UNIVERSITÀ DELLA CALABRIA

DIPARTIMENTO DI **SCIENZE AZIENDALI E GIURIDICHE**

MARKETING – CORSO PROGREDITO

ANNO ACCADEMICO 2026/2027

La realizzazione di una ricerca di mercato

a cura di
Gaetano “Nino” Miceli

In questa dispensa è raccolto materiale utile ai fini della progettazione e realizzazione di una Ricerca di Marketing, sviluppata come attività didattica proattiva nell'ambito del corso di Marketing progredito. Le attività delle ricerche di marketing saranno, inoltre, supportate da lezioni ed esercitazioni. I progetti di ricerca sul campo sono obbligatori per tutti gli studenti (incidono sul voto d'esame con un peso del 40%) e potranno essere realizzati da gruppi composti da 4 o 5 studenti.

Il lavoro necessario alla realizzazione di una ricerca di marketing è tanto; serve, quindi, organizzare un gruppo coeso e motivato, in cui i compiti siano ripartiti equamente.

Posti gli impegni esterni al corso, che si sommano ai carichi di lavoro relativi allo studio del programma, è opportuno che il processo di sviluppo della ricerca sia adeguatamente programmato; a tal proposito sono istituiti i S.A.L. (Stati di Avanzamento dei Lavori). Si tratta di scadenze intermedie volte a stimolarvi ad un lavoro graduale e "bilanciato", che potrete gestire seguendo il Gantt. Il rispetto dei SAL rappresenta un elemento di valutazione: la mancata consegna nei termini indicati del materiale richiesto può comportare una penalizzazione sul "voto" finale al rapporto di ricerca.

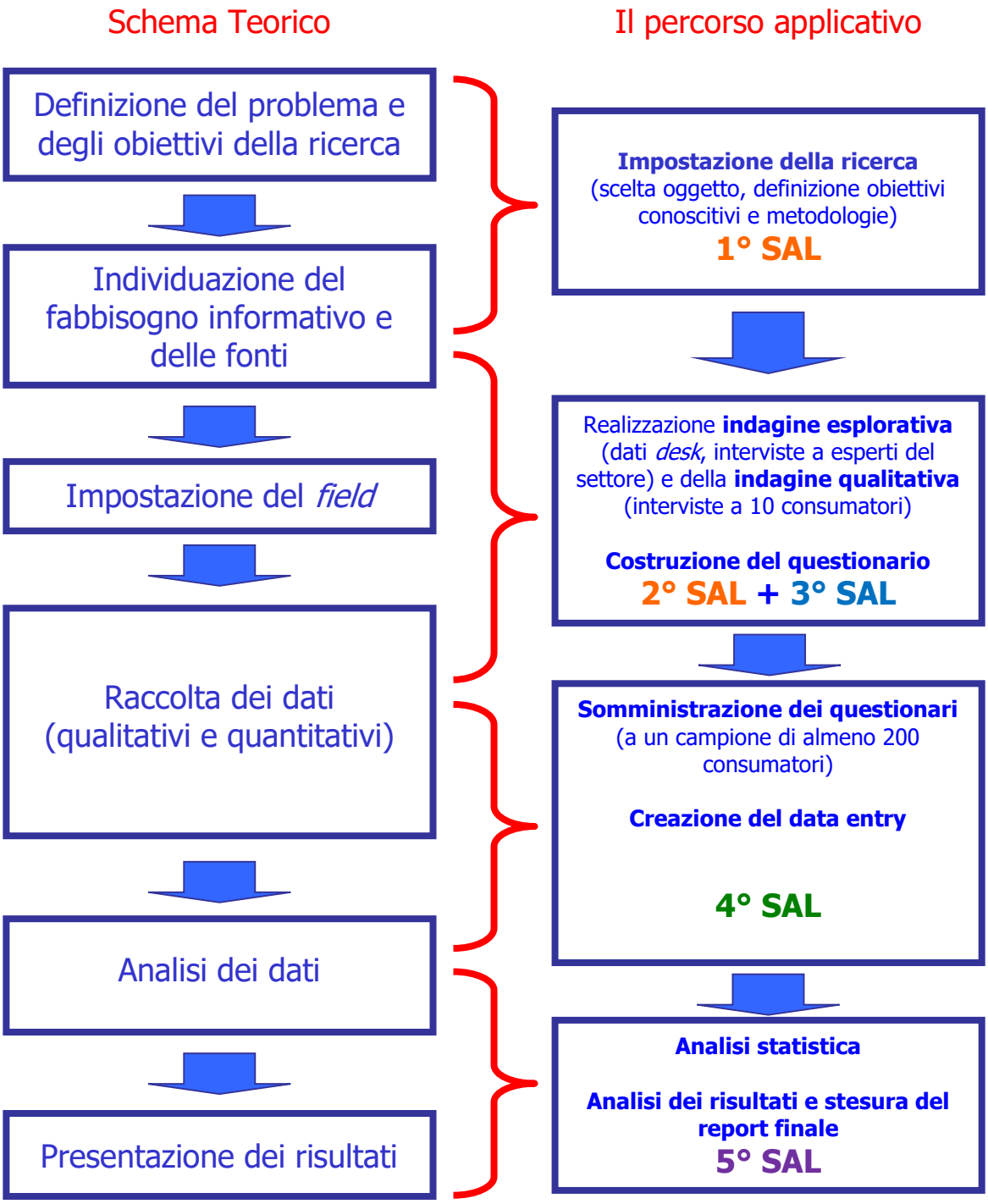
A fronte dell'impegno che vi si richiede, l'attività di ricerca vi permette di sviluppare conoscenze specialistiche fondamentali, indispensabili per la creazione di un bagaglio di competenze completo e orientato all'analisi sistematica del mercato.

Buon lavoro a tutti ed in bocca al lupo!

Indice

Il processo di sviluppo di una ricerca di mercato:

uno schema teorico e le applicazioni pratiche



SCADENZARIO

1° SAL - Iscrizione gruppi e assegnazione dei progetti di ricerca: ***29.9.2026***

2° SAL - Consegna Mappa mezzi-fini ed eventuali disegni sperimentali: ***27.10.2026***

3°SAL - Consegna questionario, eventuali stimoli sperimentali e piano di campionamento: ***10.11.2026***

4° SAL - Consegna del file con il data-entry: ***10.12.2026***

5° SAL - Consegna del rapporto di ricerca completo**:** ***26.1.2027***

GANTT per Attività di Marketing Research - Corso di Marketing Progredito - a.a. 2026/2027																														
Settembre 2026																														
ATTIVITÀ	Ma	Me	Gi	Ve	Sa	Do	Lu	Ma	Me	Gi	Ve	Sa	Do	Lu	Ma	Me	Gi	Ve	Sa	Do	Lu	Ma	Me	Gi	Ve	Sa	Do	Lu	Ma	Me
	01-set	02-set	03-set	04-set	05-set	06-set	07-set	08-set	09-set	10-set	11-set	12-set	13-set	14-set	15-set	16-set	17-set	18-set	19-set	20-set	21-set	22-set	23-set	24-set	25-set	26-set	27-set	28-set	29-set	30-set
Presentazione attività																														
Creazione gruppi																														
1° SAL: Iscrizione gruppi e assegnazione dei progetti di ricerca																														
Indagine esplorativa																														
Lezioni su ricerca qualitativa																														

GANTT per Attività di Marketing Research - Corso di Marketing Progredito - a.a. 2026/2027																																
Ottobre 2026																																
ATTIVITÀ	Gi	Ve	Sa	Do	Lu	Ma	Me	Gi	Ve	Sa	Do	Lu	Ma	Me	Gi	Ve	Sa	Do	Lu	Ma	Me	Gi	Ve	Sa	Do	Lu	Ma	Me	Gi	Ve	Sa
	01-ott	02-ott	03-ott	04-ott	05-ott	06-ott	07-ott	08-ott	09-ott	10-ott	11-ott	12-ott	13-ott	14-ott	15-ott	16-ott	17-ott	18-ott	19-ott	20-ott	21-ott	22-ott	23-ott	24-ott	25-ott	26-ott	27-ott	28-ott	29-ott	30-ott	31-ott	
Indagine esplorativa																																
Lezioni su ricerca qualitativa																																
Invio traccia di intervista (via e-mail)																																
Feedback sulla traccia																																
Indagine qualitativa: conduzione interviste, creazione mappa ed eventuali disegni sperimentali																																
2° SAL: Consegna Mappa mezzi-fini ed eventuali disegni sperimentali																																
Feedback su mappa mezzi-fini e disegni sperimentali																																
Esercitazione sul questionario e sul campionamento																																

GANTT per Attività di Marketing Research - Corso di Marketing Progredito - a.a. 2026/2027																															
Novembre 2026																															
ATTIVITÀ	Do	Lu	Ma	Me	Gi	Ve	Sa	Do	Lu	Ma	Me	Gi	Ve	Sa	Do	Lu	Ma	Me	Gi	Ve	Sa	Do	Lu	Ma	Me	Gi	Ve	Sa	Do	Lu
	01-nov	02-nov	03-nov	04-nov	05-nov	06-nov	07-nov	08-nov	09-nov	10-nov	11-nov	12-nov	13-nov	14-nov	15-nov	16-nov	17-nov	18-nov	19-nov	20-nov	21-nov	22-nov	23-nov	24-nov	25-nov	26-nov	27-nov	28-nov	29-nov	30-nov
Costruzione questionario ed eventuali stimoli sperimentali																														
3° SAL: consegna questionario, eventuali stimoli sperimentali e piano di campionamento																														
Feedback su questionario ed eventuali stimoli sperimentali																														
Somministrazione questionari e costruzione database																														
Esercitazione sul database																														

GANTT per Attività di Marketing Research - Corso di Marketing Progredito - a.a. 2026/2027																																
Dicembre 2026																																
ATTIVITÀ	Ma	Me	Gi	Ve	Sa	Do	Lu	Ma	Me	Gi	Ve	Sa	Do	Lu	Ma	Me	Gi	Ve	Sa	Do	Lu	Ma	Me	Gi	Ve	Sa	Do	Lu	Ma	Me	Gi
	01-dic	02-dic	03-dic	04-dic	05-dic	06-dic	07-dic	08-dic	09-dic	10-dic	11-dic	12-dic	13-dic	14-dic	15-dic	16-dic	17-dic	18-dic	19-dic	20-dic	21-dic	22-dic	23-dic	24-dic	25-dic	26-dic	27-dic	28-dic	29-dic	30-dic	31-dic
Somministrazione questionari e costruzione database																															
4° SAL: Consegna database (via e-mail)																															
Feedback su Database																															
Analisi statistiche																															
Redazione report e presentazione																															

GANTT per Attività di Marketing Research - Corso di Marketing Progredito - a.a. 2026/2027																																
Gennaio 2027																																
ATTIVITÀ	Ve	Sa	Do	Lu	Ma	Me	Gi	Ve	Sa	Do	Lu	Ma	Me	Gi	Ve	Sa	Do	Lu	Ma	Me	Gi	Ve	Sa	Do	Lu	Ma	Me	Gi	Ve	Sa	Do
	01-gen	02-gen	03-gen	04-gen	05-gen	06-gen	07-gen	08-gen	09-gen	10-gen	11-gen	12-gen	13-gen	14-gen	15-gen	16-gen	17-gen	18-gen	19-gen	20-gen	21-gen	22-gen	23-gen	24-gen	25-gen	26-gen	27-gen	28-gen	29-gen	30-gen	31-gen
Redazione report e presentazione																															
Invio poster per presentazione																															
Presentazioni alle imprese																															
5° SAL: Consegna report																															
Esami																															

Un format per l'indice del report

Quello che segue è un format standard per la stesura del report di ricerca; chiaramente, non si tratta di uno schema rigido, ma di un'utile guida nell'impostazione del lavoro.

1. Introduzione (circa 1.5 pagine)

Introduzione in senso stretto con motivazioni della ricerca e obiettivi; breve descrizione del processo di ricerca e una/due frasi di anticipazione dei risultati; una frase in cui si preannunciano i successivi paragrafi

2. Fase esplorativa (circa 3 pagine)

Descrizione del materiale informativo raccolto a fini esplorativi per generare conoscenza iniziale e descrizione generale del contesto d'indagine (mercato, competizione, ecc.)

3. Fase qualitativa (circa 12 pagine)

Descrizione del processo di ricerca qualitativa: obiettivi, soggetti intervistati, strumenti di raccolta, interviste, commento alla mappa, principali conclusioni (anche per la definizione dell'indagine quantitativa)

4. Fase quantitativa (circa 18-22 pagine)

Descrizione del processo di ricerca quantitativa: obiettivi, descrizione del questionario, metodo di contatto, campione; inoltre

- analisi univariate: per tutte le variabili
- analisi bivariate: per 7-8 coppie di variabili per le quali appare interessante studiare la relazione
- analisi multivariate: output, commento, interpretazione

5. Implicazioni manageriali (circa 5 pagine)

Fase fondamentale, spesso colpevolmente poco considerata: riassumere i principali risultati della ricerca e formulare implicazioni manageriali derivanti dalle evidenze ottenute

ALLEGATI:

- la mappa cognitiva (descrittiva della catena mezzi-fini);
- il questionario (e/o le cartoline utilizzate per l'applicazione della conjoint analysis).

Il percorso di ricerca, in sintesi, prevede:

1) una fase di definizione degli obiettivi informativi da perseguire;
2) una fase esplorativa, volta a creare una conoscenza di fondo sull'oggetto;
3) una fase di ricerca qualitativa;
4) una fase di ricerca quantitativa;
5) una fase di interpretazione dei risultati e di implicazioni ai fini della ricerca.

La definizione dell'oggetto della ricerca, degli obiettivi conoscitivi e delle tecniche di analisi multivariata da applicare

Oggetto della ricerca

L'oggetto della ricerca può essere essenzialmente un prodotto – un bene o un servizio, ovvero un'impresa, oppure un oggetto di ricerca di base.

Obiettivi conoscitivi e tecniche di analisi multivariata

Gli obiettivi conoscitivi di una ricerca di mercato possono essere molteplici; saranno, comunque, connessi a una problematica di marketing strategico o operativo. Per semplificare la scelta degli obiettivi da perseguire viene proposta la seguente tabella sinottica, che associa ad una serie di obiettivi di ricerca le corrispondenti tecniche di statistica multivariata idonee al perseguimento dei risultati attesi. *L'applicazione delle tecniche quantitative deve essere comunque preceduta dall'applicazione delle tecniche qualitative.*

OBIETTIVO CONOSCITIVO	TECNICHE
SEGMENTAZIONE CLASSICA Suddividere il mercato in diversi gruppi di consumatori (segmenti), caratterizzati da similarità di preferenze al loro interno; tali preferenze sono espresse rispetto ad attributi o benefici ricercati **(*chi* preferisce *cosa*)**.	**Factor Analysis + Cluster Analysis**
SEGMENTAZIONE FLESSIBILE Individuare i segmenti di mercato sulla base delle percezioni di valore (importanza e gradimento), in modo da identificare il **profilo di prodotto ideale per ciascun segmento**.	**Conjoint Analysis + Cluster Analysis**
ANALISI DELLE PERCEZIONI DI VALORE PER IL CLIENTE Con riferimento ad un determinato prodotto (nuovo o esistente), determinare l'importanza per il cliente degli attributi che lo definiscono e il gradimento associato alle specifiche di ciascun attributo, in modo da identificare il **profilo di prodotto ideale**.	**Conjoint Analysis**

ANALISI DI POSIZIONAMENTO Identificare il modo in cui una categoria di prodotti/marche (brand) concorrenti sono percepiti dai consumatori. - evidenziare le caratteristiche che differenziano i prodotti agli occhi del cliente; - comprendere i punti di forza e di debolezza dei diversi prodotti; - rappresentare graficamente il grado di sostituibilità di prodotti concorrenti; - individuare opportunità di mercato (vuoti d'offerta); - comprendere come progettare un nuovo prodotto o modificare un prodotto esistente.	**Tecniche Attribute-Based** (basate su valutazioni degli intervistati sui singoli attributi del prodotto) **Discriminant Analysis** (consigliabile per posizionare prodotti o marche che i consumatori riescono a distinguere e valutare facilmente, rispetto a una serie di attributi tecnici)
ANALISI DELLA SODDISFAZIONE DEL CLIENTE Misurare il livello di soddisfazione percepito dopo l'acquisto per un prodotto/marca e individuazione delle sue determinanti	**Analisi di Regressione Lineare Multipla**
STUDIO DELLA DIPENDENZA DI UNA VARIABILE Misurare l'incidenza di una serie di variabili indipendenti su una variabile dipendente quantitativa (misurata su scala metrica)	**ANOVA e/o Analisi di Regressione Lineare Multipla**

L’indagine esplorativa

L’indagine esplorativa consiste nella raccolta di dati secondari (desk) su pubblicazioni e riviste varie, e soprattutto sul web. La raccolta di tali dati può essere integrata da interviste a esperti del settore (colloqui informali) e interviste aperte a consumatori del prodotto: ***tale fase di indagine è comune a tutti i progetti di ricerca***.

Le finalità di questa indagine sono di tipo ESPLORATIVO (per definire più chiaramente il problema, per acquisire informazioni sul prodotto - domanda e offerta - e, in generale, per acquisire maggiore familiarità con l’argomento oggetto di studio), e di ORIENTAMENTO (per indirizzare la costruzione del questionario mediante la generazione degli attributi e dei benefici da sottoporre a valutazione).

L’indagine qualitativa

La fase qualitativa della ricerca di marketing si pone l’obiettivo della profondità dei risultati, quindi della generazione di ipotesi sulle motivazioni degli atteggiamenti e dei comportamenti, da verificare successivamente, con modalità di somministrazione estensive, in sede quantitativa. Il campione degli intervistati in sede qualitativa è piccolo (6 – 10 soggetti).

