



Quarta giornata

I modelli misti

Marcello Gallucci
Univerisità Milano-Bicocca

Modello Lineare Generale

vantaggi

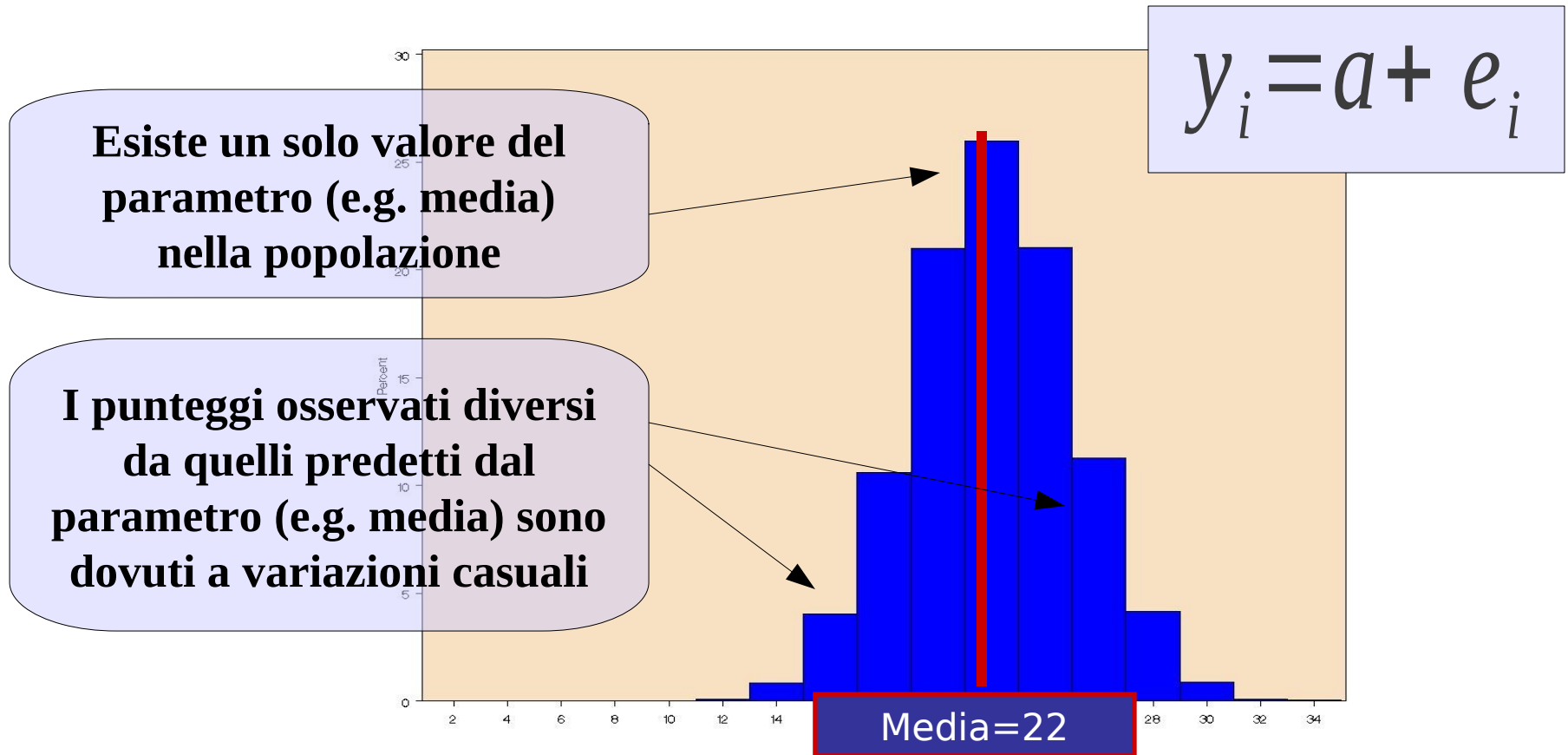
- Consente di stimare le relazioni fra due o più variabili
- Si applica ad una ampio spettro di tipi di dati
- Consente di stimare vari tipi di effetti

svantaggi

- Assume una struttura dei dati molto semplice
- Non consente di modellare una ampia serie di relazioni e dipendenza tra unità di misurazione

Assunzioni GLM

Modello Lineare Generale



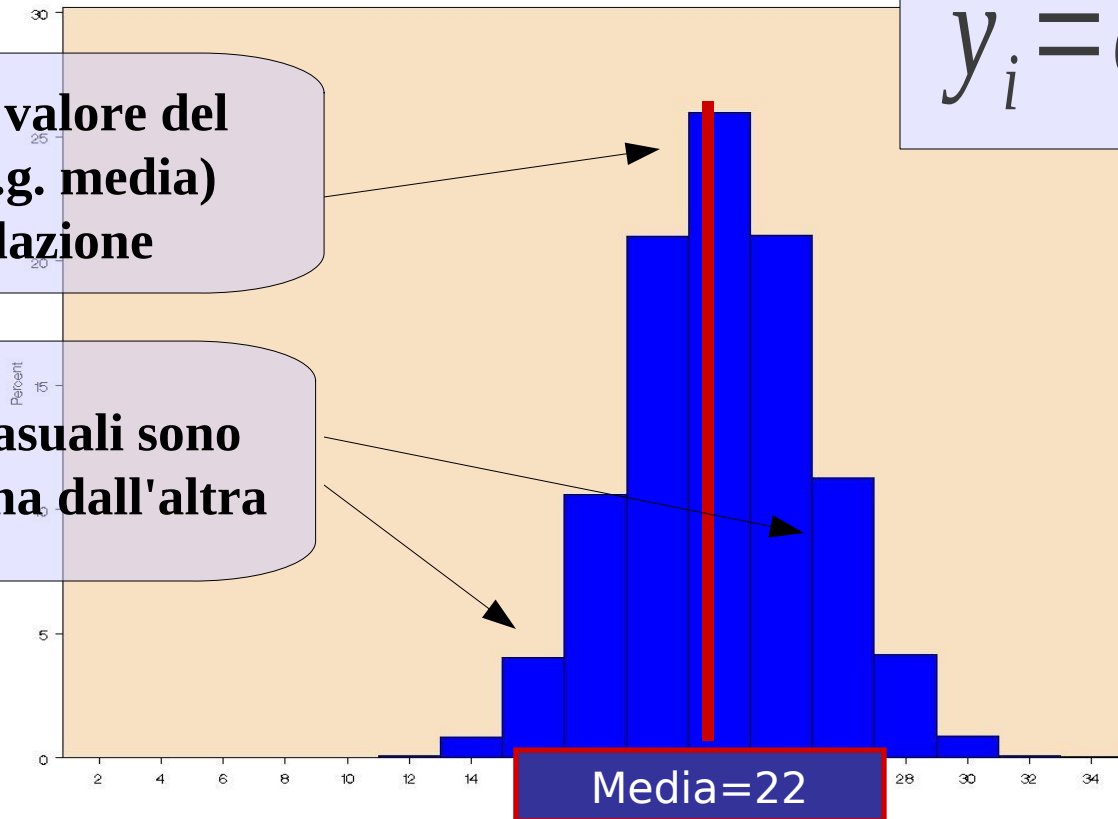
Assunzioni GLM

Modello Lineare Generale

Esiste un solo valore del parametro (e.g. media) nella popolazione

Le variazioni casuali sono indipendenti l'una dall'altra

$$y_i = a + e_i$$



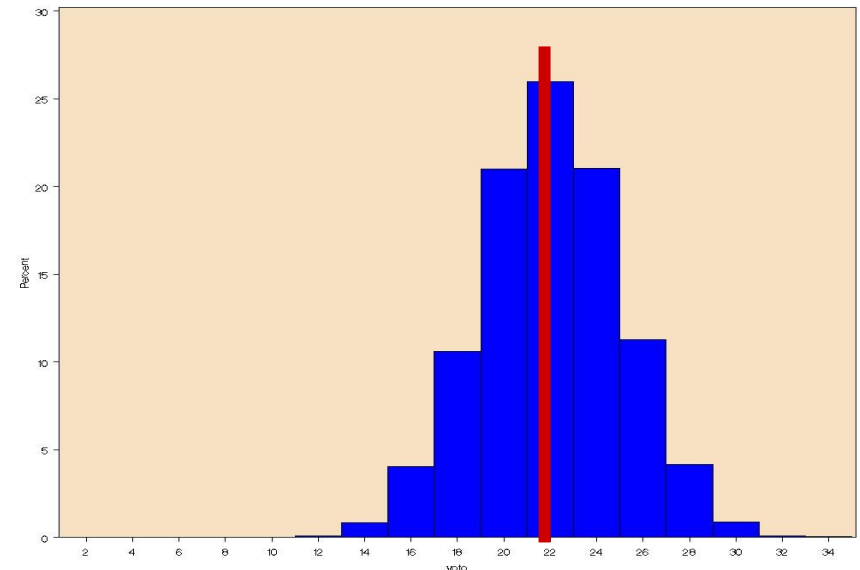
Assunzioni GLM

Modello Lineare Generale

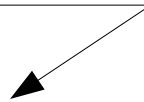
Il valore stimato della popolazione si definisce FISSO (fixed parameter)



$$y_i = a + e_i$$
$$\text{corr}(e_i, e_j) = 0$$



Le variazioni casuali sono indipendenti l'una dall'altra



Violazioni delle assunzioni

Le assunzioni di unicità degli effetti (effetti fissi) e indipendenza delle misurazioni (errori indipendenti) non sono rispettate in tutti i seguenti casi:

- Misurazioni correlate
- Disegni a misure ripetute
- Disegni longitudinali
- Dati con strutture gerarchiche
- Dati con misurazioni multi-livello

I modelli misti

**Non esiste un solo valore
fisso che intendiamo
stimare**

**Le variazioni casuali non
sono indipendenti l'una
dall'altra**

I modelli misti consentono di estendere il modello lineare generale in tutte quelle situazioni in cui le due assunzioni fondamentali del GLM non sono rispettate

I modelli misti

GLM

Regressione

T-test

ANOVA

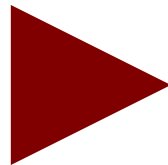
ANCOVA

Moderazione

Mediazione

Path Analysis

**Regressione
Logistica**



LMM

Random coefficients models

Random intercept regression models

One-way ANOVA with random effects

One-way ANCOVA with random effects

Intercepts-and-slopes-as-outcomes models

Generalized mixed model

Estensione del GLM al **modello misto**

Esempio “birre” 2

Consideriamo il caso in cui abbiamo ampliato il nostro campione di “bevitori di birra”, avendo raccolto ulteriori dati in diversi bar della città

