



**UNIVERSITA' DEGLI STUDI DI PADOVA**  
**DIPARTIMENTO DI SCIENZE ECONOMICHE ED AZIENDALI**  
**"M.FANNO"**

**CORSO DI LAUREA IN ECONOMIA**

**PROVA FINALE**

**"TECNICHE DI MARKETING QUANTITATIVO"**

**RELATORE:**

**CH.MO PROF. ROCCO LORENZO**

**LAUREANDO/A: CHIARION JONATHAN**

**MATRICOLA N. 1123107**

**ANNO ACCADEMICO 2019 – 2020**

## **RINGRAZIAMENTI**

Prima di procedere con la stesura della tesi ci tengo a dedicare alcune parole a tutte quelle persone che mi hanno accompagnato lungo questo percorso universitario, un viaggio che ha segnato in modo indelebile la mia crescita personale.

Ringrazio in primo luogo, la mia famiglia, che nonostante le avversità incontrate lungo questo percorso mi ha sempre sostenuto, supportandomi nei momenti in cui ho smarrito l'autostima e la speranza.

Ritengo opportuno ringraziare anche gli amici e i compagni universitari, che mi hanno aiutato in tutti i modi per giungere fino a questo giorno.

# INDICE

<b>INTRODUZIONE</b> .....	<b>1</b>
L'analisi della domanda .....	1
<b>TECNICHE DI MARKETING QUANTITATIVO</b> .....	<b>6</b>
Conjoint Analysis.....	7
Funzioni edoniche .....	11
Multidimensional Scaling .....	13
Cluster Analysis .....	18
<b>CLUSTER ANALYSIS SUL MERCATO DELL'AUTOMOTIVE</b> .....	<b>20</b>
Dendogramma.....	21
Clusterizzazione .....	22
<b>ANALISI DEI COMPONENTI PRINCIPALI</b> .....	<b>24</b>
Factor analysis.....	25
Mappa analisi componenti principali.....	26
<b>CONCLUSIONI</b> .....	<b>27</b>
<b>BIBLIOGRAFIA</b> .....	<b>28</b>

## **INTRODUZIONE**

Nella tesi di laurea si andrà ad analizzare le principali tecniche di marketing quantitativo, in seguito andremo ad osservare in modo pratico la cluster analysis una delle tecniche precedentemente esposte, verrà utilizzata nel mercato dell'automobile, prendendo un determinato numero di modelli

Come si realizza la segmentazione? Esistono diverse tecniche, le prime servono per individuare parametri rilevanti nella scelta del consumatore e sono: questionari, focus group, funzioni edoniche, conjoint analysis, multidimensional scaling. Mentre la cluster analysis e l'analisi delle componenti principali, hanno lo stesso obiettivo delle prime ma presuppongono la conoscenza dei parametri/attributi del prodotto da acquistare.

## **L'ANALISI DELLA DOMANDA**

Con l'analisi della domanda si studia il lato dei consumatori, come si comportano loro, com'è la domanda, tratta del modello struttura comportamenti.

Il ciclo di vita del prodotto è quel tipico andamento del volume delle vendite (fatturato) nel tempo.

Le fasi sono: introduzione, sviluppo, maturità e declino. Un punto debole di questo dello del ciclo di vita è quello di non chiarire con precisione quanto durano le singole fasi.

Per segmento intendiamo un pezzettino della domanda in cui i componenti del segmento sono simili tra loro. Le definizioni principali per un segmento sono:

§ “un sottoinsieme distinto di clienti, omogeneo al proprio interno, ma disomogeneo rispetto ai clienti di altri segmenti, dove ogni insieme può essere scelto come obiettivo di mercato da raggiungere con una particolare strategia.” (Valdani 1984)

§ “un segmento è definito come quell’area di consumo che mostra caratteristiche di omogeneità al proprio interno oppure come un submercato che, all’interno del mercato generale, presenta modalità simili nell’atteggiamento verso un certo prodotto, nel consumo o nell’acquisto.” (Scott e Santagostino 1984).

Un segmento è un sottoinsieme dell’intera arena competitiva che viene analizzata.

È possibile segmentare la domanda in infiniti modi ma i due principali sono per consumatori e per prodotti. I consumatori per esempio si possono segmentare a seconda che siano industriali o consumatori finali, o anche in base ai canali distributivi utilizzati e alle regioni geografiche di appartenenza.

Si ipotizza che esistano domande diverse perché nel mercato esistono prodotti diversi. I prodotti si possono quindi segmentare per tipologia, per livello di prezzo o per performance.

Esistono due criteri di valutazione per capire a posteriori se una segmentazione è rilevante oppure no:

A: Esistono due (o più) segmenti di prodotto quando più del 50% dei consumatori di ognuno dei segmenti non prende in considerazione, al momento dell’acquisto, i prodotti di tutti gli altri segmenti;

B: Esistono due segmenti di consumatori quando almeno il 50% dei consumatori ha dei criteri di scelta dei prodotti diversi da quelli dell’altro gruppo.

Se i segmenti sono auto utilitarie e auto lussuose, i due segmenti sono rilevanti e validi se almeno il 50% dei consumatori interessati al primo segmento non prende in considerazione le automobili del secondo segmento in fase di acquisto.

Capire la logica della scelta dei consumatori e capire i segmenti sono due facce della stessa medaglia.

Se si conoscono le scelte dei consumatori si possono individuare i segmenti e viceversa.

È opportuno comunque identificare pochi segmenti.

Le aree strategiche d'affari, o strategic business units, sono rappresentazioni grafiche che segmentano il mercato tenendo in considerazione: il gruppo di clienti, le tecnologie e le funzioni d'uso.

Sono quindi una porzione che raccoglie clienti precisi, tecnologie precise e funzioni d'uso precise.

Il concetto di aree strategiche d'affari è simile a quello della segmentazione perché scompone l'arena competitiva in porzioni omogenee tra loro.

Le aree strategiche d'affari sono porzioni dell'arena competitiva mentre le segmentazioni sono porzioni della domanda.

I tre criteri utilizzati per definire un'arena competitiva si possono utilizzare anche per le aree strategiche d'affari.

Per essere in grado di capire qual è la logica di scelta dei consumatori e i criteri di selezione utilizzati in fase d'acquisto, si hanno due possibilità:

1) Chiedere direttamente al consumatore.

- a. Con questionari. I questionari non possono essere sottoposti a tutti i consumatori quindi devono essere sottoposti ad un campione rappresentativo dei consumatori. Si chiede quindi al consumatore di dare un punteggio ad ogni caratteristica che gli viene chiesto riguardo un prodotto/servizio. Massimo 10 criteri possibili di logica della scelta, altrimenti il consumatore non ha interesse a procedere con il questionario.

Successivamente si crea una mappa utile per la segmentazione utilizzando le due variabili più rilevanti. Affinché la segmentazione sia rilevante, è opportuno avere sugli

assi gli attributi più rilevanti.

- b. L'altro modo è quello di non prendere in considerazione tutti ma un piccolo gruppo di consumatori interrogarli, tramite quindi i focus group, e chiedere direttamente la logica della scelta.