Tra le varie tecniche di indagine qualitative, un posto di riguardo lo merita il laddering. Tale tecnica si basa sul modello di analisi del consumatore chiamato “modello della catena mezzi-fini”. Per una descrizione della tecnica si veda l’estratto riportato di seguito.

Le tecniche di rilevazione della catena mezzi-fini: il laddering

tratto dal paper "L'analisi del valore per il cliente: il contributo dei modelli 'mezzi-fini'" (di Michele Costabile e Maria Antonietta Raimondo)

Il metodo più utilizzato per ricostruire e valutare la catena mezzi-fini è il *laddering*, che è una tecnica qualitativa di intervista in profondità *one-to-one*.
La teoria basata sulla catena mezzi-fini, e la connessa tecnica del laddering, sono state applicate con successo allo sviluppo delle strategie di comunicazione, al posizionamento del prodotto, alla segmentazione e al copy testing. Le applicazioni qualitative sono state oggi potenziate dallo sviluppo di applicazioni quantitative a partire dai dati del *laddering*.
Il laddering è una tecnica d'interazione con il cliente, utilizzabile in sede di *focus group*, o più opportunamente nel corso di interviste personali in profondità, che ha l'obiettivo ricostruire e rappresentare graficamente la mappa cognitiva delle relazioni tra prodotti, attributi, benefici e valori. Più specificamente, vengono indagati i motivi di una scelta d'acquisto, tentando di risalire dagli attributi del prodotto che l'hanno determinata, ai benefici percepiti in associazione a tali attributi e così via, sino a identificare i valori terminali che con il comportamento di consumo si ritiene di poter affermare (Reynolds e Gutman, 1988)[1]. La sequenza di connessioni cognitive (network associativo o *ladders*) fra attributi, benefici e valori così ottenute rappresenta gli orientamenti percettivi, o "modi di pensare", del consumatore con riferimento ad una certa categoria di prodotti e produce una più diretta ed efficace comprensione del suo comportamento.
La procedura basata sul *laddering* prevede tre fasi distinte: l'individuazione delle caratteristiche salienti dei prodotti o delle marche oggetto di indagine, e la risalita ai benefici e ai valori, sia nella fase di raccolta dei dati (conduzione intervista) che di analisi dei risultati (costruzione della mappa). In particolare, tale approccio prevede che i soggetti intervistati siano indotti, inizialmente, a verbalizzare i concetti che utilizzano per valutare/ differenziare i prodotti o le marche (o qualsiasi altro stimolo sia interessante) e, successivamente, gli esiti/vantaggi che ne derivano e i valori di base, esplicitando i collegamenti che formano il loro network cognitivo; infine, alle risposte verbali ottenute a livello individuale si applica un insieme di procedure analitiche che le combina in una mappa di struttura cognitiva aggregata.

La fase di raccolta dei dati
Nella fase di raccolta dei dati, l'applicazione del *laddering* richiede che ogni soggetto intervistato sia indotto a ricostruire i *ladders* (i legami fra attributi, benefici e valori) e a risalirli attraverso livelli di astrazione maggiori, partendo dal livello concreto rappresentato dalle caratteristiche fisiche del prodotto per giungere a quello più astratto dei valori, che si colloca lontano da qualsiasi referente fisico.

Esistono diverse tipologie di laddering, ossia diversi approcci utilizzabili dall'intervistatore per stimolare la risalita dagli attributi ai valori. Si sottolinea che tutte le forme di laddering partono "elicitando" l'esplicitazione degli attributi concreti e poi proseguono risalendo ai benefici e ai valori. Può accadere, però, che gli intervistati esplicitino prima gli attributi astratti o prima i benefici, qui emerge il problema sollevato da alcuni autori relativamente alla configurazione gerarchica di un network, una sorta di contraddizione implicita (gerarchia versus rete) alle

[1]La tecnica del *laddering* è, spesso, confusa con la teoria della catena mezzi-fini; in realtà, la teoria dovrebbe sempre essere separata dalla metodologia.

applicazioni che guidano alle HVM (*Hierarchical Value Map*). In pratica, però, si può ovviare facendo ladder up (risalita: dagli attributi concreti a quelli astratti, ai benefici…) e ladder down (dai benefici agli attributi...) e ricostruendo una mappa che per convenienza viene letta in prospettiva gerarchica, anche se trattasi di un network associativo.
Le principali tipologie di *laddering* sono:

- ***laddering diretto.*** Si chiede all'intervistato di esplicitare le ragioni che lo spingono all'acquisto del prodotto considerato, spingendolo a chiarire cosa cerca in esso. Si procede quindi con la risalita.
- ***laddering comparativo (scelta entro una triade).*** Consiste nel sottoporre ad ogni intervistato, una alla volta, delle triadi di prodotti o marche e nel chiedergli di enunciare le differenze e le similarità percepite tra le diverse marche o prodotti.
- ***laddering di contesto***. Viene richiesto all'intervistato di ricordare un'occasione realistica di utilizzo del prodotto in esame. Il focus di tali tipi di domande è analizzare le occasioni per il consumo del prodotto e l'uso, o il fine, ad esso associato. Successivamente si ricostruiscono i singoli *ladders* tramite il processo di risalita.
- ***laddering ipotetico*.** Si invita il soggetto a ragionare ipoteticamente, immaginando il suo comportamento in caso di assenza del prodotto in esame. Agendo in tal modo i consumatori sono incoraggiati a verbalizzare il beneficio ricercato o il valore sotteso al consumo di un bene.
- ***laddering negativo*.** A differenza delle altre tecniche di *laddering*, le domande formulate in questo caso riguardano le ragioni per cui i soggetti non fanno determinate cose o non vogliono provare certe sensazioni. Anche questa tecnica, come le precedenti, viene ad ogni modo seguita da un processo di risalita. Il *laddering* negativo è particolarmente rilevante quando gli intervistati non riescono ad articolare le loro motivazioni.
- ***laddering regressivo***. Si riporta il soggetto indietro nel tempo e lo si incoraggia ad esprimere sentimenti e comportamenti di acquisto in riferimento a precedenti occasioni di consumo. Gli intervistati potranno, in tal modo, giudicare criticamente e con obiettività le proprie passate abitudini di consumo.
- ***laddering proiettivo***. Si invita il soggetto a ricordare una situazione d'acquisto e ad esprimere le impressioni non proprie, ma di un terzo che l'intervistato sceglie di impersonare. Segue la risalita.

L'approccio più frequentemente utilizzato nel corso delle interviste prevede l'impiego la procedura, nota come "scelta entro una triade" (*triading sorting*) o *laddering* comparativo e "risalita" (Olson e Reynolds, 1990), che sebbene costituisca solo uno dei metodi utilizzabili consente, completandosi a vicenda, di pervenire alla conoscenza dell'intera struttura cognitiva del consumatore.
L'analisi della catena mezzi-fini inizia, dunque, con una procedura d'attivazione diretta che consiste nel sottoporre ad ogni intervistato, una alla volta, delle triadi di prodotti o marche e nel chiedergli di enunciare le differenze e le similarità percepite tra le diverse marche o prodotti, con una domanda del tipo "Mi dica in che modo due di questi sono simili e differiscono dal terzo". Il soggetto enuncia il maggior numero di distinzioni possibili, poi gli viene presentata la triade successiva e così via. La scelta entro una triade è utilizzata nel primo stadio dell'intervista per individuare i più bassi livelli di distinzione, cioè gli elementi cognitivi che i consumatori usano per discriminare gli stimoli in un dato campo (ad esempio, le marche in una categoria di prodotti). Sebbene queste distinzioni possono essere ad un qualsiasi livello di astrazione, la maggior parte dei consumatori effettua distinzioni al livello concreto degli

attributi del prodotto (Olson e Reynolds, 1990); l'esame delle triadi, dunque, ha lo scopo specifico di individuare, sulla base delle distinzioni ritenute esistenti tra una marca e l'altra, gli attributi rilevanti nel processo di valutazione e scelta di un determinato prodotto.
Nello stadio successivo dell'intervista, condotto senza alcun riferimento a specifiche marche, gli attributi salienti identificati nella fase precedente costituiscono il punto di partenza per il processo di risalita, che costringe i consumatori a ripercorrere la scala di astrazione rivelando gli aspetti strutturali della categorizzazione. Si chiede ai partecipanti di valutare i concetti esplicitati durante la scelta entro una triade, spesso in termini della loro importanza in relazione a una decisione d'acquisto, esplicitando così i benefici e i valori che derivano dai singoli attributi, nonché il modo in cui si articolano le associazioni esistenti tra i diversi livelli di astrazione. Durante il processo di risalita al consumatore è continuamente posta una domanda del tipo "Perché ciò è importante per lei?", fino a quando egli non è più in grado di rispondere ulteriormente. Non sempre questo percorso conduce al livello dei valori, ma ciò è quello che frequentemente accade.
Dal momento che l'intero processo di analisi del *laddering* si basa sulle parole, sulle associazioni di parole e sui significati che i consumatori utilizzano per formulare le loro risposte, è necessario ricreare il contesto di consumo, concretizzando le circostanze di scelta in cui i soggetti sono ripetutamente coinvolti. Uno dei principali ostacoli all'instaurarsi di una relazione fiduciosa d'intervista è proprio la difficoltà dei consumatori ad articolare le loro ragioni.
Per superare, in parte, tali difficoltà si ricorre a tecniche di *laddering* alternative, o più opportunamente consecutive, caratterizzate da un approccio di ricerca analitico e destrutturato, il quale permette di interagire facilmente con il soggetto, dandogli l'impressione di essere personalmente poco coinvolto. Considerando d'altra parte che il modo migliore per capire e per apprendere è chiedere ai consumatori di motivare le loro affermazioni con i loro termini risulta prioritario, nell'instaurare una precipua relazione d'intervista, considerare i soggetti intervistati dei partner o collaboratori, e l'intervista come una conversazione. I simboli, le metafore, le parole, le associazioni di parole che gli stessi utilizzano, sono assolutamente esatte nel descrivere i significati o le sequenze di significati che appartengono a ciascuno di loro. La relazione d'intervista diventa così un processo endogeno di costruzione cognitiva.
Per superare i principali problemi che possono insorgere durante la conversazione, è possibile, inoltre, utilizzare una serie di tattiche come "il silenzio ripetuto" e "la verifica delle affermazioni". Nel primo caso il silenzio da parte dell'intervistatore ha lo scopo di indurre l'intervistato a fornire una risposta più appropriata e definitiva quando non è disposto a pensare criticamente circa la domanda postagli o quando si sente a disagio per ciò che sta scoprendo su se stesso; nel secondo caso, l'intervistatore ripete all'intervistato la stessa affermazione invitandolo a spiegarla con maggiore precisione.
Nella classica procedura del *laddering* appena descritta, la spontaneità del consumatore è limitata il meno possibile. Questo genere di *laddering* è definito "*soft laddering*", distinto dall'approccio "*hard laddering*" utilizzato da alcuni ricercatori (Mulvey e al., 1994; Botschen e Hemetsberger, 1998), che dà meno libertà di risposta ai consumatori e li costringe a seguire un *ladder* alla volta, in cui ogni risposta successiva è ad un livello di astrazione più elevato. Un esempio di questo approccio è il metodo *paper-and-pencil*, consistente in un questionario *self-administered* sviluppato da Walker e Olson (1991). Tale metodo prevede che i soggetti intervistati indichino quali sono le quattro caratteristiche del prodotto considerato che essi ritengono più importanti nel processo decisionale di acquisto e, per ognuno dei quattro criteri

di scelta identificati, quali sono le ragioni per cui essi sono importanti. Questo processo consistente nello scrivere perché ogni risposta è importante continua fino a quando gli intervistati non sono più capaci di rispondere, ossia quando essi raggiungono la fine della catena mezzi-fini. Rispetto alla procedura classica del *laddering*, Grunert e Grunert (1995) ritengono che il metodo *paper-and-pencil* permetta all'intervistato di essere meno condizionato dalla presenza dell'intervistatore e di seguire più spontaneamente il proprio percorso cognitivo.

L'analisi dei risultati

L'esame delle risposte verbali generate dal *laddering* inizia con un'analisi di contenuto (*content analysis*) finalizzata a ridurre le risposte idiosincratiche dei soggetti ad un comune set di significati: ad ogni pensiero/risposta di ciascun soggetto viene attribuito un codice di categoria, eliminando le espressioni personali di pensieri di base simili e giungendo così ad identificare un set di concetti standard che riassumono tutti gli attributi, conseguenze e valori menzionati nelle risposte date durante le interviste. I concetti trasformati in codici diventano così i nodi di un'ipotetica rete associativa.

Il passo successivo consiste nell'analisi strutturale, ossia nell'analisi delle connessioni cognitive esistenti tra questi concetti, spesso chiamate implicazioni. Un'implicazione è definibile come la percezione di una relazione causale o strumentale tra due concetti. Una catena mezzi-fini è semplicemente una sequenza di implicazioni causali: un attributo implica una conseguenza che implica un'altra conseguenza che implica un valore. Tali implicazioni possono essere dirette o indirette. Ad esempio, una catena mezzi-fini del tipo:

$$A \rightarrow B \rightarrow C$$

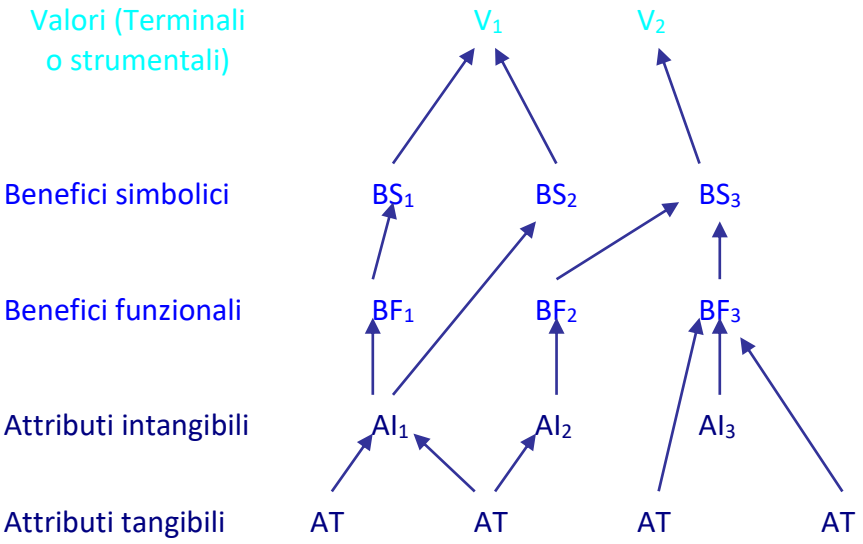
contiene due implicazioni dirette, ossia quella da A a B e da B a C, ognuna della quale è stata esplicitamente dichiarata da un consumatore nel corso dell'intervista, ed un'implicazione indiretta tra A e C, che non è stata effettivamente menzionata da un consumatore ma è implicata dalle due associazioni dirette.

Per poter costruire i collegamenti tra i concetti emersi dal *laddering*, stabilendo così gli archi della rete, si costruisce una matrice chiamata matrice delle implicazioni (*implication matrix*), che rappresenta tutte le relazioni dirette e indirette tra l'insieme di concetti opportunamente categorizzati, e quindi la struttura cognitiva aggregata dell'intero campione intervistato.

Si tratta di una matrice quadrata che presenta nelle righe e nelle colonne i concetti precedentemente individuati. I dati inseriti nelle celle indicano le frequenze, relative a tutti i soggetti del campione, con cui un attributo, conseguenza o valore (ossia un elemento sulla riga) conduce direttamente o indirettamente ad un altro attributo, conseguenza o valore (ossia un elemento sulla colonna). Un dato è, dunque, registrato in una particolare cella della matrice quando il concetto che si trova lungo la riga precede il concetto lungo la colonna nelle risposte ottenute nella risalita.

Infine, a partire dalla matrice delle implicazioni viene costruita una mappa che ricostruisce le relazioni gerarchiche a livello aggregato, ossia con riferimento all'intero campione. Tale mappa è chiamata *Hierarchical Value Map* (HVM) e mostra graficamente il contenuto e la struttura della conoscenza dei consumatori. Essa consiste in un insieme di nodi e di linee; i nodi rappresentano i significati concettuali, classificati in attributi, benefici e valori; le linee che uniscono i nodi sono le associazioni tra questi concetti. Inoltre, nella sua forma originaria

(Reynolds e Gutman, 1988), essa si presenta come un diagramma ad albero: i concetti a livello di valore sono in cima perché sono i concetti astratti finali che sono collegati attraverso una serie di associazioni a concetti meno astratti; gli attributi relativi a concetti più concreti tendono ad essere posizionati verso il basso della mappa perché sono il più delle volte gli stimoli che hanno dato inizio al processo di risalita; in una posizione intermedia si collocano i benefici percepiti in associazione agli attributi che contribuiscono a realizzare i valori finali.
La costruzione della mappa avviene tentando di mantenere un equilibrio tra le esigenze di ritenzione dei dati e quelle di sintesi. A tal fine, viene spesso selezionato un valore soglia per determinare quali relazioni devono essere rappresentate nella mappa e quali, invece, devono essere eliminate: solo le associazioni menzionate da almeno una certa percentuale di rispondenti (ad esempio il 5% su un campione di 90 individui) viene così inclusa nella rappresentazione grafica. Inoltre, le associazioni ridondanti sono eliminate: se, ad esempio, il concetto "A" è associato sia al concetto "B" che al concetto "C", e il concetto "B" è associato al concetto "C", nella mappa saranno tracciati solo i segmenti AB e BC poiché il segmento AC è ridondante, ossia già rappresentato in un percorso.
La teoria della catena mezzi-fini e la relativa tecnica del *laddering* sono state impiegate in numerose utilizzazioni: alcune applicative, altre puramente accademiche (Reynolds e Gutman, 1984; Reynolds e Braddock, 1988; Jolly, Reynolds e Slocum, 1988; Cozzi e Molinari, 1990; Olson e Reynolds, 1990; Gengler e Reynolds, 1995; Reynolds e Whitlark, 1995; Botschen e Hemetsberger, 1998; Valette-Florence, Siriex, Grunert e Nielse, 1998; Valette-Florence, 1998).
Come in precedenza descritto, l'output tipico degli studi sulla *means-end chain* è costituito da una mappa cognitiva aggregata (*Hierarchical Value Map*), che fornisce un'efficace rappresentazione grafica del contenuto e dell'organizzazione delle strutture cognitive di un insieme di consumatori con riferimento ad una certa categoria di prodotti.
Come nella maggior parte delle ricerche di marketing, l'uso di un grafico per visualizzare delle informazioni svolge principalmente tre funzioni (Gengler, Klenosky e Mulvey, 1995): immagazzinare informazione, comunicarla e facilitare il processo di elaborazione dei dati. Un grafico può, infatti, funzionare come un inventario di informazioni mediante la traduzione dei dati in un sistema di segni, ossia in forma simbolica; inoltre, esso organizza i dati, spesso disponibili in forma caotica, e li trasforma in informazioni significative; infine, un efficace impiego dei simboli può aiutare gli individui ad assimilare rapidamente complesse relazioni.



