

Segmentazione del mercato

Marketing strategico e tattico

- Scopo del marketing: rispondere a bisogni ed esigenze dei consumatori attraverso adeguata offerta
- Importanza del 'marketing planning': marketing **tattico** e **strategico**
- Prima bisogna sapere dove si vuole andare e poi si decide come farlo!
- Valore dell'analisi SWOT

Marketing strategico e tattico

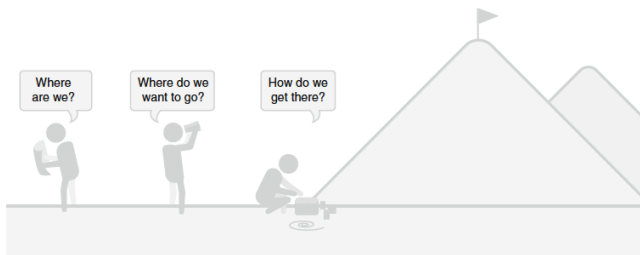


Figura: Marketing strategico e tattico

Marketing strategico e tattico

- Una volta stabiliti punti di forza ed eventuali debolezze, vanno fatte due scelte cruciali per il lungo periodo:
- Su quali consumatori focalizzare l'attenzione → **segmentazione** e **targeting**
- Quale immagine si vuole dare della propria azienda → **posizionamento**
- Solo a questo punto il marketing tattico può essere sviluppato: secondo quali modalità? 4P (product, price, place, promotion)
- Importante: **esiste una asimmetria tra marketing tattico e strategico!**

Marketing strategico e tattico



Figura: Marketing strategico e tattico: asimmetria

Definizione di segmentazione

- Segmentazione del mercato: elemento cruciale per il marketing strategico
- Aziende di successo sviluppano il proprio business sulla segmentazione
- Caratteristiche dei consumatori sono critiche per definire la segmentazione → **criteri di segmentazione**
- Strategie differenti: concentrata o differenziata . . . quali differenze?

Segmentazione: benefici e costi

La segmentazione può comportare sia benefici che costi ... quali?

- Vantaggio competitivo basato su una migliore conoscenza della propria organizzazione
- Maggiore Return on Investments: non si investe su segmenti poco raggiungibili
- Migliore livello organizzativo ... ma
- Se la segmentazione non viene realizzata bene può rivelarsi un totale spreco di risorse!

Segmentazione: analisi

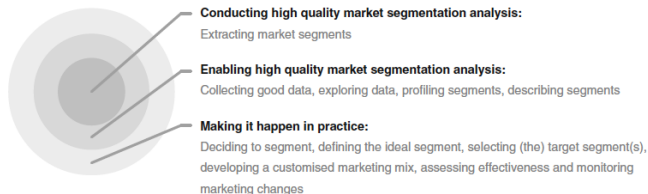


Figura: Processo di segmentazione

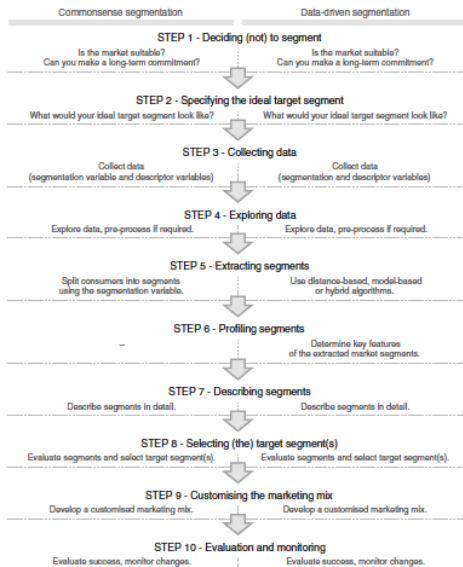
- Il processo di segmentazione è tipicamente di natura **statistica**, ma anche **esplorativa**.
- Molte decisioni devono essere prese dall'analista → conseguenze sulla segmentazione . . .
- **Importante che l'analista sia affiancato da qualcuno che comprende la 'mission' dell'organizzazione.**

Variabili di segmentazione

Variable	Dimensions	Sample survey question
Age	Unidimensional	How old are you?
Gender	Unidimensional	Are you female or male?
Country of origin	Unidimensional	Where do you live?
Prior purchase	Unidimensional	Have you booked a cruise trip before?
Benefits sought	Multidimensional	When booking flights online, do you care about <ul style="list-style-type: none">• convenience• value for money• speed• ability to compare fares
Motives	Multidimensional	When choosing a vacation, do you want to <ul style="list-style-type: none">• rest and relax• explore new things• meet new people• learn about other cultures• get away from everyday routine

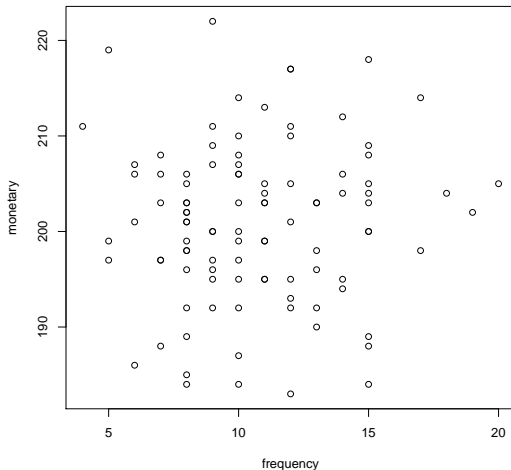
Figura: Esempi di variabili utilizzate per realizzare la segmentazione