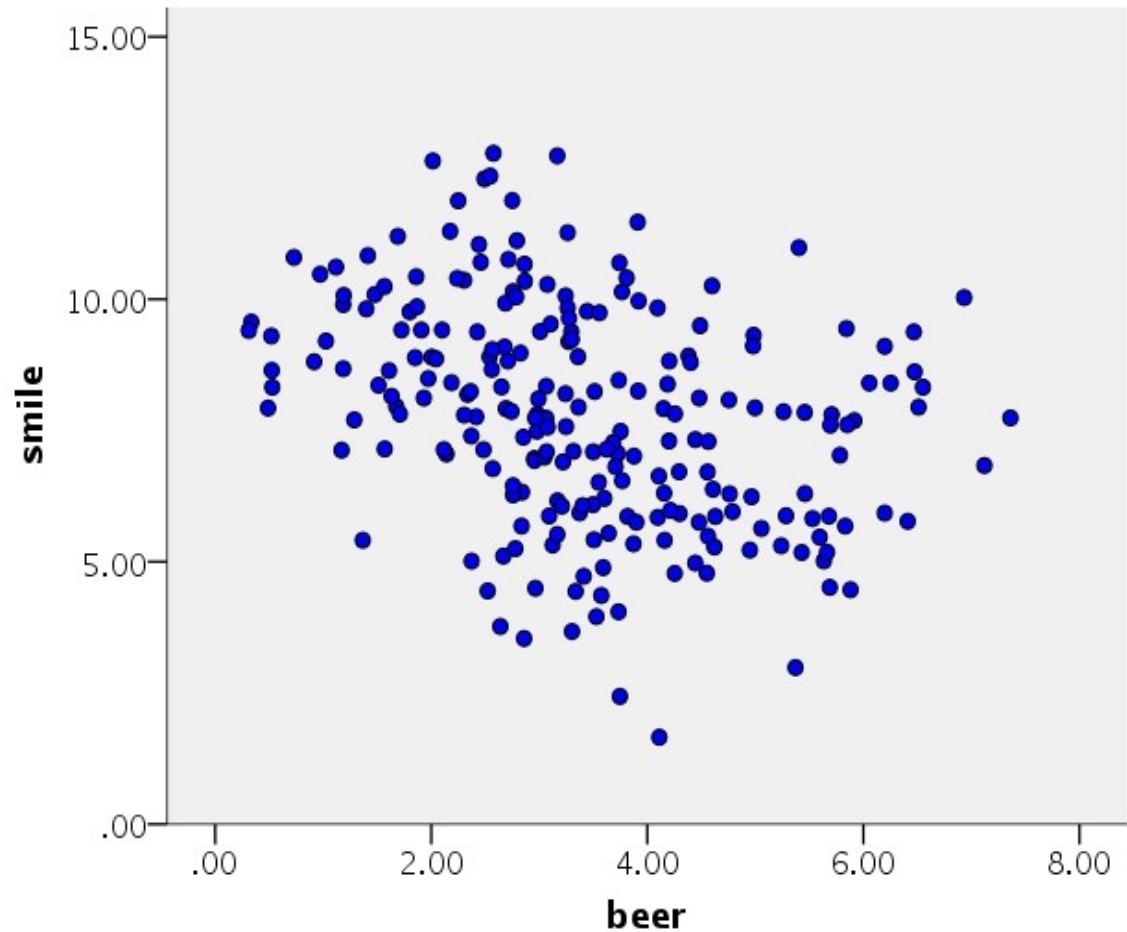
bar

	Frequency	Percent	Valid Percent	Cumulative Percent
Valid a	3	1.3	1.3	1.3
b	14	6.0	6.0	7.3
c	22	9.4	9.4	16.7
d	21	9.0	9.0	25.6
e	14	6.0	6.0	31.6
f	20	8.5	8.5	40.2
g	24	10.3	10.3	50.4
h	12	5.1	5.1	55.6
i	16	6.8	6.8	62.4
l	22	9.4	9.4	71.8
m	21	9.0	9.0	80.8
n	15	6.4	6.4	87.2
o	16	6.8	6.8	94.0
p	11	4.7	4.7	98.7
q	3	1.3	1.3	100.0
Total	234	100.0	100.0	

Totale di 234 soggetti

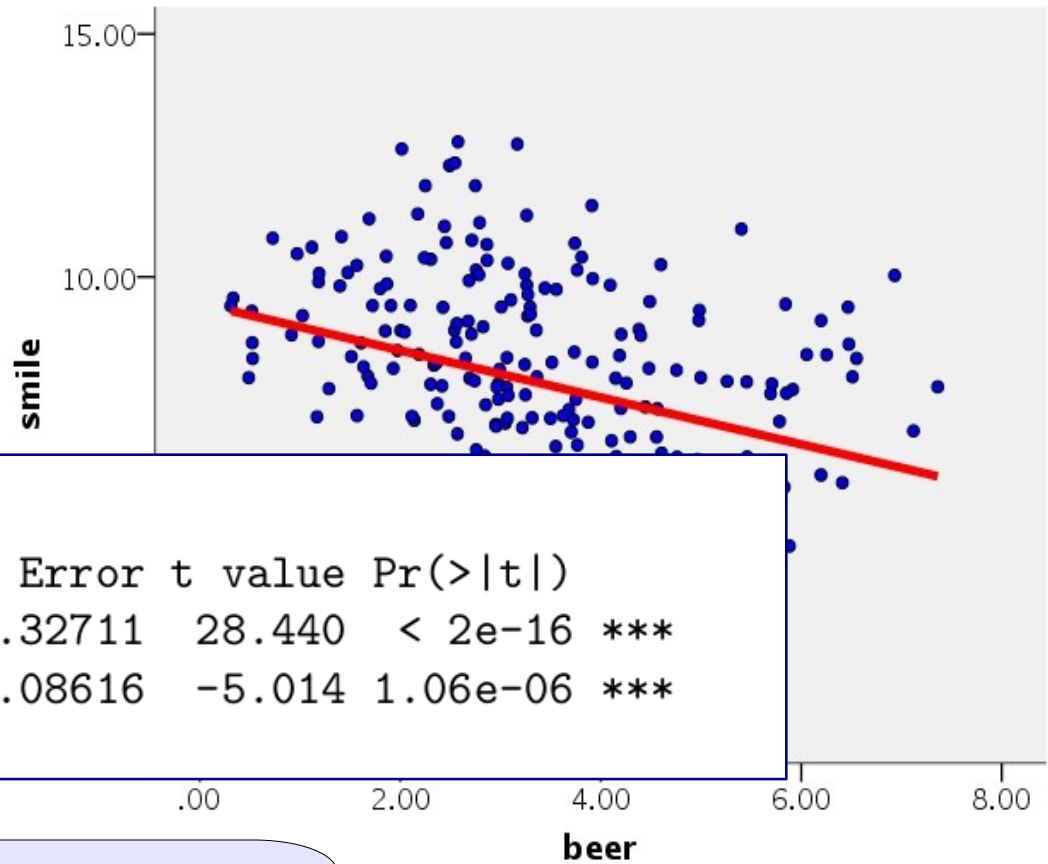
Esempio “birre” 2

Lo scatterplot mostra una distribuzione differente dall'esempio precedente



Esempio “birre” 2

La regressione semplice conferma il risultato assai differente



```
##  
## Coefficients:  
##           Estimate Std. Error t value Pr(>|t|)  
## (Intercept)  9.30306    0.32711  28.440  < 2e-16 ***  
## beer        -0.43198    0.08616  -5.014  1.06e-06 ***  
## ---
```

**Relazione
negativa**

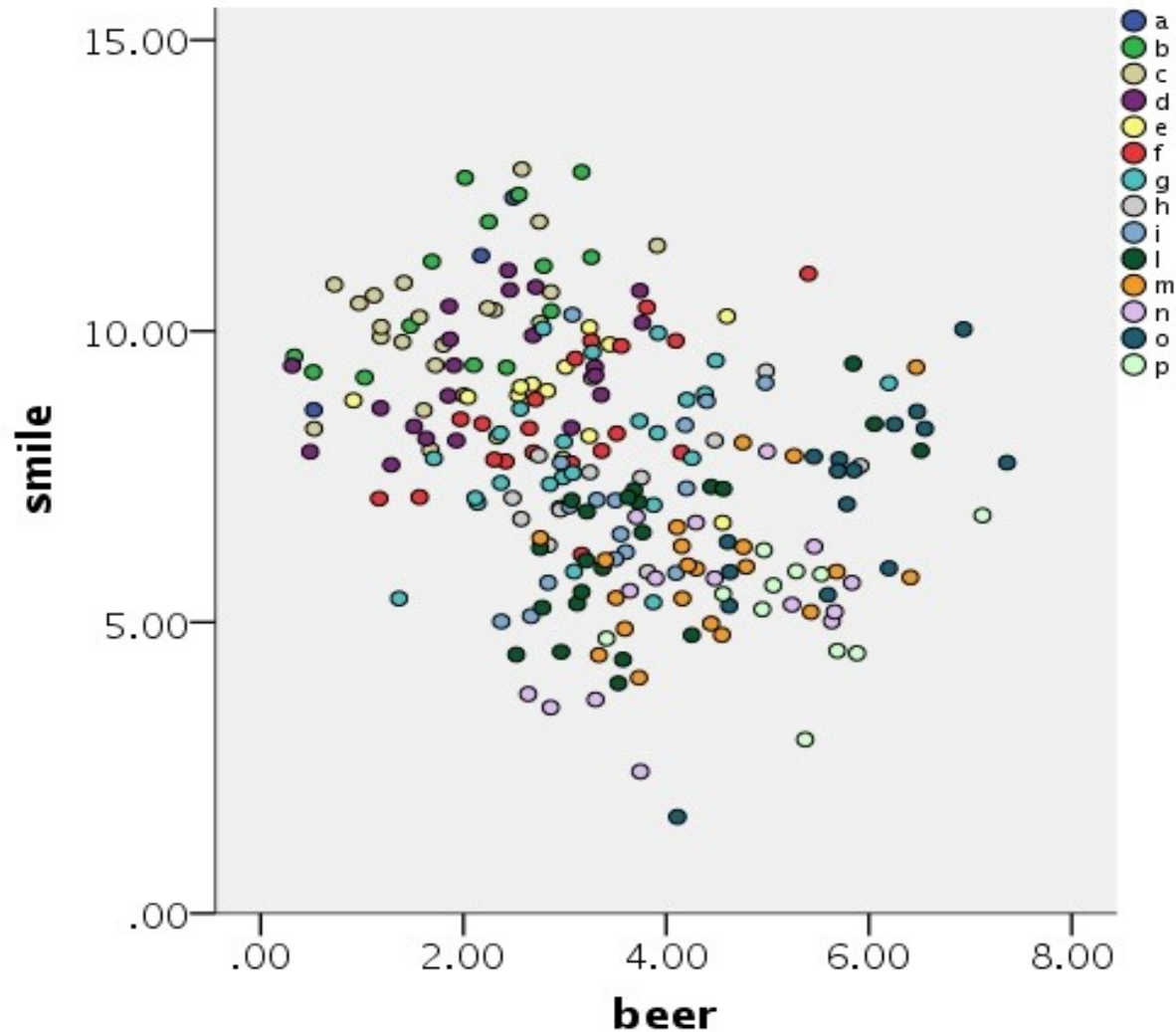
Possibili spiegazioni

I risultati potrebbero essere distorti (e ciò spiegherebbe il risultato inatteso) dal non aver considerato la struttura dei dati

