	-0,302	-0,255
VAR 20	0,593	-0,272	0,061	0,083	-0,034	-0,334

Extraction Method: Principal Component Analysis

6 components extracted.

Nel caso in cui non venga effettuata una rotazione dei fattori, la matrice dei pesi fattoriali (Tabella 6.5), che indica il valore che una variabile ha sul fattore, mostra che ogni latente non cerca di spiegare maggiormente alcune variabili

trascurando le restanti e di conseguenza l'interpretazione dei fattori risulta molto difficoltosa.

Tabella 6.6 Matrice dei pesi fattoriali per fattori ruotati secondo il criterio Varimax

### Rotated Component Matrix(a)

	Component						
	1	2	3	4	5	6	
VAR 1	0,003	0,082	0,085	0,051	0,113	<b>0,834</b>	Esami area matematico-statistica
VAR 2	0,129	0,028	<b>0,625</b>	0,033	0,002	<b>0,547</b>	Esami area statistica applicata
VAR 3	0,191	0,350	-0,179	<b>0,641</b>	-0,208	0,238	Esami area economica
VAR 4	-0,044	-0,020	0,419	<b>0,676</b>	0,092	-0,174	Argomenti attuali
VAR 5	0,295	-0,055	0,169	<b>0,564</b>	0,274	0,294	Organizzazioni e seminari
VAR 6	0,085	-0,071	<b>0,738</b>	0,302	-0,061	0,097	Pacchetti statistici
VAR 7	-0,006	0,031	0,144	<b>0,685</b>	0,086	-0,028	lingue straniere
VAR 8	0,456	0,131	0,257	0,262	0,245	-0,013	Disponibilità docenti per chiarimenti
VAR 9	0,255	0,413	<b>0,502</b>	0,009	-0,004	0,161	Chiarezza espositiva
VAR 10	-0,113	<b>0,607</b>	<b>0,499</b>	0,154	0,011	0,119	Capacità di suscitare interesse
VAR 11	0,001	0,075	0,063	0,033	<b>0,697</b>	0,217	Prestigio Università
VAR 12	0,142	0,243	0,443	0,239	0,169	-0,304	Collaborazione Università-lavoro/ricerca

VAR 13	0,213	<b>0,806</b>	-0,055	0,017	0,075	0,007	Organizzazione e orario lezioni
VAR 14	0,266	<b>0,517</b>	0,137	-0,047	<b>0,532</b>	-0,090	Disponibilità aule studio
VAR 15	0,152	<b>0,583</b>	0,097	0,058	<b>0,447</b>	0,021	Disponibilità aule computer
VAR 16	<b>0,593</b>	0,140	0,391	-0,055	0,157	-0,097	Piano di studio personalizzato
VAR 17	<b>0,617</b>	-0,028	-0,064	0,251	0,346	0,006	Biblioteca ben fornita
VAR 18	0,285	0,083	-0,130	0,139	<b>0,579</b>	-0,098	Vita universitaria
VAR 19	<b>0,795</b>	0,106	0,085	-0,115	-0,005	0,089	Costo della vita
VAR 20	<b>0,606</b>	0,382	-0,011	0,179	0,043	0,041	Servizi correlati

Extraction Method: Principal Component Analysis. Rotation Method: Varimax with Kaiser Normalization.

a Rotation converged in 18 iterations.

La rotazione dei fattori ha come obiettivo la semplificazione della matrice dei pesi fattoriali, in quanto questi diventano o prossimi all'unità o all'essere nulli, e così risulta più semplice l'interpretazione del significato dei fattori.

Nella tabella 6.6 sopra riportata, sono stati evidenziati in grassetto i pesi fattoriali più vicini all'unità, in modo da delineare il significato di ogni fattore in rapporto alle variabili da questo maggiormente spiegate.

Il *primo fattore* è principalmente composto dalle seguenti variabili:

- Costo della vita nella città universitaria
- Biblioteca ben fornita
- Possibilità e facilità di fruizione di servizi correlati (mensa, librerie..)
- Possibilità di creare un piano di studio personalizzato

E viene definito come *Presenza di Servizi correlati alla didattica*.

Il *secondo fattore* dalle variabili:

- Organizzazione dell'orario delle lezioni
- Capacità del docente di suscitare interesse
- Disponibilità e capienza aule computer
- Disponibilità e capienza aule studio.

E viene definito come *Struttura e logistica e capacità del docente nel suscitare interesse.*

Il *terzo fattore* dalle seguenti variabili:

- Utilizzo di pacchetti statistici
- Esami dell'area statistica applicata
- Chiarezza espositiva dei docenti
- Capacità del docente di suscitare interesse.

E viene definito come *Statistica applicata e Capacità didattiche del docente.*

Il *quarto fattore* dalle variabili:

- Utilizzo lingue straniere
- Approfondimento di argomenti attuali
- Esami dell'area economica
- Organizzazione seminari tenuti da docenti esterni.

E viene definito come *Didattica economica ed attuale.*

Il *quinto fattore* dalle variabili:

- Prestigio dell'Università frequentata
- Organizzazione seminari tenuti da docenti esterni
- Disponibilità e capienza aule studio
- Disponibilità e capienza aule computer.

E viene definito come *Caratteristiche dell'ateneo universitario.*

Il *sesto fattore* dalle variabili:

- Esami dell'area matematico-statistica

- Esami dell'area statistica applicata.

E viene definito come *Insegnamenti di statistica*.

Tramite la matrice dei pesi fattoriali, i dati relativi le 20 variabili analizzate, ottenuti da un campione di 135 individui, vengono trasformati nei sei fattori, passando così da una matrice di dimensioni 135 x 20 ad una di dimensioni 135 x 6.

La nuova matrice dei dati viene ora utilizzata per suddividere gli studenti intervistati in cluster omogenei al loro interno e tra loro eterogenei.

### 6.3 Cluster analysis

Il pacchetto statistico SPSS for Windows 12.0 effettua la classificazione secondo due possibili metodi: K-Means cluster (Metodo gerarchico scissorio) e Hierarchical cluster (Metodo gerarchico aggregativo) (vedi paragrafi 2.11.1 e 2.11.2). Il primo criterio si adatta meglio al nostro caso di studio, anche se sorge il problema di dover definire a priori, in quanto ci viene richiesto come dato di input, il numero dei cluster. Per stabilire il numero ottimale di segmenti nei quali suddividere il campione di individui selezionato, sono stati effettuati alcuni tentativi e quindi è stata scelta la soluzione che meglio si adattava al nostro caso, secondo la mia personale opinione.

Inizialmente, il campione è stato suddiviso in sei gruppi e i risultati ottenuti sono stati i seguenti (Tabella 6.7 e 6.8):

Tabella 6.7 Posizionamento dei centri dei sei cluster

**Final Cluster Centers**

	Cluster					
	1	2	3	4	5	6
VAR 1	15,89	19,98756	13,086	22,138154	15,53857	18,27647
VAR 2	13,085	19,15378	14,504	20,358173	17,04643	17,24706
VAR 3	15,72	18,21178	13,134	19,484481	18,17429	15,12765
VAR 4	14,695	15,59933	11,33	17,984442	15,32143	13,28941
VAR 5	7,91	13,77889	10,548	15,821	11,23214	12,58765
VAR 6	8,645	7,769333	6,22	8,1339615	7,637857	6,638235

Tabella 6.8 Numero di individui presenti in ogni cluster

<b>Number of Cases in each Cluster</b>		
<b>Cluster</b>	1	2
	2	45
	3	5
	4	52
	5	14
	6	17
<b>Valid</b>		135
<b>Missing</b>		0

Il primo e terzo cluster risultano poco significativi in quanto comprendono un numero limitato di individui. Si prova quindi a vedere come cambiano i risultati considerando un segmento in meno (Tabelle 6.9 e 6.10).

*Tabella 6.9 Posizionamento dei centri dei cinque cluster*

**Final Cluster Centers**

	Cluster				
	1	2	3	4	5
VAR 1	14,48	18,52727	20,174583	21,826567	15,08294
VAR 2	10,88	18,02303	19,314583	20,26575	15,64412
VAR 3	15,75	18,4397	16,52125	19,497717	14,75353
VAR 4	16,48	15,75424	14,074167	17,82285	12,94706
VAR 5	8,22	12,28242	14,272917	15,6707	10,95353
VAR 6	8,95	7,528788	7,5341667	8,1089333	6,823529

*Tabella 6.10 Numero di individui presenti in ogni cluster*

<b>Number of Cases in each Cluster</b>		
<b>Cluster</b>	1	1
	2	33
	3	24
	4	60
	5	17
<b>Valid</b>		135
<b>Missing</b>		0

Anche in questo caso un cluster risulta poco significativo in quanto comprende un solo individuo. Si effettua allora nuovamente l'analisi considerando solo quattro segmenti (Tabelle 6.11, 6.12, 6.13 e 6.14).