Variabili di segmentazione



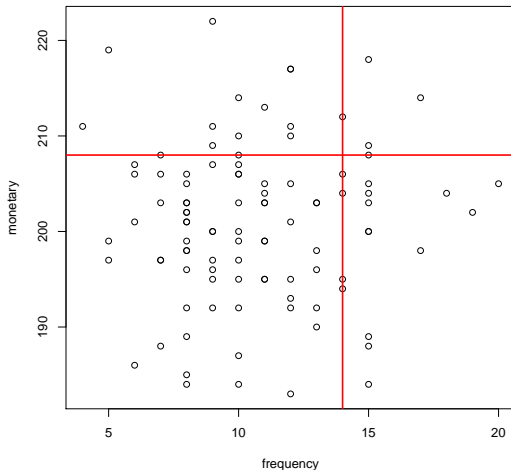
Segmentazione 'triviale'

Clienti ottimi, standard e cattivi



Segmentazione 'triviale'

Clienti ottimi, standard e cattivi



Analisi RFM (recency, frequency, monetary)

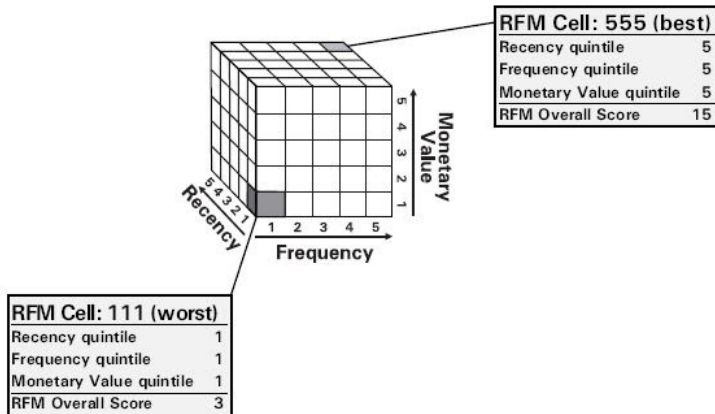
- L'analisi Recency, Frequency, Monetary è una tecnica normalmente utilizzata per condurre una segmentazione
- Ogni cliente riceve una valutazione (score)
- Si basa su 3 metriche:
- **Recency**: quanto di recente ha comprato il cliente? quando è stata l'ultima volta che il cliente ha fatto un ordine?
- **Frequency**: quanto spesso compra? con quale regolarità il cliente fa un ordine?
- **Monetary**: quanto denaro spende in un determinato periodo di tempo?
- Analisi esplorativa semplice ma molto utilizzata

Analisi RFM: tre principi

- Clienti che hanno comprato recentemente sono più ricettivi a promozioni di clienti che hanno fatto l'ultimo acquisto molto tempo fa
- Clienti che acquistano frequentemente sono più ricettivi di clienti occasionali
- Clienti che spendono di più sono più ricettivi di clienti che spendono poco
- I clienti migliori sono quelli con valori elevati di RFM
- In qualche modo collegato con il **customer lifetime value**.

- In pratica, è necessario definire un indice RFM
- Ad esempio R potrebbe essere il numero di giorni dall'ultima transazione, F il numero di transazioni negli ultimi 12 mesi, M la somma di denaro spesa negli ultimi 12 mesi.
- Bisogna quindi ordinare i clienti secondo queste tre metriche
- Per ogni ordinamento realizzato, dividere la lista in percentili: ad esempio si possono scegliere 3 o 5 livelli
- A ciascuno di questi ordinamenti bisogna dare un punteggio (score)
- I 3 punteggi forniscono le coordinate per la matrice RFM

Matrice RFM



La decisione di (non) segmentare

- La scelta di effettuare una segmentazione implica una strategia di lungo periodo. . .
- Importanti aspetti: cultura aziendale, risorse (organizzative, umane, finanziarie . . .)
- Vale la pena fare una 'checklist'

Task	Who is responsible?	Completed?
Ensure that a marketing expert is on the team.		<input type="checkbox"/>
Ensure that a data expert is on the team.		<input type="checkbox"/>
Ensure that a data analysis expert is on the team.		<input type="checkbox"/>
Set up an advisory committee representing all affected organisational units.		<input type="checkbox"/>
Ensure that the objectives of the market segmentation analysis are clear.		<input type="checkbox"/>
Develop a structured process to follow during market segmentation analysis.		<input type="checkbox"/>
Assign responsibilities to segmentation team members using the structured process.		<input type="checkbox"/>
Ensure that there is enough time to conduct the market segmentation analysis without time pressure.		<input type="checkbox"/>

Criteri di segmentazione

I segmenti devono soddisfare alcuni criteri:

- Devono essere omogenei
- Devono essere distinti
- Devono essere sufficientemente grandi
- Devono incontrare i punti di forza dell'azienda
- Devono essere formati da membri identificabili
- Devono essere accessibili

La raccolta dei dati

Le variabili di segmentazione possono essere di vario genere, e basate su alcuni criteri.