- I dati infatti:
 - I soggetti sono stati campionati in diversi bar
 - Ogni bar potrebbe avere caratteristiche particolari (ambiente, qualità della birra, etc) che condizionano la relazione tra le variabili
 - I soggetti in ogni singolo bar potrebbero essere più simili tra loro di quando lo siano soggetti in bar diversi

Scatterplot per Bar

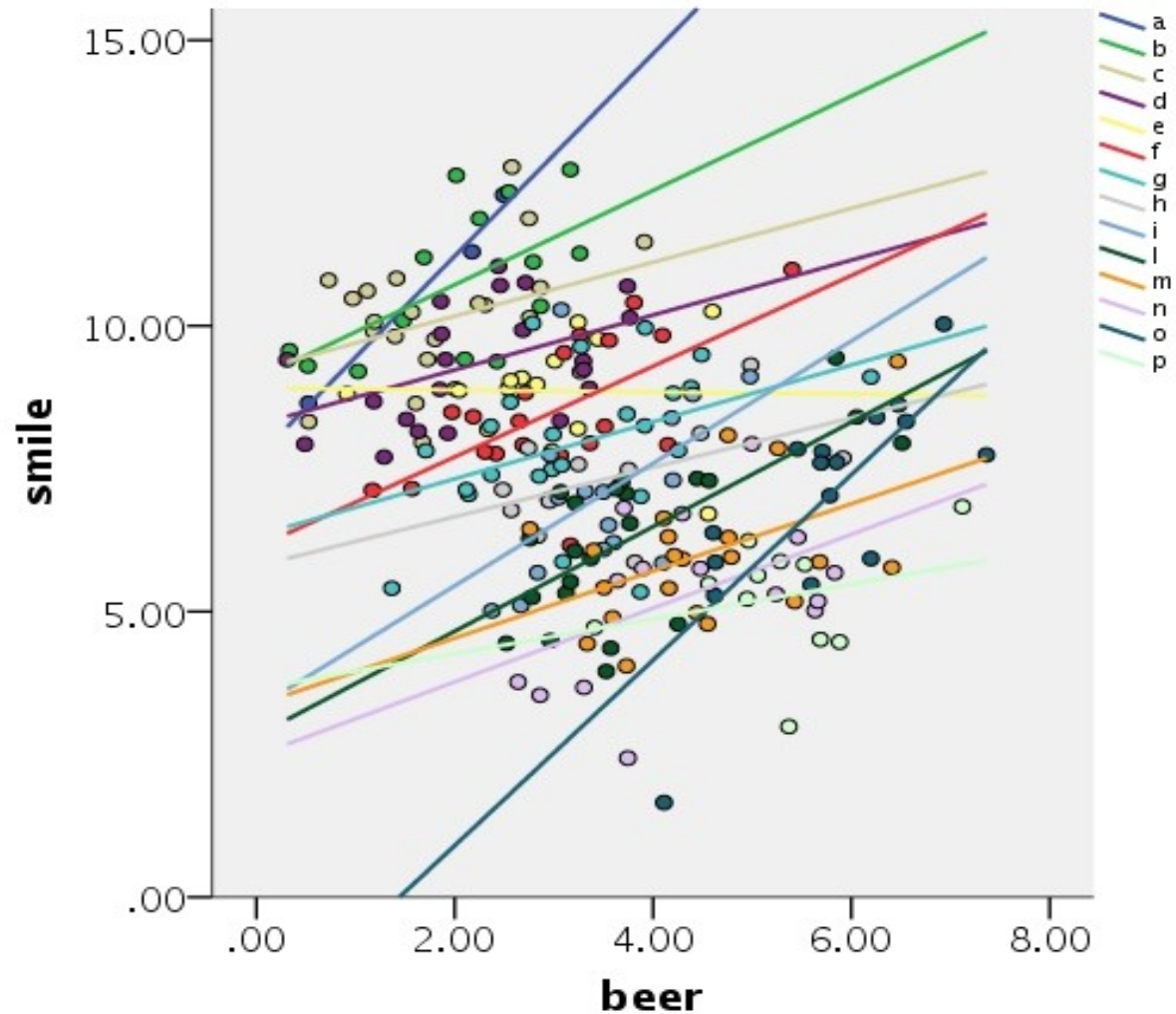
Lo scatterplot, distinguendo per bar, offre indicazioni



Bar

Scatterplot per Bar

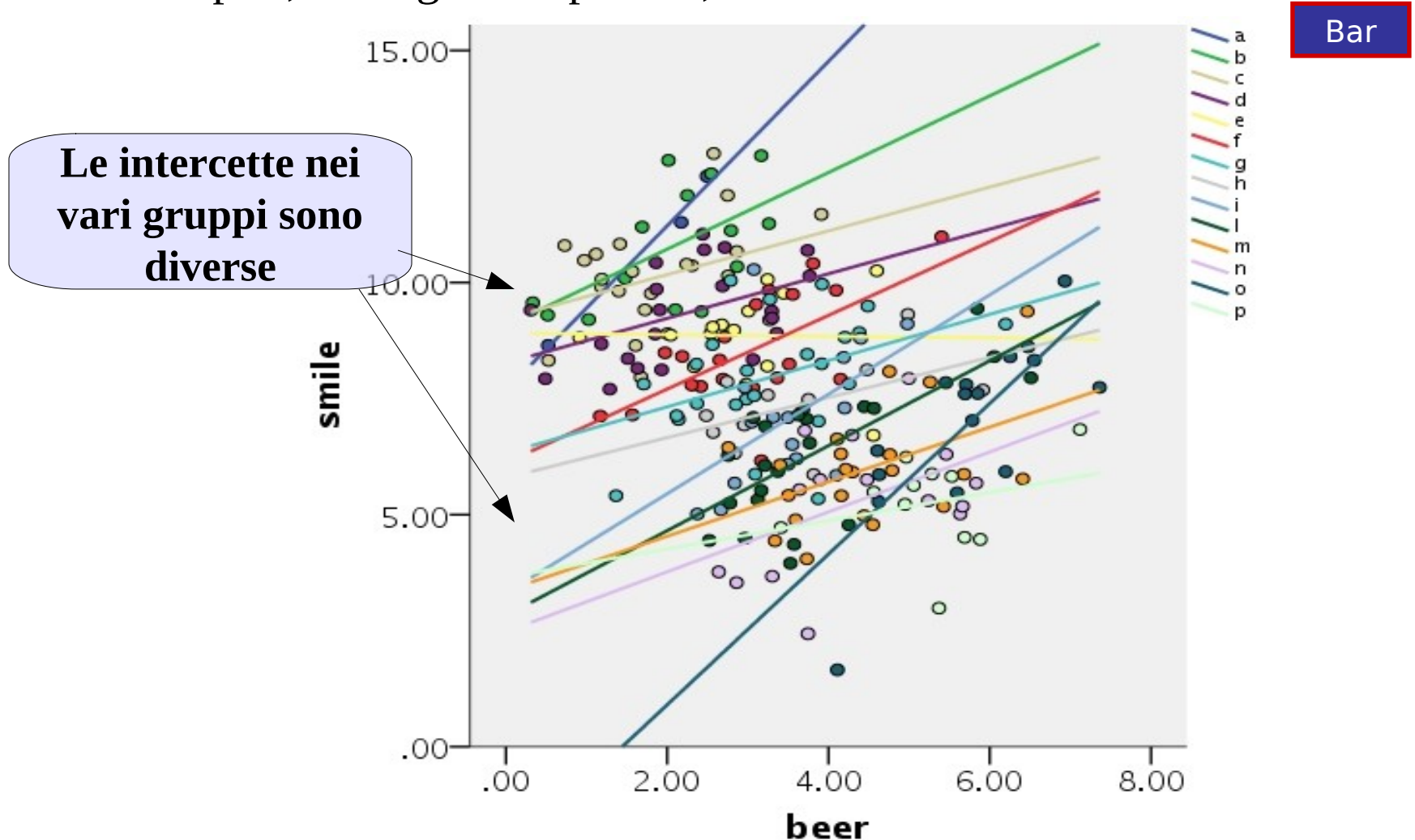
Lo scatterplot, distinguendo per bar, offre indicazioni