Ai consumatori possono essere poste domande per capire i criteri della scelta dei consumatori. Per arrivare di conseguenza ad individuare gli attributi maggiormente rilevanti nella scelta del consumatore, anche se gli attributi rilevanti non sono necessariamente gli attributi necessari.

2) Attraverso l'applicazione di tecniche statistiche.

**a) CONJOINT ANALYSIS**

Questa è una tecnica di ricerca sviluppatasi attorno agli anni '70, si basa su una simulazione del processo di valutazione e scelta dei clienti relativa ad alcune alternative di offerta (sviluppate a partire da un set di attributi e di livelli di prestazione predefiniti, prezzi compresi) rispetto ai quali i clienti esprimono la loro preferenza o intenzione d'acquisto.

Toscano G. (2010) a tale proposito afferma che per l'applicazione di questa tecnica è importante:

“ricostruire la catena mezzi – fini, per meglio specificare il valore di un prodotto è ritenuto scomponibile in un insieme di valori parziali, connessi ai singoli attributi dell'offerta la rilevazione dei livelli di prestazione connessi a ciascun attributo: è necessario richiedere agli intervistati l'esplicitazione dei livelli di prestazione attesi, quelli minimi (accettabili), e quelli ottimali (o ideali) per ciascun attributo”.

Nella mente del consumatore, egli attribuisce un certo valore a prodotti diversi

in un certo mercato e questo valore dipende dalle caratteristiche del prodotto stesso.

$$\text{Valore del bene} = \alpha_0 + \alpha_1 X_1 + \alpha_2 X_2 + \alpha_3 X_3 + \dots + \varepsilon$$

Le caratteristiche del prodotto specifico sono indicate con  $x_1, x_2, x_3$ . Gli alpha sono i pesi degli attributi, maggiore è l'alpha tanto maggiore è il valore di quell'attributo sul valore totale del bene.

$$\text{Utilità totale} = \alpha_0 + \alpha_1 UTA + \alpha_2 UTB + \alpha_3 UTC + \dots + \varepsilon$$

Il valore del bene è la nostra Y nella funzione.

Il beneficio che il consumatore trae da un prodotto

dipende dal beneficio che ha dalle singole utilità che derivano dalle specifiche caratteristiche.

L'analisi congiunta mira a scomporre il grado di apprezzamento, l'utilità totale che si ha di un prodotto, in tante utilità parziali che si sommano tra di loro e che derivano dalle caratteristiche, è una tecnica basata sulla regressione.

Secondo Arboretti Giancristofaro R. e Bonnini S. si può affermare che "l'utilità totale deriva dalla somma di utilità parziali pesate da un certo valore, le utilità parziali riferite ai singoli livelli degli attributi sono denominate part-worth.

L'utilità è un valore che indica il livello di soddisfazione che un consumatore ottiene da un certo bene o da un paniere di beni con date caratteristiche".

Lo scopo dell'analisi sarà quindi andare a vedere quali sono questi coefficienti; se si è in grado di comprendere quali sono si capirà anche quali sono quelli che pesano di più nella scelta del consumatore. Di conseguenza si sarà in grado di riconoscere quindi quali sono i pesi che il consumatore dà inconsciamente nella logica della scelta.

Nell'immagine qui in alto, un esempio riportato da Barbarito L. in Analisi competitiva (pag.70) si è individuata l'analisi congiunta per la scelta di un televisore.

### La conjoint analysis

● *Valore del bene* =  $\alpha_0 + \alpha_1 X_1 + \alpha_2 X_2 + \alpha_3 X_3 + \dots + \varepsilon$

	dimensione	tecnologia	prezzo
1	37"	Lcd	800 €
2	37"	Lcd	1200 €
3	37"	Lcd	2000 €
4	37"	Plasma	800 €
5	37"	Plasma	1200 €
6	37"	Plasma	2000 €
7	40"	Lcd	800 €
8	40"	Lcd	1200 €
9	40"	Lcd	2000 €
10	40"	Plasma	800 €
11	40"	Plasma	1200 €
12	40"	Plasma	2000 €
13	50"	Lcd	800 €
14	50"	Lcd	1200 €
15	50"	Lcd	2000 €
16	50"	Plasma	800 €
17	50"	Plasma	1200 €
18	50"	Plasma	2000 €



Gli attributi rilevanti (ovvero  $x_1$ ,  $x_2$ ,  $x_3$ ) sono la dimensione, la tecnologia e il prezzo. I livelli della dimensione, si è ipotizzato che potessero essere tre: 37, 40 o 50 pollici.

I livelli di tecnologia sono led o al plasma. I livelli di prezzo sono 800, 1200, 2000.

Con la conjoint analysis si può procedere dunque utilizzando due modalità diverse:

§ Chiedendo ai consumatori di ordinare tutte le schede dei prodotti, dal più preferibile al meno preferibile. Il consumatore esprimendo l'ordine, rivela il giudizio sul valore e quindi la  $y$  della relazione. Il prodotto che verrà indicato per primo sarà il prodotto a cui dà il punteggio più alto.

L'analisi congiunta applica una specie di regressione e cerca di capire quali sono i pesi  $a_1$ ,  $a_2$ ,  $a_3$  dei diversi attributi.

Per esempio, in base alle risposte che i consumatori danno indicano quali sono gli attributi più rilevanti.

Una volta osservato quali sono i due attributi più rilevanti, si riportano sugli assi cartesiani in una mappa di segmentazione però all'interno si dispongono i prodotti veri che ci sono nel mercato.

Non vengono poste domande dirette ai consumatori ma viene esclusivamente richiesto di ordinare i 18 prodotti potenziali che nascono dalle combinazioni tra livelli e attributi.

§ Il secondo metodo si utilizza per arrivare a capire quali siano i valori che i consumatori attribuiscono ai pesi degli attributi nella loro logica di scelta – è una modalità che avviene via computer. Il computer ha comunque l'intero elenco dei prodotti potenziali ma inizialmente ci propone, in una schermata video, un numero limitato di prodotti, a discrezione del ricercatore, scelti casualmente e chiede di selezionare il preferito di quelli presentati. Il computer propone poi un'altra

scheda con lo stesso numero di prodotti, poi un'altra ed un'altra ancora. Questa tecnica termina nel momento in cui il computer riesce a capire in base alle risposte qual è il valore che attribuiamo ai pesi di ciascun attributo. La diversità si basa spesso su fatto che il consumatore sceglie meglio fra poche cose e quindi preferisce stare magari 5 minuti a scegliere 50 volte ma sempre fra tre prodotti rispetto ad ordinare secondo la preferenza in 5 minuti, 18 prodotti diversi. Il computer arriva comunque allo stesso output che propone il metodo uno prima descritto. Una volta saputo il valore dei pesi, si procede poi comunque a creare una mappa di segmentazione ponendo negli assi cartesiani gli attributi più rilevanti.

La conjoint analysis ha una seconda utilizzazione, ovvero che i programmi normalmente ci dividono in gruppi di consumatori che si comportano con una logica radicalmente diversa.

Nella logica dell'analisi congiunta queste si chiamano le inversioni.

