

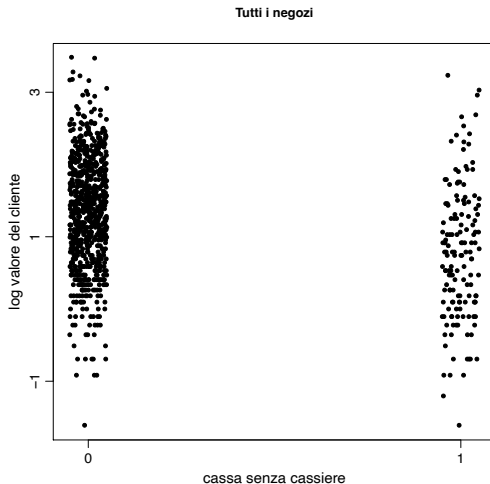
Modelli lineari gerarchici

Modelli lineari gerarchici

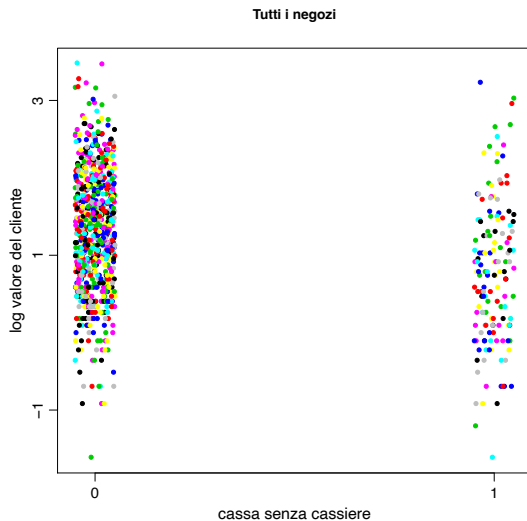
- Accanto alle informazioni che riguardano consumatori e prodotti, in ambito aziendale e di marketing è frequente raccogliere informazioni riferite alla regione geografica, al punto vendita, al prodotto, alla campagna promozionale.
- La presenza di questo **doppio livello informativo** può rendere utile o necessaria l'applicazione di un modello gerarchico (o multilivello).
- Il termine **gerarchico** deriva dal fatto che sia i dati (ad es. clienti raggruppati per punti vendita), sia il modello presentano una struttura gerarchica.

Modelli lineari gerarchici: motivazione

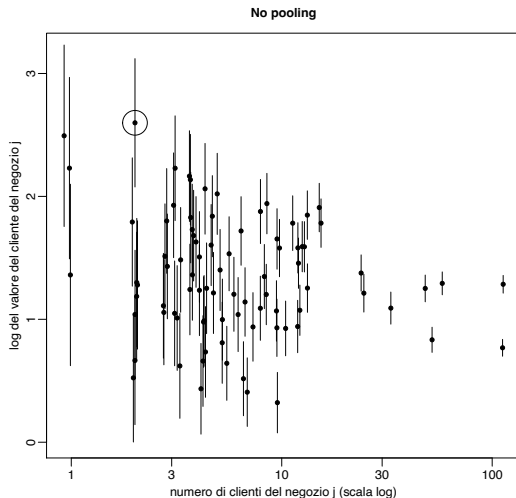
Clienti titolari di una carta fedeltà in 85 negozi



Modelli lineari gerarchici: motivazione



Modelli lineari gerarchici: motivazione



Modelli lineari gerarchici: notazione

- Unità $i = 1, \dots, n$
- Variabile risposta $y = (y_1, \dots, y_n)$
- Predittori, matrice X di dimensione $n \times k$
- $\hat{y}_i = X_i\beta$, previsione per l'unità i
- Gruppi $j = 1, \dots, J$
- Indici $j[i]$, appartenenza a un gruppo

Modelli lineari gerarchici

- I modelli gerarchici sono estensioni del modello di regressione, in cui i dati sono strutturati in gruppi e i coefficienti possono variare tra i gruppi.
- Quando i dati contengono informazioni “di gruppo”, una regressione che contiene indicatori di gruppo può essere un **modello a intercetta variabile**

$$y_i = \alpha_{j[i]} + \beta x_i + \epsilon_i$$

Modelli lineari gerarchici

Un'altra possibilità è un **modello a pendenza variabile**

$$y_i = \alpha + \beta_{j[i]}x_i + \epsilon_i$$

In questo caso la pendenza variabile indica un'interazione tra il predittore x e l'indicatore di gruppo.