Bar

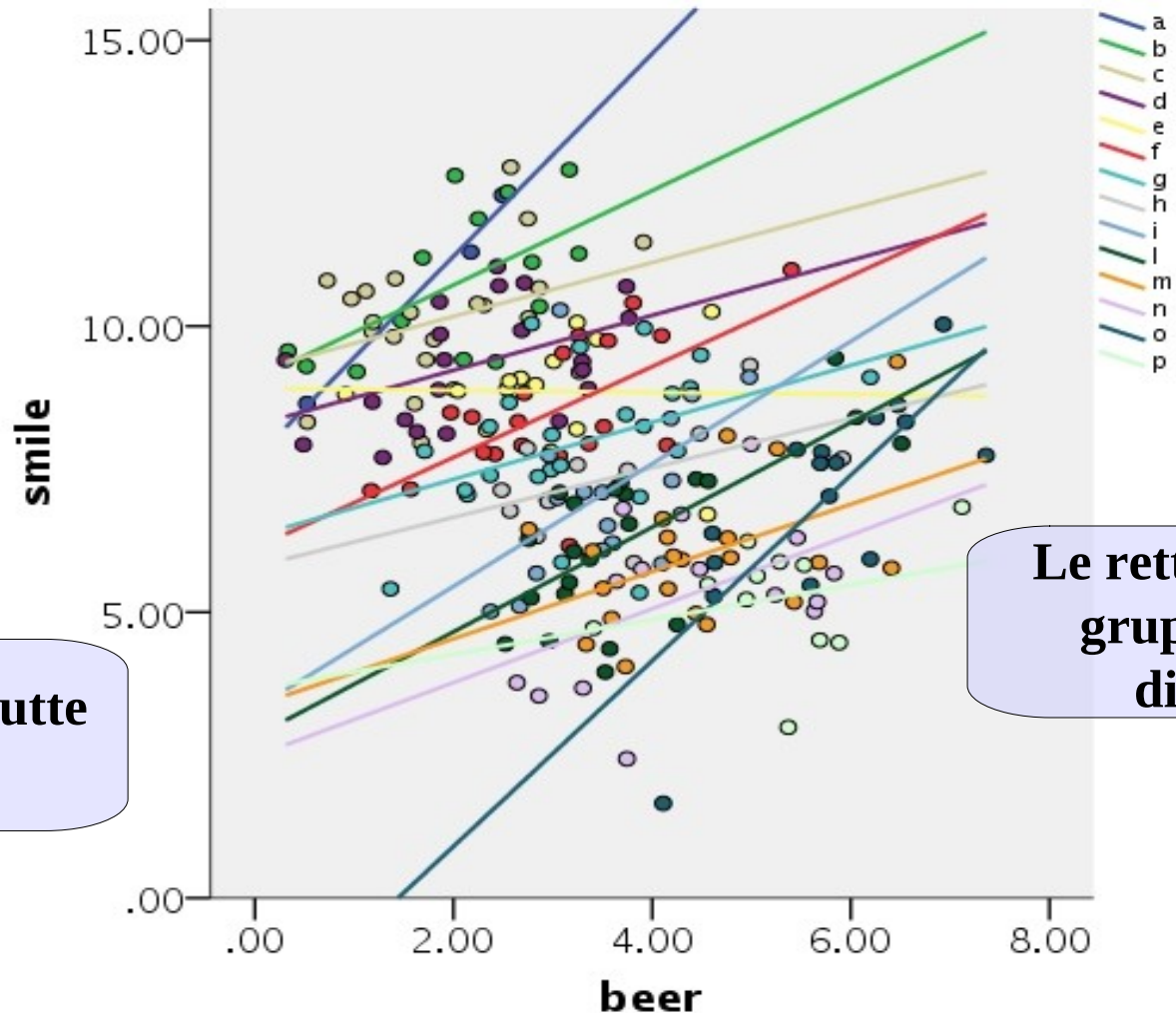
Scatterplot per Bar

Lo scatterplot, distinguendo per bar, offre indicazioni



Scatterplot per Bar

Lo scatterplot, distinguendo per bar, offre indicazioni



Bar

Le rette sono tutte positive

Le rette nei vari gruppi sono diverse

Modello

- Sembrerebbe che considerando tutti i soggetti come equivalenti ed indipendenti (assunzione della regressione) otteniamo un risultato distorto
- Se stimassimo un modello in cui la retta di regressione (intercetta e coefficiente B) sia diversa in ogni gruppo, avremmo dei risultati più soddisfacenti

Modello

- Definiamo dunque una regressione per ogni gruppo

 y_{ij}

Numero di sorrisi del soggetto i nel gruppo j

$$\hat{y}_{ia} = a_a + b_a \cdot x_{ia}$$

$$\hat{y}_{ib} = a_b + b_b \cdot x_{ib}$$

$$\hat{y}_{ic} = a_c + b_c \cdot x_{ic}$$

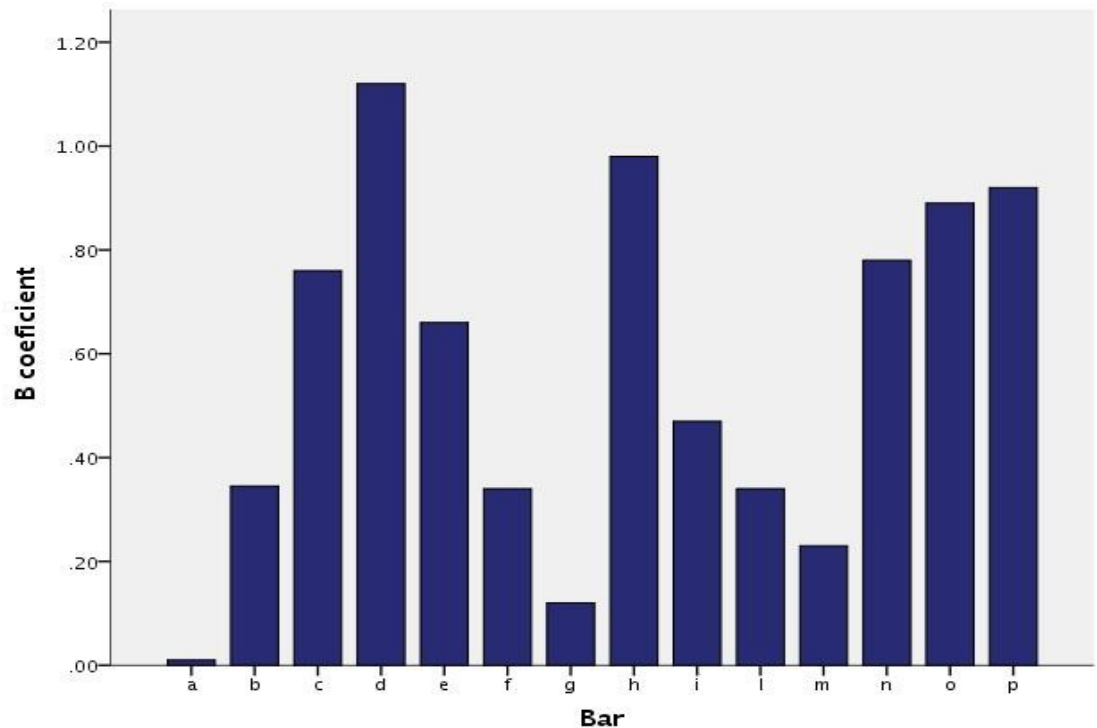
$$\hat{y}_{ij} = a_j + b_j \cdot x_{ij}$$

**In queste regressioni, sia
l'intercetta che i
coefficienti sono diversi
(non fissi) nei vari gruppi**

Coefficienti variabili

- Se i coefficienti cambiano nei vari gruppi, ovviamente non sono fissi (!!!)

I coefficienti avranno una distribuzione rispetto ai bar per i quali sono calcolati

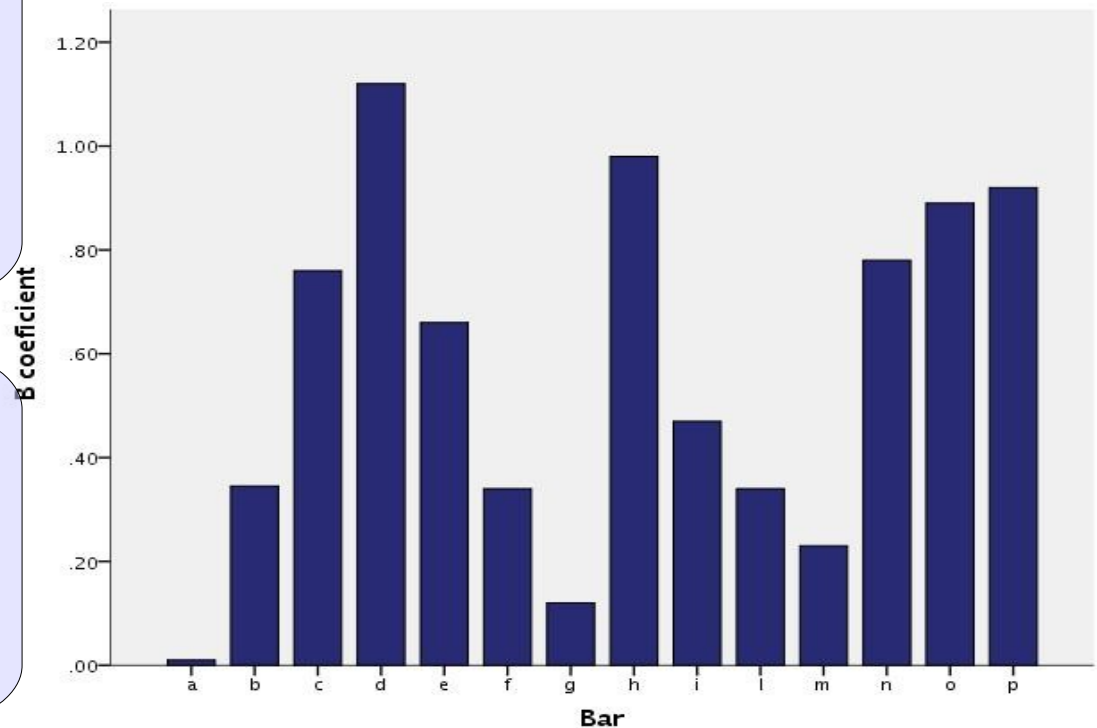


Coefficienti random

- I coefficienti che cambiano sono definiti **coefficienti random**

I coefficienti avranno una distribuzione random (cioè avranno una loro variabilità)