Le inversioni sono quindi un collaterale della conjoint analysis che ci fanno capire se ci sono porzioni di consumatori con una logica di scelta radicalmente diversa.

C'è una seconda cosa aggiuntiva sulla conjoint analysis che riguarda sempre i pesi. Se prendiamo come esempio il mercato dei telefoni, supponiamo di aver capito che avere 128GB di memoria fa la differenza che tutti preferiscono avere un telefono con la capacità maggiore. Il fatto che i consumatori preferiscono la memoria maggiore e che attribuiscono un valore maggiore al telefono con memoria elevata rispetto ad uno con memoria inferiore, in che modo le imprese potrebbero cogliere questo diverso valore dato dai consumatori in termini di altre politiche di marketing?

A cosa potrebbe essere utile questa informazione? Se si è a conoscenza

dell'incremento di valore del telefono quando c'è una capacità di memoria maggiore,

si può anche ipotizzare che ci sia un diverso prezzo che il consumatore è disposto a pagare per una memoria superiore. Quindi la conjoint analysis permette anche di determinare qual è il livello che il consumatore sarebbe disposto a pagare per un prezzo di livello superiore. Un output collaterale della conjoint sarebbe oltre a quello di capire l'inversione, anche quello di capire quanto di più è disposto a pagare il consumatore per prodotti che hanno caratteristiche del prodotto stesso che sono un po' diverse.

## **b) LE FUNZIONI EDONICHE**

La seconda tecnica statistica utilizzabile per capire la logica di scelta del consumatore è rappresentata dalle funzioni edoniche. In questo caso non si prende in considerazione il consumatore, non gli si sottopone nulla ma si lavora con i prodotti che già esistono nel mercato. Le funzioni edoniche ipotizzano che il prezzo di un prodotto dipenda/rifletta dalle caratteristiche del prodotto stesso. Nelle funzioni edoniche si hanno i prodotti reali.

$$\text{Prezzo} = \alpha_0 + \alpha_1 X_1 + \alpha_2 X_2 + \alpha_3 X_3 + \dots + \varepsilon$$

Nelle funzioni edoniche, il modello è una regressione.

La differenza tra conjoint analysis e le funzioni edoniche è che sono entrambe due tecniche basate sul modello regressivo ma nelle funzioni edoniche la y è sempre il prezzo, cioè quello che sono disposti a pagare i consumatori, il valore e nella conjoint analysis invece il prezzo è una delle x che mi determina l'utilità totale (perché nella conjoint il valore è la disponibilità soggettiva a pagare, è il beneficio complessivo, ipotetico e teorico). Nelle funzioni edoniche quello che il consumatore paga dipende da quello che ha.

Nelle funzioni edoniche devo standardizzare le variabili  $x_1$ ,  $x_2$ ,  $x_3$  perché non hanno le stesse unità di misura. Per comprendere quali sono gli attributi più rilevanti, guardo il valore dei coefficienti della mia funzione di regressione cioè i pesi che i consumatori danno alle caratteristiche del prodotto nella loro logica di scelta.

Infatti gli studi di Barbarito L. (2015) ci portano ad affermare che le funzioni edoniche sono delle regressioni in cui la variabile dipendente  $y$  è il prezzo.

La composizione del prezzo viene stimata attraverso una serie di altre variabili. In economia la regressione serve anche per studiare la relazione fra due o più fenomeni. Viene inoltre utilizzata per realizzare una previsione delle vendite, per stimare la quantità domandata di un prodotto (che a sua volta dipende dal prezzo). I modelli che si costruiscono possono avere scopo descrittivo, interpretativo e previsivo.

Una relazione semplice scritta in forma estesa è  $Y_i = b_0 + b_1 x_i + E_i$  dove  $b_0$  è l'intercetta ovvero il valore di  $Y$  quando  $X$  è uguale a zero. La differenza tra le  $y$  reali e le  $y$  stimate sono chiamate residui. La migliore retta di regressione è quella che rende minima questa differenza.

La varianza totale (SST) riflette la variabilità del carattere  $Y$  (come se non fosse considerata la variabile  $X$ ). Se fosse pari a zero  $Y$  non presenta variabilità ( $Y - Y_{me}$ ).

La varianza di regressione (SSR) è media delle distanze al quadrato tra i punti  $y_i^*$  (sulla retta di regressione) e i punti sulla retta parallela all'asse delle ascisse  $Y = Me(Y)$ : esprime la parte di variabilità "spiegata" dal legame lineare individuato del carattere  $Y$  dal carattere  $X$ .

Se fosse uguale a zero la relazione di regressione non riduce l'incertezza.

La varianza residua (SSE) è media delle distanze al quadrato tra i punti osservati  $y_i$  e la retta di regressione (ossia i punti  $y_i^*$ ): esprime ciò che rimane della variabilità di  $Y$  dopo aver utilizzato le informazioni fornite dal carattere  $X$  mediante il modello di regressione

(lineare semplice). Se fosse uguale a zero tutti i punti osservati sono sulla retta di regressione.

La correlazione misura la dipendenza tra variabili, cioè quanto “si muovono nella stessa direzione”

(es. quanto più una persona è alta tanto più pesa).

Il coefficiente di determinazione ( $R^2$ ) indica la proporzione di variabilità di

Y spiegata da X tramite il modello di regressione (LA VARIANZA “SPIEGATA” DAL

MODELLO DI REGRESSIONE)  $R^2 = SSR/SST = \sum(y^*-y)^2 / \sum(y_i - \bar{y})^2$  .

Quando  $R^2=0$ ? Quando  $y^*=\bar{y}$  Cioè la retta di regressione coincide con la retta parallela all'asse x con intercetta pari a  $\bar{y}$ .

Quando  $R^2=1$ ? Quando retta di regressione passa esattamente per ogni punto.

La funzione edonica non è altre che la stima di una funzione di regressione in cui  $y$  = prezzo e  $x$

corrisponde agli attributi del prodotto, ovvero potenziali criteri di scelta del consumatore. Trovati i coefficienti beta delle nostre  $x$  possiamo individuare le caratteristiche principali per realizzare una mappa di posizionamento e mappe percettive.

Se le variabili sono espresse in unità di misure diverse, bisogna standardizzarle.

Per standardizzare si Calcola valore / media – scarto.

I due coefficienti più grandi indicano quali sono gli attributi più rilevanti.

### c) MULTIDIMENSIONAL SCALING

La terza tecnica statistica utilizzabile è il multidimensional scaling che consiste in una tabella

a doppia entrata dove vengono messe le stesse variabili sia nelle righe che nelle colonne.

I numeri contenuti nelle singole celle sono le distanze percepite dal consumatore fra questi prodotti quindi più i numeri all'interno sono piccoli, più sono identici i prodotti.

Questa matrice viene definita matrice delle dissimilarità, delle distanze: tanto maggiore

è il numero all'interno tanto maggiore è la distanza soggettiva che il consumatore percepisce tra i due prodotti o i due brand.