Tabella 6.11 Posizionamento iniziale dei centri dei quattro cluster

**Initial Cluster Centers**

	Cluster			
	1	2	3	4
VAR 1	17,96	14,48	24,3	11,87
VAR 2	19,26	10,88	22,07	17,25
VAR 3	16	15,75	20,9	15,92
VAR 4	16,91	16,48	19,81	10,94
VAR 5	14,73	8,22	17,68	10,31
VAR 6	5,76	8,95	9,04	6,71

Tabella 6.12 Cambiamento di posizione dei centri dei quattro cluster

**Iteration History(a)**

Iteration	Change in Cluster Centers			
	1	2	3	4
1	3,456187	3,921084	3,462486	3,518379
2	0,497583	2,163433	0,756272	0,856824
3	0,370554	1,262135	0,244293	0,609844
4	0,1878	0,691705	0,149818	0,576256
5	0,071413	0	0,054144	0
6	0	0	0	0

Convergence achieved due to no or small change in cluster centers. The maximum absolute coordinate change for any center is ,000. The current iteration is 6. The minimum distance between initial centers is 9,354.

*Tabella 6.13 Posizionamento finale dei centri dei quattro cluster*

**Final Cluster Centers**

	Cluster			
	1	2	3	4
VAR 1	19,630	15,307	21,857	15,474
VAR 2	18,745	16,289	20,312	15,538
VAR 3	17,361	18,401	19,488	14,460
VAR 4	14,919	15,852	17,766	12,622
VAR 5	13,448	10,902	15,637	10,880
VAR 6	7,495	7,7590	8,120	6,6866

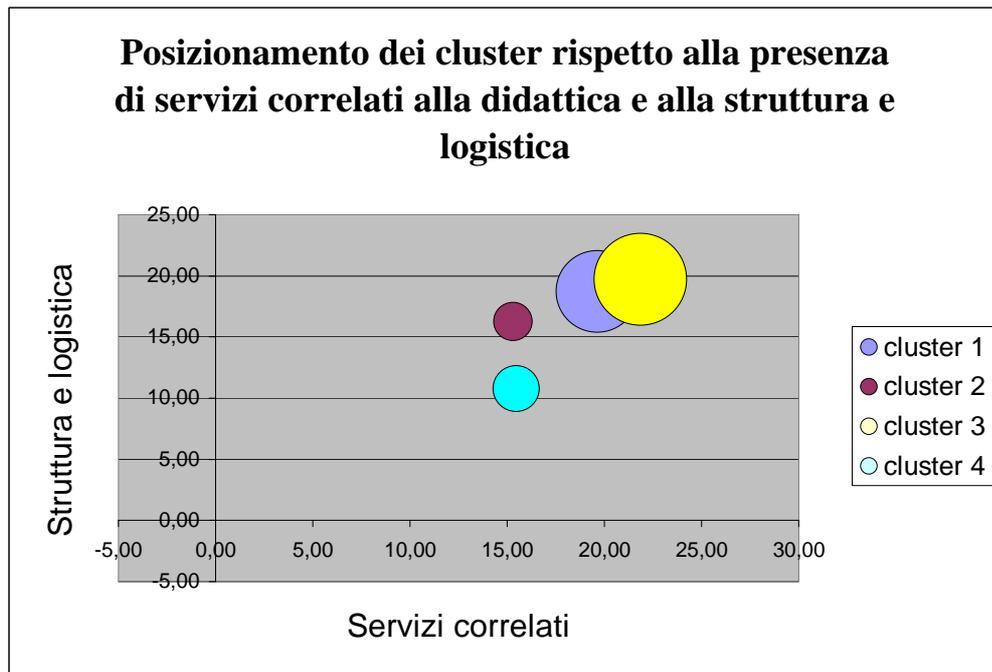
*Tabella 6.14 Numero di individui presenti in ogni cluster*

<b>Number of Cases in each Cluster</b>		
<b>Cluster</b>	1	48
	2	11
	3	61
	4	15
<b>Valid</b>		135
<b>Missing</b>		0

In questo caso tutti i segmenti risultano significativi.

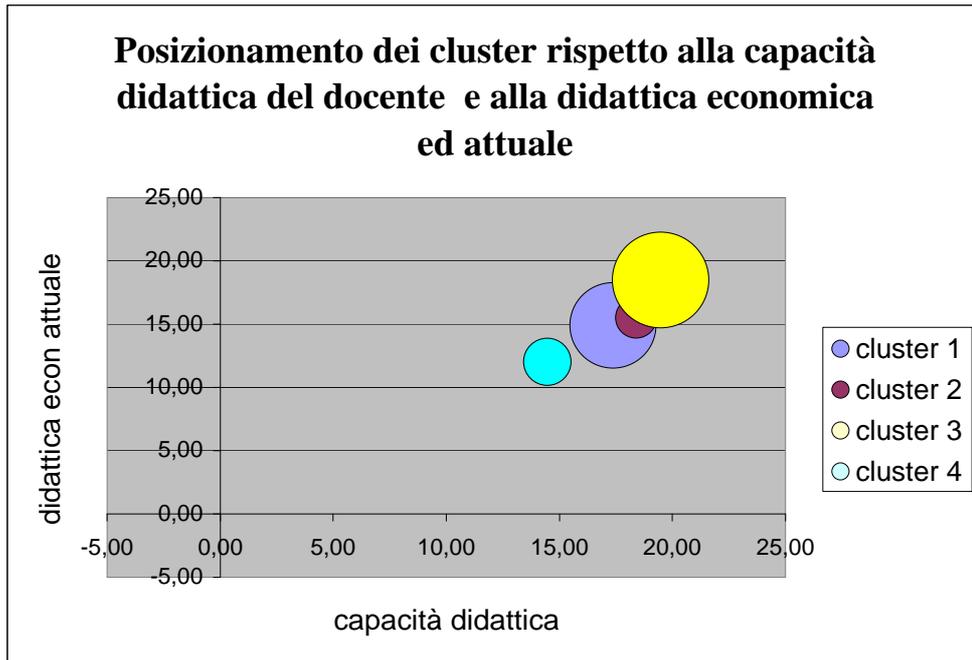
Analizziamo ora come i quattro cluster si posizionano rispetto ai sei fattori trovati con l'analisi fattoriale (Grafici 6.2, 6.3 e 6.4).

Grafico 6.2 Posizionamento dei cluster rispetto ai primi due fattori



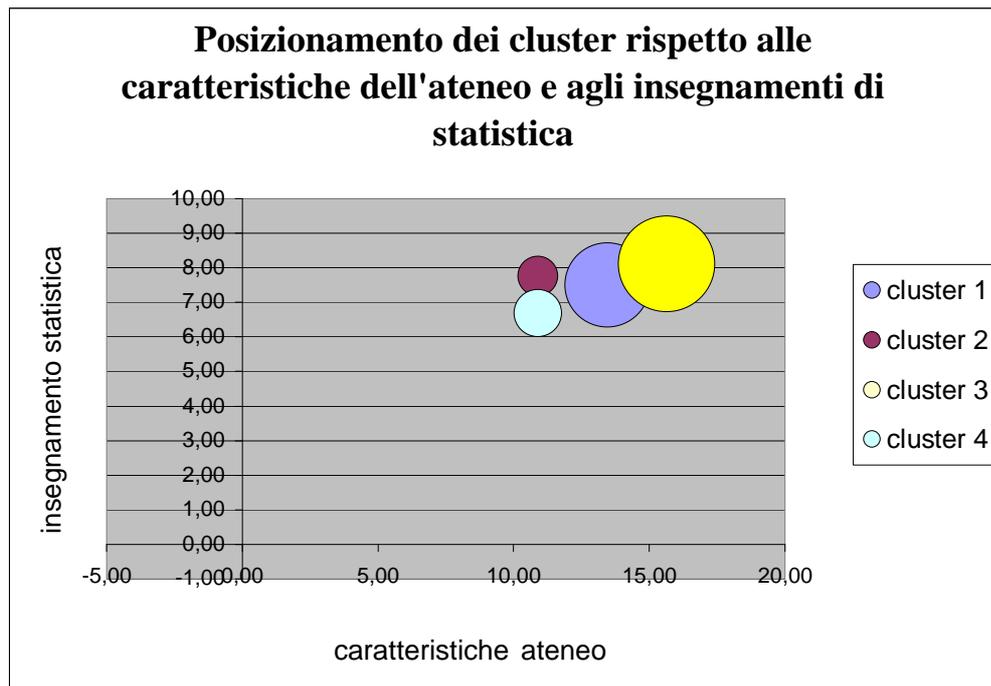
Il primo e terzo cluster, i più numerosi, danno maggiore importanza rispetto agli altri due sia ai servizi correlati alla didattica che alla struttura e logistica, mentre i restanti due attribuiscono il medesimo valore alla presenza di servizi correlati ma differiscono relativamente la struttura e logistica.

Grafico 6.3 Posizionamento dei cluster rispetto al terzo e quarto fattore



I centri dei quattro cluster sembrano appartenere ad una stessa retta, indicante che gli individui che danno maggiore importanza alle capacità didattiche del corpo docente pretendono anche la trattazione di materie economiche ed argomenti attuali e lo studio di lingue straniere nel corso di laurea frequentato.