Cioè, nella popolazione esiste una variazione random dei coefficienti

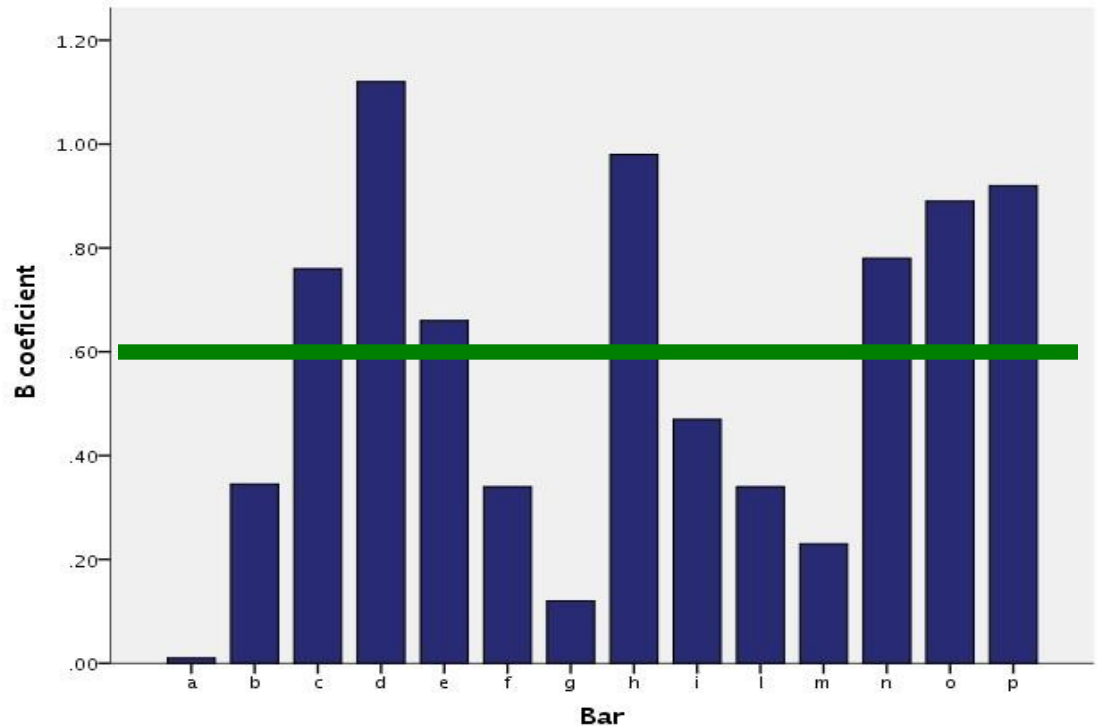


Media dei Coefficiente

- Se i coefficienti sono delle variabili, avranno una loro **media** ed una loro **varianza**

MEDIA

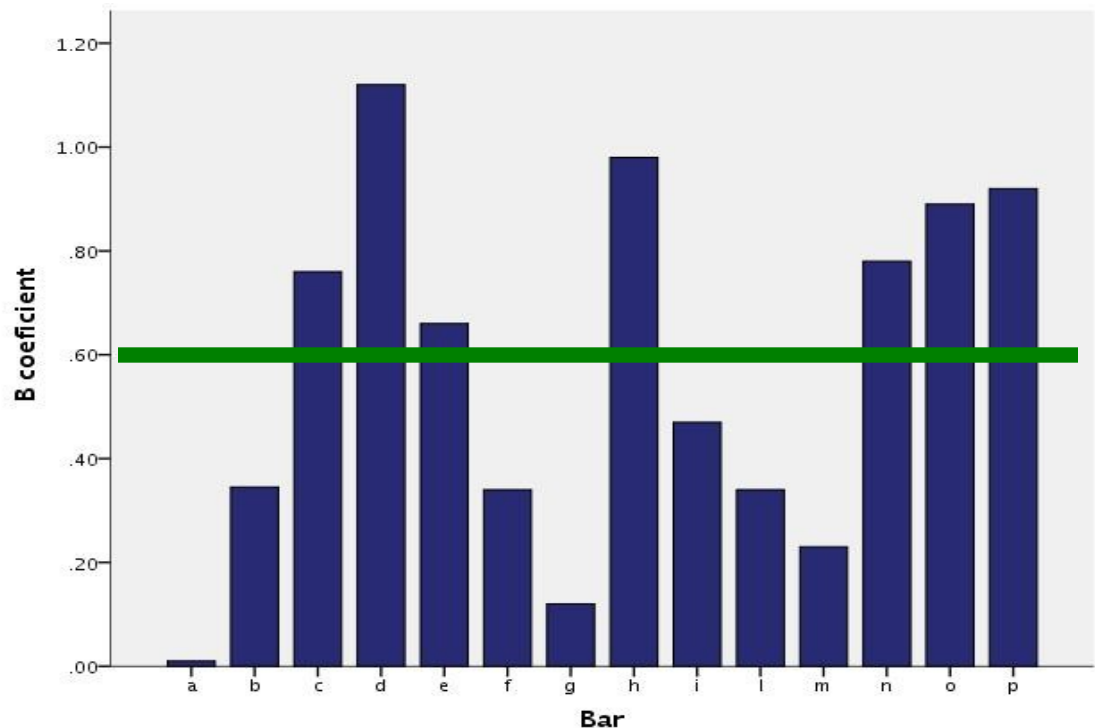
$$\bar{b} = \frac{\sum_j b_j}{k}$$



Coefficienti fissi

- La media dei coefficienti per bar indica la relazione (media) tra birre e sorrisi in tutto il campione

La media (come visto prima) è un parametro fisso del modello che descrive la distribuzione dei coefficienti nei cluster (bar)



Modello

- Definiamo ora un modello con le varie regressioni per cluster e la loro media

Una regressione per cluster

$$\hat{y}_{ij} = a_j + b_j \cdot x_{ij}$$

Ogni coefficiente è espresso come deviazione dalla media dei coefficienti

$$b'_j = b_j - \bar{b}$$

Modello generale

$$\hat{y}_{ij} = a_j + b'_j \cdot x_{ij} + \bar{b} \cdot x_{ij}$$

Modello

- Definiamo ora un modello con le varie regressioni per cluster e la loro media

Modello generale

$$\hat{y}_{ij} = a_j + b'_j \cdot x_{ij} + \bar{b} \cdot x_{ij}$$

**Coefficienti
random**

Coefficiente fisso

Modello Misto

- Definiamo ora un modello con le varie regressioni per cluster e la loro media

Modello generale

$$\hat{y}_{ij} = a_j + b'_j \cdot x_{ij} + \bar{b} \cdot x_{ij}$$

Coefficienti
random

Coefficiente fisso

I modelli che contengono coefficienti sia random
che fissi sono definiti **modelli misti**
(mixed models)

Modello Misto

- Analogamente

Una regressione per cluster

$$\hat{y}_{ij} = a_j + b_j \cdot x_{ij}$$

Ogni intercetta espressa come deviazione dalla media delle intercette

$$a'_j = a_j - \bar{a}$$

Modello generale

$$\hat{y}_{ij} = \bar{a} + a'_j + b'_j \cdot x_{ij} + \bar{b} \cdot x_{ij}$$

- Interpretazione

Modello generale

$$\hat{y}_{ij} = \bar{a} + a'_j + b'_j \cdot x_{ij} + \bar{b} \cdot x_{ij}$$

**Il punteggio della VD (i sorrisi)
di ogni soggetto in un dato
cluster (bar) è influenzato da:**

Modello Misto

- Interpretazione

Modello generale

$$\hat{y}_{ij} = \bar{a} + a'_j + b'_j \cdot x_{ij} + \bar{b} \cdot x_{ij}$$

**La media dei valori attesi di Y
per x=0**

**Per x=0, in media quanto è
grande y**

Modello Misto

- Interpretazione

Modello generale

$$\hat{y}_{ij} = \bar{a} + a'_j + b'_j \cdot x_{ij} + \bar{b} \cdot x_{ij}$$

**I valori attesi di y per x=0 in
ogni cluster (bar)**

**Per x=0, quanto devo
aggiungere o sottrarre al
valore atteso medio per un
cluster specifico**

Modello Misto

- Interpretazione

Modello generale

$$\hat{y}_{ij} = \bar{a} + a'_j + b'_j \cdot x_{ij} + \bar{b} \cdot x_{ij}$$

**L'effetto specifico di x su y per
il cluster j**

**In un dato cluster, quanto
aumenta (o diminuisce)
l'effetto di x su y**

- Interpretazione

Modello generale

$$\hat{y}_{ij} = \bar{a} + a'_j + b'_j \cdot x_{ij} + \bar{b} \cdot x_{ij}$$

L'effetto medio di x su y

**In media, quanto aumenta y
per ogni unità in più di x**

GLM come sottocaso

La corrispondenza logica tra le varie tecniche inerenti al Modello Lineare Generale con le tecniche inerenti ai Modelli Misti è data dal fatto che il GLM può essere pensato come sottocaso dei MM

MM

$$\hat{y}_{ij} = \bar{a} + a'_j + b'_j \cdot x_{ij} + \bar{b} \cdot x_{ij}$$

GLM

$$\hat{y}_{ij} = \hat{a} + \bar{b} \cdot x_{ij}$$

Notazione

Per chiarezza, useremo questa notazione

$$y_{ij} = \bar{a} + a_j + b_j \cdot x_{ij} + \bar{b} \cdot x_{ij} + e_{ij}$$

y_{ij}, x_{ij}

Variabili osservate per caso i nel cluster j

\bar{a}, \bar{b}

Effetti fissi

a_j, b_j

**Effetti random calcolati nel cluster j
espressi come deviazione dalla loro media**

e_{ij}

Errore associato al singolo caso i

Varianze

Per chiarezza, useremo questa notazione

$$y_{ij} = \bar{a} + a_j + b_j \cdot x_{ij} + \bar{b} \cdot x_{ij} + e_{ij}$$

 σ_a

Varianza dei coefficienti a

 σ_b

Varianza dei coefficienti b

 σ

Varianza di errore

 σ_{ab}

Covarianza tra i coefficienti a e b

- In sostanza, i modelli misti consentono di stimare gli effetti di VI su una VD, consentendo a tali effetti di variare in diverse unità di misurazione (cluster).
- Gli effetti che variano sono detti **effetti random**
- Gli effetti che non variano (cioè gli effetti medi uguali per tutto il campione) sono detti **effetti fissi**

- Per stimare correttamente un modello misto, si deve semplicemente capire quale siano gli effetti random, e per quali unità variano (quali sono i cluster)
- Una volta stimato il modello, gli **effetti fissi** si interpretano esattamente come nel GLM (regressione/anova etc)
- Gli **effetti random** generalmente non si interpretano, ma se ne può studiare la variabilità
- La definizione corretta del modello, consente di ottenere stime e errori standard (e dunque test inferenziali) corretti

Costruire il modello

Per costruire il modello, dobbiamo rispondere a tre semplici domande

- Quali sono i cluster su cui variano gli effetti random?
- Quali sono gli effetti fissi?
- Quali sono gli effetti random?