La segmentazione può essere, ad esempio

- Geografica
- Socio-Demografica
- Psicografica
- Comportamentale

Spesso le variabili utilizzate per la segmentazione derivano da questionari. La selezione delle variabili deve essere fatta con molta attenzione . . . tutte le variabili rilevanti per lo scopo devono essere incluse, ma variabili non necessarie devono essere escluse (noisy or masking variables).

Attenzione ai questionari lunghi e noiosi!

Dimensione del campione

Quando il campione è troppo piccolo, è impossibile determinare il corretto numero di segmenti

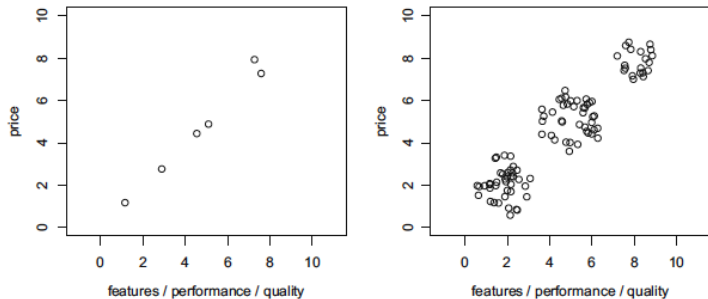


Figura: Importanza della giusta dimensione del campione

Dati per la segmentazione

Una buona segmentazione è basata su dati che:

- contengono tutta l'informazione necessaria
- non contengono informazione ridondante o 'dannosa'
- non contengono variabili correlate
- contengono risposte di alta qualità
- sono di natura metrica o binaria
- non dipendono da 'stili di risposta'
- sono di dimensioni sufficienti. . . 100 volte il numero di variabili coinvolte

Estrazione dei segmenti

Essenzialmente, possiamo distinguere due metodi principali per la segmentazione:

- **Metodi basati sulle distanze**: usano una particolare nozione di **similarità** o distanza tra osservazioni e cercano di trovare gruppi di osservazioni simili (segmenti)
- **Metodi basati su modelli** (model-based): i segmenti vengono rappresentati tramite un modello sintetico
- Spesso una buona soluzione viene ottenuta combinando i risultati ottenuti con più di un metodo

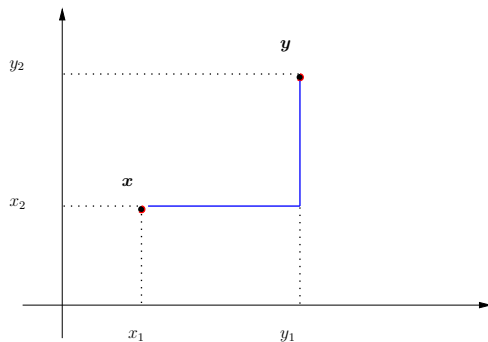
- Distanza di Manhattan (city block distance)

$$d_1(\mathbf{x}, \mathbf{y}) = \sum_{i=1}^p |x_i - y_i|$$

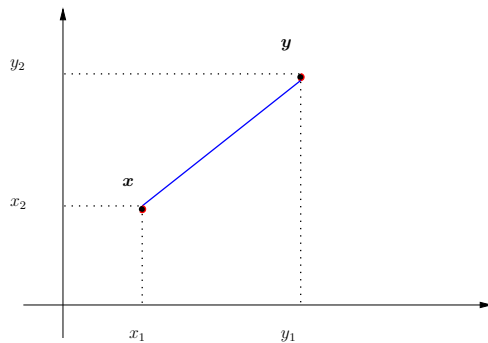
- Distanza Euclidea

$$d_2(\mathbf{x}, \mathbf{y}) = \sqrt{\sum_{i=1}^p (x_i - y_i)^2}$$

Distanza di Manhattan



Distanza Euclidea



Esempio

Consideriamo un semplice esempio: 7 persone e la percentuale di tempo speso in tre attività

	beach	action	culture
Anna	100	0	0
Bill	100	0	0
Frank	60	40	0
Julia	70	0	30
Maria	80	0	20
Michael	0	90	10
Tom	50	20	30

Calcoliamo le distanze fra tutti ...

Metodi gerarchici: esempio

Distanza Euclidea

	Anna	Bill	Frank	Julia	Maria	Michael
Bill	0.00					
Frank	56.57	56.57				
Julia	42.43	42.43	50.99			
Maria	28.28	28.28	48.99	14.14		
Michael	134.91	134.91	78.74	115.76	120.83	
Tom	61.64	61.64	37.42	28.28	37.42	88.32

Metodi gerarchici: esempio

Distanza di Manhattan

	Anna	Bill	Frank	Julia	Maria	Michael
Bill		0				
Frank	80	80				
Julia	60	60	80			
Maria	40	40	80	20		
Michael	200	200	120	180	180	
Tom	100	100	60	40	60	140

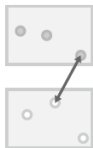
Metodi basati sulle distanze: metodi gerarchici

Consideriamo due insiemi di osservazioni X e Y .