La forma **generale** delle mappe cognitive (hierarchical map)

Un esempio di catena mezzi-fini: l'oggetto è un portale web (estratto dalla mappa cognitiva della ricerca di mercato 2001/2002 del gruppo "I ritardatari")

TASSONOMIE DI VALORI STRUMENTALI E TERMINALI

Nella costruzione della mappa (HVM) può risultare ambigua la distinzione tra benefici psicologici, valori strumentali e valori terminali. Per agevolare la categorizzazione sono di seguito riportate alcune tipologie di valori strumentali e terminali.

VALORI STRUMENTALI O DI CONSUMO *(condotta preferita, coerentemente con i valori terminali determinano la cultura, gli atteggiamenti e i comportamenti di consumo in senso lato)*	***VALORI TERMINALI*** *(principi guida dell'esistenza di un individuo, condizione di vita e/o stati psicologici che il consumatore aspira a raggiungere)*
Eleganza	Naturalismo
Ostentazione	Tradizionalismo
Risparmio di risorse ambientali	Edonismo
Risparmio di risorse economiche	Individualismo
Competenza (ambizione, immaginazione)	Armonia sociale (uguaglianza, libertà)
Compassione (solidarietà, comprensione)	Gratificazione personale (consenso sociale)
Socialità (educazione, remissività)	Sicurezza (cura per la famiglia)
Integrità (responsabilità, onestà)	Amore (intimità, amicizia)
Logica	Felicità (ilarità)
Apertura mentale	Tranquillità
Capacità	Libertà
Dolcezza	Armonia interiore
Pulizia e ordine	Salvezza (in senso religioso)
Coraggio	Ammirazione sociale
Autocontrollo	Auto realizzazione
Intelligenza	Autostima
	Indipendenza

In definitiva, i valori terminali possono essere classificati nelle seguenti tipologie:

- valori "*individualistici*" (ad esempio: "affermazione della propria personalità", "autogratificazione", "autostima");
- valori "*sociali*" (ad esempio: "sicurezza sociale", "gratificazione sociale", affermazione nel gruppo");
- valori legati a sentimenti "forti" (ad esempio: "amore", "amicizia");
- valori legati alla salute ed alla persona (ad esempio: "salute", "rispetto della vita umana");
- valori legati ad ideologie (ad esempio: "ambientalismo", "comunismo").

Altre tecniche di indagine qualitativa

Protocol Analysis

La protocol analysis è una tecnica qualitativa che consiste nell'osservazione delle operazioni svolte dall'intervistato durante l'acquisto, l'utilizzo o il consumo del prodotto oggetto d'indagine. Ad esempio, è possibile osservare e trarre indicazioni dall'osservazione di un soggetto che utilizza un software, per indagarne le difficoltà operative, gli aspetti positivi, le sensazioni. È necessario preparare una lista di operazioni da far compiere all'intervistato, e che devono stimolare l'uso e l'esplorazione del prodotto. In particolare, i soggetti intervistati vengono stimolati a pensare ad alta voce durante le operazioni che compiono ("thinking aloud"), ovvero a comunicare ciò che stanno facendo ("talking aloud").

Critical Incident Technique

La Critical Incident Technique prevede la richiesta di narrare un episodio particolarmente positivo o particolarmente negativo vissuto rispetto all'acquisto/consumo del prodotto oggetto d'indagine. Dalle risposte ottenute, si procede quindi all'analisi di motivazioni e percezioni degli intervistati.

Test Proiettivi

I test proiettivi servono per "proiettare" l'intervistato in un particolare contesto (occasione d'uso/consumo) del prodotto, per ricostruire il momento critico da indagare, o in altri soggetti ("se fossi un dirigente della Ferrari..."), al fine di superare eventuali inibizioni.
Alcuni esempi sono:

- i test di completamento di fumetti, frasi, storie a puntate: si richiede al soggetto di completare delle vignette che richiamano il contesto di acquisto, di consumo o di valutazione;
- i test associativi: si richiede di associare marche o prodotti ad animali, personaggi famosi, fiori, colori, ecc. ecc.
- la ZMET (Zaltman Metaphor Elicitation Technique): si propone di comprendere il significato profondo che gli individui provano nei confronti di uno specifico prodotto facendo ricorso alle metafore visive che questi utilizzano per descrivere il prodotto stesso tramite collage di immagini.

L'elaborazione dei risultati qualitativi può sfociare in una mappa cognitiva (simile a quella vista in precedenza), ovvero può riguardare l'analisi delle parole e delle reazioni ricorrenti.

Le interviste personali e il focus group

Le tecniche qualitative presentate possono essere applicate nell'ambito di interviste personali, oppure di uno o più focus group. Nella tabella che segue sono riportati le caratteristiche e i vantaggi che caratterizzano i due strumenti.

NB: *il focus group non è una somma di interviste, ma una discussione di gruppo deputata a far emergere le dinamiche sociali che riguardano l'acquisto e il consumo del prodotto.*

Interviste Personali	Focus Group
interviste a singoli soggetti	discussione di gruppo (6-10 persone)
maggior controllo da parte dell'intervistatore	ruolo di mediazione del conduttore
dati facilmente analizzabili	flessibilità
facilità logistica	stimola l'interazione sociale
minori problemi di inibizione	costi tendenzialmente ridotti

Nelle interviste personali (durata media: 60 – 90 minuti) vengono applicati soprattutto il laddering – nelle sue varie forme –, con l'obiettivo di ricostruire la catena mezzi-fini del soggetto intervistato, la critical incident technique e i test proiettivi.

Nell'ambito dei focus group (durata media: 2 – 3 ore) possono essere applicate le varie tecniche proposte, coinvolgendo l'intero gruppo di intervistati.

In entrambi i casi è corretto preparare dei dispositivi di intervista, non delle domande. L'intervistatore o il conduttore devono fornire stimoli, non porre delle domande troppo nette, posti gli obiettivi di profondità dei risultati.

La fase quantitativa

La fase quantitativa si pone l'obiettivo di generalizzare dei risultati ottenuti in un campione alla popolazione di riferimento. Le ipotesi formulate dopo le indagini esplorative e qualitative vengono sottoposte al test di significatività attraverso l'applicazione di tecniche quantitative. I dati su cui è possibile applicare tecniche quantitative sono di natura:

- *qualitativa*
  - nominali (non numerici, né codificabili; ad esempio: città, provincia, ecc.);
  - ordinale (è possibile dire se A è <, > o = rispetto a B; ad esempio: "poche telefonate", "tante telefonate");
- *quantitativa*
  - a intervallo (scale da 1 a n, con approssimazione della proprietà di continuità; NB: in una scala da 1 a 7 non è vero che 2 è il doppio di 4, posto la continuità non reale)
  - a rapporto (dati oggettivi; ad esempio: litri di benzina, chili di pasta; NB: continuità reale, è "vero" che 2 è il doppio di 4)

Attenzione a non confondere la distinzione tra tecniche qualitative (ad esempio il laddering) e quantitative (ad esempio il calcolo delle medie) con quella tra dati qualitativi e dati quantitativi.

Gli item del questionario, volti a raccogliere dati qualitativi sono caratterizzati da alternative di risposta (univoche o con possibilità di più risposte), comunque, seppur codificabili, non riconducibili a numeri. Ad esempio, l'item "Qual è il principale gestore di telefonia mobile da cui acquisti i servizi?" ha come possibili risposte 4 alternative codificabili, ma di natura qualitativa.

Le scale di misurazione quantitative che è possibile utilizzare sono riconducibili alle seguenti forme:

- valutazioni di performance, da 1 (estremamente negativo) a 7 (estremamente positivo);
- valutazioni dello scostamento tra performance e aspettative, da 1 (molto peggio di quanto mi aspettassi) a 7 (molto meglio di quanto mi aspettassi);
- scala di Likert, con cui si richiede il grado di accordo/disaccordo rispetto ad un'affermazione, da 1 (decisamente in disaccordo) a 7 (decisamente d'accordo);

- differenziale semantico, con cui si pongono agli estremi delle scale (da 1 a 7) due aggettivi opposti; ad esempio, "il rapporto con gli addetti al call-center di Tim è *Conflittuale (1) vs Amichevole (7)".*

Oltre alle scale di misurazione a sette punti, sono anche molto diffuse quelle a 5 e a 9 punti. Le scale a sette punti rappresentano un buon compromesso tra le esigenze di semplicità e di approssimazione della proprietà della continuità.

Le tecniche quantitative, sulla base del numero di variabili[2] analizzate simultaneamente, possono essere classificate in:

- univariate;
- bivariate;
- multivariate.

Tali analisi possono essere effettuate tramite il software SPSS. In merito è opportuno segnalare che al fine di leggere l'output SPSS anche su un computer in cui non è installato il programma, occorre incollarlo su un documento Word (sul foglio SPSS selezionare le tabelle con i risultati con Modifica/seleziona tutto/modifica/copia oggetti. Aprire un foglio di Word e incollare l'output).

Qualora ci sia una tabella con molte colonne (su Word apparirebbe molto piccola) è opportuno trasporre le righe con le colonne (doppio clic sulla tabella da trasporre in SPSS. Sul menu principale sarà comparso il comando Pivot, selezionarlo e scegliere trasponi righe e colonne). Qualora ci sia una tabella con molte righe è opportuno incollarla su un foglio Excel che non presenta i problemi di visualizzazione che potrebbero presentarsi in questi casi su Word.

Le analisi univariate

Le analisi univariate vengono effettuate sulle singole variabili. È possibile distinguere le tecniche da applicare sulla base della natura dei dati a disposizione.

[2] Si ricorda che una variabile è tale in quanto rappresenta un evento che cambia tra unità di osservazione (e.g., individui, tempo); ad esempio sono variabili la temperatura, i chili di gelato consumati in Italia, i gol realizzati in una giornata di campionato; viceversa le costanti rappresentano eventi immutabili (ad esempio i giorni del mese di marzo sono sempre 31).

Dati qualitativi: tali tipi di dati permettono esclusivamente di calcolarne le distribuzioni di frequenza. Ad esempio, è possibile *contare* quanti tra gli intervistati abitano a Cosenza, quanti a Rende, ecc. ecc. Si calcoleranno quindi le frequenze assolute, le frequenze relative, le frequenze cumulate (solo per dati ordinali), le percentuali dei dati validi. ATTENZIONE! Su questi tipi di dati non è corretto calcolare la media aritmetica e la deviazione standard.

Ancora, è possibile descrivere sinteticamente i risultati con i *grafici*. È corretto rappresentare distribuzioni di frequenza con poche categorie con il grafico a torta (ad esempio: "gestore di telefonia mobile da cui si acquista i servizi", esistono poche categorie); altresì è possibile rappresentare tali distribuzioni con grafici a barre o tramite istogrammi (se le modalità sono espresse in classi di valori).

Dati quantitativi: su questa tipologia di dati è consigliabile calcolare, al posto della distribuzione di frequenza, degli indici sintetici. Tra gli *indici di posizione centrale* è richiesto il calcolo, per ogni variabile quantitativa, della media e della mediana[3]. Circa gli indici di variabilità deve essere calcolata, per ogni variabile quantitativa, la deviazione standard.

Le analisi quantitative sono facilmente gestibili tramite il pacchetto SPSS (anche Excel può essere utilizzato per questo livello di analisi). Dal menù *Statistica* (o *Analizza*) bisogna selezionare il sotto-menù *Riassumi*, che permette alla voce "Frequenze" di calcolare le distribuzioni di frequenza e di ottenere i relativi grafici e alla voce "Descrittive" di calcolare medie, mediane e deviazione standard.

Per valutare in termini comparativi la variabilità delle variabili è possibile calcolare il *coefficiente di variazione* tramite il rapporto tra la deviazione standard e la media (su Excel o su una semplice calcolatrice); valori sotto lo 0,5 denotano bassa variabilità; valori tra 0,5 e 1 indicano una variabilità media; valori che tendono a 2 indicano una variabilità molto alta.

N.B.: le analisi univariate devono essere applicate su tutte le variabili osservate, avendo cura di utilizzare per i dati qualitativi solo le distribuzioni di frequenza e i grafici, mentre per i dati quantitativi si farà ricorso solo alle medie, alle mediane e alle deviazioni standard.

Tecniche univariate per il posizionamento

L'indice di Fishbein

[3] Per le implicazioni di tali analisi si rinvia alle esercitazioni; utili approfondimenti possono essere riscontrati in qualsiasi buon libro di statistica descrittiva.

L'indice di Fishbein consiste in una misura sintetica calcolata come media ponderata delle valutazioni di performance di una marca/prodotto rispetto ad una serie di *n* attributi/benefici, con questi ultimi moderati da un peso d'importanza. Quindi:

$$\text{Indice di Fishbein della marca "Y"} = \Sigma_i \ (P_i \times I_i) \quad \text{con i = da 1 a "n"}$$

È possibile calcolare per ogni marca/prodotto esaminata un indice di Fishbein, per definire le posizioni competitive nelle percezioni dei consumatori.

Per il calcolo dell'indice di Fishbein sono necessari, come dati in input, le valutazioni di performance medie di ogni marca/prodotto su ogni attributo/beneficio, oltre che i pesi di importanza per questi ultimi su un totale prefissato. Di seguito, viene presentato un esempio di calcolo dell'indice di Fishbein rispetto ad una serie di marche di produttori di un software di grafica:

Attributi rilevanti	Importanza degli attributi	Performance "NetK"	Performance "WebY"	Performance "InterX"	Performance "E-Z"
Semplicità	0,10	70	60	50	30
Flessibilità	0,30	30	80	90	60
Completezza	0,25	40	80	90	90
Aggiornamenti on-line	0,25	60	80	70	90
Tempi di consegna	0,10	90	70	60	80
Indice di Fishbein		50	77	78	74

dove, a mo' di esempio, l'indice di Fishbein dell'impresa NetK è calcolato come (0,10*70)+(0,30*30)+(0,25*40)+(0,25*60)+(0,10*90) = 50

I dati in input raccolti per il calcolo dell'indice possono essere utilizzati anche per l'applicazione della *quadrant analysis*, consistente nella rappresentazione sugli assi cartesiani per ogni marca/prodotto, rispetto al principale concorrente, delle dimensioni "differenziale di performance" (Performance Impresa "X" – Performance Impresa "Y") e "importanza degli attributi"; si identificheranno, incrociando tali dimensioni e definendo i punteggi, i quadranti di "eccellenza" (differenziale positivo, importanza dell'attributo sopra la media), "critico" (differenziale negativo, importanza dell'attributo sopra la media), di

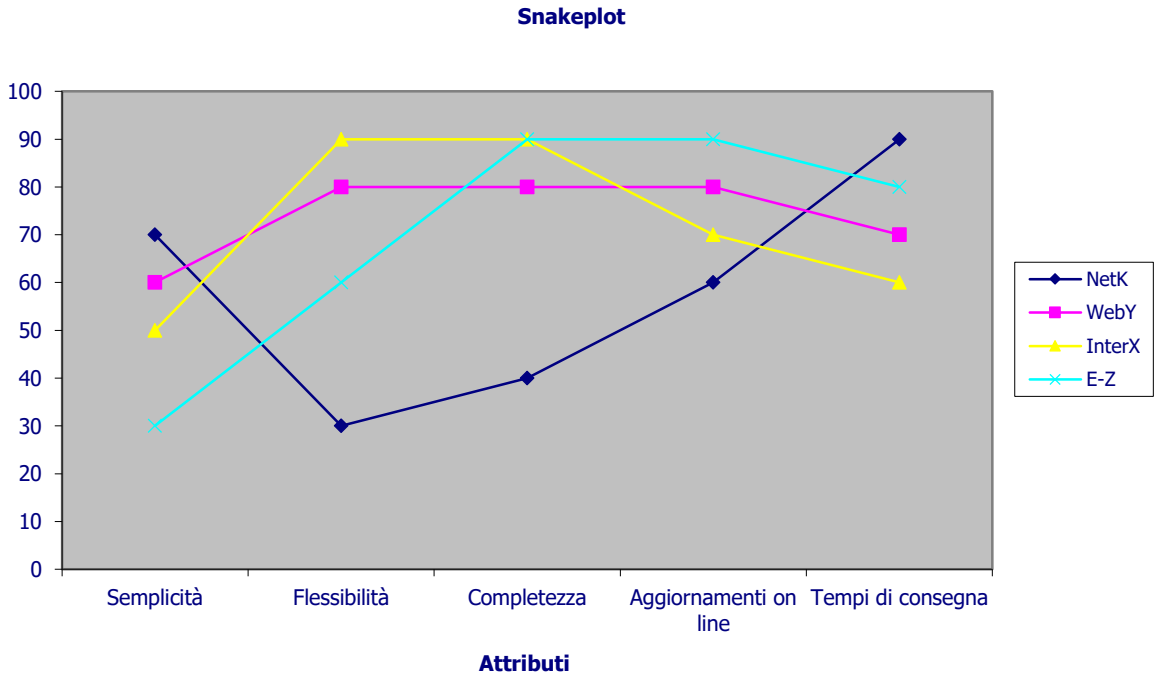
"debolezza" (differenziale negativo, importanza dell'attributo sotto la media), e "relativo" (differenziale positivo, importanza dell'attributo sotto la media).