Variabili cluster

- Quali sono i cluster su cui variano gli effetti random?
- Quali sono gli effetti fissi?
- Quali sono gli effetti random?
 - Qualunque variabile che raggruppa le osservazioni (i casi o le osservazioni) in modo che i punteggi possano essere più simili entro i gruppi che tra gruppi
 - Una variabili i cui livelli rappresentano un campione casuale di gruppi estratti da una popolazione più ampia di gruppi

Effetti Fissi

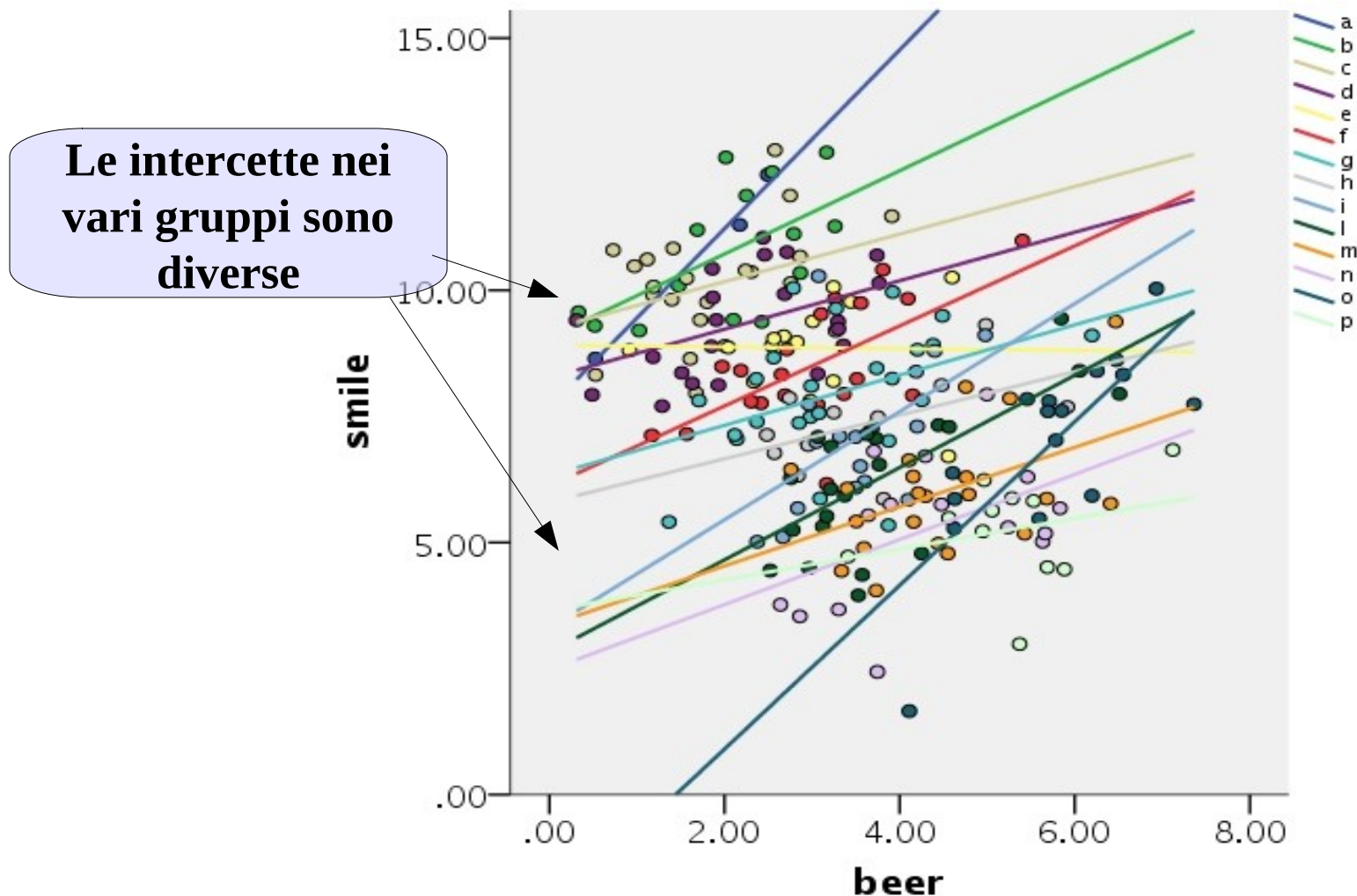
- Quali sono i cluster su cui variano gli effetti random?
- **Quali sono gli effetti fissi?**
- Quali sono gli effetti random?
 - Qualunque effetto che ci interessa in generale (equivalenti agli effetti nel GLM)
 - Esempio: L'effetto di birra sui sorrisi

Effetti Random

- Quali sono i cluster su cui variano gli effetti random?
- Quali sono gli effetti fissi?
- **Quali sono gli effetti random?**
 - Qualunque effetto che può variare da cluster a cluster
 - (Dunque:) **Qualunque effetto (coefficiente) che può essere calcolato dentro ogni cluster**
 - Esempio: le intercette e il B di birre su sorrisi

Birre al Bar

Definiamo il modello, iniziando dal più semplice



Birre al Bar

Definiamo un modello dove le intercette possono variare nei diversi bar, mentre il coefficiente b è fisso ed uguale per tutti i bar

$$y_{ij} = \bar{a} + a_j + \bar{b} \cdot x_{ij} + e_{ij}$$

- Quali sono gli effetti fissi?
- Quali sono gli effetti random?
- Quali sono i cluster su cui variano gli effetti random?

Birre al Bar

Definiamo un modello dove le intercette possono variare nei diversi bar, mentre il coefficiente b è fisso ed uguale per tutti i bar

$$y_{ij} = \bar{a} + a_j + \bar{b} \cdot x_{ij} + e_{ij}$$

- Quali sono gli effetti fissi? Intercetta e effetto di birre
- Quali sono gli effetti random? Intercetta
- Quali sono i cluster su cui variano gli effetti random? bar

Vari autori e libri definiscono questomodello:

Random-intercepts regression

altri

Intercepts-as-outcomes model

- In jamovi “mixed model” è nel modulo GAMLj

The screenshot displays the jamovi software interface with the 'Mixed Model' analysis module selected. The top menu bar includes 'Data' and 'Analyses'. The 'Analyses' menu is open, showing various statistical tests: Exploration, T-Tests, ANOVA, Regression, Frequencies, Factor, Base R, TOSTER, MAJOR, medmod, Linear Models, and Modules. The 'Mixed Model' module is highlighted.

The 'Mixed Model' panel on the left contains a list of variables (A, case, smile, beer, bar) and four input fields for 'Dependent Variable', 'Factors', 'Covariates', and 'Cluster variables'. Below these fields are checkboxes for 'REML' and 'Confidence intervals', with a dropdown for 'Interval' set to '95 %'. A list of options is visible at the bottom: Fixed Effects, Random Effects, Factors Coding, Covariates Scaling, Post Hoc Tests, Fixed Effects Plots, Simple Effects, and Estimated Marginal Means.

The right panel, titled 'Mixed Model', displays the 'Model Info' section with a table of instructions:

Model Info	
Get started	Select the dependent variable
Get started	Select at least one cluster variable
Get started	Select at least one term in Random Effects
Optional	Select factors and covariates

Below this is the 'Fixed Effect ANOVA' section with a table:

	F	Num df	Den df	p

Next is the 'Fixed Effects Parameter Estimates' section with a table:



Effect	Contrast	Estimate	SE	Lower	Upper	df	t	p


Finally, the 'Random Components' section has a table:

Groups	Name	SD	Variance

Definiamo il ruolo delle variabili


Mixed Model


 A
 case

→  smile

→

Covariates

→  beer

→  bar

Estimation Confidence Intervals

☒ REML ☒ Confidence intervals Interval %

Definiamo gli effetti fissi

▼ Fixed Effects

Components		Model Terms
beer	→	beer
	→ ▼	

☒ Fixed Intercept

Definiamo la componente random

▼ | Random Effects

Components

beer | bar

→

Random Coefficients

Intercept | bar

☒ Correlated Effects

Risultati

- Una volta definita la componente random, otteniamo i risultati

Mixed Model

Model Info

Info	
Estimate	Linear mixed model fit by REML
Call	smile ~ 1 + (1 bar) + beer
AIC	811.1613
R-squared Marginal	0.0894
R-squared Conditional	0.8172

**R-squared Marginal: Quanta
varianza è spiegata dai fixed
effects da soli**

**R-squared Conditional: quanta
varianza è spiegata dai fixed e
dai random effects tutti insieme**

Componente random

La varianza delle intercette è diversa da zero, dunque le intercette variano, dunque ok che siano random

Random Components

Groups	Name	SD	Variance
bar	(Intercept)	2.40	5.77
	Residual	1.20	1.45

Note. Numer of Obs: 234 , groups: bar , 15

F-test per l'effetto fisso di beer

Fixed Effect ANOVA

	F	Num df	Den df	p
beer	46.0	1	229	< .001

Note. Satterthwaite method for degrees of freedom

Output

Se tutto è ok, guarderemo gli effetti fissi per valutare ed interpretare la relazione tra VD e VI

Fixed Effects Parameter Estimates

Effect	Contrast	Estimate	SE	95% Confidence Interval		df	t	p
				Lower	Upper			
(Intercept)	Intercept	7.778	0.6276	6.548	9.008	13.2	12.39	< .001
beer	beer	0.548	0.0808	0.390	0.706	229.4	6.79	< .001

Output

Se tutto è ok, guarderemo gli effetti fissi per valutare ed interpretare la relazione tra VD e VI

Fixed Effects Parameter Estimates

Effect	Contrast	Estimate	SE	95% Confidence Interval		df	t	p
				Lower	Upper			
(Intercept)	Intercept	7.778	0.6276	6.548	9.008	13.2	12.39	< .001
beer	beer	0.548	0.0808	0.390	0.706	229.4	6.79	< .001