Grafico 6.4 Posizionamento dei cluster rispetto al quinto e sesto fattore



I primi tre cluster danno circa la stessa importanza al fattore “insegnamenti di statistica”, mentre si differenziano significativamente per l’importanza data alle varie caratteristiche dell’ateneo. I due cluster di dimensioni inferiori non sono molto interessati alla possibilità di svolgere vita universitaria o al prestigio dell’Università frequentata, mentre gli altri due danno una maggiore importanza a queste variabili, in particolare il segmento più numeroso.

#### *6.4 Un confronto tra l'importanza data ad ogni variabile ed il livello di soddisfazione ottenuta*

Per avere una fotografia della situazione attuale delle tre Facoltà di Scienze Statistiche ed Economiche analizzate, si possono confrontare i valori medi dell'importanza data ad ogni variabile e del livello di soddisfazione ottenuto.

Risulta difficile che gli studenti siano pienamente soddisfatti della proposta universitaria offerta loro, ma si può notare che in alcuni casi i due valori non sono significativamente diversi, mentre si è voluto evidenziare con il colore blu i casi in cui il livello di soddisfazione sia di almeno un punto inferiore all'importanza data.

Nelle tabelle 6.15, 6.16, 6.17, 6.18 vengono riassunte tali medie suddivise per argomento (Contenuti, Corpo docente, Organizzazione e logistica, Rapporto Università e ambiente) e specificate per ogni singola variabile, di cui si può avere maggiori informazioni al paragrafo 5.4.

Con il colore rosso sono state evidenziate le variabili alle quali il campione intervistato ha dato un'importanza media inferiore al valore quattro e si nota che sono sei le variabili sottolineate in tal modo e che ben tre di queste riguardano la didattica. In particolare sono l'approfondimento di argomenti attuali come CRM, e-commerce...(3,76), l'organizzazione di seminari tenuti da docenti esterni, che è la variabile che suscita minor interesse nei giovani ottenendo il valore minimo (3,1), e l'utilizzo di lingue straniere (3,93) che però è al limite dell'essere ritenuto significativo.

Anche il prestigio dell'Università frequentata (3,64) sembra non aver molto influito sulla scelta fatta e così la possibilità offerta dalle goliardiche città di Padova e Bologna di vivere una vita universitaria divertente (3,43); lo stesso avviene per i costi che devono essere affrontati durante gli anni di studio (3,76).

Inoltre, per ogni variabile è stata sottolineata la Facoltà che ha ottenuto il maggiore livello di soddisfazione.

Tabella 6.15 Confronto tra l'importanza data ed il livello di soddisfazione ottenuta per le variabili relative i contenuti

	Importanza	Soddisfazione Padova	Soddisfazione Bologna	Soddisfazione Milano
<b>Esami dell'area matematico-statistica</b>	4,29	3,60	<u>3,96</u>	3,51
<b>Esami dell'area statistica applicata</b>	4,45	<u>3,58</u>	3,49	<b>3,27</b>
<b>Esami dell'area economica</b>	4,13	3,84	3,84	<u>4,04</u>
<b>Approfondimento argomenti attuali</b>	<b>3,76</b>	<u>3,16</u>	2,80	<b>2,33</b>
<b>Organizzazione seminari</b>	<b>3,10</b>	<u>3,11</u>	2,20	2,44
<b>Utilizzo pacchetti statistici</b>	4,18	3,27	<u>3,36</u>	<b>2,84</b>
<b>Utilizzo lingue straniere</b>	<b>3,93</b>	<b>2,00</b>	<b>1,98</b>	<u>2,76</u>

Gli studenti delle tre Facoltà danno un'elevata importanza a tutte le materie affrontate durante i tre anni di corso di laurea, prediligendo in maniera particolare tutti i corsi di Statistica applicata (4,45), dei quali però non sono molto soddisfatti soprattutto i giovani milanesi (3,27), che però risultano pienamente soddisfatti degli esami e del quantitativo di ore di lezione dedicate a materie di tipo economico (4.04).

Suscitano non troppo interesse gli argomenti più attuali, ma forse per il semplice motivo che non sono mai stati approfonditi durante i corsi e di questo si lamentano in particolare i giovani della Bicocca (2,33).

L'utilizzo di pacchetti statistici è da tutti ritenuto fondamentale ma nessuna Facoltà riesce ad essere all'altezza delle aspettative e sono nuovamente i milanesi a lamentarsi di questa inefficienza (2,84). Come detto prima, l'insegnamento di lingue straniere è al limite della non importanza, forse perché un giovane può coltivare tale conoscenza anche da autodidatta, ma sicuramente il livello di soddisfazione in questo caso è molto basso. Ormai la conoscenza di almeno una lingua straniera è di fondamentale importanza e viene richiesto come requisito necessario per quasi tutti i tipi di assunzione.

*Tabella 6.16 Confronto tra l'importanza data ed il livello di soddisfazione ottenuta per le variabili relative il corpo docenti*

	<b>Importanza</b>	<b>Soddisfazione Padova</b>	<b>Soddisfazione Bologna</b>	<b>Soddisfazione Milano</b>
<b>Disponibilità docenti per chiarimenti</b>	4,40	3,49	3,93	<u>4,07</u>
<b>Chiarezza espositiva docenti</b>	4,72	<b>3,36</b>	<u>3,51</u>	<u>3,51</u>
<b>Capacità di suscitare interesse</b>	4,47	<b>3,22</b>	<u>3,29</u>	<b>3,16</b>
<b>Prestigio Università</b>	<b>3,64</b>	3,84	<u>4,16</u>	3,29
<b>Collaborazione Università-lavoro/ricerca</b>	4,59	<b>3,11</b>	<b>3,20</b>	<u>3,64</u>

Le variabili relative i docenti sono quelle alle quali viene data maggiore importanza e che in generale non si rivelano all'altezza delle aspettative.

I docenti sono disponibili per chiarimenti sulle spiegazioni o per spostamenti di orario, in particolare a Milano (4,07) e solo Padova si lamenta un po' di questo tipo di servizio (3,49); mentre all'unanimità tutti gli studenti accusano una scarsa chiarezza espositiva durante le ore di lezione (3,46) e una incapacità di suscitare interesse per gli argomenti trattati (3,22).

Si lamentano anche di una scarsa collaborazione tra Università e mondo del lavoro e della ricerca: l'offerta di stage è limitata e poco orientata all'apprendimento di nuove conoscenze.

Il livello di soddisfazione per il prestigio dell'Università frequentata è elevata, soprattutto nel caso di Bologna (4,16).

*Tabella 6.17 Confronto tra l'importanza data ed il livello di soddisfazione ottenuta per le variabili relative l'organizzazione e la logistica*

	<b>Importanza</b>	<b>Soddisfazione Padova</b>	<b>Soddisfazione Bologna</b>	<b>Soddisfazione Milano</b>
<b>Organizzazione orario lezioni</b>	4,30	3,33	<b>3,11</b>	<u>3,56</u>
<b>Disponibilità aule studio</b>	4,06	3,47	<u>3,76</u>	3,38
<b>Disponibilità aule computer</b>	4,32	3,38	<u>3,73</u>	3,67
<b>Piano di studio personalizzato</b>	4,17	<u>3,80</u>	3,33	3,20
<b>Biblioteca ben fornita</b>	4,04	<u>4,09</u>	3,60	3,40

Per riuscire a studiare nel modo migliore e ridurre al minimo i tempi morti, gli studenti vorrebbero che l'orario delle lezioni fosse organizzato in modo

efficiente, ma né l'Università di Bologna (3,11) né quella di Padova (3,33) riescono a gestire in modo adeguato questo problema.

Le aule studio sono in tutti e tre i casi sufficientemente capienti e disponibili e solo gli studenti patavini lamentano una limitata disponibilità dell'aula computer (3,38), non tanto perché poco capiente ma perché spesso occupata dallo svolgersi di alcune lezioni. Proprio questi ultimi vantano però una ben fornita biblioteca (4,09), invidiata in particolar modo dagli studenti milanesi (3,4).

Una rilevante importanza acquisisce anche la possibilità di creare un piano di studio personalizzato in base alle proprie esigenze ed attitudini, possibilità data ai giovani padovani (3,80) e limitata per i restanti (3,33 e 3,20).

*Tabella 6.18 Confronto tra l'importanza data ed il livello di soddisfazione ottenuta per le variabili relative il rapporto Università e ambiente*

	<b>Importanza</b>	<b>Soddisfazione Padova</b>	<b>Soddisfazione Bologna</b>	<b>Soddisfazione Milano</b>
<b>Vita universitaria</b>	<b>3,43</b>	<u>2,93</u>	<u>2,93</u>	2,84
<b>Costo della vita</b>	<b>3,76</b>	<u>2,58</u>	<b>2,11</b>	<b>2,33</b>
<b>Servizi correlati</b>	4,01	3,24	<b>2,67</b>	<u>3,53</u>

Gli studenti non danno molta importanza alla possibilità di trascorrere una vita universitaria “godereccia” e comunque reputano all'altezza delle loro aspettative l'offerta proposta da discoteche ed altri locali.