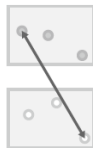
Per calcolare la distanza tra questi insiemi, possiamo considerare:

- **Legame singolo:** $l(X, Y) = \min d(x, y)$
- **Legame completo:** $l(X, Y) = \max d(x, y)$
- **Legame medio:** $l(X, Y) = \frac{1}{k_1 k_2} \sum \sum d(x, y)$

Single linkage



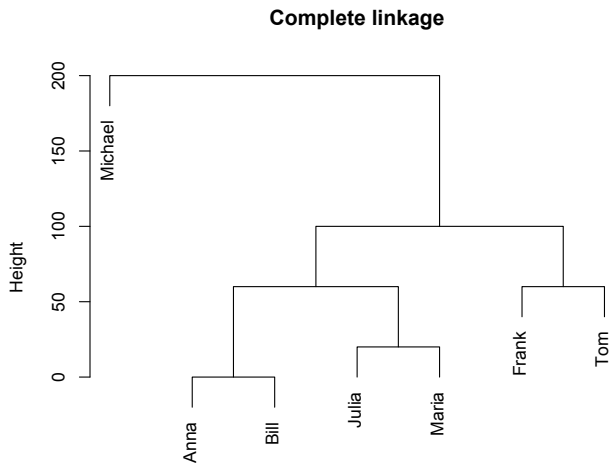
Complete linkage



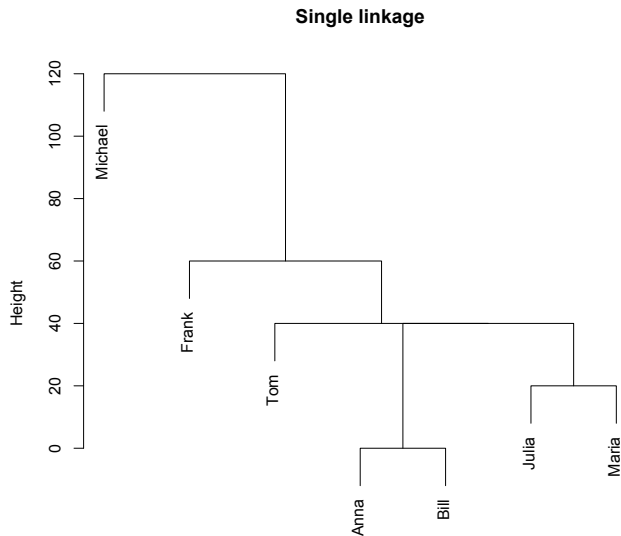
Average linkage



Metodi gerarchici: esempio



Metodi gerarchici: esempio



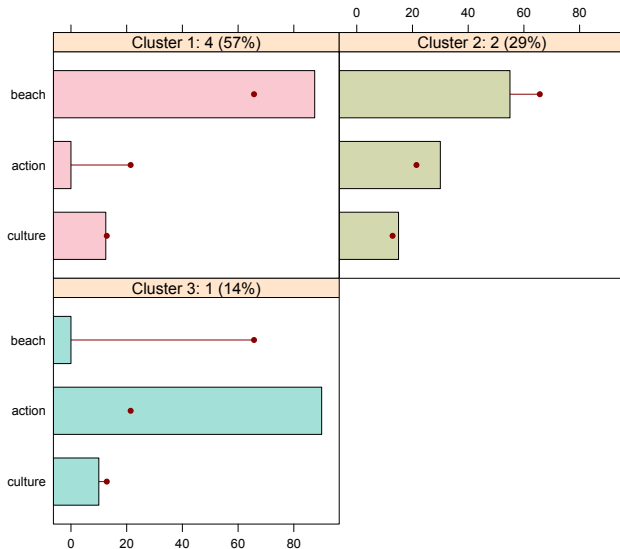
Metodi gerarchici: esempio

- Distanza Euclidea, Legame Completo
- Riteniamo ragionevole la formazione di 3 gruppi.
- Tagliamo il dendrogramma in modo da formare i tre gruppi.

```
c3<- cutree(c11, k = 3)
```

Anna	Bill	Frank	Julia	Maria	Michael	Tom
1	1	2	1	1	3	2

Metodi gerarchici: esempio



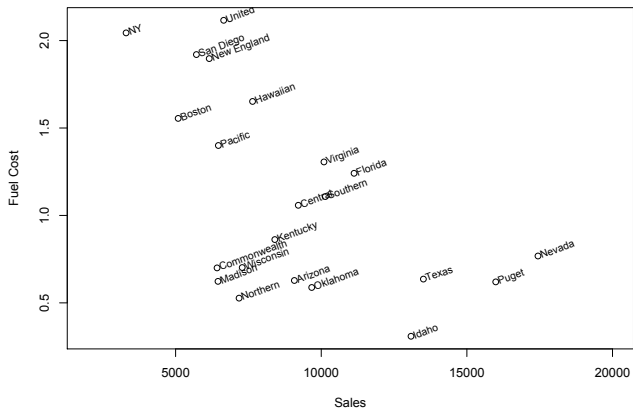
Metodi gerarchici: esempio

- Consideriamo il caso di 22 aziende americane quotate in borsa che operano nel settore dell'energia
- Si dispone per ciascuna azienda di 8 variabili
- Sulla base di queste variabili si vuole fare un'analisi di raggruppamento
- L'analisi di raggruppamento potrebbe servire a identificare una azienda 'tipo' per ciascun gruppo e costruire un modello di costo