Quindi l'obiettivo del multidimensional scaling è quello di sottoporre a tanti consumatori diversi – anche qui campione rappresentativo – la compilazione della matrice e gli viene chiesto sostanzialmente chi è simile a chi. Si suggerisce una scala di dissimilarità 0-10, dove ovviamente lo zero è sinonimo di uguaglianza dei prodotti e il dieci di totale dissimilarità.

La scala non è fissa, può essere anche 0-7 o 0-100.

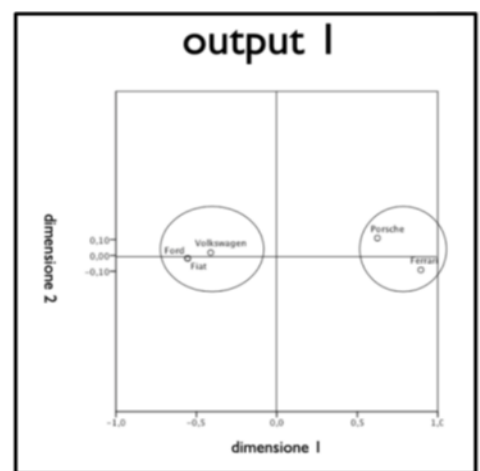
I valori all'interno li inseriscono ovviamente i consumatori secondo la loro opinione soggettiva; il consumatore dà delle distanze sul suo percepito delle differenze.

La dissimilarità è un indice di diversità, di lontananza: tanto più alto è il valore, tanto più diversi sono i valori.

Il multidimensional scaling poi, dopo aver sottoposto la compilazione della matrice ai consumatori, arriva ad una matrice media in cui sono indicate le distanze medie del percepito da parte dei consumatori e le mette su una mappa bidimensionale. La mappa è creata dal computer ed è il primo output del multidimensional scaling. Se, come riportato nell'esempio qui affianco, nella matrice delle dissimilarità la distanza tra Porsche e Ferrari era 2 allora nell'output 1 la distanza grafica dev'essere pari a 2.

Ed ancora, se nella matrice la distanza Ferrari-Fiat era pari ad 8 allora nell'output la distanza grafica tra le due dev'essere quattro volte maggiore rispetto alla distanza grafica tra Porsche-Ferrari.

I cerchi vengono messi poi ex-post per individuare i diversi segmenti di prodotti.



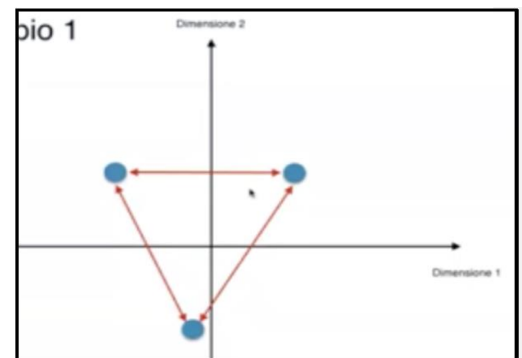
Intuitivamente in base alla posizione dei brand lungo l'asse x si potrebbe dire che la dimensione 1 indica il prezzo o la potenza.

La dimensione 2 invece non è chiara: non c'è una caratteristica che mi permetta di inserire così i brand di automobili.

Il multidimensional scaling quindi intuisce quale sia la logica di scelta dei consumatori sulla base delle distanze soggettive attribuite ai prodotti dai consumatori

stessi. La tecnica del multidimensional scaling a volte può non funzionare quando una mappa a due dimensioni non è sufficiente.

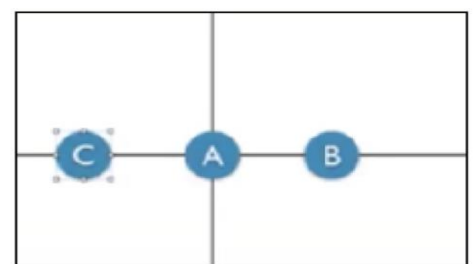
In realtà non è sempre possibile avere delle distanze stimate nella mappa che corrispondono esattamente alle distanze di una



matrice. Per esempio, se si suppone di avere una matrice con 3 prodotti e che ciascuno di essi sia distante 2 dagli altri due, allora ci aspetteremmo un output in cui i 3 prodotti formano un triangolo equilatero di lato 2.

La posizione del triangolo nel piano cartesiano non è predefinita, l'importante è che le distanze fra i tre punti siano rispettate.

Se invece fossimo nel caso in cui i 3 prodotti avessero le seguenti distanze: A è distante sia da B che da C 2, mentre B e C sono distanti 4. L'output non presenterebbe più un



piano cartesiano con un triangolo ma un piano cartesiano in cui i 3 prodotti sono disposti semplicemente in una linea in cui disposti secondo l'ordine B-A-C o C-A-B, come rappresentato in figura.

Ultimo caso: qualora A è distante sia da B che da C 2, mentre B e C sono distanti 6.

Da questo esempio capisco che non è possibile trasferire questa matrice in una mappa bidimensionale perché la mappa non rispecchierà mai le distanze reali inserite nella matrice.

Questo dice che il multidimensional scaling a volte può non funzionare, quando non si riesce a trasportare bene le distanze da una matrice iniziale ad una mappa bidimensionale.

Per identificare la qualità di una trasposizione di una matrice su di una mappa

posso utilizzare lo stress di Kruskal. Lo stress di Kruskal è una misura che della bontà

del Multidimensional scaling. Lo stress di Kruskal è il rapporto fra la sommatoria delle

distanze sulla matrice rispetto alle distanze stimate sulla

mappa e la sommatoria delle distanze sulla matrice

rispetto alle distanze medie fra le osservazioni.

$d_{ij}$  = distanza media =  $\sum d_{ij} / n$

$d^*_{ij}$  = distanza stimata tra due osservazioni

$$\text{Stress} = \left( \frac{\sum_{i < j} [d_{ij} - d^*_{ij}]^2}{\sum_{i < j} d_{ij}^2} \right)^{1/2}$$

Quando lo stress è uguale a zero vuol dire che il modello è ottimo e la mia mappa rappresenta al 100% la mia matrice.

Tanto maggiore è il denominatore – quindi tanto più le distanze sulla mappa sono diverse dalle distanze sulla matrice – tanto più aumenta la misura dello stress.

Il valore massimo che può assumere lo stress è pari ad 1 ed in tal caso significa che

la mappa è inaffidabile rispetto alla matrice.

Non c'è un valore soglia dello stress di Kruskal per decidere se l'analisi va bene o meno; se la mappa non rispecchia bene, non utilizziamo quest'analisi.

Si può ottenere la matrice dell'output del multidimensional scaling anche in un altro modo, non partendo dall'opinione dei consumatori ma partendo da dei dati oggettivi dei prodotti.

Per esempio, prendo 5 automobili e considero 5 variabili. Dopo aver standardizzato, ossia ripulito per



unità di misura, nella cluster analysis è possibile calcolare le misure delle distanze euclidee.

Le distanze euclidee sono anch'esse delle misure di dissimilarità, quindi quando il valore della distanza euclidea fra due prodotti è pari a zero significa che entrambi riportano gli stessi valori delle variabili considerate. Tanto maggiore è la distanza, tanto più i due prodotti sono diversi. Le distanze euclidee sono l'equivalente oggettivo della distanza soggettiva inserita dai consumatori in precedenza con il multidimensional scaling.

