

Segmentazione del mercato

Metodi basati su modelli

- Diversamente dai metodi basati sulle distanze, i metodi model-based non usano misure di similarità
- La soluzione al problema di segmentazione ha due proprietà:
 - ogni segmento ha una certa dimensione
 - se un consumatore appartiene a un certo segmento A , avrà caratteristiche tipiche dei membri di quel segmento
- Usiamo **Modelli a mistura finita**
- Il numero di segmenti è finito e il modello generale è una mistura di modelli per specifici segmenti

Metodi basati su modelli: Modelli a mistura finita

Possiamo definire un modello a mistura finita come

$$\sum_{h=1}^k \pi_h f(y|\theta_h), \quad \pi_h > 0, \quad \sum_{h=1}^k \pi_h = 1$$

- π_h sono i pesi, ovvero le dimensioni dei segmenti
- $f(y|\theta_h)$ sono i modelli specifici per ciascun segmento e corrispondono a determinate distribuzioni
- una scelta tipica per $f()$ è la distribuzione normale multivariata

Metodi basati su modelli: Modelli a mistura finita

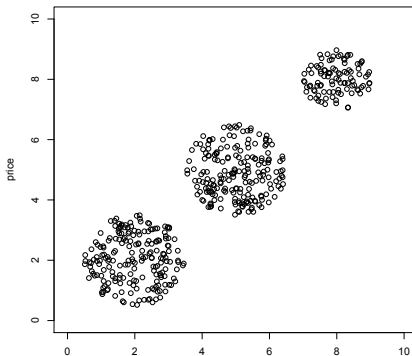
- I parametri da stimare per ciascun segmento sono le dimensioni π e le caratteristiche specifiche del segmento θ
- Tipicamente la stima può essere fatta via massima verosimiglianza
- Tuttavia anche per le misture più semplici questo può rivelarsi complicato
- Si può quindi usare un metodo iterativo, ad esempio l'algoritmo EM (Expectation-Maximization)

Metodi basati su modelli: Modelli a mistura finita

- L'uso dell'algoritmo EM richiede di specificare a priori il numero di segmenti da estrarre
- Siccome il numero di segmenti di solito non è noto ...
- Una strategia standard è di estrarre vari numeri di segmenti e confrontarli. Come?
- Possiamo usare i criteri di informazione che conosciamo, AIC e BIC

Metodi basati su modelli: Modelli a mistura finita

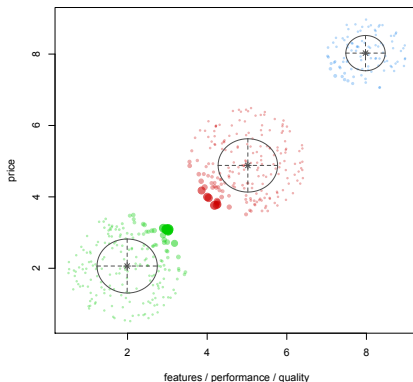
- Consideriamo un caso artificiale: mercato dei telefoni.
- Due variabili: numero di caratteristiche del telefono, x , e prezzo che i consumatori sono disposti a pagare, y .
- Tre segmenti ben definiti.



Metodi basati su modelli: Modelli a mistura finita

Stimiamo con EM modelli per differenti numeri di segmenti (da 2 a 8). Possiamo visualizzare i risultati tramite l'[uncertainty plot](#).

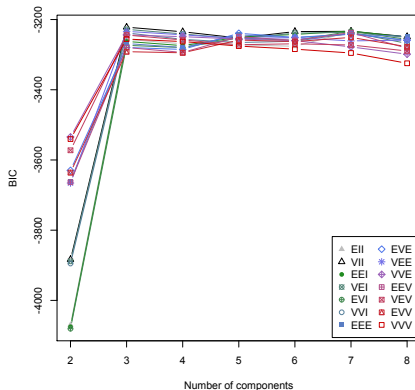
I punti più grandi sono quelli in cui è presente maggiore incertezza.



Metodi basati su modelli: Modelli a mistura finita

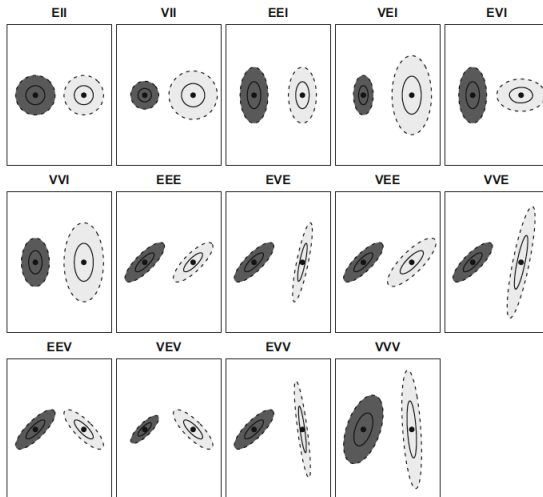
Selezione tramite BIC: il pacchetto 'mclust' considera i valori negativi di BIC. Vogliamo quindi **massimizzare il BIC**.