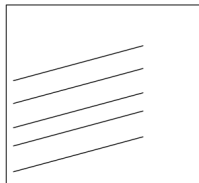
Modelli lineari gerarchici

Inoltre è anche possibile un modello a intercetta e pendenza variabile

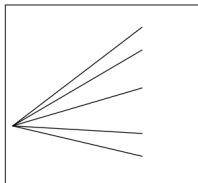
$$y_i = \alpha_{j[i]} + \beta_{j[i]}x_i + \epsilon_i$$

Modelli lineari gerarchici

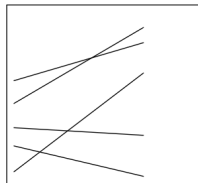
Varying intercepts



Varying slopes



Varying intercepts and slopes



- modello a intercetta variabile
- modello a pendenza variabile
- modello a intercetta e pendenza variabile

Benefici della modellazione gerarchica

Quando un modello gerarchico fa la differenza?

- L'usuale alternativa al modello gerarchico è la regressione classica: in alcuni casi limite questi due approcci coincidono.
- Quando c'è una variazione a livello di gruppo molto limitata, il modello gerarchico si riduce al modello di regressione classico.
- Diversamente, quando i coefficienti a livello di gruppo variano molto, il modello gerarchico coincide con il **modello di regressione con indicatori di gruppo**.

Modelli lineari gerarchici: le basi

La modellazione gerarchica può essere pensata in due modi equivalenti:

- Una generalizzazione del modello di regressione lineare, in cui intercetta e pendenza possono variare tra gruppi;
- Una regressione che include una variabile categoriale che rappresenta l'appartenenza al gruppo; l'indicatore di gruppo è un fattore con J livelli, che corrispondono a J predittori nel modello di regressione.

Il punto cruciale della modellazione gerarchica è che anche questi J coefficienti sono a loro volta modellati.

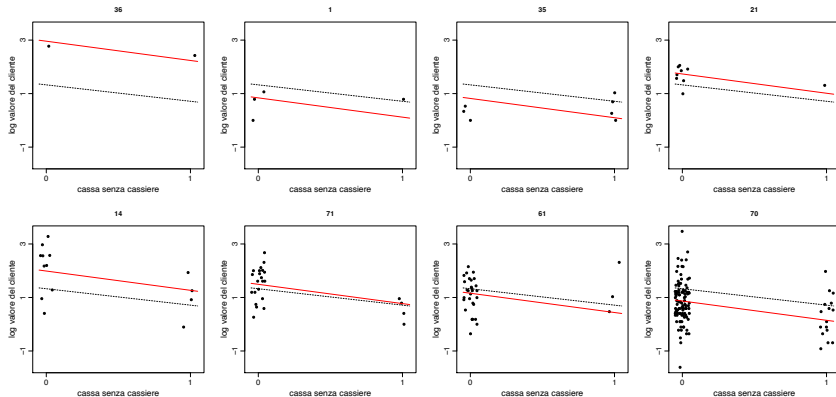
Modelli lineari gerarchici: un compromesso

I modelli lineari gerarchici possono essere considerati un compromesso tra due estremi:

- **complete pooling**: esclusione della variabile categoriale dal modello
- **no pooling**: stima di modelli separati per ogni livello della variabile categoriale

Si osservi che il modello *no pooling* tende a sovrastimare la variabilità all'interno dei singoli gruppi, facendoli apparire più diversi di quanto non siano realmente.