Metodi gerarchici: esempio

Company	Fixed_charge	RoR	Cost	Load	Δ Demand	Sales	Nuclear	Fuel_Cost
Arizona	1.06	9.2	151	54.4	1.6	9077	0	0.628
Boston	0.89	10.3	202	57.9	2.2	5088	25.3	1.555
Central	1.43	15.4	113	53	3.4	9212	0	1.058
Commonwealth	1.02	11.2	168	56	0.3	6423	34.3	0.7
Con Ed NY	1.49	8.8	192	51.2	1	3300	15.6	2.044
Florida	1.32	13.5	111	60	-2.2	11127	22.5	1.241
Hawaiian	1.22	12.2	175	67.6	2.2	7642	0	1.652
Idaho	1.1	9.2	245	57	3.3	13082	0	0.309
Kentucky	1.34	13	168	60.4	7.2	8406	0	0.862
Madison	1.12	12.4	197	53	2.7	6455	39.2	0.623
Nevada	0.75	7.5	173	51.5	6.5	17441	0	0.768
New England	1.13	10.9	178	62	3.7	6154	0	1.897
Northern	1.15	12.7	199	53.7	6.4	7179	50.2	0.527
Oklahoma	1.09	12	96	49.8	1.4	9673	0	0.588
Pacific	0.96	7.6	164	62.2	-0.1	6468	0.9	1.4
Puget	1.16	9.9	252	56	9.2	15991	0	0.62
San Diego	0.76	6.4	136	61.9	9	5714	8.3	1.92
Southern	1.05	12.6	150	56.7	2.7	10140	0	1.108
Texas	1.16	11.7	104	54	-2.1	13507	0	0.636
Wisconsin	1.2	11.8	148	59.9	3.5	7287	41.1	0.702
United	1.04	8.6	204	61	3.5	6650	0	2.116
Virginia	1.07	9.3	174	54.3	5.9	10093	26.6	1.306

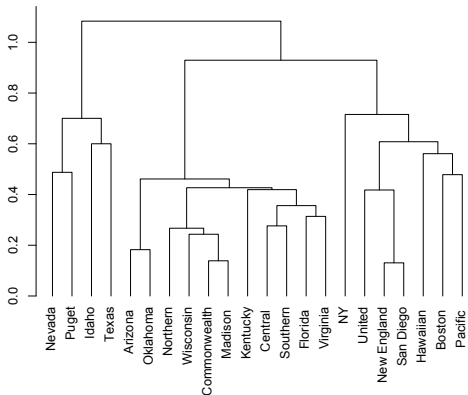
Metodi gerarchici: esempio



La possibile esistenza di gruppi viene già colta da un grafico che descrive la relazione tra vendite (sales) e costo del combustibile (fuel cost)

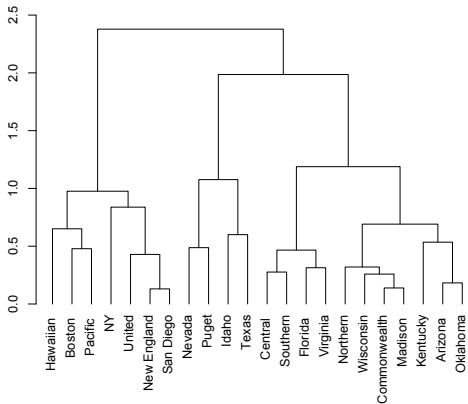
Metodi gerarchici: esempio

Legame singolo



Metodi gerarchici: esempio

Legame medio



Metodi gerarchici: esempio

Legame medio, con $k = 6$

Arizona	Boston	Central	Commonwealth	NY	
1	2	3	1	4	
Florida	Hawaiian	Idaho	Kentucky	Madison	
3	2	5	1	1	
Nevada	New England	Northern	Oklahoma	Pacific	
6	4	1	1	2	
Puget	San Diego	Southern	Texas	Wisconsin	
6	4	3	5	1	
United	Virginia				
4	3				

Metodi gerarchici: esempio

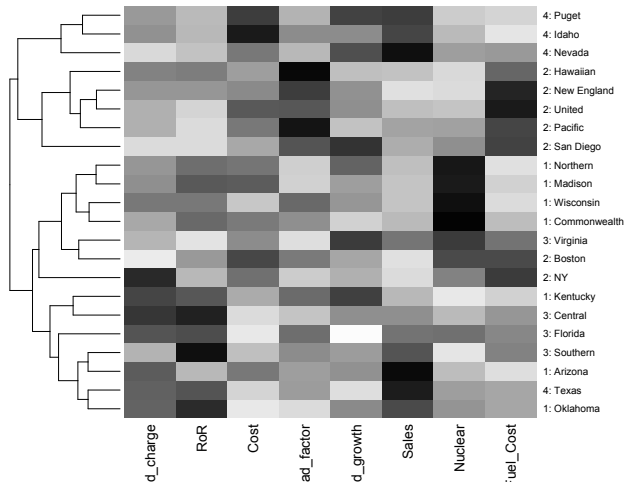
Legame medio, con $k = 4$

Arizona	Boston	Central	Commonwealth	NY
1	2	3	1	2
Florida	Hawaiian	Idaho	Kentucky	Madison
3	2	4	1	1
Nevada	New England	Northern	Oklahoma	Pacific
4	2	1	1	2
Puget	San Diego	Southern	Texas	Wisconsin
4	2	3	4	1
United	Virginia			
2	3			