Il modello migliore è il VII (spherical varying volume model).



Metodi basati su modelli: Modelli a mistura finita

Vari modelli presenti in 'mclust'



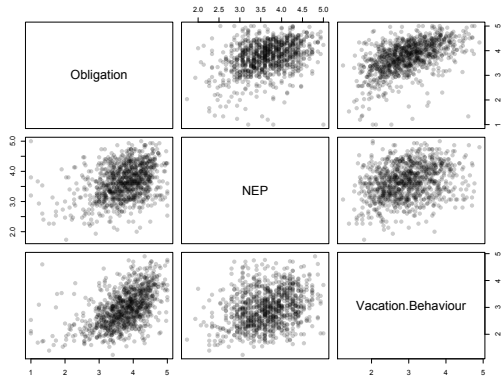
Esempio

- Consideriamo i dati provenienti da un questionario sulle ragioni per cui viene scelta un certo tipo di vacanza, con particolare focus sul comportamento attento all'ambiente degli intervistati.
- 1000 osservazioni
- 20 variabili di tipo sociodemografico

Esempio

Consideriamo le variabili:

- Moral obligation
- NEP score
- Environmental behavior on vacation

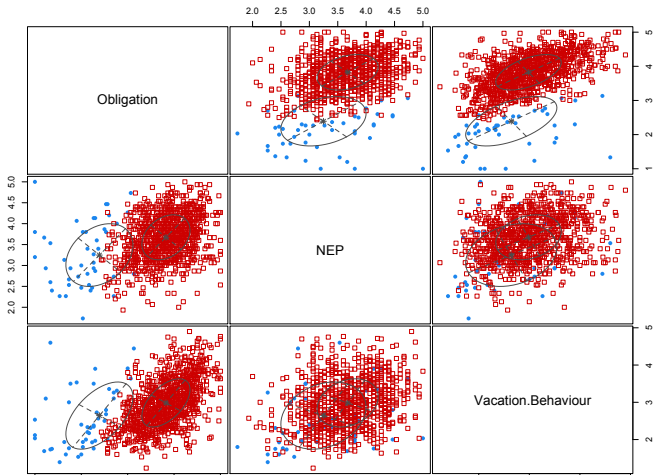


Abbiamo a questo punto due possibilità:

- Stimare i 14 modelli presenti in 'mclust' e selezionare il miglior modello rispetto al BIC
- Selezionare solo alcuni modelli, vincolando la matrice di covarianze ad avere determinate forme. Per esempio vogliamo solo matrici di covarianza con uguale forma, volume e orientamento per tutti i segmenti.
- Uguale forma e orientamento implicano che la struttura di correlazione tra variabili sia la stessa nei vari segmenti.

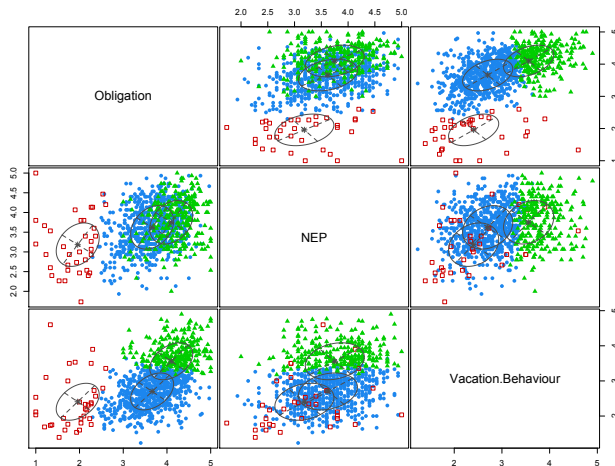
Esempio

Nel primo caso viene selezionato un modello con 2 segmenti, ellissoidale, con uguale forma e uguale orientamento (VEE)



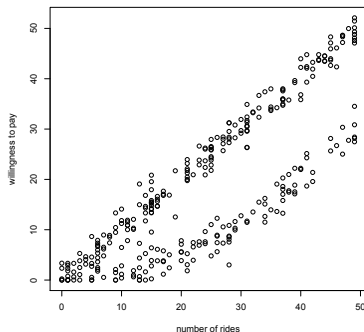
Esempio

Nel caso vincolato viene selezionato un modello con 3 segmenti, ellissoidale, uguale volume, forma e orientamento (EEE)