Queste misure oggettive riportate dalle distanze euclidee derivano dall'aver standardizzato i valori ed aver calcolato un'unica misura della distanza di tutti con tutti che metto su questa matrice.

La conseguenza della matrice delle distanze euclidee è un output che, seppur essendo molto simile rispetto a quello prodotto dal multidimensionale scaling, è basata sulle caratteristiche oggettive.

L'output 1 di prima era basato sulle opinioni dei consumatori.

	Prezzo (€)	Potenza (Kw)	Consumo (lx100Km)	Emissioni CO2 (g/km)	Bagagliaio (dm <sup>3</sup> )
Ferrari	195.193	360	13,1	299	340
Fiat	19.350	62	4,2	110	1310
Ford	22.250	110	6	139	1148
Porsche	92.495	257	9	212	135
Volkswagen	21.300	90	5,2	120	1270

**le distanze euclidee**

	Ferrari	Fiat	Ford	Porsche	Volkswagen
Ferrari	0				
Fiat	3,6	0			
Ford	3,1	0,5	0		
Porsche	1,6	2,4	1,9	0	
Volkswage	3,3	0,2	0,3	2,1	0

#### **d) CLUSTER ANALYSIS**

La Cluster Analysis è una tecnica statistica di marketing quantitativo, che consente di creare gruppi da un insieme di osservazioni.

Spesso utilizzata per identificare gruppi di consumatori, la Cluster analysis viene considerata un raggruppamento di dati con caratteristiche simili: i membri sono simili tra loro e diversi dagli altri.

Il primo passo da compiere per realizzare una cluster analysis è effettuare la standardizzazione delle variabili per colonna affinché si abbia varianza 1 e media 0, per cui i numeri che si osservano sono numeri di deviazioni standard rispetto alla media. Se l'unità di misura è la stessa non è necessario standardizzare.

Non si può affermare che la cluster analysis crei dei segmenti totalmente oggettivi perché dipende dalle variabili che vengono prese in considerazione per l'analisi, e la scelta è già soggettiva.

Questa tecnica statistica è apparentemente oggettiva, ma in realtà è soggettivo il modo in cui vengono considerati simili i consumatori.

La cluster analysis procede standardizzando e la cluster gerarchica, che prende i consumatori vicini tra loro e li considera un gruppo perché complessivamente sono simili, utilizzano principalmente la distanza euclidea.

Quando si hanno a disposizione solo due variabili, A e B, si può calcolare con il teorema di Pitagora: spiega che il quadrato costruito sopra l'ipotenusa è uguale alla somma dei quadrati costruiti sui due cateti. Se aumento però le variabili, non posso più utilizzare il teorema di Pitagora ma applico la stessa logica.

La cluster si conclude con un dendrogramma che ha lo scopo di individuare i gruppi di consumatori simili fino ad arrivare ad un unico grande gruppo.

Il numero di cluster da scegliere è soggettivo ma bisogna osservare la distanza media intra-cluster,

quest'ultima viene considerata una misura della variabilità media delle osservazioni che appartengono allo stesso gruppo. Per trovare il numero di gruppi ottimale è necessario osservare come e quanto aumenta e dove si nota un salto rilevante dell'incremento percentuale, cioè quando si aggregano gruppi largamente diversi tra loro.

Il Cluster che viene considerato il migliore è quello che ha distanza minima tra i suoi membri e distanza massima con i membri degli altri cluster.

Alcune dimensioni strategiche su cui basare la similitudine, possono essere il livello di prezzo, il livello della qualità, il grado di penetrazione distributiva, il grado di penetrazione geografica (locale, nazionale, internazionale), il numero di prodotti offerti o i segmenti serviti, il livello di spese pubblicitarie e il livello di tecnologia utilizzata.

Non è negativo se nell'analisi della cluster utilizzo delle dimensione correlate, perché se le performance sono rilevanti, posso tenerle in considerazione.

La cluster analysis non ha il problema della multicollinearità come nella regressione.

Le variabili correlate non influenzano negativamente ma bisogna prestare attenzione se le dimensioni vengono misurate due volte.

## CLUSTER ANALYSIS SUL MERCATO DELL'AUTOMOTIVE

Ora andremo a prendere in considerazione una determinato numero di autovetture di diversi brand automobilistici, successivamente effettueremo delle cluster analysis per identificare diversi raggruppamenti utilizzando variabili quantitative specifiche per ciascun campione di auto preso in considerazione.

BRAND	VOTO MEDIO UTENTI 0a5	PREZZO BASE	PESO (Kg)	EMISSIONI CO2 G/KM	CARBURANTE	CONSUMO MEDIO L/100 KM	CILINDRATA Cm	AUTO IMMATRICOLATE 2019 (GENNAIO/MAGGIO)
FIAT 500	3,9	16050	900	115	Benzina		5 1200	19699
TOYOTA Aygo	3,1	12000	873	93	Benzina		4 1000	6483
RENAULT Twingo	3,3	11500	914	100	Benzina		4 1000	3349
VOLKSWAGEN Up	2,9	13600	915	100	Benzina		5 1000	7864
PEUGEOT 108	2,9	12530	865	87	Benzina		3 1000	3005
RENAULT Clio	3,5	15000	1062	112	Benzina		4 1000	23058
VOLKSWAGEN Polo	3	16950	1045	106	Benzina		5 1000	19369
PEUGEOT 208	3,9	15150	980	94	Benzina		5 1200	15797
VOLKSWAGEN T-Cross	2,6	19200	1170	116	Benzina		5 1000	3859
FORD Fiesta	3,5	17050	1082	102	Benzina		4 1100	15050
FIAT 500X	3,5	20500	1320	122	Benzina		5 1000	21128
JEEP Renegade	3,5	23300	1320	130	Benzina		5 1000	19135
VOLKSWAGEN T-Roc	2,7	24000	1237	119	Benzina		5 1000	20402
PEUGEOT 3008	3,6	27830	1320	113	Benzina		5 1200	12387
FORD Ecosport	2,6	20500	1274	116	Benzina		5 1000	13875
VOLKSWAGEN Tiguan	2,7	31150	1435	132	Benzina		6 1500	10870
AUDI Q3	3	35000	1495	132	Benzina		6 1500	5217
BMW X1	3,3	33000	1400	123	Benzina		5 1500	5415
TOYOTA Rav4	3,1	36400	1590	104	Ibrida-Benzina		4 2500	4824
ALFA ROMEO Stelvio	3,8	47500	1660	138	Diesel		5 2200	5242
MERCEDES GLE	3	70312	2090	161	Diesel		6 2000	309
AUDI A6	3,3	52750	1645	107	Ibrida-Gasolio		5 2000	1952
BMW Serie 5	3,7	53550	1530	129	Benzina		5 2000	2055
PORSCHE Cayenne	2,9	80030	1985	210	Benzina		11 3000	458
JAGUAR F-Pace	3,6	48275	1685	140	Diesel		6 2000	988
PORSCHE 911	4,2	110139	1505	206	Benzina		9 3000	336
MASERATI Ghibli	4,2	72200	1800	184	Diesel		7 3000	293
FERRARI 488	4,6	296000	1385	282	Benzina		12 4000	90
JAGUAR F-Type	4,1	66400	1520	184	Benzina		9 2000	85
BMW Serie 7	3,6	96200	1840	143	Diesel		5 3000	161

Le variabili scelte come riportate qui sopra sono: il voto medio degli utenti, il prezzo di base, il peso, le emissioni, il consumo medio, la cilindrata, e il numero di auto immatricolate nel periodo tra gennaio e Maggio 2019.