Modelli lineari gerarchici: un compromesso



Pooling completo (linea tratteggiata) e no pooling (linea **rossa**)

Modelli lineari gerarchici: pooling parziale senza predittori

Nel caso in cui non ci siano predittori, la stima del modello gerarchico per un certo gruppo j , $\hat{\alpha}_j$, può essere pensata come una approssimazione di media pesata della media di gruppo $\bar{y}_j$ e della media totale $\bar{y}$

$$\hat{\alpha}_j \approx \frac{\frac{n_j}{\sigma_y^2} \bar{y}_j + \frac{1}{\sigma_\alpha^2} \bar{y}}{\frac{n_j}{\sigma_y^2} + \frac{1}{\sigma_\alpha^2}}$$

dove n_j è il numero di unità all'interno del gruppo j , σ_y^2 è la varianza **all'interno del gruppo** –per semplicità assunta costante fra tutti i gruppi– e σ_α^2 è la varianza **tra gruppi**.

Modelli lineari gerarchici: pooling parziale senza predittori

La media pesata riflette la quantità di informazione disponibile a livello di singolo gruppo da un lato, e la media fra tutti i gruppi dall'altro, per cui:

- Le medie di gruppi di dimensioni più piccole apportano minor informazione
- Più è piccolo il gruppo, più la stima del modello gerarchico $\hat{\alpha}_j$ si avvicina alla media globale, $\bar{y}$. Nel caso in cui $n_j = 0$ la stima del modello gerarchico è semplicemente la media globale, $\bar{y}$
- Le medie di gruppi di dimensioni più grandi apportano maggior informazione
- Più è grande il gruppo, più la stima del modello gerarchico $\hat{\alpha}_j$ si avvicina alla media di gruppo, $\bar{y}_j$.

Pooling parziale senza predittori

Esempio

```
lmer(formula = y ~ 1 + (1 | negozio))
coef.est  coef.se
    1.31    0.05
```

Error terms:

Groups	Name	Std.Dev.
negozio	(Intercept)	0.31
Residual		0.80

number of obs: 919, groups: negozio, 85

AIC = 2265.4

Modelli lineari gerarchici: pooling parziale con predittori

Possiamo scrivere il modello di **pooling completo** come

$$y_i = \alpha + \beta x_i + \epsilon_i$$

Il modello **no pooling** invece è del tipo

$$y_i = \alpha_{j[i]} + \beta x_i + \epsilon_i$$

Entrambe queste soluzioni possono presentare dei problemi:

- nel primo caso viene ignorata ogni variazione tra i gruppi,
- nel secondo si dà troppo peso a differenze che possono derivare da gruppi poco numerosi.

Modelli lineari gerarchici: pooling parziale con predittori

Il più semplice modello gerarchico con un predittore è del tipo

$$y_i \sim N(\alpha_{j[i]} + \beta x_i, \sigma_y^2) \quad i = 1, \dots, n$$

in cui α_j ha una **distribuzione di probabilità**

$$\alpha_j \sim N(\mu_\alpha, \sigma_\alpha^2) \quad j = 1, \dots, J$$

Si osservi che

con $\sigma_\alpha \rightarrow \infty$: **no pooling**

con $\sigma_\alpha \rightarrow 0$: **pooling completo**

Modelli lineari gerarchici: pooling parziale con predittori

Anche in questo caso, la stima di α_j può essere pensata come media pesata della stima no pooling, $(\bar{y}_j - \beta \bar{x}_j)$, e della media, μ_α

$$\hat{\alpha}_j \approx \frac{\frac{n_j}{\sigma_y^2}}{\frac{n_j}{\sigma_y^2} + \frac{1}{\sigma_\alpha^2}} (\bar{y}_j - \beta \bar{x}_j) + \frac{\frac{1}{\sigma_\alpha^2}}{\frac{n_j}{\sigma_y^2} + \frac{1}{\sigma_\alpha^2}} \mu_\alpha$$