La formazione di 4 gruppi sembra la più ragionevole in questo caso

Metodi gerarchici: esempio

Heatmap: celle più scure indicano valori più elevati di una variabile



Metodi gerarchici: esempio

Consideriamo il caso di 563 persone adulte che hanno risposto a un questionario, riguardante la frequenza con cui assumono comportamenti rischiosi in queste categorie:

- rischi collegati ad attività ricreative (climbing, scuba-diving)
- rischi collegati alla salute (fumo, alcol, dieta ricca di grassi)
- rischi collegati al lavoro (lasciare il lavoro senza avere l'alternativa)
- rischi finanziari (scommesse, investimenti)
- rischi collegati alla sicurezza (alta velocità in auto)
- rischi sociali (manifestazioni pubbliche)

Esempio

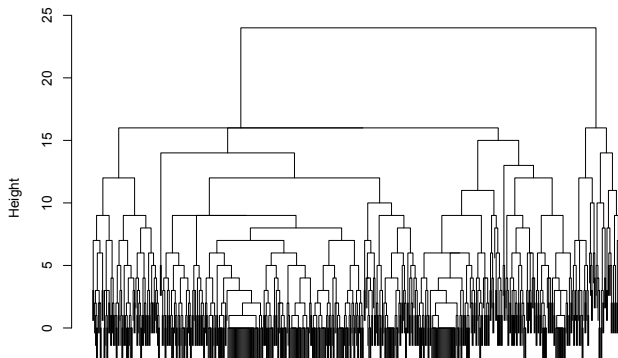
La risposta per ciascuna categoria è data su scala ordinale con 5 livelli (1: mai, 5: molto spesso)

Recreational	Health	Career	Financial
2.190053	2.396092	2.007105	2.026643
Safety	Social		
2.266430	2.017762		

I rispondenti, in media, si mostrano piuttosto avversi al rischio

Metodi gerarchici: esempio

Viene applicato un metodo gerarchico basato sulla **distanza di Manhattan** e il **legame completo**



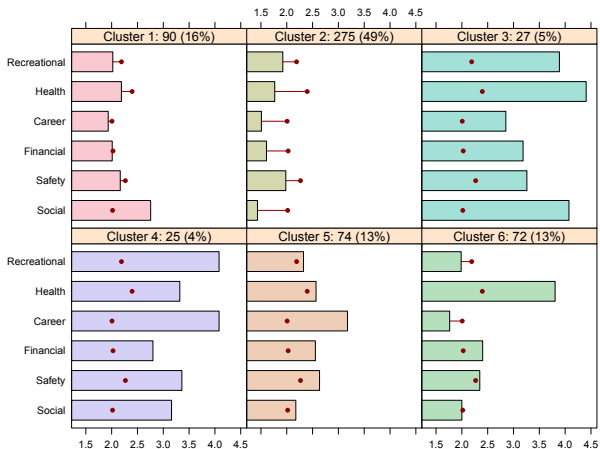
Metodi gerarchici: esempio

Un modo semplice per valutare le caratteristiche dei gruppi identificati è di visualizzare le medie di ogni variabile per ciascuno di essi

	Recreational	Health	Career	Financial	Safety	Social
1	2.0	2.2	1.9	2.0	2.2	2.8
2	1.9	1.8	1.5	1.6	2.0	1.4
3	3.9	4.4	2.9	3.2	3.3	4.1
4	4.1	3.3	4.1	2.8	3.4	3.2
5	2.3	2.6	3.2	2.6	2.6	2.2
6	2.0	3.8	1.8	2.4	2.3	2.0

Metodi gerarchici: esempio

Le stesse caratteristiche possono essere valutate per mezzo di un grafico a barre (i punti rossi indicano le medie per l'intera popolazione)



Limiti dei metodi gerarchici

- Richiedono il calcolo di una matrice di distanze: con dataset molto grandi può risultare dispendioso e lento
- Bassa stabilità dei gruppi: un ri-ordinamento dei dati può portare a una soluzione differente
- Sensibile alla presenza di outliers

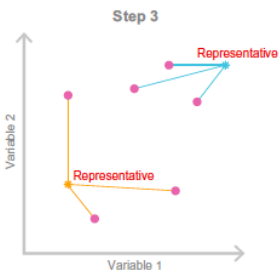
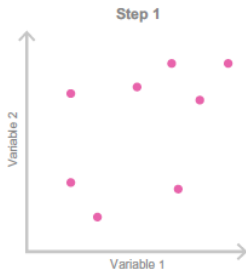
Metodo delle K-medie

- Sia $X = x_1, x_2, \dots, x_n$ un insieme di osservazioni in un dataset
- I metodi di partizione assegnano ciascuna osservazione al segmento al quale somigliano di più
- Il rappresentante di ciascun segmento viene chiamato **centroide**
- Il metodo delle K-medie si basa su un algoritmo iterativo, che migliora ad ogni passo.