Le variabili pesano tutte in egual modo, e dunque variabili influenti sono importanti quanto quelle più maggiormente influenti, riuscendo così a interferire nel risultato finale.

E' presente quindi anche nella cluster una certa discrezionalità nella scelta delle variabili, con la differenza che se ritenessimo rilevanti anche più di due, potremmo utilizzarle nella creazione dei gruppi. Le auto che sono state prese in considerazione toccano tutti i segmenti presenti nel mercato,

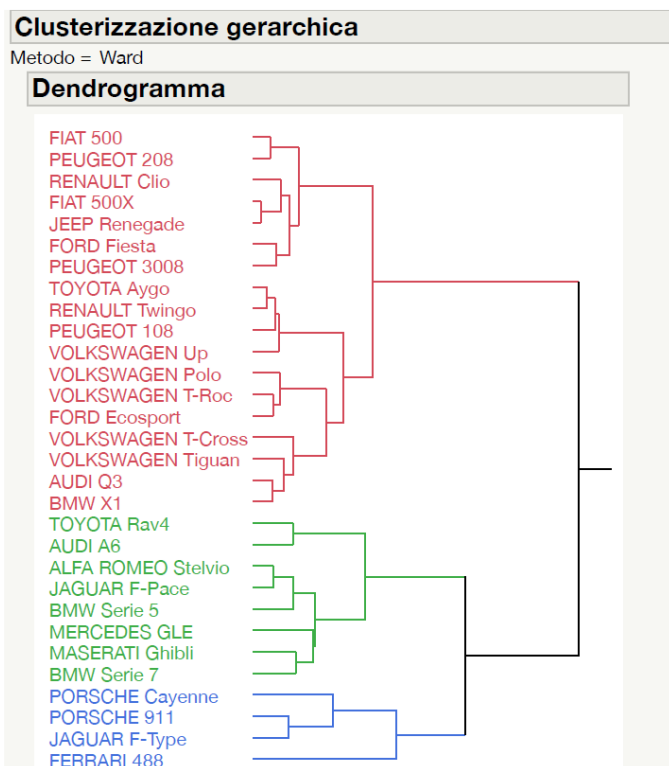
sottolineando che sono tra i modelli più venduti dalle rispettive case nel periodo che va da gennaio a maggio 2019. Abbiamo utilizzato poi un prezzo medio dei suoi modelli base, l'immagine è espressa in una scala da 1 a 5 dove il valore più alto corrisponde un miglior apprezzamento qualitativo da parte dei consumatori, I valori trovati derivano da una ricerca su un campione di consumatori.

Abbiamo poi convertito le diverse tipologie di carburante con i valori 1-2-3-4, nell'ordine di benzina, diesel, ibrida-benzina e ibrida-diesel poichè si tratta di una variabile qualitativa ed era necessario trasformarla in quantitativa.

In seguito abbiamo standardizzato le variabili, perchè in questo caso sono espresso in unità di misura diverse. Allo stesso tempo abbiamo scelto la misura della distanza, quella euclidea che è preferibile quando le variabili sono continue.

La procedura gerarchica parte da un numero di gruppi uguale al numero di osservazioni, ad ogni stadio dell'analisi, diminuisce di un unità il numero dei gruppi.

Nello stadio finale tutti i modelli sono raggruppati in unico gruppo.



Come possiamo osservare al primo stadio vengono aggregate le due osservazioni più vicine che dal

dendogramma sono 500X e Renegade, poi si aggregano Toyota Aygo e Renault Twingo, il passaggio successivo è tra fiat 500 e Peugeot 208, si continuano aggregare come riportato nella tabella sottostante.

<b>Cronologia di clusterizzazione</b>			
<b>Numero di cluster</b>	<b>Distanza</b>	<b>Leader</b>	<b>Subordinato</b>
29	0,22770460	FIAT 500X	JEEP Renegade
28	0,41981071	TOYOTA Aygo	RENAULT Twingo
27	0,52502038	FIAT 500	PEUGEOT 208
26	0,59136934	AUDI Q3	BMW X1
25	0,61382198	ALFA ROMEO Stelvio	JAGUAR F-Pace
24	0,61723820	VOLKSWAGEN T-Roc	FORD Ecosport
23	0,67913208	TOYOTA Aygo	PEUGEOT 108
22	0,70960447	FORD Fiesta	PEUGEOT 3008
21	0,78508509	TOYOTA Aygo	VOLKSWAGEN Up
20	0,81810510	VOLKSWAGEN Polo	VOLKSWAGEN T-Roc
19	0,85008492	RENAULT Clio	FIAT 500X
18	0,94392929	VOLKSWAGEN Tiguan	AUDI Q3
17	1,09760569	PORSCHE 911	JAGUAR F-Type
16	1,13234145	RENAULT Clio	FORD Fiesta
15	1,22209142	TOYOTA Rav4	AUDI A6
14	1,24019564	VOLKSWAGEN T-Cross	VOLKSWAGEN Tiguan
13	1,24172791	ALFA ROMEO Stelvio	BMW Serie 5
12	1,31105649	MASERATI Ghibli	BMW Serie 7
11	1,39852339	FIAT 500	RENAULT Clio
10	1,85418034	MERCEDES GLE	MASERATI Ghibli
9	1,89732528	ALFA ROMEO Stelvio	MERCEDES GLE
8	2,24948891	VOLKSWAGEN Polo	VOLKSWAGEN T-Cross
7	2,47757426	PORSCHE Cayenne	PORSCHE 911
6	2,79259591	TOYOTA Aygo	VOLKSWAGEN Polo
5	3,47207007	TOYOTA Rav4	ALFA ROMEO Stelvio
4	3,70524858	FIAT 500	TOYOTA Aygo
3	4,43309247	PORSCHE Cayenne	FERRARI 488
2	6,55460376	TOYOTA Rav4	PORSCHE Cayenne
1	10,06753696	FIAT 500	TOYOTA Rav4

Il primo cluster è quello a cui appartengono i modelli di auto più comuni con un maggior numero di auto immatricolate, con dimensioni minori e con prezzi medio bassi, infatti è composta da un numero di osservazioni ben superior rispetto alle alter cluster poichè prende in considerazione una quota di mercato ben maggiore.

A differenza di quelli più borderline come Audi Q3 e BMW X1 che riportati alla reale suddivisione possono appartenere anche alla seconda cluster.