All'unanimità tutti si lamentano del costo elevato della vita nelle tre città, in particolare gli studenti frequentanti l'Università di Bologna (2,11), anche se ci si aspettava che fossero i milanesi a lamentarsi maggiormente (2,33) dato che

sono tutti costretti a fare i pendolari e solo pochi possono permettersi di prendere un appartamento in affitto.

La città di Bologna offre pochi servizi correlati, quali mense, copisterie, librerie ben fornite, che possono agevolare i giovani nel loro studio (2,67), mentre Milano riesce ad essere molto più efficiente (3,53).

Da questa prima analisi si evince che la Facoltà di Statistica di Padova è quella che riesce meglio a soddisfare le esigenze degli studenti, in quanto solo in cinque casi non riesce ad essere all'altezza delle loro aspettative. I suoi punti deboli sono relativi il corpo docenti, incapace di suscitare interesse e di esporre in modo chiaro gli argomenti trattati, e il rapporto tra Università e mondo del lavoro e della ricerca. La Facoltà inoltre non offre la possibilità di coltivare le lingue già imparate negli anni precedenti e di impararne di nuove: forse dovrebbe proporre delle lezioni di lingua, magari anche solo una volta la settimana, in modo che i giovani laureati non siano costretti a pagarsi un corso privato o non siano le aziende che li assumono a dover sopperire a questa loro mancanza.

Altro punto debole, presente in tutte e tre le Facoltà, è il costo della vita troppo elevato e in questo caso l'unica soluzione sarebbe quella di offrire agli studenti un maggior quantitativo di borse di studio o agevolazioni per i servizi correlati come la mensa.

La Facoltà di Statistica di Bologna presenta le stesse debolezze di Padova, ma inoltre possiede altri due limiti: è disorganizzata in termini di orario sia per tempi morti sia per il fatto che ha a disposizione più sedi e costringe gli studenti a continui spostamenti ed inoltre scarseggiano i servizi correlati quali mense, copisterie e librerie.

Anche l'Università Bicocca di Milano raccoglie le stesse lamentele degli studenti patavini, alle quali si aggiungono quelle relative la didattica. Le ore dedicate alla statistica applicata sono insufficienti e di conseguenza vengono utilizzati pochi pacchetti statistici; inoltre i corsi non comprendono nel

programma la trattazione di argomenti attuali, che potrebbero risultare interessanti e stimolanti per i giovani.

Se invece si confrontano i vantaggi competitivi delle tre Facoltà, ci si accorge che ognuna valorizza aspetti diversi, ma non si può concludere che una Università sia migliore di un'altra in quanto tali punti di forza sono presenti in ugual quantità nei tre casi (vedi valori sottolineati nelle tabelle precedenti). Padova eccelle relativamente gli insegnamenti proposti, in particolare per gli argomenti attuali trattati e lo svolgersi di seminari, ma non riesce ad eguagliare le capacità didattiche degli insegnanti milanesi e bolognesi, ossia le variabili ritenute più importanti dagli studenti.

L'Università di Bologna inoltre vanta capienti aule studio ed aule computer e un considerevole utilizzo di pacchetti statistici, mentre Milano mette a disposizione degli studenti mense, librerie ed altri servizi correlati.

### *6.5 Il posizionamento della Facoltà di Scienza Statistiche ed Economiche dell'Università di Padova, Bologna e Milano Bicocca in una mappa delle percezioni*

Come già spiegato nel terzo capitolo, i metodi principalmente utilizzati per il posizionamento del prodotto sono tre: l'analisi discriminante, la corrispondence analysis e il multidimensional scaling. Data la struttura del questionario somministrato, la tecnica che maggiormente si addice al nostro caso di studio è l'analisi discriminante, in quanto per effettuare un'analisi del tipo multidimensional scaling dovremmo avere la valutazione di ogni individuo per le tre Facoltà, ma uno studente è in grado di giudicare solo il corso di laurea da lui frequentato e non i restanti. Per quanto riguarda invece

l'analisi delle corrispondenze, essa si basa su variabili dicotomiche, quindi presenza o assenza di una certa caratteristica, e si potrebbero trasformare i dati rilevati stabilendo che un punteggio da 1 a 3 indica l'assenza della caratteristica mentre da 4 a 5 la sua presenza. In questo caso però non utilizzeremmo tutte le informazioni a disposizione.

### *6.6 L'analisi discriminante*

Tramite l'utilizzo del pacchetto statistico R 1.6.1 si è potuto effettuare l'analisi discriminante o più precisamente definire le funzioni discriminanti che meglio separano i quattro gruppi analizzati. Questo tipo di analisi non solo vuole confrontare le Facoltà di Padova, Bologna e Milano ma vuole rapportarle anche con l'Università "ideale" descritta dagli studenti intervistati.

Le variabili considerate sono ventuno: le prime venti quantitative relative al questionario e l'ultima di tipo qualitativo indicante l'appartenenza al gruppo di riferimento. I quattro gruppi sono stati definiti equiprobabili anche se con dimensione campionaria differente, in modo da non dare maggiore importanza al "prodotto ideale" che conta 135 individui e non 45.

Dall'analisi risulta che le funzioni discriminanti sono tre e con i seguenti coefficienti (Tabella 6.19):

Tabella 6.19 I coefficienti delle funzioni discriminanti

**Coefficients of linear discriminants:**

	LD1	LD2	LD3
V1	-0.1106570131	-0.25602988	-0.39437502
V2	-0.2607416901	-0.09447033	0.17554750
V3	0.1671732322	<b>0.41804496</b>	-0.07897174
V4	-0.0413259916	<b>-0.52440822</b>	0.04763629
V5	0.1784550902	-0.08309396	<b>0.58142452</b>
V6	0.0031378619	-0.19977802	-0.35981409
V7	<b>-0.4590928469</b>	0.25989532	0.08749498
V8	-0.0002652012	<b>0.53831667</b>	-0.39483716
V9	<b>-0.6334767587</b>	-0.09497359	-0.33238755
V10	-0.1562027677	-0.28884314	-0.10261679
V11	0.2880036292	-0.16914715	-0.29772446
V12	<b>-0.4948916393</b>	0.29669115	0.01772913
V13	-0.2366169695	0.10682728	<b>0.46682442</b>
V14	-0.0134077871	-0.06392223	-0.13197197
V15	-0.0389006774	0.27273881	-0.36860436
V16	0.0450984678	-0.34474523	0.15920033
V17	0.1097779168	-0.31919408	0.13156309
V18	-0.0097476592	-0.08395189	-0.08618640
V19	-0.2239212143	-0.15548332	0.02471995
V20	0.0990562764	0.21446386	<b>0.42071079</b>

Le variabili che danno un maggiore contributo nella determinazione della prima funzione discriminante sono tre: l'utilizzo di lingue straniere, la chiarezza espositiva dei docenti durante le lezioni e la collaborazione tra Università e mondo del lavoro e della ricerca. Relativamente la seconda funzione, le variabili che meglio riescono a discriminare i quattro gruppi sono la presenza di esami dell'area economica e l'approfondimento di argomenti attuali e la disponibilità da parte dei docenti per chiarimenti, spostamenti d'orario...Nella terza funzione discriminante invece, le variabili che

contribuiscono maggiormente sono l'organizzazione di seminari tenuti da docenti esterni, la possibilità e facilità di fruizione di servizi correlati quali mense, librerie e copisterie e un efficiente organizzazione dell'orario di lezione.

Notiamo che, a differenza di quanto avvenuto nell'analisi fattoriale, in questo caso troviamo difficile assegnare un nome ad ogni funzione discriminante in quanto le variabili considerate sembrano apparentemente poco correlate tra loro. Questo ovviamente renderà difficoltosa l'interpretazione delle mappe di percezione ottenute e il conseguente posizionamento delle varie facoltà.

Osserviamo inoltre che i coefficienti ottenuti non sono tutti concordi, il che implica che il contributo dato da ogni variabile non sempre si somma ai restanti ma a volte si sottrae: questo comunque non altera i risultati ottenuti.

La tabella seguente 6.20 indica il potere discriminante delle tre funzioni: la prima riesce a discriminare più efficacemente i quattro gruppi rispetto alle altre due funzioni.