Pooling parziale con predittori

Esempio

```
lmer(formula = y ~ x + (1 | negozio))
      coef.est coef.se
(Intercept)  1.46    0.05
x            -0.69    0.07
```

Error terms:

Groups	Name	Std.Dev.
negozio	(Intercept)	0.33
Residual		0.76

```
number of obs: 919, groups: negozio, 85
AIC = 2179.3
```

Modelli lineari gerarchici: pooling parziale con predittori

Iperparametri

$$\hat{\mu}_\alpha = 1.46, \hat{\beta} = -0.69, \hat{\sigma}_y = 0.76, \hat{\sigma}_\alpha = 0.33$$

Effetti fissi

$$y = 1.46 - 0.69x$$

Effetti casuali

1	-0.27
2	-0.53
3	0.02
...	...
85	-0.08

Modelli lineari gerarchici: pooling parziale con predittori

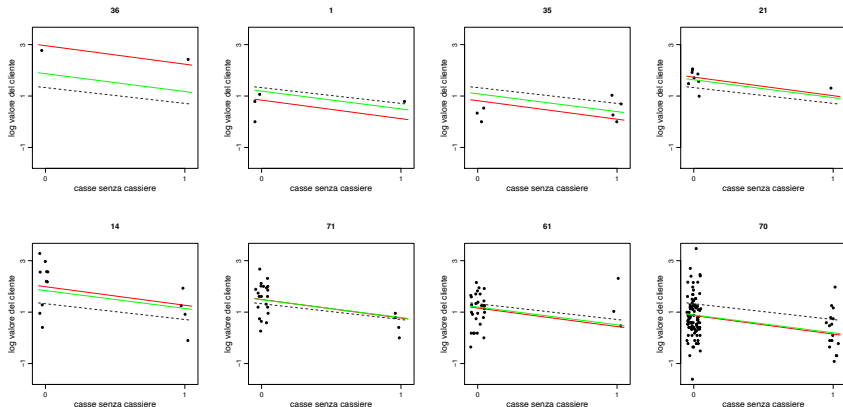
Alternativamente, è possibile analizzare il modello stimato per ciascun negozio

negozio	α_j	β
1	1.19	-0.69
2	0.93	-0.69
3	1.48	-0.69
...	...	...
85	1.39	-0.69

Quindi, ad esempio, il modello per il negozio 1 è

$$y = 1.19 - 0.69x = (1.46 - 0.27) - 0.69x$$

Modelli lineari gerarchici: pooling parziale con predittori



Modello stimato per alcuni negozi: pooling parziale con predittore (linea verde)

Modelli lineari gerarchici: aggiunta di un predittore a livello di gruppo

Aggiungendo un predittore u_j a livello di gruppo, possiamo esprimere il modello gerarchico come

$$y_i \sim N(\alpha_{j[i]} + \beta x_i, \sigma_y^2), \quad i = 1, \dots, n$$

$$\alpha_j \sim N(\gamma_0 + \gamma_1 u_j, \sigma_\alpha^2), \quad j = 1, \dots, J$$

Nel caso considerato, viene inserita la variabile ' $u =$ **reddito medio dell'area in cui si trova il negozio**'

Pooling parziale con predittori per negozio

u.full è la variabile u=reddito medio dell'area del negozio su tutte le unità statistiche

```
lmer(formula = y ~ x + u.full + (1 | negozio))
```

	coef.est	coef.se
(Intercept)	1.47	0.04
x	-0.67	0.07
u.full	0.72	0.09

Error terms:

Groups	Name	Std.Dev.
negozio	(Intercept)	0.16
Residual		0.76

number of obs: 919, groups: negozio, 85

AIC = 2144.2

Modelli lineari gerarchici: pooling parziale con predittori per negozio

Effetti fissi

$$y_i = 1.47 - 0.67x_i + 0.72u_{j(i)}$$

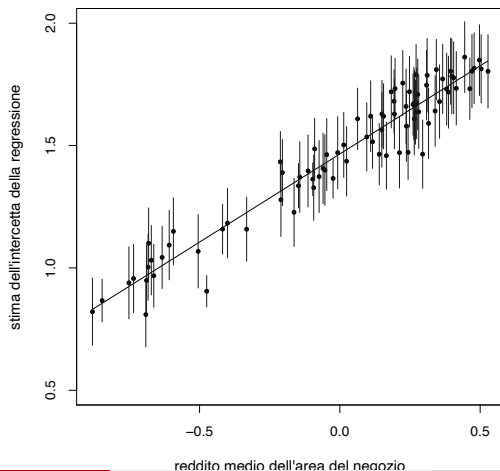
Effetti casuali

1	-0.02
2	0.01
3	0.02
...	...
85	-0.04

Quindi ad esempio il modello per il negozio 85 sarà

$$\begin{aligned} y_i &= 1.42 - 0.67x_i + 0.72u_{85} = (1.42 + 0.72 \cdot 0.36) - 0.67x_i \\ &= 1.68 - 0.67x_i \end{aligned}$$

Modelli lineari gerarchici: pooling parziale con predittori per negozio



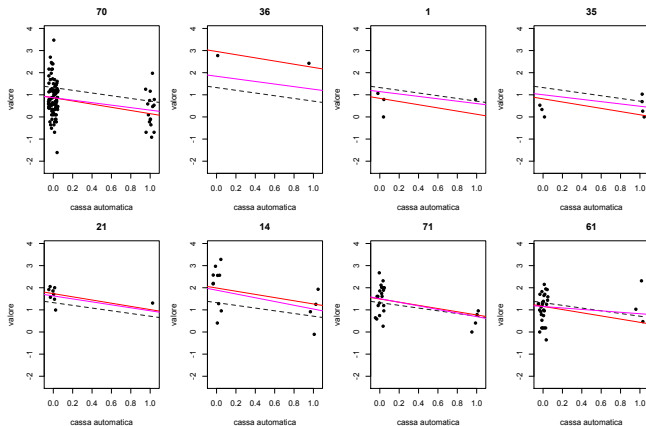
Modelli lineari gerarchici: intercette e pendenze variabili

Il modello può essere esteso per comprendere la presenza di intercetta e pendenza variabili

$$y_i \sim N(\alpha_{j[i]} + \beta_{j[i]}x_i, \sigma_y^2), \quad i = 1, \dots, n$$

Interazione dell'indicatore di gruppo con il predittore a livello individuale.

Modelli lineari gerarchici: intercette e pendenze variabili



Pooling parziale con intercetta e pendenza variabili (linea **rosa**)

Modelli lineari gerarchici: intercette e pendenze variabili

Analizziamo il modello stimato per ciascun negozio

negozio	α_j	β_j
1	1.14	-0.54
2	0.93	-0.77
3	1.47	-0.67
...	...	...
85	1.38	-0.65

Quindi, ad esempio, il modello per il negozio 1 è

$$y = 1.14 - 0.54x = (1.46 - 0.32) + (-0.68 + 0.14)x$$

Modelli lineari gerarchici: aggiunta di un predittore a livello di gruppo

Il modello può essere esteso aggiungendo un predittore u_j a livello di gruppo, sia per intercetta che per pendenza

$$y_i \sim N(\alpha_{j[i]} + \beta_{j[i]}x_i, \sigma_y^2), \quad i = 1, \dots, n$$

$$\alpha_j \sim N(\gamma_0^\alpha + \gamma_1^\alpha u_j, \sigma_\alpha^2), \quad j = 1, \dots, J$$

$$\beta_j \sim N(\gamma_0^\beta + \gamma_1^\beta u_j, \sigma_\beta^2), \quad j = 1, \dots, J$$

Nel caso considerato, viene inserita la stessa variabile ' u = reddito medio dell'area in cui si trova il negozio'