Nella seconda cluster appunto troviamo, un numero inferiore di osservazioni rispetto alla prima, le caratteristiche infatti si differenziano dalla prima, la quota di mercato rappresentata da questi modelli

è ben più bassa poichè composta da auto considerate sportive, con prezzi medio alti e con dimensioni più elevate.

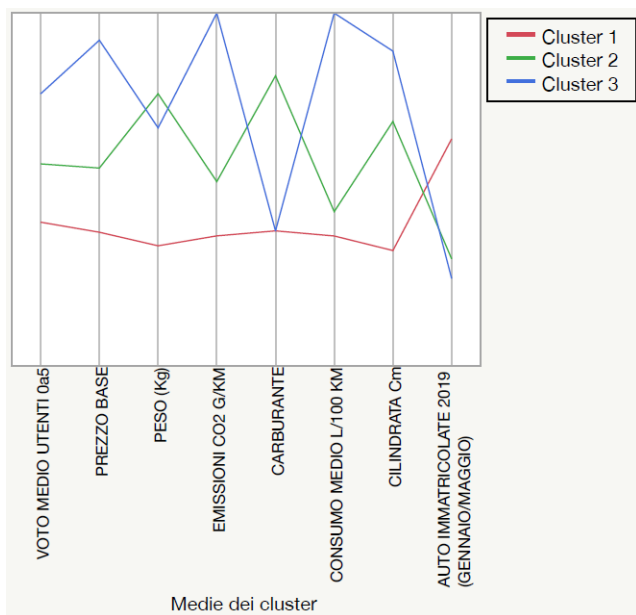
Nella terza cluster si riduce ulteriormente il numero di osservazioni, rappresenta un segmento elitario e di lusso nel mercato dell'automobile, la sua quota è molto più bassa rispetto a quelle precedenti, visto l'esponenziale aumento dei prezzi.

Si ricorda che la distanza media intra-cluster è una misura dell'eterogeneità media dei gruppi.

Più si procede nell'aggregazione, scendendo con il numero di gruppi, tanto più aumenta l'eterogeneità media intra-gruppo, la misura dell'incremento da uno stadio all'altro esprime il fatto che abbiamo aggiunto un'osservazione più o meno vicino a un gruppo precedente. Tanto più uniamo cose simili e tanto minore sarà l'incremento coefficiente.

Possiamo osservare qui sotto, le medie trovate delle variabili prese in considerazione per ogni cluster eseguita, e l'andamento grafico di queste.

Medie dei cluster										
Cluster	Conteggio	VOTO MEDIO		PREZZO BASE	PESO (Kg)	EMISSIONI		CONSUMO MEDIO		AUTO IMMATICOLATE 2019 (GENNAIO/MAGGIO)
		UTENTI 0a5				CO2 G/KM	CARBURANTE	L/100 KM	CILINDRATA Cm	
1	18	3,2		20239,4	1144,83	111,778	1,00	4,8	1122,2	12553,4
2	8	3,5		59648,4	1730,00	138,250	2,25	5,4	2337,5	1978,0
3	4	4,0		138142,3	1598,75	220,500	1,00	10,3	3000,0	242,3



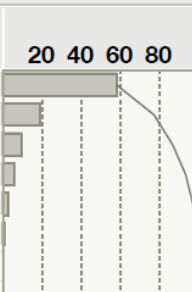
## ANALISI DEI COMPONENTI PRINCIPALI

Successivamente è stata eseguita un'analisi dei componenti principali, questa tecnica statistica viene usata per studiare le relazioni tra un ampio numero di variabili e per spiegare queste variabili in termini di dimensioni comuni, ovvero altre variabili, diverse dalla matrice iniziale.

Queste variabili hanno la capacità di ricostruire quelle originali, le componenti principali sono quindi delle variabili latenti in grado di riprodurre le variabili reali della matrice iniziale. Bisogna appuntare però che non è esattamente possibile ricostruire perfettamente la matrice di partenza ma l'analisi sarà tanto migliore quanto più la matrice che verrà rigenerata sarà simile a quella reale.

Con questa tecnica è stato possibile sintetizzare le 8 variabili iniziali in 2 componenti principali, se Venisse utilizzato un solo componente si riuscirebbe a spiegare il 58,36% della varianza complessiva, ma utilizzandone due siamo in grado di raggiungere il 77,76% della varianza della matrice iniziale, per considerare un'analisi dei componenti principali efficace è consigliato prendere in considerazione modelli che spieghino almeno il 65% della varianza totale.

Come possiamo infatti notare dalla tabella qui sotto, i primi due autovalori superano il valore 1, e sono i componenti da tenere in considerazione.

Autovalori								
Numero	Autovalore	Percentuale	20 40 60 80	Percentuale cumulativa	Chi-quadrato	DF	Prob>ChiQu	
1	4,6688	58,360		58,360	235,492	26,219	<,0001*	
2	1,5519	19,399		77,759	135,090	26,227	<,0001*	
3	0,7683	9,603		87,363	89,932	20,718	<,0001*	
4	0,4862	6,078		93,440	63,960	15,023	<,0001*	
5	0,3031	3,789		97,229	41,893	9,073	<,0001*	
6	0,1370	1,713		98,942	20,221	4,954	0,0011*	
7	0,0660	0,825		99,767	8,464	1,993	0,0144*	
8	0,0186	0,233		100,000	0,000	.	.	

Nell'ultima colonna è stato calcolato il test di Bartlett, il quale misura la probabilità che le variabili iniziali siano correlate tra loro.



Se la probabilità risulta  $<0,05$  le variabili sono correlate e quindi osservando la colonna sono accettabili.

Successivamente è stato sviluppato un diagramma dei pesi ruotato, dove è stato diviso in fattore 1 e fattore 2, il primo indica l'accessibilità mentre il secondo indica l'esclusività questo diagramma attribuisce ciascuna variabile iniziale al primo o al secondo fattore osservando il valore più alto come si può notare qui in basso.

<b>Diagramma dei pesi ruotato</b>		
	Fattore 1	Fattore 2
EMISSIONI CO2 G/KM	0,954509	0,207311
CONSUMO MEDIO L/100 KM	0,914811	0,128374
PREZZO BASE	0,892578	0,242525
CILINDRATA Cm	0,771421	0,584418
VOTO MEDIO UTENTI 0a5	0,674740	-0,003725
PESO (Kg)	0,348977	0,791705
CARBURANTE	-0,206998	0,869588
AUTO IMMATRICOLATE 2019 (GENNAIO/MAGGIO)	-0,383476	-0,686462