*Tabella 6.20 Potere discriminante delle funzioni*

**Proportion of trace:**

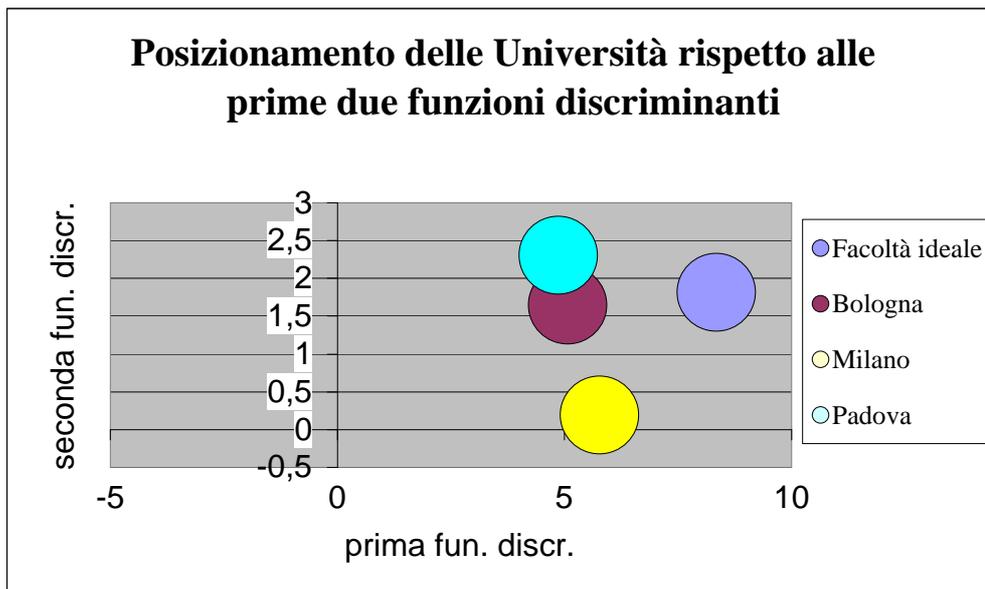
LD1	LD2	LD3
0.6502%	0.2107%	0.1390%

Dopo aver definito i coefficienti delle tre funzioni discriminanti, è stato calcolato il punteggio ottenuto da ogni individuo e quindi fatta la media per ogni gruppo, in modo da poter così posizionare le quattro Facoltà su una mappa avente come assi due funzioni discriminanti.

Tabella 6.21 Punteggio medio per ogni gruppo relativamente le tre funzioni discriminanti

	<b>Prima fun. Discriminante</b>	<b>Seconda fun. discriminante</b>	<b>Terza fun. Discriminante</b>
<b>Ideale</b>	8,34	1,82	2,61
<b>Bologna</b>	5,07	1,65	3,6
<b>Milano</b>	5,77	0,19	2,27
<b>Padova</b>	4,86	2,31	1,88

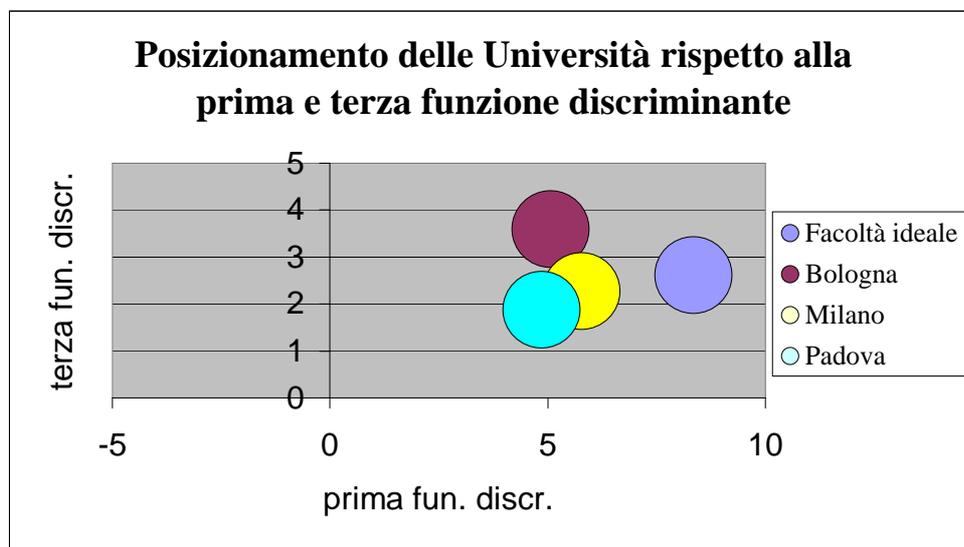
Grafico 6.5 Posizionamento delle quattro Facoltà rispetto alla prima e seconda funzione discriminante



Considerando la prima funzione discriminante, l'Università Bicocca di Milano si avvicina maggiormente a quella ideale, quindi per l'insegnamento di lingue straniere, la chiarezza espositiva dei docenti e la collaborazione tra il mondo imprenditoriale e quello universitario. La Facoltà patavina e bolognese invece

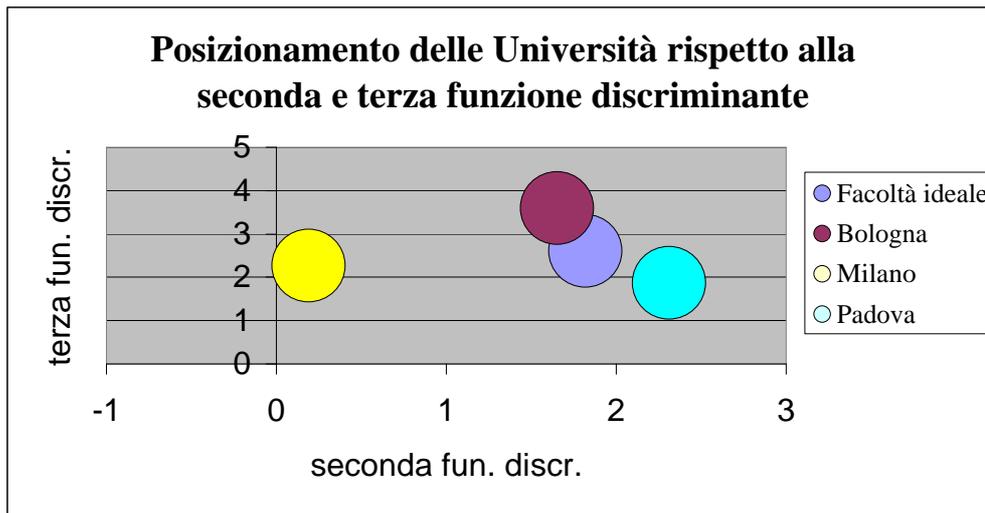
si avvicinano di più al prodotto immaginato dagli studenti relativamente alla seconda funzione discriminante, quindi per la trattazione di argomenti attuali e materie economiche e per la disponibilità del corpo docenti (Grafico 6.5)

*Grafico 6.6 Posizionamento delle quattro Facoltà rispetto alla prima e terza funzione discriminante*



Relativamente la terza funzione discriminante, è ancora la Facoltà milanese quella che si avvicina maggiormente a quella ideale questo grazie all'ottima offerta di servizi correlati e a una buona organizzazione dell'orario delle lezioni (Grafico 6.6).

*Grafico 6.7 Posizionamento delle quattro Facoltà rispetto alla seconda e terza funzione discriminante*



Relativamente alla prima e terza funzione discriminante, l'Università Bicocca di Milano si avvicina maggiormente alla Facoltà ideale, mentre relativamente alla seconda Padova e Bologna risultano le migliori (Grafico 6.7). Ma le variabili che contribuiscono maggiormente nella definizione della prima e terza funzione discriminante sono proprio quelle che hanno ottenuto un maggior livello di soddisfazione da parte degli studenti milanesi. Per cui osservando i grafici 6.5, 6.6 e 6.7 si potrebbe dedurre che l'Università Bicocca sia quella meglio posizionata; se però consideriamo anche quanto detto al paragrafo 6.4, capiamo che una conclusione simile risulterebbe in parte errata. Le variabili che maggiormente discriminano i quattro gruppi sono quelle che hanno ottenuto un maggior livello di soddisfazione negli studenti milanesi: questi ultimi vantano un buon corpo docenti, anche se non sempre all'altezza delle aspettative, e la presenza di servizi correlati, ma dichiarano di non essere pienamente soddisfatti degli argomenti trattati nel loro corso di laurea. Per quanto riguarda invece la Facoltà di Bologna, vanta anche lei un valido corpo docenti e capienti aule studio ed informatiche, mentre l'Università

patavina propone lezioni ricche di contenuti anche se a volte le capacità didattiche degli insegnanti sono limitate.

Bisogna osservare però che, dato il significato delle funzioni discriminanti, complessivamente l'Università milanese si distingue in maniera significativa dalle rimanenti, mentre quella patavina e bolognese sono tra loro più simili.



## CONCLUSIONI

In seguito è stata effettuata una trasformazione dei dati originali relativi la soddisfazione degli studenti per l'Università di Padova, Bologna e Milano nei sei fattori definiti al paragrafo 6.2, in modo tale che i valori medi per gruppo potessero essere confrontati con quelli ottenuti tramite la clusterizzazione al paragrafo 6.3.

I risultati vengono riportati nella Tabella C.1 dove vengono evidenziati con il colore blu le Università che, relativamente a quel fattore, hanno registrato un livello di soddisfazione maggiore. I risultati ottenuti vengono poi rappresentati nei grafici C.1, C.2 e C.3 dove sono raffigurati i quattro segmenti definiti al paragrafo 6.3 e le tre Università analizzate.