Lo snakeplot

Lo snakeplot è una tecnica che prevede la rappresentazione grafica delle performance medie delle marche/prodotti esaminate rispetto a una serie di attributi/benefici. I dati in input sono quindi costituiti da tali valutazioni medie. Lo snakeplot prevede la determinazione sugli assi delle posizioni ottenute dalle marche/prodotti esaminate; sull'asse delle ascisse saranno posizionati gli attributi/benefici, sull'asse delle ordinate sarà posta la scala di misurazione delle valutazioni (ad esempio, da 1 a 7).

Identificati i punti, le valutazioni della marca "Y" saranno congiunti da una linea spezzata colorata che rappresenta lo snakeplot della detta marca (il grafico è facilmente costruibile su Excel).

Si propone di seguito un esempio, utilizzando come dati in input quelli dell'esempio precedente.



Coloro che applicano come tecnica multivariata la discriminant analysis (vedi oltre), hanno automaticamente i dati in input per applicare anche lo snakeplot.

N.B.: l'indice di Fishbein e lo snakeplot sono tecniche univariate perché, pur considerando più attributi, questi partecipano ai calcoli presi singolarmente, senza considerare, quindi, le interazioni tra gli attributi.

Le analisi bivariate

Le analisi bivariate consistono nell'analisi delle relazioni esistenti tra *coppie* di variabili. Considerando la possibile duplice natura dei dati (qualitativi e quantitativi) è necessario individuare in una matrice 2x2 le tecniche da applicare:

		Variabile 2	
		Qualitativa	Quantitativa
Variabile 1	Qualitativa	Cross-Tab e analisi di connessione	t-test o One-way ANOVA
	Quantitativa	t-test o One-way ANOVA	Correlazione lineare

Cross-tab e analisi di connessione

Le cross-tab sono delle tabelle a doppia entrata in cui vengono rappresentate le frequenze incrociate considerando due variabili qualitative. Ad esempio, è possibile incrociare le frequenze riscontrate sulle domande "Da quale gestore acquista in prevalenza i servizi di telefonia mobile?" (variabile qualitativa, con modalità 1=Tim, 2=Vodafone, 3=Wind, 4=Tre) e "Ha pensato di cambiare gestore?" (variabile qualitativa, con modalità 1=Si, 2=No).

principale gestore * pensato di cambiare gestore
Crosstabulation

Count

		pensato di cambiare gestore		
		si	no	Total
principale gestore	tim	55	160	215
	vodafone	49	148	197
	wind	17	32	49
	tre		4	4
Total		121	344	465

Le cross tabulation possono essere facilmente calcolate tramite SPSS; dal menù *Statistica* (o *Analizza*) bisogna selezionare il sotto-menù *Riassumi*, che permette, alla voce "Tabelle di contingenza", di calcolare le cross tab. Cliccando sul pulsante "Statistiche" è possibile richiedere il calcolo del Test del chi-quadrato e del Phi di Cramer. Il primo esprime *la significatività* della relazione tra le due variabili (nell'esempio, il gestore da cui si acquista e la decisione di cambiare sono variabili connesse?): se il valore della significatività (p-value, che SPSS chiama "Sig.") del test è inferiore allo 0,05 si dovrà concludere che le variabili sono connesse, altrimenti si accerterà che le variabili sono indipendenti una dall'altra. Se - e solo se - il test del chi-quadrato risulterà significativo si osserverà il valore del Phi/V di Cramer (ignorandone il segno) che segnala *l'intensità* della relazione: valori sotto lo 0,1 indicano una relazione debole, valori tra 0,1 e 0,25 segnalano una relazione di media intensità, valori sopra lo 0,25 caratterizzano relazioni molto intense.

Analisi di correlazione lineare

L'analisi di correlazione si applica a coppie di variabili quantitative; i dati devono essere quindi assolutamente numerici, anche se con unità di misura diverse. Non è corretto applicare l'analisi di correlazione lineare a dati qualitativi codificati in ordine crescente (ad esempio, 1=poco, 2=medio, 3=tanto): è necessario utilizzare scale di misura ad almeno 5 punti con intervalli equivalenti.

Un esempio di applicazione della correlazione lineare riguarda la verifica della relazione tra il tempo da cui si è clienti di un gestore (espresso in anni) ed il grado di soddisfazione (espresso su una scala da 1 a 7).

Correlations

		tempo di utilizzo	soddisfazione
tempo di utilizzo	Pearson Correlation	1,000	,201**
	Sig. (2-tailed)	,	,003
	N	219	219
soddisfazione	Pearson Correlation	,201**	1,000
	Sig. (2-tailed)	,003	,
	N	219	219

**. Correlation is significant at the 0.01 level (2-tailed).

L'indice di correlazione lineare può assumere valori compresi tra –1 (perfetta correlazione negativa) ed 1 (perfetta correlazione positiva). I valori che tendono allo 0 indicano correlazione lineare nulla. In merito alla significatività dell'indice di correlazione sono

richiesti, anche in questo caso, valori di significatività inferiori allo 0,05. Se si ottiene una correlazione non significativa (sig. > 0,05), anche se elevata, non è corretto assumere che questa sia diversa da zero nella popolazione.

Su SPSS le correlazioni lineari possono essere calcolate selezionando dal menu *Statistica* (o *Analizza*) il sotto-menù "Correlazione" alla voce "Bivariata".

Analisi della varianza a una via (one-way ANOVA)

La one-way ANOVA consiste nella verifica della significatività delle differenze tra le medie di una variabile quantitativa nei gruppi di osservazione definiti da una variabile qualitativa. In particolare, la variabile qualitativa deve prevedere almeno tre categorie (se la variabile qualitativa forma 2 gruppi si usa il t-test, che si basa sulla stessa logica della one-way ANOVA). Consideriamo, ad esempio, come variabile quantitativa il grado di soddisfazione (espresso su una scala da 1 a 7), valutando le differenze tra le medie nei gruppi identificati dalla variabile qualitativa "gestore da cui si acquistano in prevalenza i servizi di telefonia mobile" (i gruppi sono 1=Tim, 2=Vodafone, 3=Wind, 4=Tre). L'analisi della varianza esprime, quindi, la significatività delle eventuali differenze tra le medie del grado di soddisfazione dei clienti di Tim, Vodafone, Wind e Tre. SPSS permette di applicare l'analisi della varianza, selezionando dal menù *Statistica* (o *Analizza*) il sotto-menù "*Confronta Medie"*, alla voce "ANOVA". Come variabile dipendente va selezionata la variabile quantitativa (nell'esempio, la soddisfazione), mentre come fattore (o variabile indipendente) va selezionata la variabile qualitativa di raggruppamento (nell'esempio, il gestore). Dal sotto-menù "Post-hoc" è preferibile selezionare i test di Bonferroni, mentre dal sotto-menù opzioni è necessario richiedere le statistiche descrittive e i grafici delle medie.

Nell'output (si veda l'esempio in basso), bisogna focalizzare l'attenzione su:

- le medie dei gruppi (nell'esempio, la soddisfazione per ogni gestore);
- il ratio F e sul relativo valore di significatività: un valore di significatività minore di 0,05 indica che esistono significative differenze tra le medie dei gruppi; questo test ha natura generale;
- i test di Bonferroni, che includono i test della significatività delle differenze tra le medie per ogni possibile coppia di modalità (ad esempio: Tim vs Vodafone, Tim vs Wind, ecc.); anche in questo caso, un valore di significatività minore di 0,05 indica che esistono significative differenze tra le medie dei gruppi in questione.

Descriptives

soddisfazione

	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimum	Maximum
					Lower Bound	Upper Bound		
tim	166	4,9699	1,1304	8,77E-02	4,7966	5,1431	1,00	7,00
vodafone	222	5,0225	1,1153	7,49E-02	4,8750	5,1700	1,00	7,00
wind	67	4,6269	1,0421	,1273	4,3727	4,8810	2,00	7,00
tre	6	5,6667	1,5055	,6146	4,0867	7,2466	4,00	7,00
Total	461	4,9544	1,1229	5,23E-02	4,8517	5,0572	1,00	7,00

ANOVA

soddisfazione

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	11,302	3	3,767	3,027	,029
Within Groups	568,742	457	1,245		
Total	580,043	460			

Multiple Comparisons

Dependent Variable: soddisfazione
Tukey HSD

(I) principale gestore	(J) principale gestore	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
tim	vodafone	-5,2643E-02	,1145	,968	-,3467	,2414
	wind	,3430	,1615	,145	-7,1802E-02	,7578
	tre	-,6968	,4636	,436	-1,8878	,4942
vodafone	tim	5,264E-02	,1145	,968	-,2414	,3467
	wind	,3957	,1555	,053	-3,8316E-03	,7951
	tre	-,6441	,4615	,502	-1,8299	,5416
wind	tim	-,3430	,1615	,145	-,7578	7,180E-02
	vodafone	-,3957	,1555	,053	-,7951	3,832E-03
	tre	-1,0398	,4754	,127	-2,2611	,1815
tre	tim	,6968	,4636	,436	-,4942	1,8878
	vodafone	,6441	,4615	,502	-,5416	1,8299
	wind	1,0398	,4754	,127	-,1815	2,2611

Attenzione! Si rinvia alle slide del corso per una più dettagliata discussione dei test di differenze tra medie in caso di violazione dell'assunzione di varianze uguali nei gruppi.

N.B.: le analisi bivariate devono essere presentate nel report in riferimento non a tutte le possibili coppie di variabili, ma rispetto a coppie di variabili per le quali sia presumibile o interessante l'esistenza di un legame associativo che si vuole sottoporre a verifica.

Le analisi multivariate

Le tecniche di statistica multivariata permettono di effettuare analisi complesse, funzionali al perseguimento di specifici obiettivi informativi per il marketing management. Le sette tecniche multivariate, che verranno presentate di seguito, sono riconducibili a obiettivi espressi nella tabella sinottica di pag. 9.

La Factor Analysis

La Factor Analysis si pone l'obiettivo di ridurre i dati, sintetizzando l'informazione contenuta in molte variabili in pochi fattori sintetici. Lo scopo è quello di identificare una struttura più astratta sottostante ad un insieme di variabili osservate.

Il suo uso implica lo studio delle correlazioni tra le variabili, con l'obiettivo di trovare un nuovo insieme di dimensioni, i *fattori*, meno numeroso rispetto a quello delle *variabili originarie*. I fattori devono esprimere ciò che le variabili originarie hanno in comune, perdendo il minor numero di informazioni rilevanti.

La Factor Analysis consente, infatti, di studiare le correlazioni di un elevato numero di variabili (osservate, "note"), raggruppandole intorno ai fattori (non osservabili, "nascosti"). Le variabili osservate che risultano altamente correlate tra loro tenderanno a essere rappresentate da un fattore. La factor analysis, infatti, produce in output i cosiddetti coefficienti fattoriali, che indicano quanto ogni variabile osservata è spiegata dai fattori. Ogni fattore viene interpretato sulla base delle variabili che mostrano i coefficienti fattoriali più alti rispetto a esso. Il ricercatore "battezzerà" i fattori sulla base di tali variabili osservate con coefficienti fattoriali più alti.

Le applicazioni di marketing più ricorrenti:

- *Uso esplorativo:* F.A. per ridurre il numero di variabili originarie al fine di facilitare la loro lettura ed interpretazione, perdendo la minor quantità possibile di informazione;
- *Per analisi di posizionamento* attribute-based (per analisi di posizionamento attribute-based, la tecnica più indicata è la Discriminant Analysis, in quanto l'algoritmo che utilizza consente di cogliere informazioni rilevanti che una Factor non è in grado di cogliere);
- *Base preliminare* per l'utilizzazione di altre tecniche (soprattutto cluster analysis e regressione lineare) al fine di ridurre il numero di variabili da analizzare.

I dati da raccogliere come input dell'analisi fattoriale, attraverso il questionario, devono essere di natura *quantitativa;* in particolare, è opportuno raccogliere valutazioni su scala metrica a 7 o 9 punti (es. valutazioni di importanza dei diversi attributi di un certo tipo di prodotto nel processo di scelta). Ad esempio:

Potrebbe indicare in che misura i seguenti attributi sono importanti nelle sue valutazioni del servizio?

	Per niente importante						Molto importante
1. Qualità della trasmissione	1	2	3	4	5	6	7
2. Copertura del territorio	1	2	3	4	5	6	7
3. Varietà delle tariffe	1	2	3	4	5	6	7
4. Convenienza delle offerte promozionali	1	2	3	4	5	6	7
5. Comprensione delle esigenze dei clienti	1	2	3	4	5	6	7
6. Innovazione ed offerta di nuovi servizi	1	2	3	4	5	6	7
7. Risoluzione tempestiva di qualsiasi problema	1	2	3	4	5	6	7
8. Cortesia e competenza dell'assistenza offerta ai clienti	1	2	3	4	5	6	7
9. Chiarezza della comunicazione ai clienti	1	2	3	4	5	6	7
10. Accuratezza delle informazioni fornite	1	2	3	4	5	6	7
11. Altro (*specificare*) _____	1	2	3	4	5	6	7

Il data entry per la factor analysis

Su un documento SPSS (o in alternativa su un foglio excel da importare in SPSS al momento delle analisi) si costruisce una tabella con M righe (M = n° di intervistati) ed N colonne (N = n° di variabili + 1). Ogni intervistato genera una riga (se gli intervistati sono 100, le righe saranno 100). La prima colonna è nominata "soggetti" (oppure "osservazioni") e serve a contare i rispondenti (cioè le righe). Ogni variabile indagata (attributo, beneficio, ecc...) genera poi una colonna (se le variabili sono 20, le colonne, in totale, saranno 21). In ogni cella si inserisce il numero corrispondente alla valutazione fatta dagli intervistati per ogni variabile.

L'applicazione con SPSS: i comandi

Aprire il file editor dei dati (la tabella a doppia entrata contenente i dati) quindi:

1. Selezionare STATISTICA (o ANALIZZA) dal menu principale;
2. Selezionare il comando RIDUZIONE DATI ed il sottocomando FATTORIALE;
3. Specificare nella finestra di dialogo risultante le VARIABILI da analizzare (es.: gli attributi);
4. Selezionare il sottocomando DESCRITTIVE e scegliere, nella finestra di dialogo aperta, le voci: SOLUZIONE INIZIALE (si chiede cioè di visualizzare le informazioni iniziali sull'analisi, come gli autovalori e la percentuale di variabilità spiegata da tutti i fattori potenzialmente ottenibili); COEFFICIENTI (si richiedono le correlazioni tra le variabili) e TEST KMO E TEST DI SFERICITÀ DI BARTLETT;
5. Selezionare il sottocomando ESTRAZIONE e scegliere: METODO - COMPONENTI PRINCIPALI (come metodo di estrazione dei fattori); ESTRAI - AUTOVALORI MAGGIORI DI 1 (si chiede cioè di determinare il numero di fattori utilizzando il criterio degli autovalori maggiori di 1); VISUALIZZA - SOLUZIONE FATTORIALE NON RUOTATA (si chiede di visualizzare la soluzione ottenuta con l'analisi fattoriale);
6. Selezionare il sottocomando ROTAZIONE e scegliere METODO VARIMAX; alzare il NUMERO MASSIMO DI ITERAZIONI almeno a 100;
7. Selezionare il sottocomando OPZIONI e selezionare VALORI MANCANTI – SOSTITUISCI CON LA MEDIA, VISUALIZZAZIONE COEFFICIENTI ORDINATI PER DIMENSIONE e SOPPRIMI VALORI ASSOLUTI INFERIORI A 0,30.

Tali operazioni dovranno essere effettuate solo durante la prima analisi; i tentativi successivi – sulla base degli output, vedi oltre – potranno essere lanciati semplicemente aggiornando le variabili in input; le altre istruzioni saranno confermate automaticamente da SPSS. Quando si giunge alla soluzione ideale è possibile salvare i punteggi fattoriali da utilizzare come input per altre applicazioni. È necessario, a tal fine, selezionare dal sotto-menù OPZIONI il sottocomando PUNTEGGI e scegliere SALVA COME VARIABILI (si chiede cioè di aggiungere nella tabella editor dei dati, ossia dei dati in input, delle nuove variabili contenenti i punteggi fattoriali (input di un'eventuale Cluster Analysis o di una Regressione Lineare), ossia i punteggi che ogni intervistato avrebbe assegnato a ogni fattore (variabile teorica).