Coefficiente b: In media, per ogni birra in più i sorrisi aumentano di .548

Intercetta: In media, per zero (media di) birre ci attendiamo 7.7 sorrisi

- Plots

▼ | Fixed Effects Plots

→

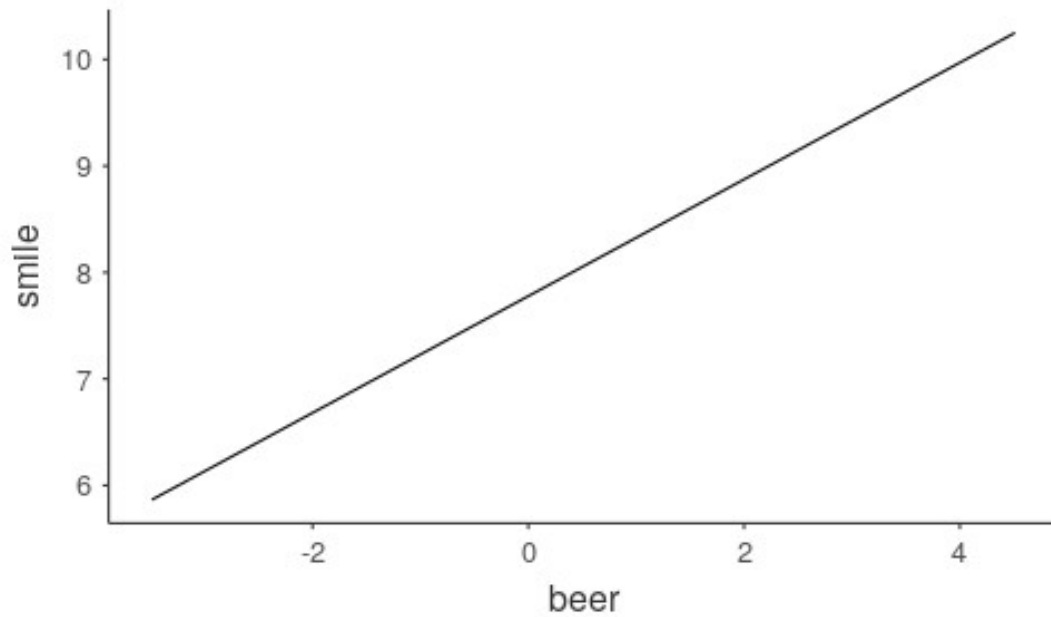
beer

→

→

Plot dell'effetto fisso

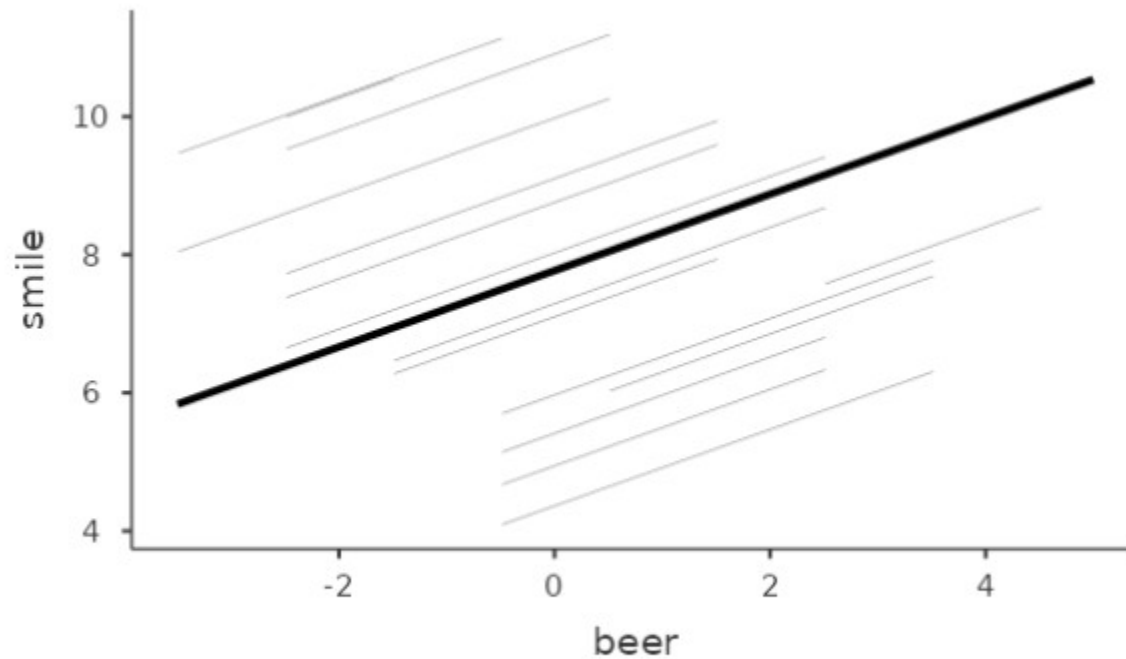
Fixed Effects Plots



R Plot

Possiamo anche plottare gli effetti random

Effects Plots



Note: Random effects are plotted by bar

Birre al Bar

Definiamo un modello dove le intercette e i coefficienti di regressione possono variare nei diversi bar,

$$y_{ij} = \bar{a} + a_j + \bar{b} \cdot x_{ij} + b \cdot x_{ij} + e_{ij}$$

- Quali sono gli effetti fissi? Intercetta e effetto di birre
- Quali sono gli effetti random? Intercetta ed effetto di birre
- Quali sono i cluster su cui variano gli effetti random? bar

Vari autori e libri definiscono questo modello:

Random-coefficients regression

Altri come

Intercepts- and Slopes-as-outcomes model

Effetti random

Definiamo un modello dove le intercette possono variare nei diversi bar, mentre il coefficiente b è fisso ed uguale per tutti i bar

$$y_{ij} = \bar{a} + a_j + \bar{b} \cdot x_{ij} + b \cdot x_{ij} + e_{ij}$$

The screenshot shows a software interface for defining random effects. At the top, there is a dropdown menu labeled "Random Effects" with a downward arrow. Below this, the interface is divided into two main sections: "Components" on the left and "Random Coefficients" on the right. A central button with a right-pointing arrow is positioned between these two sections. The "Random Coefficients" section contains two lines of text: "Intercept | bar" and "beer | bar". At the bottom of the interface, there are two tabs: "Effects visualization" and "Tests".

Output

Poi guarderemo la variabilità degli effetti random, per capire se è
abbiamo fatto bene a settarli come tali

Random Components

Groups	Name	SD	Variance	ICC
bar	(Intercept)	2.417	5.8417	0.803
	beer	0.167	0.0278	
Residual		1.196	1.4314	

Note. Number of Obs: 234 , groups: bar 15

Random Parameters correlations

Groups	Param.1	Param.2	Corr.
bar	(Intercept)	beer	-0.766

**La varianza dei b è piccola
dunque i b variano molto
poco, dunque potremmo
tenere il modello precedente**

Varianze

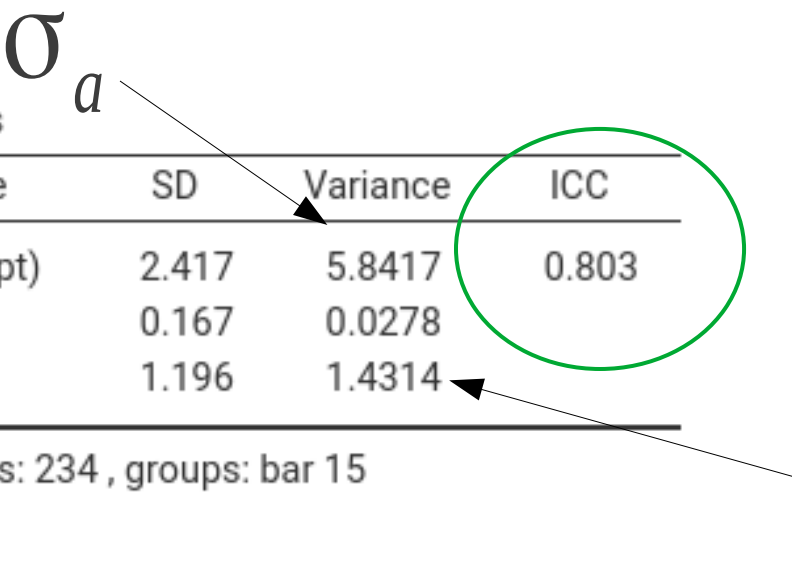
La varianza degli effetti random ci indica quanta variabilità c'è tra i cluster nell'effetto

- L'effetto random lo lasciamo anche se la varianza è molto piccola
- Se è zero (esattamente), l'effetto random deve essere tolto dal modello

Coefficiente di dipendenza

Possiamo quantificare la dipendenza tra punteggi mediante il **coefficiente di correlazione intraclass**

$$ICC = \frac{\sigma_a}{\sigma_a + \sigma}$$



Random Components				
Groups	Name	SD	Variance	ICC
bar	(Intercept)	2.417	5.8417	0.803
	beer	0.167	0.0278	
Residual		1.196	1.4314	

Note. Number of Obs: 234 , groups: bar 15

Output

Guarderemo gli effetti fissi per valutare ed interpretare la relazione tra VD e VI

Fixed Effects Parameter Estimates

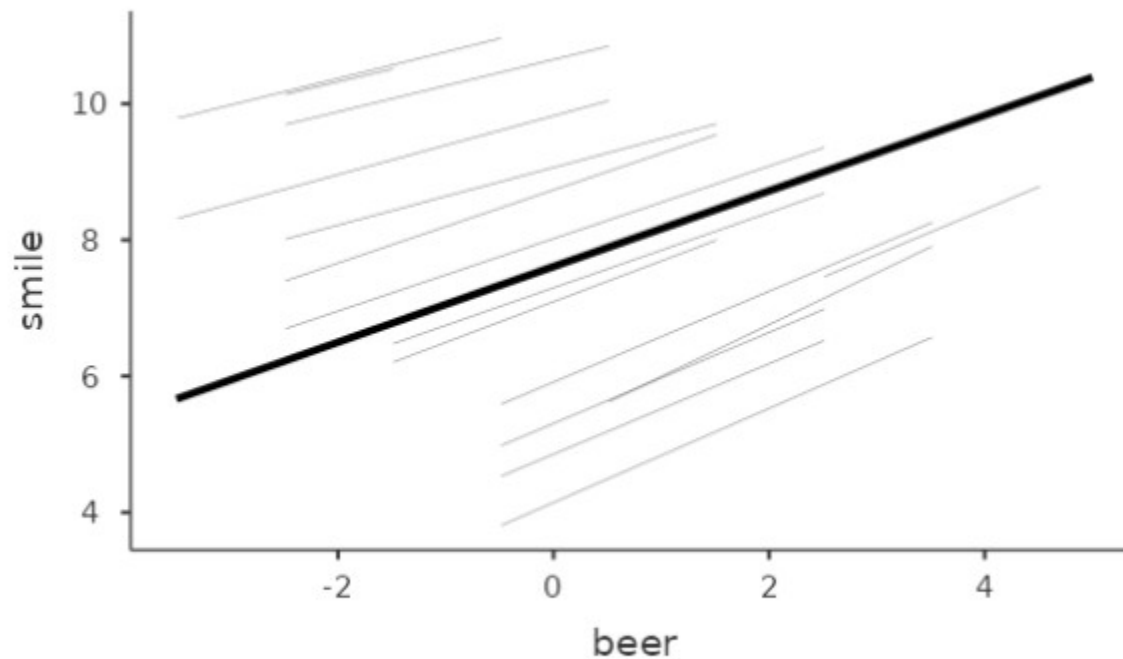
Names	Estimate	SE	95% Confidence Interval		df	t	p
			Lower	Upper			
(Intercept)	7.610	0.6335	6.368	8.851	12.93	12.01	< .001
beer	0.555	0.0925	0.374	0.737	7.23	6.00	< .001