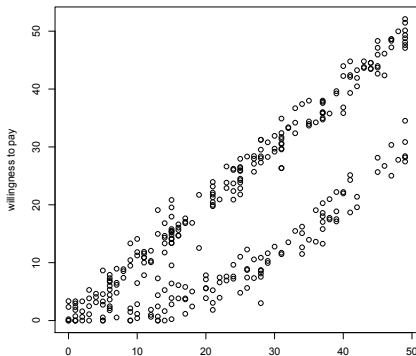
Metodi basati su modelli: Misture finite di regressioni

- I modelli basati su misture finite di regressioni assumono l'esistenza di una variabile dipendente y che può essere spiegata da una serie di variabili x
- Ma la relazione funzionale tra variabile dipendente e variabili indipendenti è diversa tra i diversi segmenti del mercato



Metodi basati su modelli: Misture finite di regressioni

- Esempio artificiale: disponibilità a pagare per entrare in un parco tematico in funzione di quanti giri si possono fare
- Segmento 1: $y = x + \epsilon$
- Segmento 2: $y = 0.0125x^2 + \epsilon$



Metodi basati su modelli: Misture finite di regressioni

- Vengono stimati simultaneamente due modelli di regressione per le due componenti

	Comp.1	Comp.2
coef.(Intercept)	1.60901610	0.3171846123
coef.rides	-0.11508969	0.9905130420
coef.(rides ²)	0.01439438	0.0001851942

Esempio

Possiamo voler analizzare separatamente le due componenti

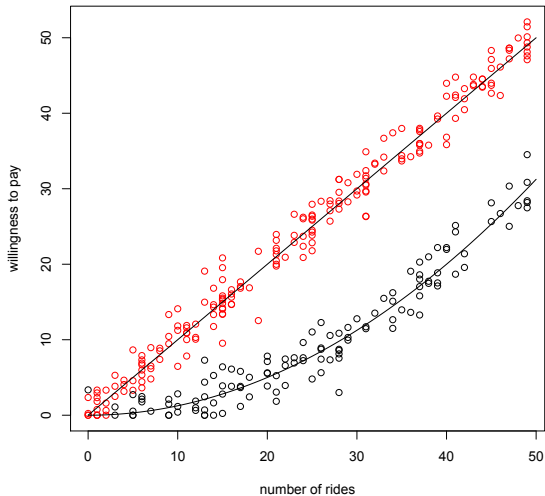
\$Comp.1

	Estimate	Std. Error	z value	Pr(> z)
Intercept	1.6090161	0.6614587	2.4325	0.01499 *
rides	-0.1150897	0.0563449	-2.0426	0.04109 *
rides^2	0.0143943	0.0010734	13.4104	< 2e-16 ***

\$Comp.2

	Estimate	Std. Error	z value	Pr(> z)
Intercept	0.31718461	0.48268972	0.6571	0.5111
rides	0.99051304	0.04256232	23.2721	<2e-16 ***
rides^2	0.00018516	0.00080704	0.2294	0.8185

Metodi basati su modelli: Misture finite di regressioni



Esempio

- Consideriamo nuovamente il dataset relativo alle motivazioni legate alle scelte di vacanza
- Prima di tutto stimiamo un modello di regressione 'generale'

Call:

```
lm(formula = Vacation.Behaviour ~ Obligation + NEP)
```

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	2.96280	0.01821	162.680	< 2e-16	***
Obligation	0.32357	0.01944	16.640	< 2e-16	***
NEP	0.06599	0.01944	3.394	0.000718	***

Multiple R-squared: 0.2775, Adjusted R-squared: 0.276

Esempio

Ci chiediamo dunque se una mistura di regressioni lineari possa aiutare a identificare gruppi diversi

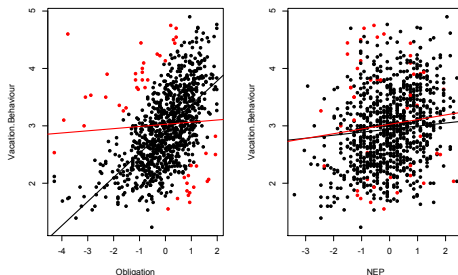
\$Comp.1

	Estimate	Std. Error	z value	Pr(> z)	
(Intercept)	2.944634	0.032669	90.1342	< 2e-16	***
Obligation	0.418934	0.030217	13.8641	< 2e-16	***
NEP	0.053489	0.027023	1.9794	0.04778	*

\$Comp.2

	Estimate	Std. Error	z value	Pr(> z)	
(Intercept)	3.023214	0.139161	21.7246	<2e-16	***
Obligation	0.018619	0.145845	0.1277	0.8984	
NEP	0.082207	0.105744	0.7774	0.4369	

Esempio



- Esiste quindi una forte associazione tra comportamento in vacanza e obbligo morale, per il primo segmento (colore nero)
- Questa associazione non è invece presente nel secondo segmento (linea rossa praticamente orizzontale)
- Non risulta una associazione degna di nota tra comportamento in vacanza e la variabile NEP