Gli output [4]e i test statistici:
Matrice delle correlazioni. Contiene le correlazioni fra tutte le variabili osservate (punto di partenza della factor analysis) e permette di avere un'indicazione sui legami fra le variabili osservate.
Valutare quindi i valori assunti dai 2 indicatori:

Indice KMO (Kaiser-Meyer-Olkin): costruito comparando i coefficienti di correlazione con quelli di correlazione parziale. Questo rapporto varia tra 0 e 1. Valori bassi dell'indice suggeriscono la potenziale inadeguatezza dell'analisi dei fattori, perché le correlazioni fra coppie di variabili non possono essere spiegate dalla varianza condivisa dall'insieme delle variabili (ovvero non possono essere individuati fattori comuni). Kaiser suggerisce che valori al di sopra di 0,7 sono da ritenersi soddisfacenti, mentre valori al di sotto di 0,5 sono sostanzialmente inaccettabili.

Test di sfericità di Bartlett: utilizzato per verificare l'ipotesi che la matrice delle correlazioni sia una matrice identità (con 1 sulla diagonale principale e 0 altrove), ossia che le variabili siano indipendenti. Valori bassi di questo test, e di conseguenza valori elevati di significatività (maggiori di 0,05), indicano che questa ipotesi non può essere esclusa e che l'utilizzo del modello fattoriale potrebbe non essere adeguato.

Dai successivi output, si possono ottenere indicazioni sulla opportunità di "purificazione", cioè di eliminare alcune delle variabili in input che non sono spiegate bene dal modello ad "n" fattori e che potrebbero limitarne la bontà descrittiva.

Comunalità. Tabella che indica quanto la varianza di ogni variabile considerata è spiegata dai fattori comuni. Questa misura può variare da un minimo di 0 (i fattori comuni non spiegano niente della variabilità della variabile considerata) ad un massimo di 1 (tutta la variabilità è spiegata dai fattori comuni).
Nel caso della soluzione iniziale, nella quale sono considerati tutti i possibili fattori (tanti quante sono le variabili originarie), le comunanze sono uguali ad 1. Nella soluzione finale le comunanze saranno generalmente minori di 1, in quanto ci sarà sicuramente una perdita di informazione. Si potrebbe valutare, in fase di purificazione delle variabili in input, la possibilità di eliminare dal modello quelle variabili che presentino comunità estratta molto bassa (minore di 0,40) e, successivamente, di rilanciare l'analisi.

Varianza totale spiegata. Sono riportati in tabella: gli autovalori iniziali (indicanti fattore per fattore la varianza complessiva spiegata ed usati per stabilire quali fattori o componenti mantenere nella soluzione); i pesi dei fattori non ruotati (indicano quanta varianza ogni fattore è in grado di spiegare prima della rotazione); i pesi dei fattori ruotati (indicano quanta varianza ogni fattore è in grado di spiegare dopo la rotazione; mediante la rotazione non muta la % cumulata di varianza spiegata, ma migliora la leggibilità dell'output). Da ognuna delle 3 colonne sopra indicate (autovalori iniziali, i pesi dei fattori non ruotati e i pesi dei fattori

[4] Qualora attraverso la Factor Analysis venissero estratti pochi o addirittura un solo fattore, vorrebbe dire che la variabilità del fenomeno risulta spiegata da pochi fattori e quindi che le variabili originarie sono altamente correlate tra loro (non colgono aspetti molto diversi del fenomeno indagato, sono ridondanti). La causa potrebbe anche essere costituita dal fatto che gli intervistati valutino l'intero fenomeno in modo univoco, benché le variabili siano state scelte accuratamente dal ricercatore.

ruotati) è possibile leggere quanto ogni fattore estratto è capace di spiegare la variabilità del fenomeno indagato.
Quanti fattori prendere in considerazione? Premesso che con la "sintesi" effettuata attraverso una factor analysis è chiaramente opportuno perdere il minor numero di informazioni, la scelta cadrà sul numero di fattori che cumulativamente sono in grado di spiegare almeno la metà (50%) della varianza, anche se sono preferibili soglie più alte (60-70%). SPSS seleziona per default i fattori con autovalori maggiori di uno, cioè capaci di spiegare una quantità sufficiente di varianza.

Matrice dei componenti (o fattori). Indica la correlazione tra i fattori estratti e le variabili osservate e permette l'interpretazione dei fattori. Componente per componente si individuano le variabili con un elevato indice di correlazione (si considera solo il valore assoluto, poiché interessa l'entità, ma non il "verso" della correlazione) ossia un indice maggiore di 0,5. In questa soluzione, però, vengono esposte le correlazioni in ordine gerarchico: il primo fattore sarà quindi quello che "attirerà" la maggior parte dei coefficienti espressivi delle correlazioni tra le variabili e lo stesso fattore. Per avere una visione più chiara e per potere definire il nome di ognuno dei fattori è preferibile analizzare la matrice dei componenti ruotata, molto più leggibile e utile a fini interpretativi.

Matrice dei componenti ruotata. Si tratta della matrice dei componenti sopra esaminata ruotata al fine di migliorare l'interpretazione dei fattori. Dopo la rotazione, infatti, è generalmente più marcata la correlazione di una variabile rispetto ad uno ed un solo fattore. La stessa variabile rispetto agli altri fattori estratti presenterà invece indici di correlazione più contenuti. Al fine di dare un nome ai fattori estratti risulta dunque più agevole da consultare la matrice dei componenti ruotata. Se, ad esempio, su un fattore i coefficienti più alti sono quelli relativi alle variabili "prezzo" e "consumi" di un'automobile quel fattore potrà essere definito "convenienza".
In sede di purificazione, si potranno eliminare dalle analisi le variabili che non "saturano" su nessun fattore (cioè denotano coefficienti tutti minori di 0,4) oppure che "saturano" su più fattori in modo ambiguo (ad esempio, una variabile con coefficiente 0,6 su un fattore e 0,5 su un altro). Nei successivi tentativi è opportuno rilanciare l'analisi escludendo una variabile per volta. Basta infatti che il modello cambi di una sola variabile per essere potenzialmente diverso nei risultati.

L'analisi di regressione lineare
L'analisi di regressione lineare multipla ha l'obiettivo di descrivere la relazione di causalità (dipendenza) fra variabili; una di queste è ipotizzata essere dipendente da più variabili ipotizzate come indipendenti.
Con le tecniche di regressione, quindi, si tenta di descrivere in termini analitici (ricorrendo a una funzione) la relazione che esiste tra le variabili quantitative considerate, tentando di spiegare e predire la variabilità della variabile dipendente (quale valore ci possiamo aspettare nella variabile dipendente dati certi valori delle variabili indipendenti?). La funzione esplicativa alla quale ricorre l'analisi di regressione lineare è la funzione retta (interpolazione statistica).
I coefficienti della retta di regressione, calcolati mediante l'analisi di regressione, descrivono quanto le singole variabili indipendenti (un coefficiente per ogni variabile indipendente) spiegano la variabilità della variabile dipendente.

In termini analitici:

$Y = a + b_1X_1 + b_2X_2 + \ldots + b_kX_k + e_i$ (forma non standardizzata)

$Y = \beta_1X_1 + \beta_2X_2 + \ldots + \beta_kX_k + e_i$ (forma standardizzata)

I parametri nella forma non standardizzata esprimono di quanto varia la variabile dipendente all'aumentare di un'unità della variabile indipendente, *poste costanti le altre variabili indipendenti*. Nella forma standardizzata (in cui tutte le variabili sono espresse come differenza dalla media e divise per la loro deviazione standard: in tal modo si eliminano le diverse unità di misura) non esiste la costante "a". Se le variabili indipendenti sono misurate con scale diverse o con indicatori percettivi (come le scale di misurazione della soddisfazione) è preferibile utilizzare i parametri standardizzati. In questo caso, i parametri standardizzati indicano di quante deviazioni standard cambia la variabile dipendente all'aumentare di *una deviazione standard della variabile indipendente*, *poste costanti le altre variabili indipendenti*.

Le applicazioni di marketing più ricorrenti:

- misurazione della customer satisfaction e analisi delle sue determinanti;
- analisi della dipendenza di una variabile o un fenomeno (es.: vendite di un prodotto, la fiducia negli addetti di vendite) da altre variabili (es.: prezzo, percezione di affidabilità).

I dati da raccogliere come input attraverso il questionario da utilizzare nella fase di rilevazione sono valutazioni su scala metrica a sette (o a nove) punti delle variabili indagate (dipendente e indipendenti) in termini, per esempio, di performance globale (variabile dipendente) e di performance per specifici attributi (variabili indipendenti); oppure in termini di soddisfazione globale (variabile dipendente) e di conferma/disconferma delle aspettative per specifici attributi (i gap tra le performance e le aspettative rappresentano le variabili indipendenti).

Un esempio – variabili indipendenti:
Potrebbe indicare in che misura le prestazioni ricevute sono state al di sotto, in linea o al di sopra delle sue aspettative per ognuno dei seguenti aspetti caratterizzanti l'offerta di questo gestore dei servizi di telefonia mobile?

	Al di sotto delle aspettative		In linea con le aspettative			Al di sopra delle aspettative	
1. Qualità della trasmissione	1	2	3	4	5	6	7
2. Copertura del territorio	1	2	3	4	5	6	7
3. Varietà delle tariffe	1	2	3	4	5	6	7
4. Convenienza delle offerte promozionali	1	2	3	4	5	6	7
5. Comprensione delle esigenze dei clienti	1	2	3	4	5	6	7
6. Innovazione ed offerta di nuovi servizi	1	2	3	4	5	6	7
7. Risoluzione tempestiva di qualsiasi problema	1	2	3	4	5	6	7
8. Cortesia e competenza dell'assistenza offerta a clienti	1	2	3	4	5	6	7
9. Chiarezza della comunicazione ai clienti	1	2	3	4	5	6	7
10. Accuratezza delle informazioni fornite	1	2	3	4	5	6	7
11. Altro (*specificare*) _____	1	2	3	4	5	6	7

Variabile dipendente:
Complessivamente, quanto è soddisfatto dei servizi ricevuti dal suo principale gestore di telefonia mobile?

Estremamente insoddisfatto			Ne' soddisfatto, ne' insoddisfatto			Estremamente soddisfatto
1	2	3	4	5	6	7

Il data entry per l'analisi di regressione lineare
Su un documento SPSS (o in alternativa su un foglio excel da importare in SPSS al momento dell'analisi) si costruisce una tabella con M righe (M = n° di intervistati) ed N colonne (N = n° di variabili + 1). La prima colonna è nominata "soggetti" (oppure "osservazioni") e serve a contare i rispondenti (cioè le righe). Ogni variabile indagata (attributo, beneficio, ecc.) genera poi una colonna (se le variabili sono 20, le colonne, in totale, saranno 21). In ogni cella si inserisce il numero contrassegnato dall'intervistato (valutazione).

L'applicazione con SPSS: i comandi
Aprire il file editor dei dati (la tabella M*N contenente i dati) quindi:
1. Selezionare STATISTICA (o ANALIZZA) dal menu principale;
2. Selezionare il comando REGRESSIONE ed il sottocomando LINEARE;
3. Specificare nella finestra di dialogo risultante la VARIABILE DIPENDENTE (es.: la soddisfazione globale) e le VARIABILI INDIPENDENTI (es.: le conferme/disconferme delle aspettative sugli specifici attributi);
4. Selezionare il sottocomando OPZIONI e scegliere, nella finestra di dialogo aperta, la voce: SOSTITUISCI (I VALORI MANCANTI) CON LA MEDIA (si chiede cioè di inserire al posto dei valori mancanti, ossia delle valutazioni non effettuate dagli intervistati, la media calcolata sugli altri valori raccolti). Mantenere gli altri elementi delle opzioni in posizione di default.

Gli output ed i test statistici:
Focalizzare l'attenzione su:
Riepilogo del modello, per verificare che il modello di regressione lineare, ossia la relazione di dipendenza ipotizzata (teoricamente) tra la variabile dipendente e le variabili indipendenti, sia verificata, ossia per accertare che vi sia una dipendenza effettiva tra le variabili indagate.

Leggere il valore che assume R^2. R^2 è una misura della bontà di adattamento di un modello lineare, è detto anche coefficiente di determinazione. Indica la proporzione di variabilità della variabile dipendente spiegata dal modello di regressione. Può variare tra 0 e 1. Valori inferiori a 0,50 indicano che il modello non si adatta bene ai dati, dal momento che spiega meno della metà della varianza. Ad ogni modo il contesto di analisi può influenzare la soglia di R^2 da assumere come limite. Se si misurano percezioni e atteggiamenti (quindi fenomeni soggettivi), è possibile accettare soglie più basse. Certamente, valori inferiori a 0,2 segnalano modelli non coerenti.

Parametri di regressione: descrivono la capacità delle diverse variabili indipendenti di determinare la variabilità della variabile dipendente. La verifica sui parametri di regressione deve considerare due elementi: il valore numerico dei parametri – ossia il *segno* (+/-) della

relazione che lega le rispettive variabili indipendenti alla variabile dipendente e l'*intensità* – e la loro significatività statistica, ovvero la loro proprietà di risultare validi nella popolazione di riferimento, seppur calcolati su un campione ridotto di consumatori. Circa segno e intensità della relazione occorre *leggere il valore che assume ogni BETA.* I coefficienti BETA sono i parametri di regressione standardizzati (i B sono invece i coefficienti di regressione non standardizzati); si tratta dei coefficienti di regressione che si ottengono quando tutte le variabili prese in considerazione vengono espresse in forma standardizzata. La trasformazione di variabili indipendenti in forma standardizzata produce coefficienti più facili da confrontare in quanto, come già detto, le differenze dovute alle diverse o ignote unità di misura (utilizzate per misurare le variabili) vengono eliminate (la standardizzazione consente, cioè, di fare un confronto tra la capacità esplicativa delle diverse variabili indipendenti). Tuttavia, va precisato che i coefficienti BETA non riflettono in senso assoluto l'importanza delle diverse variabili indipendenti poiché essi dipendono anche dalle altre variabili nell'equazione.

Per verificare la significatività dei parametri di regressione occorre *leggere il t-value di ogni parametro.* Se un parametro risulta statisticamente significativo il suo valore può essere esteso a tutta la popolazione di riferimento, sebbene la sua stima derivi dal campione osservato (che, naturalmente, ha dimensioni ridotte rispetto alla popolazione intera). I t-value sono calcolati come rapporto tra il parametro non standardizzato di regressione in esame e il suo errore standard. Per verificarne la significatività deve essere controllato il relativo valore *Sig.*, che esprime, intuitivamente la percentuale di errore che si deve "sopportare" assumendo che il parametro – stimato sui dati campionari – sia corretto e generalizzabile all'intera popolazione. La soglia convenzionalmente fissata per un Sig. accettabile è il 5% (ovvero 0,05). Con valori di Sig. inferiori si assumerà che il parametro osservato sia *significativo;* con valori di Sig. superiori a 0,05 si assumerà il parametro *non significativo*. Più elasticamente, la soglia "critica" può essere fissata al 10%. Le variabili indipendenti che esprimono parametri non significativi non devono essere escluse dal modello. Si rinvia alle lezioni per ulteriori dettagli sui check delle assunzioni del modello.

L'analisi della varianza (ANOVA)

L'ANOVA, ossia l'analisi della varianza, è un modello lineare (della stessa famiglia del modello di regressione lineare multipla) avente come obiettivo principale quello di identificare le fonti della variabilità (varianza) di una variabile dipendente d'interesse. Nello specifico, tale tecnica permette di studiare gli effetti di 2 o più variabili indipendenti (di natura qualitativa e che quindi formano dei gruppi) su una variabile dipendente (di natura quantitativa).
L'ANOVA è tipicamente utilizzata negli studi sperimentali per analizzare gli effetti diretti delle variabili indipendenti e gli effetti di interazione tra le stesse variabili indipendenti sulla variabile dipendente.

Le applicazioni di marketing più ricorrenti:

- analisi degli effetti di variabili manipolabili (tramite la creazione di gruppi sperimentali e di controllo) sull'atteggiamento verso prodotti, marche, comportamenti;

- analisi della dipendenza di una variabile o di un fenomeno (es.: scelta, valutazione, intenzione di acquisto) da altre variabili (es.: prezzo, percezione di affidabilità), nell'ambito di disegni sperimentali.

I dati da raccogliere come input attraverso il questionario da utilizzare nella fase di rilevazione sono valutazioni su scala metrica a sette (o a nove) punti della variabile dipendente.

Un esempio – variabile dipendente:
In che misura le piace la campagna pubblicitaria del prodotto x?

Per niente						Moltissimo
1	2	3	4	5	6	7

Le variabili indipendenti, invece, possono essere manipolate ricorrendo a "scenari" costruiti incrociando le modalità delle variabili indipendenti. Ad esempio, immaginiamo di valutare gli effetti di coinvolgimento del consumatore (prima variabile indipendente, che forma i gruppi basso vs. alto) e tipo di argomento in un avviso pubblicitario (seconda variabile indipendente, che forma i gruppi "basato sulla qualità" vs. "basato sull'immagine"), oltre che della loro interazione, sull'atteggiamento verso uno spot pubblicitario (variabile dipendente). Incrociando le due variabili indipendenti[5], dunque, è possibile identificare 4 differenti condizioni sperimentali (o gruppi): basso coinvolgimento – avviso pubblicitario basato sull'immagine; basso coinvolgimento – avviso pubblicitario basato sulla qualità; alto coinvolgimento – avviso pubblicitario basato sull'immagine; alto coinvolgimento – avviso pubblicitario basato sulla qualità.