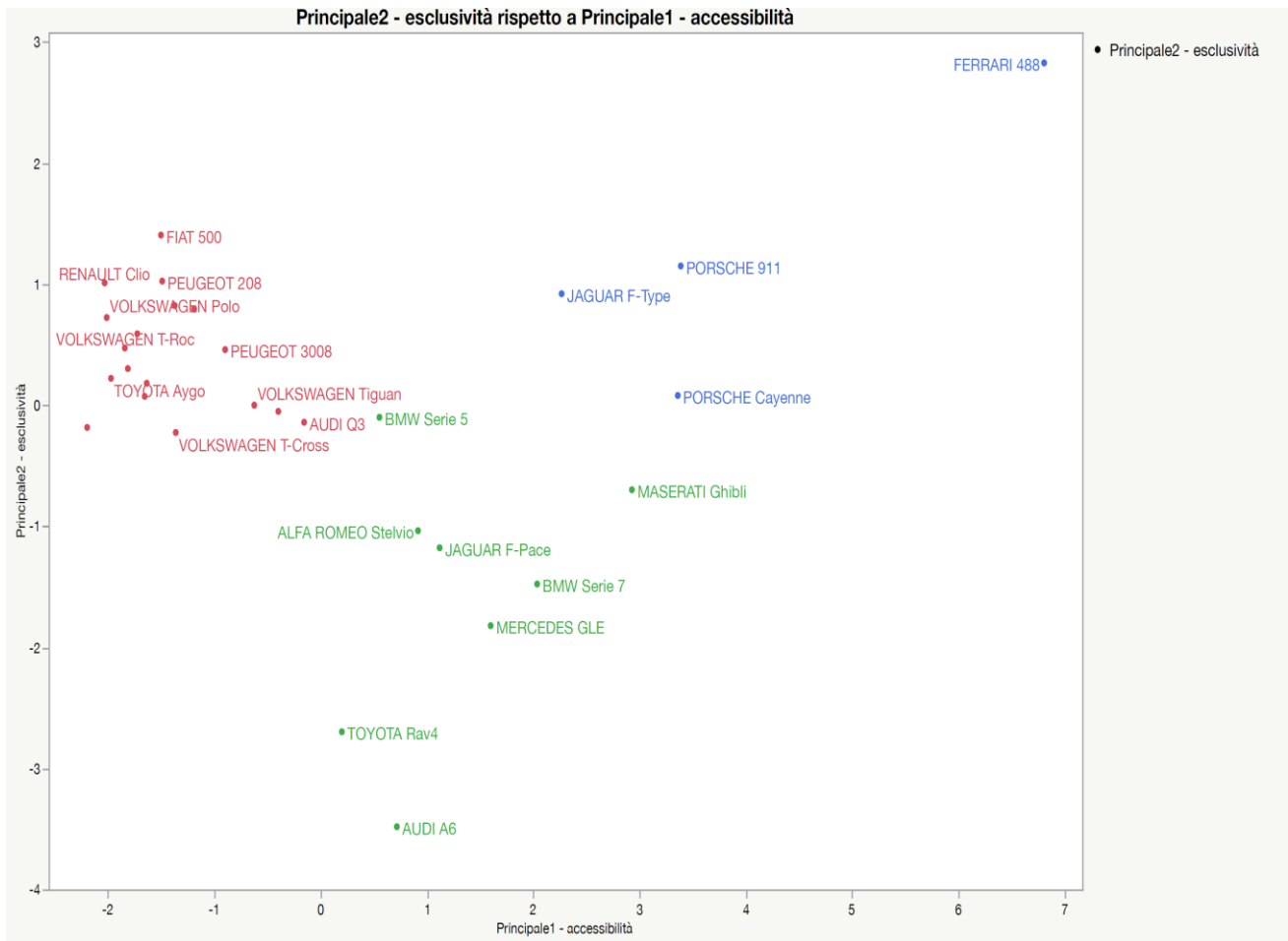
Infine è stato costruito un grafico, si può osservare sul quale è stato posizionato sull'asse delle X il componente principale 1 di accessibilità e sull'asse delle Y il componente principale 2 di esclusività.

Sono stati posizionati poi, in base al punteggio ottenuto, i vari campioni di modelli d'auto.

I valori che sono stati attribuiti ai diversi modelli d'automobile non sono di natura economica, ma hanno utilità per costruire il grafico.

Si può notare che alcune osservazioni nonostante appartengano a gruppi differenti sono molto vicine tra loro, questo perchè non esiste una perfetta distinzione e alcuni modelli hanno molte similitudini con altri appartenenti ad un gruppo strategico diverso.

Il gruppo "verde" che rappresenta in modo sufficientemente realistico il segmento di mercato E, composto da auto di grosse dimensioni, solitamente berline, appartenenti ad una fascia di prezzo decisamente alta.



L'ultimo gruppo strategico, quello "blu" che rappresenta il segmento di auto di lusso è più distanziato rispetto agli altri due, questo evidenzia che i modelli presi in considerazione hanno caratteristiche appartenenti ad un mercato esclusivo, infatti la quota di mercato è molto più bassa rispetto agli altri.

Un particolare da non tralasciare è l'isolamento di un modello d'auto, "Ferrari 488", questo viene denominato "outlier, termine usato per definire un campione dal valore molto diverso rispetto alla popolazione da cui è estratto", osservando le variabili "Prezzo" e "Auto immatricolate" si riscontra che il primo è molto più alto rispetto alla media degli altri modelli e il secondo è tra i più bassi rispetto alla media ottenuta.

Si può affermare, in conclusione che anche l'analisi dei componenti principali ha dato buoni risultati, questo è avvenuto grazie anche ai risultati ottenuti con la cluster analysis, con la quale siamo stati in grado di suddividere i gruppi.

## **CONCLUSIONE**

Questo studio è nato con l'obiettivo di sperimentare un modello statistico in particolare, ovvero la cluster analysis nel settore automobilistico. E' stata applicata questa tecnica di marketing quantitativo per ottenere una suddivisione in gruppi di modelli di autovetture con caratteristiche simili tra loro.

L'esperimento è riuscito con successo poichè i gruppi ottenuti rispecchiano i reali macro-gruppi dei diversi segmenti di mercato dell'automobile.

Successivamente è stata eseguita un'analisi dei componenti principali, per spiegare le variabili iniziali in termini di dimensioni comuni.

La Cluster Analysis e l'analisi dei componenti principali sono state eseguite utilizzando un software statistico chiamato "JMP", terminando il tutto con una mappa di posizionamento strategico che integra Cluster analysis e analisi dei componenti principali.

## **BIBLIOGRAFIA**

Barbarito Luca, “L'analisi competitiva. Metodologia e applicazioni”, Franco Angeli,

10 marzo 2015

Scott W.G.- Santagostino M., “Impresa e mercato”, I.S.U.- Università Cattolica, Milano, 1984

Valdani Enrico, “Definizione segmentazione del mercato per i beni industriali e di largo consumo”,

Giuffrè Editore, Milano 1984

## **SITOGRAFIA**

Alvolante, Maggio (2020)

[https://www.alvolante.it/listino\\_auto/](https://www.alvolante.it/listino_auto/)

Arboretti Giancristofaro R., Bonnini S.,

<http://static.gest.unipd.it/~salmaso/Lez-CJA.pdf>

Toscano Giuseppe, “Conjoint Analysis”, 2010, Università Cattaneo Castellanza (VA)

<http://my.liuc.it/MatSup/2010/F83016/12%20-%20Conjoint%20Analysis.pdf>

Unrae (2020) “Dati statistici immatricolazioni autovetture per segmento”, Maggio 2020

<http://www.unrae.it/dati-statistici/immatricolazioni/5036/top-10-per-segmen-to-maggio-2020>