Metodo delle K-medie

- 1 Specificare il numero di segmenti k
- 2 Selezionare casualmente k osservazioni dal dataset X e utilizzarle come centroidi $C = c_1, c_2, \dots, c_k$
- 3 Se devono essere estratti 5 segmenti, allora 5 osservazioni vengono estratte dal dataset e vengono elette come rappresentanti dei 5 segmenti (centroidi)
- 4 Ogni osservazione x_i viene assegnata al gruppo con il centroide più vicino, in modo da creare k segmenti S_1, S_2, \dots, S_k
- 5 Il risultato è una prima -subottimale- soluzione di segmentazione
- 6 A questo punto è necessario ri-calcolare i centroidi (ri-calcolando le medie all'interno di ciascun segmento)
- 7 L'operazione di ri-calcolo dei centroidi e di assegnazione delle osservazioni ai segmenti viene ripetuta fino a convergenza (o fino al raggiungimento del numero massimo di iterazioni)

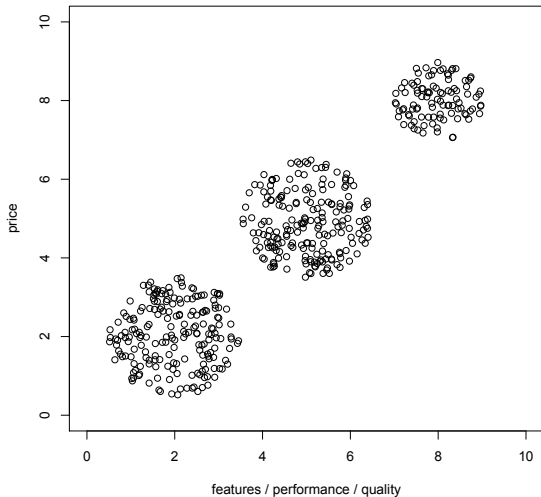
Metodo delle K-medie



Metodo delle K-medie: osservazioni

- L'algoritmo arriva sempre a una soluzione: ovviamente questo può richiedere più tempo con dataset grandi e molti segmenti
- Il punto di partenza è **casuale**
- Quindi punti di partenza diversi inevitabilmente potranno portare a soluzioni diverse
- La chiave per un segmentazione di buona qualità è la **ripetizione**
- Inoltre, l'algoritmo richiede di **specificare il numero di segmenti da creare...**
- Operazione non sempre semplice

Metodo delle K-medie: esempio



Metodo delle K-medie: esempio

Estraiamo 3 segmenti

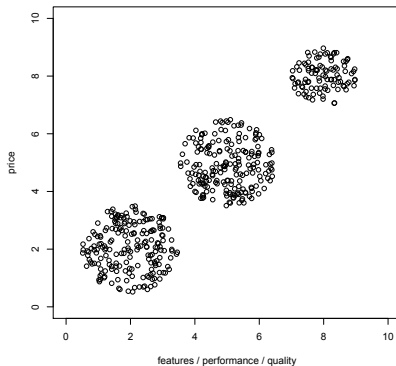


Figura: Esempio: 3 segmenti distinti per telefoni mobili

Metodo delle K-medie: esempio

Estraiamo 3 segmenti

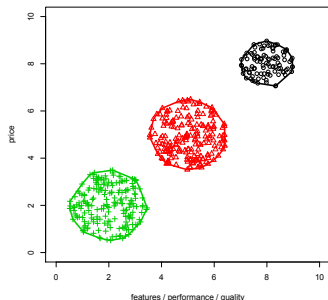
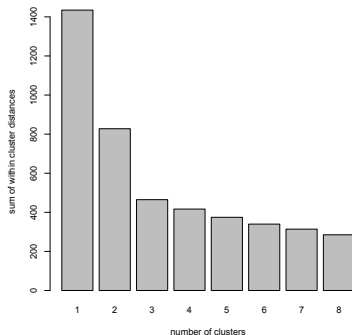


Figura: Esempio: 3 segmenti distinti per telefoni mobili

Chiaramente in questo caso l'operazione è semplice perchè i 3 segmenti sono molto ben definiti. Ma se così non fosse?

Metodo delle K-medie: esempio

Possiamo provare a estrarre da 2 a 8 segmenti
Come valutare i risultati?



Screeplot per $k = 1, \dots, 8$: il miglior numero di segmenti sia ha dove viene osservato un "gomito" nello screeplot

Metodo delle K-medie: esempio

- Torniamo all'esempio delle aziende di pubblica utilità
- Prendiamo in considerazione le due variabili "sales" e "fuel cost"
- Standardizziamo le due variabili
- Consideriamo le prime 5 aziende: *Arizona, Boston, Central, Commonwealth, Consolidated*

Metodo delle K-medie: esempio

- Immaginiamo che $k = 2$ e che i segmenti che vengono creati inizialmente siano:
- $A = (\text{Arizona}, \text{Boston})$
- $B = (\text{Central}, \text{Commonwealth}, \text{Consolidated})$
- Calcoliamo i centroidi e le distanze da essi ...