Una volta identificati i quattro gruppi, quindi, si definiscono i 4 differenti scenari. Immaginiamo, ad esempio, che il prodotto oggetto di indagine sia un dentifricio: ai soggetti che vengono assegnati al gruppo basso coinvolgimento – avviso pubblicitario basato sulla qualità verrà mostrato un avviso pubblicitario in cui un dentista (stimolo scelto per manipolare la condizione "avviso pubblicitario basato sulla qualità") illustra le peculiarità del dentifricio; successivamente, i rispondenti leggeranno un articolo in cui viene enfatizzata la maggiore rilevanza dello spazzolino rispetto al dentifricio per la cura dei denti (stimolo scelto per manipolare la condizione "basso coinvolgimento"). Ai soggetti che vengono assegnati al gruppo basso coinvolgimento – avviso pubblicitario basato sull'immagine, invece, verrà mostrato un avviso pubblicitario in cui un famoso/a attore/attrice riveste il ruolo di testimonial (stimolo scelto per manipolare la condizione "avviso pubblicitario basato sull'immagine"); successivamente, i rispondenti leggeranno un articolo in cui viene enfatizzata la maggiore rilevanza dello spazzolino rispetto al dentifricio per la cura dei denti (stimolo scelto per manipolare la condizione "basso coinvolgimento"). Ai soggetti che vengono assegnati al terzo scenario, ossia alto coinvolgimento - avviso pubblicitario basato sulla qualità verrà mostrato un avviso pubblicitario in cui un dentista (stimolo scelto per manipolare la

[5] A tal proposito, è opportuno specificare che le variabili indipendenti possono essere manipolate *between-subjects* (ciascun rispondente viene assegnato, in maniera completamente *random*, a una sola condizione sperimentale), oppure *within-subjects* (ciascun rispondente viene sottoposto a tutte le condizioni sperimentali, invece che ad una sola di esse). Fatta questa distinzione, è opportuno precisare che durante il corso si utilizzeranno prevalentemente manipolazioni *between-subjects*.

condizione “avviso pubblicitario basato sulla qualità”) illustra le peculiarità del dentifricio; successivamente, i rispondenti leggeranno un articolo in cui viene enfatizzata la maggiore rilevanza del dentifricio rispetto allo spazzolino per la cura dei denti (stimolo scelto per manipolare la condizione “alto coinvolgimento”). Infine, ai soggetti che vengono assegnati al gruppo alto coinvolgimento – avviso pubblicitario basato sull’immagine verrà mostrato un avviso pubblicitario in cui un famoso/a attore/attrice riveste il ruolo di testimonial (stimolo scelto per manipolare la condizione “avviso pubblicitario basato sull’immagine”); successivamente, i rispondenti leggeranno un articolo in cui viene enfatizzata la maggiore rilevanza del dentifricio rispetto allo spazzolino per la cura dei denti (stimolo scelto per manipolare la condizione “alto coinvolgimento”).

Inoltre è possibile raccogliere dati su scala metrica a sette (o a nove) punti su cosidette “covariate”, cioè altre variabili indipendenti numeriche di cui si vuole controllare gli effetti diretti. In presenza di covariate nel modello si parla più correttamente di ANCOVA (Analisi della Covarianza). Ad esempio, è possibile analizzare/studiare se e come aspetti della personalità di un individuo (es.: la autostima) influenzano la relazione esistente tra la variabile dipendente e le variabili indipendenti.

Un esempio – covariate:

Per favore, potrebbe indicare il suo grado di disaccordo/accordo con le seguenti affermazioni:

	Decisamente in disaccordo						Decisamente d’accordo
In generale, sono soddisfatto di me stesso	1	2	3	4	5	6	7
In generale, penso di avere un certo numero di buone qualità.	1	2	3	4	5	6	7
In generale, ho un atteggiamento positivo verso me stesso.	1	2	3	4	5	6	7

Il data entry per l’ANOVA
Su un documento SPSS (o in alternativa su un foglio Excel da importare in SPSS al momento dell'analisi) si costruisce una tabella con M righe (M = n° di intervistati) ed N colonne (N = n° di variabili + 1). La prima colonna è nominata “soggetti” (oppure “osservazioni”) e serve a contare i rispondenti (cioè le righe). Ogni variabile indagata (attributo, beneficio, ecc.) genera poi una colonna (se le variabili sono 20, le colonne, in totale, saranno 21). In ogni cella si inserisce il numero contrassegnato dall’intervistato (valutazione), ovvero il codice assegnato alle modalità delle variabili indipendenti. Nello specifico, per ciascuna variabile indipendente si dovranno codificare, rispettivamente con i valori 0 e 1, le differenti modalità formate (es.: basso coinvolgimento = 0; alto coinvolgimento = 1).

L'applicazione con SPSS: i comandi
Aprire il file editor dei dati (la tabella M*N contenente i dati) quindi:

1. Selezionare STATISTICA (o ANALIZZA) dal menù principale;
2. Selezionare il comando MODELLO LINEARE GENERALIZZATO ed il sottocomando UNIVARIATA;
3. Specificare nella finestra di dialogo risultante la VARIABILE DIPENDENTE, le VARIABILI INDIPENDENTI e, se previste, le VARIABILI COVARIATE;
4. Selezionare il sottocomando OPZIONI e, nella finestra di dialogo aperta, spostare dalla sezione "fattori e interazioni tra fattori" alla sezione "medie marginali" tutte le variabili presenti. Inoltre, da questa interfaccia bisogna selezionare la voce STATISTICHE DESCRITTIVE;
5. Selezionare il sottocomando GRAFICI e, nella finestra di dialogo aperta, inserire la variabile indipendente più rilevante nella finestra dell'asse orizzontale e la variabile indipendente che si presume possa moderare l'effetto della prima variabile indipendente nella finestra per ottenere linee separate nel grafico e premere AGGIUNGI;
6. In questa fase, inoltre, è richiesta anche l'analisi dei contrasti, che consente di testare le differenze tra le medie di specifici gruppi. Nello specifico, dall'interfaccia dell'ANOVA selezionare il sottocomando INCOLLA: si aprirà una nuova finestra con la sintassi all'interno della quale si dovrà individuare la riga dell'interazione tra le 2 variabili indipendenti;
7. Una volta individuata la riga bisogna inserire manualmente, nella stessa riga e senza lasciare spazi, la seguente sintassi: compare(nome variabile indipendente1);
8. Premere INVIO, copiare e incollare la stessa riga di sintassi [ossia: /EMMEANS=TABLES(nome variabile indipendente 1*nome variabile indipendente 2)compare(nome variabile indipendente 1] e cambiando manualmente il comando compare(nome variabile indipendente 2);
9. Terminati i punti 7 e 8 cliccare sull'icona con il comando PLAY (▸) sulla barra degli strumenti per avviare l'analisi.

Gli output ed i test statistici:
Focalizzare l'attenzione su:
Statistiche descrittive e la significatività dei test F nella tabella ANOVA. Nello specifico, dalla tabella "statistiche descrittive" è possibile osservare il valore medio di ciascun gruppo creato dalle variabili indipendenti (quindi, le medie dei 2 gruppi formati dalla variabile indipendente 1, le medie dei 2 gruppi formati dalla variabile indipendente 2 e le medie dei 4 gruppi formati dall'interazione delle due variabili indipendenti); dalla tabella ANOVA è possibile osservare il valore del test F per verificare se le relazioni tra le variabili considerate sono realmente significative. In particolare, per verificarne la significatività deve essere controllato il relativo valore Sig. dei rispettivi test F per verificare:

- l'effetto diretto della variabile indipendente 1 (che sarà significativo se la differenza tra le medie dei gruppi formati dalla variabile indipendente 1 è elevata);
- l'effetto diretto della variabile indipendente 2 (che sarà significativo se la differenza tra le medie dei gruppi formati dalla variabile indipendente 2 è elevata);

- l'effetto di interazione tra variabile indipendente 1 e variabile indipendente 2 (che sarà significativo se la differenza tra le medie dei gruppi formati da una variabile indipendente dipende dall'altra variabile indipendente).

La soglia convenzionalmente fissata per un Sig. accettabile è il 5% (ovvero 0,05). Con valori di Sig. inferiori a 0,05 si assumerà che la differenza tra le medie sia significativa; con valori superiori di Sig. si assumerà che la differenza tra le medie sia non significativa. Più elasticamente, la soglia "critica" può essere fissata al 10%. Le variabili indipendenti che esprimono test F non significativi non devono essere escluse dal modello.

Analisi dei contrasti (da richiedere tramite uso di sintassi ad hoc). Dall'output dell'analisi dei contrasti è possibile osservare la significatività delle differenze tra le medie di specifici gruppi leggendo il *t-value* relativo a ognuna delle comparazioni tra essi; in particolare, deve essere controllato il relativo valore Sig. La soglia convenzionalmente fissata per un Sig. accettabile è il 5% (ovvero 0,05). Con valori di Sig. inferiori a 0,05 si assumerà che la differenza tra le medie sia statisticamente significativa; con valori superiori di Sig. si assumerà che la differenza tra le medie sia non significativa. Più elasticamente, la soglia "critica" può essere fissata al 10%.

La Conjoint Analysis

La conjoint analysis – o analisi congiunta – ha come obiettivo la determinazione della configurazione ideale del prodotto, tale che risponda alle esigenze specifiche di un segmento di consumatori. Tramite tale tecnica è possibile misurare l'importanza relativa di una serie di fattori critici nella definizione del prodotto (ad esempio, la localizzazione di una sala cinematografica, il livello di prezzo di un paio di scarpe, la marca di un pacco di pasta), inteso come paniere di attributi, e il gradimento dei possibili livelli alternativi dei fattori stessi (ad esempio, la localizzazione della sala cinematografica in città/entro 30 km/entro 60 km; 30/50/80 euro quali possibili prezzi delle scarpe; Barilla/Buitoni/De Cecco quali marche alternative).
La principale caratteristica della conjoint analysis consiste nella richiesta ai rispondenti di espressioni delle loro scelte con modalità simili a quelle adottate nell'ambito dei reali processi d'acquisto, ossia confrontando le diverse caratteristiche di un prodotto e fornendo un giudizio complessivo sulle offerte proposte.

Le applicazioni di marketing più ricorrenti:

- per individuare il profilo di prodotto ideale;
- per individuare il profilo di prodotto ideale di particolari segmenti di mercato e descriverli sulla base delle loro preferenze.

I dati da raccogliere come input nella fase di rilevazione sono costituiti da valutazioni di preferenza – su scala a 9 punti (1 = per niente preferito; 9 = estremamente preferito) – di diverse alternative di prodotto, che vengono rappresentate graficamente e ottenute mediante la definizione del *disegno ortogonale*, tramite il software SPSS; il disegno ortogonale

consiste nella definizione di un numero ridotto di combinazioni di prodotto; tipicamente si cerca di proporre non più di 16 cartoline raffiguranti ognuna una delle possibili combinazioni di prodotto. Tali combinazioni sono estratte rispetto all'intera gamma di possibili configurazioni di prodotto ottenibili combinando, appunto, i diversi livelli dei fattori identificati in sede di ricerca qualitativa.
Ad esempio, è possibile, pensare che per un dentifricio i fattori determinanti delle preferenze possano essere il tipo di *azione*, la *consistenza*, il *materiale del tubetto*, il tipo di *chiusura* e il *prezzo*. Ognuno di questi fattori può avere dei *livelli alternativi:* l'azione potrà essere *anticarie, generica, gengiprotettiva, o sbiancante;* la consistenza potrà prevedere il *gel*, la *pasta* o una forma *mista;* il tubetto sarà in *plastica* o in *metallo;* la chiusura sarà *a scatto*, *avvitata*, *con tappo grande* o *con dispenser;* il prezzo avrà come possibili livelli *1,60 euro*, *2,50 euro* e *3,40 euro*. Posti questi fattori e i relativi livelli sarebbe possibile ricavare 288 combinazioni (quindi 288 possibili configurazioni di prodotto). Tramite il disegno ortogonale è possibile estrarre un numero ridotto di combinazioni che, proposte nella forma di cartoline, possono essere proposte ai consumatori per ottenerne una valutazione globale per ogni alternativa di prodotto.

Per costruire il disegno ortogonale con SPSS è essenziale preparare *un elenco con i fattori e con i relativi livelli*. Nell'esempio proposto in precedenza è possibile notare come i fattori possano essere degli attributi di prodotto, con potenziali diversi livelli di prestazione. È possibile utilizzare come fattori degli elementi più generali i cui livelli saranno degli attributi – anche eterogenei – del prodotto. Ad esempio, se il prodotto in esame è una sala cinematografica, uno dei fattori potrebbe essere "Elementi organizzativi", con livelli alternativi identificati dalla presenza nella struttura di *postazioni Internet*, della possibilità di *prenotare telefonicamente i posti* e dalla presenza della *galleria*.
Definito l'elenco dei fattori e degli attributi è possibile procedere con l'elaborazione del disegno ortogonale sull'elaboratore:

1. Selezionare DATI dal menu principale;
2. Selezionare il comando DISEGNO ORTOGONALE ed il sottocomando GENERA;
3. Specificare nella finestra di dialogo risultante il *nome del primo fattore*, utilizzando massimo 8 lettere (ad esempio, "mat_tub") e quindi *l'etichetta del fattore,* senza alcuna restrizione sul numero di lettere (inserendo nuovamente il nome del fattore, ma per esteso; nell'esempio, "materiale del tubetto") e cliccare su AGGIUNGI;
4. Ripetere l'operazione descritta al passaggio 3 fino all'inserimento di tutti i fattori;
5. Cliccare sul primo fattore inserito e quindi sulla casella DEFINISCI VALORI (che si sarà annerita);
6. Nella finestra di dialogo aperta inserire come *valore* i numeri progressivi (da 1 a N) indicanti i livelli del fattore che si sta definendo (nel nostro esempio, per il fattore "azione" occorre inserire i valori da 1 a 4) e quindi specificare i nomi dei livelli (es. anticarie, generica, gengiprotettiva, sbiancante); cliccare quindi su CONTINUA;
7. Ripetere l'operazione descritta al passaggio 6 fino alla definizione dei livelli di tutti i fattori;
8. Cliccare su FILE; nella finestra di dialogo aperta inserire il nome del file con le combinazioni da salvare (il nome può essere scelto a piacere, mentre è necessario annotare dove lo si salva) e cliccare su SALVA;

9. Chiudere la procedura cliccando su OK; si apre così la finestra di output che comunica la generazione delle combinazioni;
10. Per visualizzare le combinazioni, aprire il file precedentemente salvato, specificando nella finestra FILE/APRI il tipo di file cercato ossia un documento con estensione *.sav* (che rappresenta l'estensione tipica dei file dati di SPSS);
11. Si aprirà un foglio di dati nel quale compariranno nelle righe le combinazioni di prodotto selezionate e nelle colonne i fattori, oltre alla colonna "status" e a quella "card" nella quale è visualizzato il numero progressivo delle card ottenute. La composizione di ogni card si legge nelle celle, nelle quali sono riportati per ogni fattore il livello di attributo che caratterizza la card. Per visualizzare il nome per esteso del livello di attributo cliccare su VISUALIZZA e selezionare ETICHETTE DEI VALORI.

Le combinazioni ottenute dovranno essere presentate ai rispondenti in forma di cartoline, da rendere vivaci, colorate e interessanti; su ogni cartolina – ogni combinazione di prodotto – sarà quindi richiesto un "voto" da 1 a 9. Tutte le card dovranno essere valutate dai rispondenti. Per rendere più agevole la definizione delle valutazioni da parte dei rispondenti, è possibile chiedere all'intervistato di fare una prima suddivisione tra cartoline "buone", "pessime" e "medie", e di fornire successivamente una valutazione da 1 a 9 su ognuna di esse. Di seguito, viene presentata una possibile cartolina, ottenuta con i fattori e i livelli del nostro esempio.

È opportuno verificare che tra le combinazioni fornite dal disegno ortogonale non ci siano delle ripetizioni (cartoline uguali) o casi troppo simili (le cartoline devono differenziarsi almeno rispetto ai livelli di due fattori); ancora, è bene escludere cartoline irrealistiche; ad esempio, un televisore con attributi superaccessoriati non può essere proposto con livelli di prezzo troppo bassi.
ATTENZIONE!!! Il file che contiene il disegno ortogonale deve essere salvato e conservato *gelosamente*, in quanto necessario per la successiva elaborazione.

Il *data entry* che si genera dopo la somministrazione delle cartoline deve contenere, su un documento SPSS (o in alternativa su un foglio excel da importare in SPSS al momento dell'analisi), una tabella con M righe (M = n° di intervistati) ed N colonne (N = n° di card + 1).

La prima colonna (da nominare con il termine SOGGETTI) contiene i numeri progressivi (da 1 a M) indicanti gli intervistati (numeri da 1 a 100 se gli intervistati sono 100). Ogni intervistato (ossia ogni osservazione o caso) genera una riga (se gli intervistati sono 100, le righe saranno 100). Ogni card origina poi una colonna[6] (se le card sono 16, in totale ci saranno 17 colonne: 1 per i soggetti, 16 per le card). In ogni cella si inserisce il numero corrispondente alla valutazione fatta dagli intervistati per ogni card.

La conjoint analysis non prevede un vero e proprio questionario – gli intervistatori devono "armarsi" fondamentalmente delle cartoline e di *fogli risposte* su cui prendere nota delle valutazioni espresse; è opportuno, però, che vengano raccolti – con un questionario sintetico o raccogliendo direttamente le valutazioni su un foglio risposte – anche alcuni dati descrittivi e socio-demografici (soprattutto se alla conjoint segue l'applicazione della cluster analysis). Tali dati dovranno essere inseriti in un foglio SPSS *distinto* da quello in cui sono riportate le valutazioni delle cartoline.