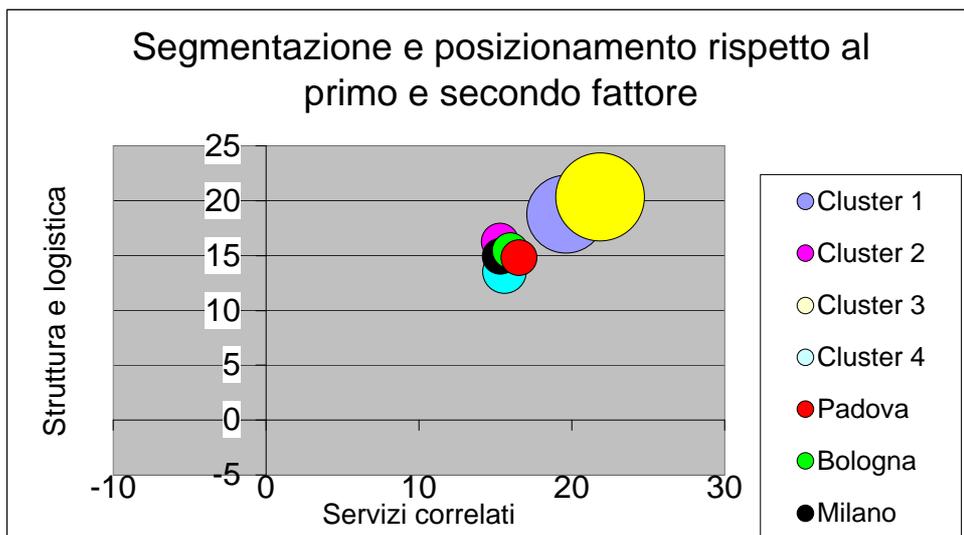
La dimensione dei cluster varia proporzionalmente alla numerosità del segmento, mentre quella delle Facoltà è uguale per tutti e tre gli atenei.

*Tabella C.1 Posizionamento dei cluster e delle Facoltà universitarie rispetto ai sei fattori*

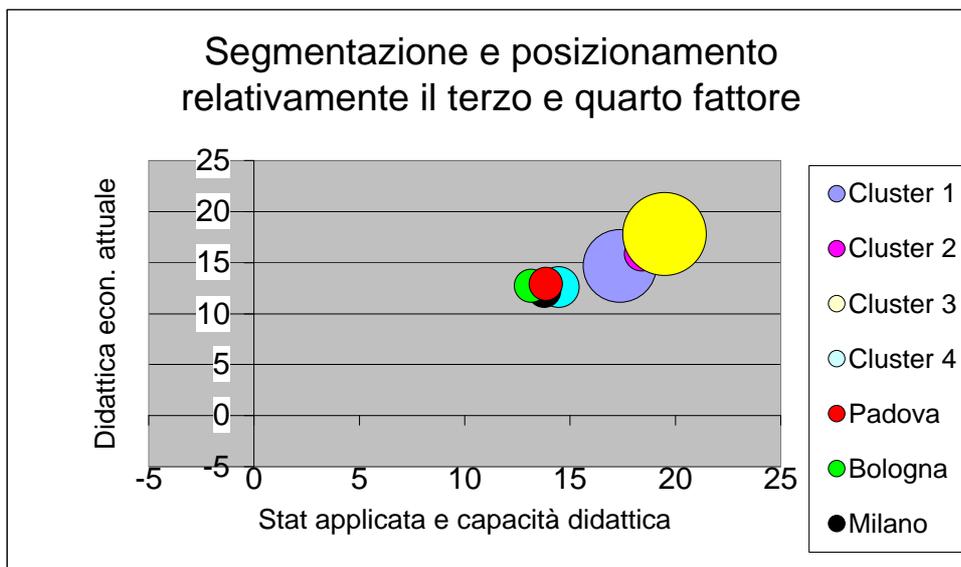
	<b>Cluster 1</b>	<b>Cluster 2</b>	<b>Cluster 3</b>	<b>Cluster 4</b>
<b>Presenza di servizi correlati alla didattica</b>	19,63	15,31	21,86	15,47
<b>Struttura e logistica</b>	18,75	16,29	20,31	15,54
<b>Statistica applicata e capacità didattiche</b>	17,36	18,40	19,49	14,46
<b>Didattica economica ed attuale</b>	14,92	15,85	17,77	12,62
<b>Caratteristiche ateneo</b>	13,45	10,90	15,64	10,88
<b>Insegnamenti di statistica</b>	7,50	7,76	8,12	6,69

	<b>Padova</b>	<b>Bologna</b>	<b>Milano</b>
<b>Presenza di servizi correlati alla didattica</b>	<b>16,55</b>	15,97	15,32
<b>Struttura e logistica</b>	14,80	<b>15,46</b>	14,94
<b>Statistica applicata e capacità didattiche</b>	<b>13,85</b>	13,15	13,77
<b>Didattica economica ed attuale</b>	<b>12,91</b>	12,69	12,20
<b>Caratteristiche ateneo</b>	12,25	11,61	<b>12,38</b>
<b>Insegnamenti di statistica</b>	6,78	6,26	<b>6,85</b>

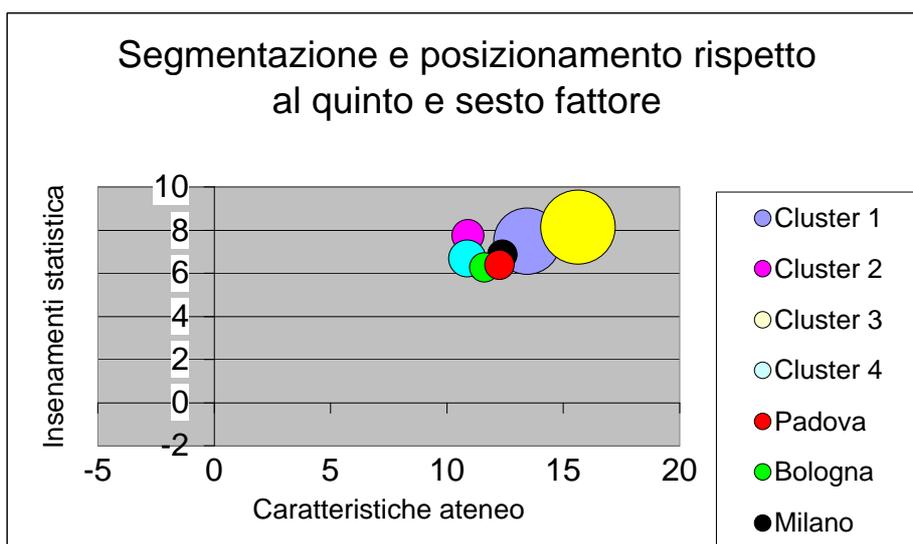
*Grafico C.1 Posizionamento dei cluster e delle Facoltà universitarie rispetto ai primi due fattori*



*Grafico C.2 Posizionamento dei cluster e delle Facoltà universitarie rispetto al terzo e quarto fattore*



*Grafico C.3 Posizionamento dei cluster e delle Facoltà universitarie rispetto al quinto e sesto fattore*



Osservando i grafici C.1, C.2, C.3 e ricordando che i sei fattori definiti al paragrafo 6.2 sintetizzano le venti variabili iniziali, possiamo dedurre che complessivamente non vi sono sostanziali differenze tra le tre Facoltà analizzate, in quanto i cerchi che rappresentano le tre Università risultano quasi concentrici. Inoltre nessuna delle tre Facoltà analizzate riesce ad avere lo stesso posizionamento dei cluster più numerosi: il primo e il terzo, ma solo dei segmenti di dimensioni inferiori. Questo sta ad indicare che le Facoltà di Statistica di Padova, Bologna e Milano non riescono a soddisfare le esigenze della maggior parte degli studenti iscritti ai rispettivi corsi di laurea.

Se comunque analizziamo la Tabella C.1 e quelle riportate al paragrafo 6.4, ovviamente osserviamo che le tre Facoltà offrono proposte per certi aspetti diverse. Ma prima di vagliare tali offerte, cerchiamo di definire sinteticamente quale sia la domanda.

Gli studenti intervistati richiedono principalmente un corpo docenti preparato, in grado di suscitare interesse e di rendere anche gli argomenti più ostici di facile comprensione. Inoltre sono interessati alla trattazione di argomenti di statistica applicata e vorrebbero che i corsi proposti comprendessero una parte pratica di utilizzo di pacchetti statistici. Ma i giovani danno molta importanza anche al loro percorso post laurea ed è per questo che chiedono vi sia una collaborazione tra Università e mondo del lavoro e della ricerca, che permetta loro di migliorare la propria formazione. Sono invece poco interessati alla trattazione di argomenti attuali o all'organizzazione di seminari tenuti da docenti esterni e al prestigio dell'Università frequentata.

Se ora osserviamo l'offerta attualmente proposta dalle tre Facoltà notiamo che l'Università patavina presenta come punti di forza corsi con contenuti approfonditi ad attuali, sia per quanto riguarda argomenti dell'area statistica applicata che di quella economica, ed inoltre offre la possibilità di partecipare a seminari e convegni. Risultano essere di ottima qualità anche i servizi correlati, quali una biblioteca ben fornita e la possibilità di creare un piano di studi personalizzato. Gli studenti frequentanti la Facoltà patavina lamentano solo una limitata chiarezza espositiva dei docenti e una certa incapacità nel

suscitare interesse, ma questo probabilmente è da attribuirsi anche alla complessità degli argomenti trattati e al fatto che materie dell'area statistico-matematica, proprio per la loro difficile comprensione, possono risultare un po' noiose.