Metodo delle K-medie: esempio

$$\bar{x}_A = [-0.516, -0.020]$$

$$\bar{x}_B = [-0.733, 0.296]$$

Azienda	distanza da \bar{x}_A	distanza da \bar{x}_B
Arizona	1.0052	1.3887
Boston	1.0052	0.6216
Central	0.6029	0.8995
Commonwealth	0.7281	1.0207
Consolidated	2.0172	1.6341

Notiamo che *Boston* risulta più vicina al segmento B, mentre *Central* e *Commonwealth* sono più vicine al segmento A.

Metodo delle K-medie: esempio

Facciamo quindi una nuova assegnazione, creando i segmenti

- A = (Arizona, Central, Commonwealth)
- B = (Consolidated, Boston).

A questo punto dobbiamo calcolare nuovamente i centri $\bar{x}_A = [-0.191, -0.553]$

$\bar{x}_B = [-1.33, 1.253]$

E le distanze

Azienda	distanza da \bar{x}_A	distanza da \bar{x}_B
Arizona	0.3827	2.5159
Boston	1.6289	0.5067
Central	0.5463	1.9432
Commonwealth	0.5391	2.0745
Consolidated	2.6412	0.5067

Qui ci fermiamo perchè ogni azienda risulta assegnata al giusto gruppo.

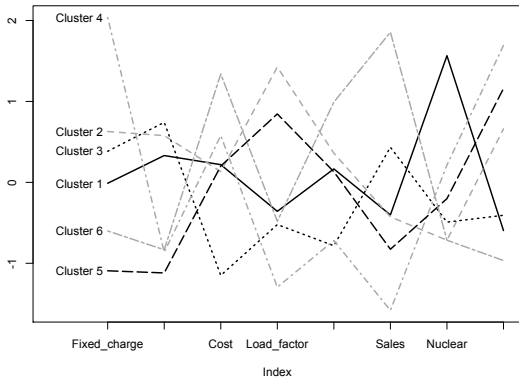
Metodo delle K-medie: esempio

Possiamo ora utilizzare l'intero dataset (22 aziende) e tutte le variabili. Usiamo il l'algoritmo 'k-means' con $k = 6$, ottenendo

Arizona	Boston	Central	Commonwealth
3	5	3	1
NY	Florida	Hawaiian	Idaho
4	3	2	6
Kentucky	Madison	Nevada	New England
2	1	6	2
Northern	Oklahoma	Pacific	Puget
1	3	5	6
San Diego	Southern	Texas	Wisconsin
5	3	3	1
United	Virginia		
5	1		

Metodo delle K-medie: esempio

Per ciascun segmento possiamo analizzare i centroidi
Come interpretare i risultati?

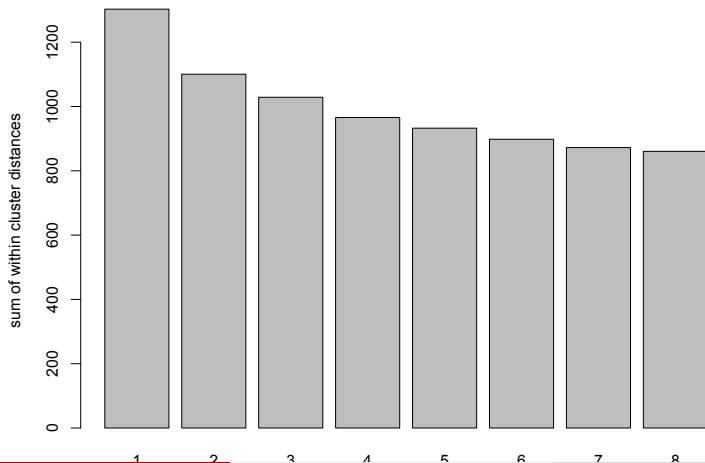


Metodo delle K-medie: esempio

- Consideriamo nuovamente l'esempio del questionario sulla propensione-avversione al rischio delle persone
- Proviamo a generare soluzioni tra 2 e 8 segmenti
- Attraverso lo screeplot cerchiamo di capire quale potrebbe essere il numero ideale di segmenti

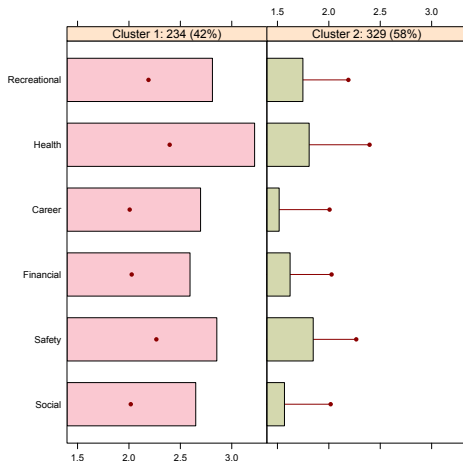
Metodo delle K-medie: esempio

Quanti segmenti?



Metodo delle K-medie: esempio

Potremmo selezionare due segmenti...avversi e propensi al rischio



Metodo delle K-medie: esempio

Oppure selezionare vari segmenti: quale interpretazione?