Dopo la raccolta dei dati e la creazione del data entry, la fase di analisi su SPSS prevede la creazione di una sintassi (la conjoint analysis è una tecnica che SPSS non gestisce in automatico).
Per creare la sintassi e lanciare le analisi è necessario aprire preliminarmente il file con le valutazioni delle card; quindi:

1. Selezionare FILE dal menu principale;
2. Selezionare il comando NUOVO ed il sottocomando SINTASSI;
3. *Scrivere quanto segue adattando i comandi ai propri dati, in particolare rispetto alle istruzioni da fornire in ogni riga dopo il segno "=", prestando molta attenzione a lasciare spazi solo se indicato e a riportare i nomi di file e fattori esattamente come sono scritti nei file originari:*

```
conjoint plan='c:\ortopc.sav'
/data=*
/score=card1 to card16
/subject=soggetti
/factors=azione (discrete)
                consistenza (discrete)
                gusto (discrete)
                chiusura (discrete)
                prezzo (linear less)
/plot=summary.
```

In questa istruzione va indicato il percorso che indica la posizione sul pc del file con il disegno ortogonale

Se sui risultati della conjoint analysis si applicherà successivamente la cluster analysis per la segmentazione flessibile, prima dell'ultima linea di comando (/plot=summary.) bisognerà aggiungere la seguente istruzione:
/utility='nome.sav'
Per "nome" si intende il percorso che su cui salvare il file che conterrà i dati per applicare la cluster analysis; ad esempio: 'd:\marketing.sav' se si decide di salvare il file su disco esterno; oppure: 'c:\marketing.sav' se se decide di salvare il file sul disco rigido; oppure:

[6] Ogni colonna deve essere nominata scrivendo senza lasciare spazio CARDn, dove n è il numero della card (CARD1; CARD2…).

'c:\documenti\marketing.sav' se si decide di salvare il file in una cartella del disco rigido, in questo caso "documenti". Il nome "marketing", naturalmente, deve essere considerato solo un esempio. In tal modo verrà creato un file dati in cui saranno salvati i punteggi di utilità virtuali che ogni intervistato avrebbe fornito su ognuno dei singoli livelli dei fattori proposti. Sulle variabili salvate in questo file si applicherà la cluster analysis, secondo le modalità proposte nella sezione ad essa dedicata.

Nella sintassi sopra indicata i comandi utilizzati hanno il seguente significato:
Nella prima riga (comando "conjoint plan") viene inserito il percorso del file contenente il disegno ortogonale (ossia si indica dove si trova il file costruito con la prima applicazione in SPSS per la costruzione delle card); in questo caso "e" indica che il file si trova sul disco "e", mentre ortopc.sav è il nome di questo file.
L'asterisco inserito dopo il comando "data" indica che il file dati, ossia quello contenente le valutazioni delle card, è attivo (aperto).
Il comando inserito nella terza riga riguarda il tipo di scala utilizzata per raccogliere le valutazioni: in questo caso è stata utilizzata una scala ad intervallo (score) e sono state valutate 16 card (le colonne del file con le valutazioni sulle card sono state nominate card1, card2…).
"Subject" indica la prima colonna del file dati con le valutazioni sulle card, in cui sono stati numerati i soggetti intervistati, e che è stata nominata, appunto, soggetti.
"Factors" indica gli attributi così come sono stati definiti nel file con il disegno ortogonale. Le parentesi riguardano le ipotesi che è possibile formulare sulla relazione esistente tra i fattori e le preferenze (con "*discrete"* non si fa alcuna ipotesi, con *"linear less"* si ipotizza che la preferenza per il prodotto diminuisca se il fattore considerato – ad esempio il prezzo – aumenta e viceversa).
Il comando "plot=summary" consente di visualizzare i grafici dei risultati dell'applicazione, consistenti negli istogrammi espressivi dell'importanza relativa dei fattori e dei valori di utilità parziali corrispondenti ai livelli di ogni fattore.

Dopo aver inserito la sintassi, selezionarla integralmente trascinando con il tasto destro del mouse e cliccare sull'icona con il comando PLAY (▸) sulla barra degli strumenti per avviare l'analisi.

Gli output ed i test statistici:
Nell'output sono presenti sia i risultati disaggregati per singolo intervistato che quelli generali (considerando tutti i dati raccolti); è su quest'ultimi che è importante focalizzare l'attenzione (se nella sintassi non viene specificato il sottocomando /subject, si hanno solo i dati complessivi). ATTENZIONE!!! La lunghezza dell'output non permette di leggere immediatamente i risultati generali, posti in fondo alla pagina di output. Per poterli visualizzare è necessario cliccare con il tasto destro sulla pagina con i risultati e selezionare il comando "apri", che permette di generare una nuova finestra in cui è possibile leggere i risultati integrali.
Le informazioni più importanti dell'output sono racchiuse nelle tabelle e nei grafici con:

- *l'importanza relativa di ogni fattore:* nel grafico è contenuta l'importanza percentuale relativa del fattore;

- *l'utilità (part worth) per ogni livello di fattore:* tali part worth indicano l'utilità per ogni livello del fattore, inferita dalle valutazioni dei rispondenti;
- *Statistiche R di Pearson e Tau di Kendall:* queste statistiche forniscono un'indicazione del grado di adattamento del modello ai dati osservati; rappresentano, infatti, la correlazione tra le preferenze stimate e quelle osservate; questi indici variano tra 0 (correlazione nulla) e 1 (correlazione perfetta); ovviamente, sono richiesti valori alti di tali indici, tendenti all'unità. Il valore di significatività al quale si può ritenere valida l'ipotesi che l'adattamento sia buono, ovvero che i dati si adattano al modello individuato, deve essere inferiore a 0,05. Più elasticamente il valore soglia può essere alzato tuttavia a 0,1.

Identificando i livelli preferiti per ogni fattore è possibile identificare *il profilo ideale* del prodotto.

La conjoint analysis permette, inoltre, di simulare le valutazioni che avrebbero fornito gli intervistati rispetto a combinazioni che in realtà non sono state proposte, partendo dai risultati ottenuti in sede di analisi. Potrebbe essere interessante, ad esempio, verificare la bontà di alcune combinazioni – non proposte in sede di somministrazione perché non incluse nel disegno ortogonale – quali quella *ideale;* ancora si potrebbe simulare l'utilità fornita da combinazioni "hi-tech", "popolari", "femminili", ecc. ecc, costruendo a tavolino le combinazioni, ma utilizzando sempre i livelli dei fattori testati. Per lanciare la simulazione è necessario riaprire il file con il disegno ortogonale ed aggiungere le combinazioni che si intendono testare, indicando nella colonna "status" il codice "2" corrispondente a "simulazione" ed il numero progressivo. Salvare quindi il nuovo disegno ortogonale e rilanciare la sintassi – che non va modificata.
Oltre all'output precedente, in tal modo sarà possibile ottenere la valutazione media che i rispondenti avrebbero assegnato alle combinazioni simulate, oltre a dei valori probabilistici. Questi ultimi sono appunto delle percentuali che esprimono la probabilità di essere scelte tra le altre delle card simulate. Nella prassi, queste percentuali vengono assimilate a delle quote di mercato virtuali. SPSS fornisce le stime di queste quote calcolate con tre algoritmi diversi:

- il *First Choice* calcola le quote assumendo che venga scelta la combinazione che manifesta l'utilità totale maggiore; il difetto di questo metodo risiede nella mera considerazione del *ranking*, senza tener conto dell'intensità delle preferenze; fornisce, generalmente, risultati in cui il divario tra la combinazione simulata migliore e quella peggiore è amplificato;
- il *BTL* calcola le quote secondo un algoritmo probabilistico che tende ad appiattire le distanze tra la combinazione simulata migliore e quella peggiore;
- il *Logit* calcola le quote tramite la trasformazione "logit", mediando tra le soluzioni offerte dai due precedenti metodi in termini di distanza tra la combinazione simulata migliore e quella peggiore.

Con campioni sufficientemente ampi (almeno 100 osservazioni) è consigliabile l'adozione del metodo First Choice; con campioni piccoli il metodo Logit risulta più affidabile.

La Cluster Analysis

La cluster analysis ha l'obiettivo di raggruppare in sotto-insiemi o classi (cluster) elementi appartenenti ad un insieme più ampio, in modo tale che gli elementi appartenenti ad ogni

gruppo siano il più possibile omogenei tra loro e che i diversi gruppi siano invece il più possibile eterogenei. Tipicamente, si applica per analisi di segmentazione, quindi per identificare gruppi di consumatori con preferenze simili al loro interno e diverse nel confronto intergruppo. È possibile adottare algoritmi di cluster analysis di tipo *gerarchico* (applicabili su dati qualitativi) e *non gerarchico* (applicabili su dati quantitativi). In questa sede verrà applicato l'algoritmo non gerarchico delle "K-medie".

Le applicazioni di marketing più ricorrenti:

analisi di segmentazione

- *classica*, fondata sui giudizi di importanza degli attributi di un prodotto (in questo caso, la cluster analysis è preceduta dalla factor analysis);
- *flessibile*, fondata sulle valutazioni complessive fornite su diverse configurazioni di prodotto (in questo caso, la cluster analysis è preceduta dalla conjoint analysis).

I dati da raccogliere attraverso il questionario come input dell'analisi di segmentazione classica consistono in *valutazioni di importanza,* su scala a 7 o 9 punti (1 = per niente importante; 7 o 9 = estremamente importante), data ad ogni attributo nel processo di scelta di un determinato prodotto.
I dati da raccogliere attraverso il questionario come input dell'analisi di segmentazione flessibile coincidono con i risultati dell'applicazione della conjoint analysis (si rinvia, in merito alla relativa sezione).

Il data entry della cluster analysis.
È possibile identificare tre situazioni alternative di partenza:

- i dati in input sono direttamente (tutte o alcune) le variabili originarie (gli attributi proposti nel questionario): in questo caso bisogna selezionare dal menu STATISTICA (o ANALIZZA) il comando RIASSUMI ed il sottocomando "DESCRITTIVE", scegliendo l'opzione SALVA VALORI STANDARDIZZATI COME VARIABILI; in tal modo sul database originario vengono create delle nuove variabili che rappresentano le trasformazioni standardizzate delle variabili originarie, che saranno utilizzate per applicare la cluster analysis;
- i dati in input sono i punteggi fattoriali salvati in seguito all'applicazione precedente della factor analysis e che saranno utilizzati per applicare cluster analysis (si tratta di punteggi già standardizzati);
- i dati in input sono i punteggi di utilità di ognuno dei livelli salvati in seguito alla precedente applicazione della conjoint analysis (si tratta di punteggi già standardizzati).

Nel primo caso è necessario creare un database "originale"; su un documento SPSS, deve essere costruita una tabella con M righe (M = n° di intervistati) ed N colonne (N = n° di variabili + 1). Ogni intervistato (ossia ogni osservazione o caso) genera una riga (se gli intervistati sono 100, le righe saranno 100). La prima colonna è nominata "soggetti" (oppure "osservazioni") e serve a contare i rispondenti (cioè le righe). Ogni variabile indagata (attributo, beneficio, ecc.) genera poi una colonna (se le variabili sono 20, le colonne, in totale, saranno 21). In ogni cella si inserisce il numero contrassegnato dall'intervistato (valutazione). Sullo stesso foglio occorre inserire le colonne contenenti i dati socio-demografici raccolti, da utilizzare per la descrizione dei cluster.

L'applicazione con SPSS della cluster analysis a k-medie: i comandi

1. Aprire il file editor dei dati e selezionare STATISTICA (o ANALIZZA) dal menu principale;
2. Selezionare il comando CLASSIFICAZIONE ed il sottocomando CLUSTER K-MEDIE;
3. Specificare nella finestra di dialogo risultante le *variabili* sulle quali deve essere basato il raggruppamento (sulla base dei tre casi citati in precedenza si tratterà di variabili standardizzate ricalcolate da SPSS, di punteggi fattoriali, o di punteggi di utilità, sulla base dei quali si vuole effettuare la clusterizzazione/segmentazione);
4. Specificare il NUMERO DI CLUSTER che devono formare la soluzione finale (partendo da un numero minimo, ad esempio 2);
5. Selezionare il sottocomando OPZIONI, e richiedere la *tabella ANOVA;* cliccare su OK;
6. Dopo aver visualizzato il 1° output, ritornare all'editor dei dati e ripetere le analisi (dal passaggio 2 al passaggio 5) più volte, cambiando ogni volta il *numero di cluster* che devono formare la soluzione finale; ad esempio, porre nella prima analisi il n° di cluster = 2; nell'analisi successiva porre il n° di cluster = 3; nell'analisi successiva porre il n° di cluster = 4, ecc. ecc., fino a un numero congruo di potenziali cluster (6-7).

A questo punto è necessario identificare la soluzione ottimale a "n" cluster. La scelta deve tenere conto di tre output:

- confrontare le colonne con il ratio F (F = varianza tra gruppi / varianza nei gruppi) contenute nelle tabelle ANOVA. È preferibile la clusterizzazione che complessivamente (considerando tutte le variabili) esprima F con i valori più elevati (clusterizzazione preferibile in quanto a valori maggiori di F corrisponde una maggiore eterogeneità tra i cluster e/o una minore eterogeneità nei cluster) e contemporaneamente una buona significatività (Sig. < 0,05);
- valutare la tabella in output con *il numero di casi di ogni cluster* per verificare l'omogeneità della numerosità dei diversi cluster. Ad esempio, nel caso di una clusterizzazione in 3 cluster, se la numerosità di quest'ultimi presentasse 35 osservazioni afferenti al cluster 1, 25 osservazioni afferenti al cluster 2, e 40 osservazioni afferenti al cluster 3 allora la numerosità dei 3 cluster potrebbe ritenersi equa; se invece la numerosità dei tre cluster presentasse 80 osservazioni afferenti al cluster 1, 15 osservazioni afferenti al cluster 2, e 5 osservazioni afferenti al cluster 3, allora la numerosità dei tre cluster risulterebbe squilibrata; in quest'ultimo caso, sarebbe preferibile scartare la soluzione;
- valutare le tabelle con i *centri dei cluster finali.* Più precisamente può ritenersi preferibile la clusterizzazione che complessivamente presenti valutazioni eterogenee tra i gruppi rispetto alle variabili in input; è auspicabile ottenere, cioè, una soluzione che preveda valutazioni alte sugli attributi "a" e "b" per il cluster "x", valutazioni alte sugli attributi "c" e "d" per il cluster "y", valutazioni alte sull'attributo "e" per il cluster "z", ecc.

In sintesi, la scelta del numero di cluster ottimale per una clusterizzazione k-medie dipende dal confronto dei risultati ottenuti mediante i diversi tentativi di clusterizzazione in termini di (in ordine di importanza decrescente) F e Sig. nella tabella ANOVA, n° di casi in ogni cluster e centri dei cluster finali.
Una volta scelta la clusterizzazione più adatta, ri-effettuare l'analisi (dal punto 2 al punto 6) selezionando, inoltre, il sottocomando SALVA e scegliendo, nella finestra di dialogo aperta, l'opzione CLUSTER DI APPARTENENZA (si chiede, cioè, di salvare nell'editor dei dati anche la

colonna nella quale, caso per caso, viene indicato il numero corrispondente al cluster di appartenenza del rispondente).
I cluster dovranno quindi essere descritti in termini di *preferenze* e sulla base dei *dati socio-demografici*.
In merito alle analisi di segmentazione classica (dati in input: punteggi fattoriali o variabili originarie standardizzate), i cluster saranno descritti sulla base della tabella dei centri finali: valori positivi indicano utilità maggiori della media per i corrispondenti fattori o attributi, valori negativi indicano utilità minori della media per i corrispondenti fattori o attributi. Sulla base di tali valori si "battezzeranno" i cluster; ad esempio, se un cluster presenta centri finali alti su attributi del tipo "convenienza", "disponibilità dei ricambi" e "vicinanza del PDV" si potrebbe definirlo "*gli economi*", o più simpaticamente, "*gli avari"*.
Ai fini della descrizione dei cluster in termini di indicatori socio-demografici, è necessario invece costruire tante tavole di contingenza per quanti sono gli indicatori socio–demografici rilevati mediante il questionario (es.: genere, età, professione, hobby, ecc.). In ogni tavola di contingenza si incrocerà la variabile CLUSTER DI APPARTENENZA (salvata nell'editor dei dati) con tali indicatori socio-demografici. Più precisamente, bisogna selezionare dal menu STATISTICA (o ANALIZZA) il sottocomando RIASSUMI e scegliere TAVOLE DI CONTINGENZA; si selezioneranno poi le due variabili da incrociare, inserendo come VARIABILE DI RIGA la variabile CLUSTER DI APPARTENENZA e come variabile di colonna ogni singolo indicatore socio-demografico. Si otterranno delle tavole di contingenza descrittive delle caratteristiche socio-demografiche di ogni cluster.

Qualora la cluster venga effettuata sui risultati di una conjoint analysis – segmentazione flessibile, l'interpretazione dei cluster avviene in maniera diversa. Bisognerà infatti riapplicare la conjoint analysis su ogni singolo cluster usando, dal menu DATI, il comando SELEZIONA CASI. Selezionare nella finestra di dialogo "*se la condizione è soddisfatta"* e cliccare sul tasto "*se"*, tramite cui è possibile fissare la condizione di filtro dei casi. E' necessario: indicare il gruppo di appartenenza come variabile di selezione, annerendola e cliccando sul tasto con la freccia a destra per porla nella finestra della formula; aggiungere, quindi "= *n"*, dove *n* sarà, di volta in volta il numero del cluster di cui si vuole selezionare gli appartenenti; dopo aver confermato con il comando CONTINUA, selezionare "*i casi non selezionati saranno filtrati*" e cliccare su OK; in tal modo dal database principale verranno selezionati solo i casi con gruppo di appartenenza uguale a quello *"n"*, indicato nelle istruzioni. Ripetendo l'operazione per ogni cluster, si potrà riapplicare la conjoint analysis – secondo le istruzioni citate nella relativa sezione, ottenendo i risultati delle preferenze sui fattori e sui livelli relativi ai singoli cluster. La descrizione dei cluster in termini di preferenze si baserà, quindi, sui risultati delle singole applicazioni della conjoint analysis, coniugandoli con le elaborazioni sui dati socio-demografici.