L'Università di Bologna invece presenta come vantaggi competitivi un'attrezzata struttura universitaria, con aule studio e aule computer capienti, e un corpo docenti che eccelle, rispetto alle altre due Facoltà, per chiarezza espositiva e capacità di suscitare interesse, ma non all'altezza delle aspettative degli studenti.

L'Università Bicocca in fine si distingue per l'offerta di alcuni servizi, come un efficiente ufficio stage che coordina i rapporti tra Università e mondo imprenditoriale. Anche il corpo docenti riesce a spiccare, non solo per le proprie capacità ma anche per la disponibilità dimostrata agli studenti. Milano però risulta carente per quanto riguarda l'offerta di corsi che trattino argomenti attuali o insegnino l'utilizzo di pacchetti statistici.

Ma allora come dovrebbero agire le tre facoltà di Statistica, in particolare quella patavina? Quali leve dovrebbe usare affinché il proprio prodotto soddisfi efficacemente le esigenze degli studenti? Le politiche di marketing utilizzate fino ad ora devono essere modificate?

Personalmente, penso che in questi ultimi anni la Facoltà patavina abbia investito molto nella realizzazione di seminari tenuti da docenti di altri atenei o da imprenditori, ma sembra che questa scelta non sia strategicamente vincente in quanto l'interesse dimostrato dagli studenti per questi argomenti è limitato e di conseguenza tali investimenti non hanno un ritorno adeguato in termini di soddisfazione.

Dovrebbe invece incrementare il numero di ore dedicate all'utilizzo di pacchetti statistici, in particolare i software che vengono utilizzati dalle aziende, e dovrebbe cercare di migliorare il rapporto Università - mondo imprenditoriale e della ricerca, magari offrendo un maggior numero di stage formativi. Inoltre dovrebbe riuscire, attraverso un'efficace politica di marketing, a comunicare ai potenziali neo iscritti la validità dei corsi tenuti, sia

## Conclusioni

dell'area statistica che in quella economica, che risultano essere vari ed approfonditi.

# APPENDICE

## Scheda di valutazione della Facoltà di Scienze Statistiche ed Economiche Università di Padova – Laurea di primo livello

Il candidato è pregato di dare una valutazione da 1 a 5 ad ogni singola voce sotto riportata. Nel primo caso si vuole misurare il **livello di importanza** data ad ogni singola voce e il valore **1** indica che si dà all'elemento un' **importanza minima** mentre il valore **5** un' **importanza massima**. Nel secondo caso invece si vuole misurare il **livello di soddisfazione** e il valore **1** indica una **soddisfazione minima** mentre il valore **5** una **soddisfazione massima**.

### Contenuti:

- 1 Esami dell'area matematico-statistica
- 2 Esami dell'area statistica applicata
- 3 Esami dell'area economica
- 4 Approfondimento di argomenti "attuali"(es.e-commerce)
- 5 Organizzazione seminari tenuti da docenti esterni
- 6 Utilizzo di pacchetti statistici
- 7 Utilizzo lingue straniere

### Importanza      Soddisfazione

1	2	3	4	5	1	2	3	4	5
1	2	3	4	5	1	2	3	4	5
1	2	3	4	5	1	2	3	4	5
1	2	3	4	5	1	2	3	4	5
1	2	3	4	5	1	2	3	4	5
1	2	3	4	5	1	2	3	4	5
1	2	3	4	5	1	2	3	4	5
1	2	3	4	5	1	2	3	4	5

### Corpo docente:

- 8 Disponibilità docenti per chiarimenti, spostamento di orario..
- 9 Chiarezza espositiva dei docenti durante le lezioni
- 10 Capacità del docente di suscitare interesse
- 11 Prestigio dell'Università frequentata
- 12 Collaborazione tra Università e mondo del lavoro/ricerca

1	2	3	4	5	1	2	3	4	5
1	2	3	4	5	1	2	3	4	5
1	2	3	4	5	1	2	3	4	5
1	2	3	4	5	1	2	3	4	5
1	2	3	4	5	1	2	3	4	5
1	2	3	4	5	1	2	3	4	5

### Organizzazione e logistica:

- 13 Organizzazione dell'orario delle lezioni
- 14 Disponibilità e capienza aule studio
- 15 Disponibilità e capienza aule computer
- 16 Possibilità di creare un piano di studio personalizzato
- 17 Biblioteca ben fornita

1	2	3	4	5	1	2	3	4	5
1	2	3	4	5	1	2	3	4	5
1	2	3	4	5	1	2	3	4	5
1	2	3	4	5	1	2	3	4	5
1	2	3	4	5	1	2	3	4	5
1	2	3	4	5	1	2	3	4	5

### Rapporto università e ambiente:

- 18 Possibilità di svolgere vita universitaria  
(organizzazione di feste e ritrovi per universitari)
- 19 Costo della vita a Padova
- 20 Possibilità e facilità di fruizione di servizi correlati(mensa,librerie...)

1	2	3	4	5	1	2	3	4	5
1	2	3	4	5	1	2	3	4	5
1	2	3	4	5	1	2	3	4	5
1	2	3	4	5	1	2	3	4	5

### Dati anagrafici dell'intervistato:

Sesso  M  F      Età..... Sei fuori corso  si  no      Hai partecipato ad un Erasmus  si  no

La tua residenza dista dalla facoltà

Meno di 10 Km	Meno di 20 Km	Meno di 30 Km	Meno di 40 Km	Meno di 50 Km	Più di 50 Km
---------------	---------------	---------------	---------------	---------------	--------------

Hai preso in affitto un appartamento nella città universitaria?  si  no

## Scheda di valutazione della Facoltà di Scienze Statistiche ed Economiche Università di Bologna – Laurea di primo livello

Il candidato è pregato di dare una valutazione da 1 a 5 ad ogni singola voce sotto riportata. Nel primo caso si vuole misurare il **livello di importanza** data ad ogni singola voce e il valore **1** indica che si dà all'elemento un' **importanza minima** mentre il valore **5** un' **importanza massima**. Nel secondo caso invece si vuole misurare il **livello di soddisfazione** e il valore **1** indica una **soddisfazione minima** mentre il valore **5** una **soddisfazione massima**.

### Contenuti:

- 7 Esami dell'area matematico-statistica
- 8 Esami dell'area statistica applicata
- 9 Esami dell'area economica
- 10 Approfondimento di argomenti "attuali"(es.e-commerce)
- 11 Organizzazione seminari tenuti da docenti esterni
- 12 Utilizzo di pacchetti statistici
- 7 Utilizzo lingue straniere

### Importanza      Soddisfazione

1	2	3	4	5	1	2	3	4	5
1	2	3	4	5	1	2	3	4	5
1	2	3	4	5	1	2	3	4	5
1	2	3	4	5	1	2	3	4	5
1	2	3	4	5	1	2	3	4	5
1	2	3	4	5	1	2	3	4	5
1	2	3	4	5	1	2	3	4	5

### Corpo docente:

- 12 Disponibilità docenti per chiarimenti, spostamento di orario..
- 13 Chiarezza espositiva dei docenti durante le lezioni
- 14 Capacità del docente di suscitare interesse
- 15 Prestigio dell'Università frequentata
- 12 Collaborazione tra Università e mondo del lavoro/ricerca

1	2	3	4	5	1	2	3	4	5
1	2	3	4	5	1	2	3	4	5
1	2	3	4	5	1	2	3	4	5
1	2	3	4	5	1	2	3	4	5
1	2	3	4	5	1	2	3	4	5
1	2	3	4	5	1	2	3	4	5

### Organizzazione e logistica:

- 20 Organizzazione dell'orario delle lezioni
- 21 Disponibilità e capienza aule studio
- 22 Disponibilità e capienza aule computer
- 23 Possibilità di creare un piano di studio personalizzato
- 24 Biblioteca ben fornita

1	2	3	4	5	1	2	3	4	5
1	2	3	4	5	1	2	3	4	5
1	2	3	4	5	1	2	3	4	5
1	2	3	4	5	1	2	3	4	5
1	2	3	4	5	1	2	3	4	5
1	2	3	4	5	1	2	3	4	5

### Rapporto università e ambiente:

- 25 Possibilità di svolgere vita universitaria  
(organizzazione di feste e ritrovi per universitari)
- 26 Costo della vita a Bologna
- 20 Possibilità e facilità di fruizione di servizi correlati(mensa,librerie...)

1	2	3	4	5	1	2	3	4	5
1	2	3	4	5	1	2	3	4	5
1	2	3	4	5	1	2	3	4	5
1	2	3	4	5	1	2	3	4	5

### Dati anagrafici dell'intervistato:

Sesso  M  F      Età..... Sei fuori corso  si  no      Hai partecipato ad un Erasmus  si  no

La tua residenza dista dalla facoltà

Meno di 10 Km	Meno di 20 Km	Meno di 30 Km	Meno di 40 Km	Meno di 50 Km	Più di 50 Km
---------------	---------------	---------------	---------------	---------------	--------------

Hai preso in affitto un appartamento nella città universitaria?  si  no

## Scheda di valutazione della Facoltà di Scienze Statistiche ed Economiche Università Bicocca di Milano – Laurea di primo livello

Il candidato è pregato di dare una valutazione da 1 a 5 ad ogni singola voce sotto riportata. Nel primo caso si vuole misurare il **livello di importanza** data ad ogni singola voce e il valore **1** indica che si dà all'elemento un' **importanza minima** mentre il valore **5** un' **importanza massima**. Nel secondo caso invece si vuole misurare il **livello di soddisfazione** e il valore **1** indica una **soddisfazione minima** mentre il valore **5** una **soddisfazione massima**.

### Contenuti:

- 13 Esami dell'area matematico-statistica
- 14 Esami dell'area statistica applicata
- 15 Esami dell'area economica
- 16 Approfondimento di argomenti "attuali"(es.e-commerce)
- 17 Organizzazione seminari tenuti da docenti esterni
- 18 Utilizzo di pacchetti statistici
- 7 Utilizzo lingue straniere

#### Importanza      Soddisfazione

1	2	3	4	5	1	2	3	4	5
1	2	3	4	5	1	2	3	4	5
1	2	3	4	5	1	2	3	4	5
1	2	3	4	5	1	2	3	4	5
1	2	3	4	5	1	2	3	4	5
1	2	3	4	5	1	2	3	4	5
1	2	3	4	5	1	2	3	4	5
1	2	3	4	5	1	2	3	4	5

### Corpo docente:

- 16 Disponibilità docenti per chiarimenti, spostamento di orario..
- 17 Chiarezza espositiva dei docenti durante le lezioni
- 18 Capacità del docente di suscitare interesse
- 19 Prestigio dell'Università frequentata
- 12 Collaborazione tra Università e mondo del lavoro/ricerca

1	2	3	4	5	1	2	3	4	5
1	2	3	4	5	1	2	3	4	5
1	2	3	4	5	1	2	3	4	5
1	2	3	4	5	1	2	3	4	5
1	2	3	4	5	1	2	3	4	5
1	2	3	4	5	1	2	3	4	5

### Organizzazione e logistica:

- 27 Organizzazione dell'orario delle lezioni
- 28 Disponibilità e capienza aule studio
- 29 Disponibilità e capienza aule computer
- 30 Possibilità di creare un piano di studio personalizzato
- 31 Biblioteca ben fornita

1	2	3	4	5	1	2	3	4	5
1	2	3	4	5	1	2	3	4	5
1	2	3	4	5	1	2	3	4	5
1	2	3	4	5	1	2	3	4	5
1	2	3	4	5	1	2	3	4	5
1	2	3	4	5	1	2	3	4	5

### Rapporto università e ambiente:

- 32 Possibilità di svolgere vita universitaria (organizzazione di feste e ritrovi per universitari)
- 33 Costo della vita a Milano
- 20 Possibilità e facilità di fruizione di servizi correlati(mensa,librerie...)

1	2	3	4	5	1	2	3	4	5
1	2	3	4	5	1	2	3	4	5
1	2	3	4	5	1	2	3	4	5

### Dati anagrafici dell'intervistato:

Sesso  M  F      Età..... Sei fuori corso  si  no      Hai partecipato ad un Erasmus  si  no

La tua residenza dista dalla facoltà

Meno di 10 Km	Meno di 20 Km	Meno di 30 Km	Meno di 40 Km	Meno di 50 Km	Più di 50 Km
---------------	---------------	---------------	---------------	---------------	--------------

Hai preso in affitto un appartamento nella città universitaria?  si  no

## RIFERIMENTI BIBLIOGRAFICI

- RUSSEL WINER [2002], *Marketing Management*, APOGEO, Milano.
- COZZI G. FERRERO G. [2000], *Marketing principi, metodi, tendenze evolutive*, Giapichelli editore, Torino.
- COLLESEI U.[1994], *Marketing*, CEDAM, Padova.
- MOLTENI L. [1993], *L'analisi multivariata nelle ricerche di marketing*, Egea, Milano.
- LEEFLANG P. WITTINK D. WEDEL M. NEART P.[2000], *Building models for marketing decisions*, Kluwer Academic Publishers, Boston.
- ZIKMUND W.[2000], *Exploring marketing research*, Harcourt College Publishers.
- AAKER D. KUMAR V. DAY G. [1998], *Marketing Research*, John Wiley & Sons.
- MARBACH G. [1996], *Le ricerche di mercato*, UTET, Torino
- CHISNALL P. [1996], *Le ricerche di marketing*, McGraw-Hill, Milano
- BARILE S. METALLO G. [2002], *Le ricerche di mercato Aspetti metodologici ed applicativi*, Giappichelli Editore, Torino
- DE LUCA A.[1995], *Le applicazioni dei metodi statistici alle analisi di mercato*, FrancoAngeli, Milano.
- KOTLER P. SCOTT W., *Marketing management*, ISEDI, Torino.
- CALDANI E.[1984], *Definizione e segmentazione del mercato*, Giuffrè Editore
- AL RIES JACK TROUT FERRARI G. ZEPPA M.[1988], *Positioning La conquista della posizione vincente*, McGraw-Hill, Milano
- DEL CIELO N. DULLI S. SACCARDI A. [2000], *Metodi di Data Mining per il Customer Relationship Management*, FrancoAngeli, Milano

FABBRIS L.[1994], *Analisi esplorativa di dati multidimensionali*, CLEUP editrice, Padova

CORBETTA P. [2002], *Metodi di analisi multivariata per le scienze sociali*, Il Mulino, Bologna

LEWIS-BECK M.S. [1994], *Factor analysis & related techniques*, SAGE Publications

HELMUTH SPATH [1980], *Cluster analysis algorithms for data reduction and classification of objects*, Ellison Horwood Limited Publishers.

JAMBU M. LEBEAUX M. [1983], *Cluster analysis and data analysis*, North Holland publishing company.

MC LACHLAN G. [1992], *Discriminant analysis and statistical pattern recognition*, John Wiley & sons.

NAVA G. [1990], *I fattori di crisi e di successo nelle imprese italiane: i risultati dell'analisi discriminante applicata a tre settori manifatturieri*, CEDAM, Padova.

GRIMALDI R. [2001], *Valutare l'università*, UTET Libreria, Torino.

HOWITT D. CRAMER D. [1999], *A guide to computing statistics with SPSS for WINDOWS*, Prentice Hall.

CAPIZZI G. [2001], *Appunti di Analisi Multivariata – Corso di Statistica (campionamento, analisi multivariata)*

JARVIS C. MACKENZIE S. PODSAKOFF P.[2003], *A Critical Review of Construct Indicators and Measurement Model Misspecification in Marketing and Consumer Research*, Journal of consumer research, Vol. 30 settembre

HOFFMAN L. NOVAK T.[1999], *La trasformazione delle attività di marketing*, Dossier : Internet marketing, CUOA Rivista, n.1 settembre.

TOSI P. [2003], *Prima relazione sullo stato delle Università italiane*, Conferenza dei Rettori delle Università Italiane, Roma 25 settembre 2003.

Facoltà di Scienze Statistiche di Padova:

<http://www.stat.unipd.it>

Facoltà di Scienze Statistiche di Bologna:

<http://www.stat.unibo.it>

Facoltà di Scienze Statistiche Bicocca di Milano:

<http://www.statistica.unimib.it>

Software SPSS

<http://www.spss.com>



*Un sincero grazie alla mia famiglia, che mi ha trasmesso la voglia di impegnarsi in tutto ciò che si fa e ad affrontare le difficoltà incontrate sempre con un sorriso, senza mai darsi per vinti.*

*Un grazie alla prof.ssa Pertile perché mi ha sempre incoraggiata con il suo splendido sorriso ed entusiasmo, alla prof.ssa Bassi per la sua disponibilità, alla prof.ssa Filosofo, il prof. Di Fonzo, il prof. Brasini, il prof. Donzelli, la prof.ssa Chiogna, , gli studenti della Facoltà di Scienze Statistiche di Padova, Bologna e Milano e quanti altri hanno collaborato nella stesura di questa tesi.*

*Un grazie ad Alessandro per essermi stato vicino in tutti questi anni e per tutti gli attimi di serenità che mi ha donato.*

*Un grazie a Mauro, il miglior compagno di viaggio che potessi sperare di incontrare, perché ha reso questa esperienza meravigliosa e mi ha fatto riscoprire il vero valore dell'amicizia.*

*Un grazie a Francesca perché la sua dolcezza mi ha scaldato il cuore e grazie a tutti gli altri compagni di Università: Monia, Martina, Ale, Sara, Nik, Dany, Fausto...e a tutti gli amici che in questi anni hanno camminato insieme a me.*



This document was created with Win2PDF available at <http://www.daneprairie.com>.  
The unregistered version of Win2PDF is for evaluation or non-commercial use only.