La Discriminant Analysis
L'analisi discriminante ha l'obiettivo di identificare gli "elementi" (funzioni discriminanti) che meglio distinguono gli oggetti (marche o prodotti) nelle percezioni del consumatore, ossia gli aspetti che spiegano meglio le differenze nelle valutazioni dei diversi oggetti. La discriminant analysis è utilizzata, quindi, per analisi di posizionamento basate sugli attributi *(attribute-based).*

Le applicazioni di marketing più ricorrenti:

- analisi di posizionamento attribute-based;
- analisi delle determinanti delle scelte dei consumatori.

In questa sede, verrà approfondito solo l'applicazione della discriminant analysis per le analisi di posizionamento.

I dati da raccogliere come input delle analisi di posizionamento, attraverso il questionario, coincidono con valutazioni di performance degli oggetti da posizionare (marche o prodotti) rispetto a diversi elementi di comparazione (attributi o benefici) su scala a 7 o a 9 punti.
La forma del questionario sarà come quella del seguente esempio, utilizzato per posizionare una serie di merendine rispetto ad una serie di attributi/benefici (normalmente si testano 6-8 marchi/prodotti rispetto a 6-8 attributi/benefici).

	Farcitura							**Gusto**							**Genuinità**						
Fiesta	1	2	3	4	5	6	7	1	2	3	4	5	6	7	1	2	3	4	5	6	7
Flauti	1	2	3	4	5	6	7	1	2	3	4	5	6	7	1	2	3	4	5	6	7
Saccottini	1	2	3	4	5	6	7	1	2	3	4	5	6	7	1	2	3	4	5	6	7
Plumcake	1	2	3	4	5	6	7	1	2	3	4	5	6	7	1	2	3	4	5	6	7
Tegolini	1	2	3	4	5	6	7	1	2	3	4	5	6	7	1	2	3	4	5	6	7
Nastrine	1	2	3	4	5	6	7	1	2	3	4	5	6	7	1	2	3	4	5	6	7
Krapfen	1	2	3	4	5	6	7	1	2	3	4	5	6	7	1	2	3	4	5	6	7
Buondì	1	2	3	4	5	6	7	1	2	3	4	5	6	7	1	2	3	4	5	6	7

Il data entry da costruire per applicare la discriminant analysis
Su un documento SPSS (o in alternativa su un foglio excel da importare in SPSS al momento dell'analisi) si costruisce una tabella con M righe (M = n° di prodotti/marche indagati x n° di intervistati) ed N colonne (N = n° di variabili + 2). La prima colonna è nominata "soggetti" (oppure "osservazioni") e serve a contare i rispondenti (cioè le righe); per ogni rispondente ci sono tante righe quanti sono i prodotti/marche da valutare: se gli oggetti in esame sono 7 e gli intervistati sono 100, il numero di righe sarà: 7x100=700). La seconda colonna è quella degli oggetti valutati (marche o prodotti) – *ATTENZIONE: deve trattarsi di una variabile ordinale di tipo "numerico", cioè i cui livelli siano codificati da 1 ad n.* Nelle celle della seconda colonna, quindi, questionario dopo questionario verranno ripetuti i codici numerici assegnati ai prodotti/marche valutati. Ogni variabile indagata (attributo/beneficio) genera poi una nuova colonna (se le variabili sono 20, le colonne in totale saranno 20 + 2).
Nelle celle delle altre colonne si inserisce, invece, il numero corrispondente alla valutazione fatta dagli intervistati per ogni prodotto/marca e rispetto ad ogni variabile. Il seguente esempio serve a chiarire la particolare conformazione del database necessario per applicare la discriminant analysis:

Soggetti	Prodotto	Farcitura	Gusto	Genuinità	.....
1	Fiesta	1	6	5	....
1	Flauti	3	2	1	....
1	Saccottini	7	3	2	....
1	Plumcake	4	1	2	....
1	Tegolini	4	7	7	....
1	Nastrine	6	1	1	....
1	Krapfen	5	2	6	....
1	Buondì	1	1	1	....
2	Fiesta	3	2	3	....
2	Flauti	5	3	1	....
2	Saccottini	6	4	5	....
2	Plumcake	3	4	7	....
2	Tegolini	2	5	5	....
2	Nastrine	1	3	4	....
2	Krapfen	5	7	5	....
2	Buondì	6	4	5	....
....	....	....	....	....	....

L'applicazione con SPSS: i comandi

Aprire il file editor dei dati raccolti; quindi:

1. Selezionare il menu STATISTICA (o ANALIZZA);
2. Selezionare il comando CLASSIFICA ed il sottocomando DISCRIMINANTE;
3. Specificare nella finestra di dialogo risultante la VARIABILE DI RAGGRUPPAMENTO (cioè la variabile indicante i prodotti/marche; *ATTENZIONE!!! Deve trattarsi di una variabile ordinale o a scala e di tipo "numerico", cioè i cui livelli siano codificati numericamente da 1 ad n)*, selezionare quindi il comando DEFINISCI INTERVALLO e specificare l'intervallo di oscillazione dei valori che contrassegnano le modalità della variabile di raggruppamento (ad esempio, valore minimo = 1 e valore massimo = 6 qualora si proponga agli intervistati la valutazione di 6 marche contrassegnate con codici numerici da 1 a 6); specificare, infine, le VARIABILI INDIPENDENTI (gli attributi o i benefici proposti);
4. Selezionare il sottocomando STATISTICHE e contrassegnare, nella finestra di dialogo aperta, le voci *ANOVA UNIVARIATE* (sub-descrittive), FISHER (sub-coefficienti di funzione);
5. Selezionare il sottocomando CLASSIFICA e contrassegnare: *USA MATRICE DI COVARIANZA ENTRO GRUPPI, TUTTI GRUPPI UGUALI, VISUALIZZA TABELLA RIASSUNTIVA, SOSTITUISCI VALORI MANCANTI CON MEDIA*.

Gli output ed i test statistici

Focalizzare l'attenzione sui seguenti output:

La tabella con gli autovalori: dalla lettura di questi dati è possibile valutare l'efficacia relativa delle funzioni discriminanti nel cogliere la variabilità del fenomeno (ossia la varianza spiegata) e quindi l'abilità predittiva del modello discriminante (affidabilità dei dati ottenuti). Nella colonna della varianza è possibile leggere la percentuale di varianza spiegata dalle funzioni discriminanti estratte. Per una rappresentazione a due dimensioni (mappa di posizionamento dei prodotti/marche valutati), la varianza cumulata spiegata dalle prime due funzioni discriminanti deve raggiungere almeno la soglia del 50%. Chiaramente maggiore è la varianza cumulata e più attendibile risulterà la rappresentazione. Dall'*indice di correlazione canonica* si ricavano informazioni sulla capacità predittiva del modello. Più l'indice è alto (maggiore di 0,5)

è migliore risulta la capacità predittiva (le variabili indagate sono cioè molto correlate con gli oggetti valutati).

La matrice di struttura. Dalla lettura della matrice (così come dalla lettura della tabella con i coefficienti standardizzati: la matrice di struttura rappresenta il risultato "ruotato", quindi più leggibile della matrice dei coefficienti standardizzati) è possibile valutare le relazioni tra le funzioni discriminanti e le variabili esaminate (o *predittori:* si tratta degli attributi/benefici testati su ogni prodotto), in modo tale da interpretare il significato delle funzioni stesse. Per ognuna delle prime due funzioni – le più esplicative, che verranno utilizzate per creare la mappa di posizionamento – devono essere individuate le variabili che presentano gli indici di correlazione più alti (in caso di valori negativi l'influenza dell'attributo sulla funzione discriminante andrà considerato in termini inversi; ad esempio se l'attributo "convenienza" esprime una correlazione negativa sulla funzione discriminante = - 0.654, significa che tale funzione è descrittiva dell'onerosità del prodotto). Analizzando le sole variabili altamente correlate con la singola funzione discriminante, si inferisce il nome da assegnare a tale funzione. Ad esempio, se la prima funzione risulta correlata con gli attributi "Gusto" e "Farcitura" sarà intuitivo denominarla "Golosità" o "Dimensione edonistica".

Funzione ai baricentri di gruppo. Tale matrice contiene le *coordinate* degli oggetti (marche o prodotti) valutati rispetto alle funzioni discriminanti estratte.

Trascrivendo i dati delle funzioni ai baricentri di gruppo e quelli della matrice di struttura in una tabella excel è possibile costruire un grafico a dispersione consistente nella mappa di posizionamento degli oggetti valutati e degli attributi. All'uopo, nell'area download del sito del corso è disponibile il file "Positioning.xls" tramite cui è possibile la generazione della mappa in automatico.

Test di uguaglianza delle medie di gruppo. Si tratta della verifica di significatività dei singoli attributi utilizzati. Il test si propone di smentire l'ipotesi nulla che le medie tra i gruppi (i prodotti/marche) siano uguali, cioè che gli attributi proposti non discriminino gli oggetti da posizionare. Vengono considerati congiuntamente gli indicatori lambda di Wilks e il ratio F, cui è associato il valore di significatività (sig.) dell'ipotesi nulla: quest'ultima può essere respinta (espressione della bontà discriminante dei singoli attributi) per valori inferiori a 0,05 (o più elasticamente, inferiori a 0,1).

Lambda di Wilks. Questo test verifica la significatività delle funzioni discriminanti; bisogna accertarsi che il valore di Sig. associato riguardante le prime due funzioni (che saranno utilizzate per creare la mappa di posizionamento) sia inferiore alla soglia di 0,05 (o più elasticamente, inferiori a 0,1).

La mappa di posizionamento generata tramite l'applicazione dell'analisi discriminante permette sia una visione generale delle percezioni dei consumatori rispetto alle due più importanti funzioni discriminanti individuate, e, allo stesso tempo, l'opportunità di verificare le posizioni rispetto ai singoli attributi; in merito a quest'ultima opportunità, basta ruotare l'asse delle ordinate rispetto al vettore che ha origine nel punto (0,0) e direzione definita dal punto di un attributo, per verificare le posizioni dei prodotti/marche rispetto a tale attributo.

Il Multidimensional Scaling

Il Multidimensional Scaling – MDS – è una tecnica che si applica per analisi di posizionamento *non-attribute based*. L'MDS si pone l'obiettivo di rappresentare sotto forma di punti, in uno spazio multidimensionale, determinati "oggetti", sulla base dei giudizi di similarità/dissimilarità tra gli stessi percepita dagli intervistati. Si prescinde, quindi, da riferimenti a set predeterminati di attributi.

Attraverso particolari algoritmi, le valutazioni di similarità espresse dagli intervistati sono convertite in distanze tra gli oggetti e rappresentate graficamente mediante mappe a due dimensioni.

Le applicazioni di marketing più ricorrenti:

- per analisi di posizionamento non attribute-based;
- per analisi di interdipendenza concorrenziale (definizione gruppi competitivi).

I dati da raccogliere attraverso il questionario come input dell'analisi consistono nelle valutazioni di similarità/dissimilarità, su scala metrica a 7 punti (1 = estremamente simili; 7 = estremamente dissimili), di *tutte* le coppie (non uguali) di oggetti da posizionare (ad esempio, marche) che è possibile combinare a partire dal set di oggetti in esame. Se le marche che si vogliono posizionare sono 8, le possibili coppie sono 28 (da 7+6+5+4+3+2+1 = 28). Dal momento che le dimensioni analizzate tramite l'MDS non sono definite, è preferibile inserire tra gli oggetti da posizionare *"l'oggetto ideale"*, chiedendo ai rispondenti quanto ognuno degli oggetti da analizzare sia simile/dissimile al proprio ideale.

L'MDS viene utilizzato soprattutto nell'ambito di analisi di posizionamento di prodotti-marche ad alto contenuto emozionale (ad esempio i profumi e i gioielli), rispetto ai quali le valutazioni sugli attributi potrebbero non essere agevoli, discriminanti o esplicitabili.

Il data entry per l'MDS.

Su un documento SPSS (o in alternativa su un foglio excel da importare in SPSS al momento delle analisi) si costruisce una tabella con M righe (M = n° di intervistati) ed N colonne (N = n° di coppie di oggetti + 1). Ogni intervistato genera una riga (se gli intervistati sono 100, le righe saranno 100). La prima colonna è nominata "soggetti" (oppure "osservazioni") e serve a contare i rispondenti (cioè le righe). Ogni coppia di oggetti indagata genera poi una colonna (se le coppie possibili sono 28, le colonne, in totale, saranno 29). In ogni cella si inserisce il numero corrispondente alla valutazione di similarità/dissimilarità fatta dagli intervistati per ogni coppia di oggetti.

È preferibile poi verificare di aver utilizzato nel questionario gli ancoraggi: 1 = simile e 7 = dissimile. In caso contrario, è necessario ricodificare le variabili, selezionando dal menu TRASFORMA il comando RICODIFICA; dopo aver inserito nella finestra di dialogo le variabili da ricodificare, selezionare "vecchi e nuovi valori"; occorrerà quindi inserire le istruzioni per sostituire i valori vecchi con i simmetrici della scala di misurazione (ad esempio, 7 diventerà 1 e viceversa, 6 diventerà 2 e viceversa, 5 diventerà 3 e viceversa; 4, in quanto baricentro, rimane invariato).

Successivamente, sarà necessario creare un nuovo file dati seguendo le seguenti istruzioni:

1. per ogni coppia di oggetti del file originario di dati, calcolare la media (vedi analisi univariate);
2. creare un nuovo file dati, con M righe e M colonne (M = n° di oggetti). Ogni oggetto (prodotto/marca) genera una colonna (se gli oggetti sono 8, le colonne saranno 8); tale file dati rappresenterà una matrice di dissimilarità, in cui la diagonale conterrà valori uguali a 0; i valori da inserire nelle celle sono i giudizi *medi* di dissimilarità calcolati precedentemente per ogni coppia di oggetti; trattandosi di una matrice simmetrica è sufficiente riempire solo la metà bassa della matrice.

L'applicazione con SPSS: i comandi

Aprire il file editor dei dati (la matrice simmetrica con i giudizi *medi* di dissimilarità); quindi:

1. selezionare il menu STATISTICA (o ANALIZZA);
2. selezionare il comando SCALING ed il sottocomando SCALING MULTIDIMENSIONALE;
3. specificare nella finestra di dialogo risultante le VARIABILI dell'editor dei dati da analizzare (es.: le marche) e contrassegnare I DATI SONO DISTANZE (si indica al programma di calcolo che i dati rappresentano una matrice di distanze);
4. selezionare il sottocomando MODELLO e specificare il LIVELLO DI MISURAZIONE delle variabili (INTERVALLO);
5. selezionare il sottocomando OPZIONI, scegliere la visualizzazione dei GRAFICI DI GRUPPO e indicare 0,002 come VALORE S-STRESS MINIMO.

Gli output ed i test statistici:

Focalizzare l'attenzione sui seguenti test:

Test S-STRESS. L'indicatore S-stress è una misura del grado di adattamento del modello, prodotto con l'MDS, rispetto ai dati. Varia da 1 (valore peggiore) a 0 (valore migliore).

Altre due misure del grado di adattamento, che in linea di massima confermano la bontà dell'adattamento del modello, già appurata con il test S-stress, sono:

- l'indice Stress di Kruskal (calcolato sulle distanze, mentre l'indice S-stress è calcolato sui quadrati delle distanze; può anch'esso assumere valori compresi tra 0 e 1); di seguito sono forniti alcuni suggerimenti relativi alla valutazione dell'adattamento mediante l'indice di Stress (dati indicati da Kruskal sulla base delle sue esperienze empiriche):

Valore indice di stress	**Valutazione adattamento**
S=0	perfetto
0<S<0,02	Eccellente
0,02<S<0,05	Buono
0,05<S<0,10	Sufficiente
0,10<S<0,20	Scarso
s>0,20	Non accettabile

- l'indice RSQ indica la proporzione media di varianza spiegata, più alto è il valore che assume (valore massimo = 1) è maggiore è la varianza spiegata dal modello.

Grafico modello distanza euclidea. È la rappresentazione grafica degli oggetti valutati, ossia la mappa di posizionamento. Dalla lettura della mappa, sulla base della vicinanza/lontananza dei punti (gli oggetti valutati), si può inferire la similarità/dissimilarità degli oggetti percepita dagli intervistati. Quanto più due oggetti sono vicini, tanto più sono percepiti come simili. Con l'applicazione in SPSS non è possibile dare un nome agli assi della mappa. Indicazioni sulle posizioni più o meno preferibili sono date dal posizionamento sulla mappa dell'oggetto ideale (posizionamento del prodotto ideale). Quanto più un oggetto è vicino all'ideale, tanto più può ritenersi preferibile dal consumatore.

Alcuni riferimenti bibliografici sull'analisi statistica dei dati

1. Barbaranelli C., 2007, *Analisi dei dati con SPSS vol. I e II*, LED Edizioni Universitarie.
2. Corbetta P., 1992, *Metodi di analisi multivariata per le scienze sociali*, Bologna, Il Mulino.
3. Corbetta P., Gasperoni G., Pisati M., 2001, *Statistica per la ricerca sociale*, Bologna, Il Mulino.
4. De Luca A., 1995, *Le applicazioni dei metodi statistici alle analisi di mercato*, Milano, F. Angeli.
5. Molteni L., 1993, *Analisi multivariata e ricerche di marketing*, Milano, EGEA.
6. Molteni L., 1998, *L'analisi di dati aziendali*, Milano, F. Angeli.
7. Molteni L. e Troilo G., 2022, Ricerche di marketing, Milano, McGraw-Hill.