Effetti non diversi da prima

Plots

- Plottando gli effetti random vediamo che non sono più tutti paralleli

Effects Plots



Note: Random effects are plotted by bar

Morale

- Il modello misto consente di estendere il modello lineare generale a cui problemi di analisi dei dati in cui la struttura dei dati non si adatta naturalmente
- I semplici concetti visti oggi, combinati alle conoscenze relative al GLM, ci consentiranno di stimare modelli misti per (quasi) tutte i problemi di ricerca (plausibili)

Il disegno a misure ripetute

Disegno a misure ripetute

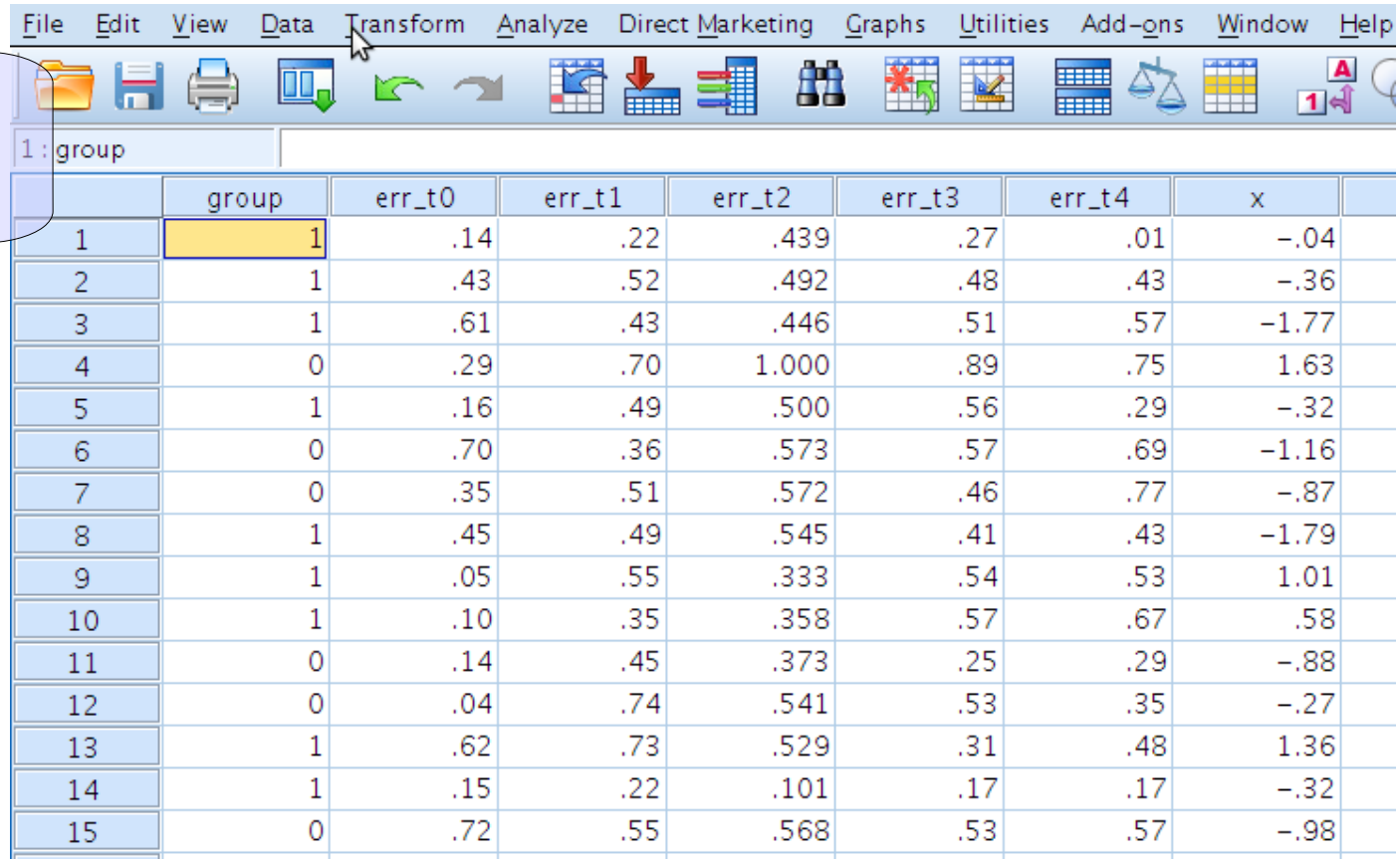
- Consideriamo un disegno a misure ripetute classico (within-subjects) in cui i livelli del fattore WS (5 differenti trials) sono misurati sulle stesse persone

		trial				
		1	2	3	4	5
Soggetti	1	Y11	Y21	Y31	Y41	Y51
	2	Y12	Y22	Y32	Y42	Y52
	3	Y13	Y23	Y33	Y43	Y53
					
	N	Y1n	Y2n	Y3n	Y4n	Y5n

Formato file Standard

- Spesso (in SPSS) il file è organizzato nel formato “wide”, una riga un soggetto

Una riga, un
soggetto



	group	err_t0	err_t1	err_t2	err_t3	err_t4	x
1	1	.14	.22	.439	.27	.01	-.04
2	1	.43	.52	.492	.48	.43	-.36
3	1	.61	.43	.446	.51	.57	-1.77
4	0	.29	.70	1.000	.89	.75	1.63
5	1	.16	.49	.500	.56	.29	-.32
6	0	.70	.36	.573	.57	.69	-1.16
7	0	.35	.51	.572	.46	.77	-.87
8	1	.45	.49	.545	.41	.43	-1.79
9	1	.05	.55	.333	.54	.53	1.01
10	1	.10	.35	.358	.57	.67	.58
11	0	.14	.45	.373	.25	.29	-.88
12	0	.04	.74	.541	.53	.35	-.27
13	1	.62	.73	.529	.31	.48	1.36
14	1	.15	.22	.101	.17	.17	-.32
15	0	.72	.55	.568	.53	.57	-.98

Formato file “Long”

- Per l'utilizzo dei modelli misti è necessario un formato “una misura, una riga”

Ogni riga
rappresenta una
misurazione

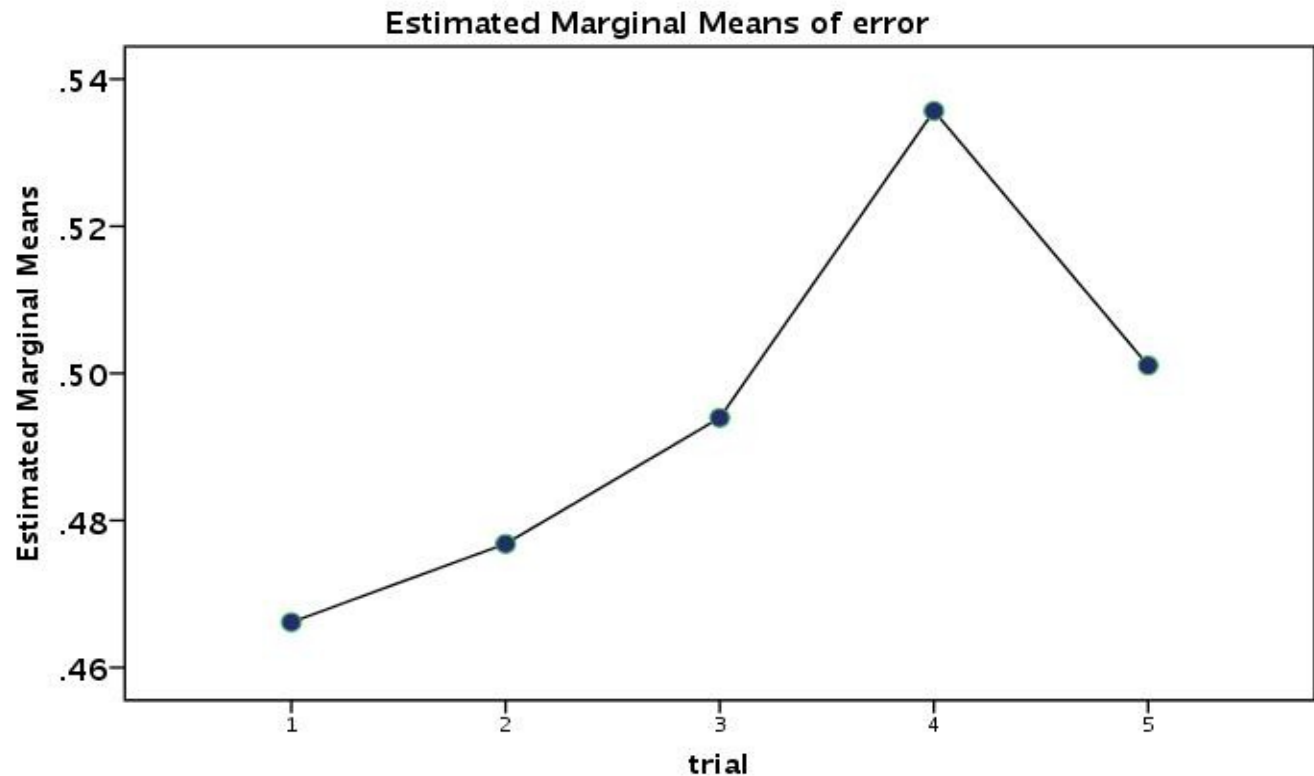
	id	group	x	trial	error
1	1	Cognitive Load	-0.03721233	1	0.139644148
2	1	Cognitive Load	-0.03721233	2	0.219611404
3	1	Cognitive Load	-0.03721233	3	0.439030338
4	1	Cognitive Load	-0.03721233	4	0.270522729
5	1	Cognitive Load	-0.03721233	5	0.009309587
6	2	Cognitive Load	-0.36044158	1	0.431302228
7	2	Cognitive Load	-0.36044158	2	0.518326274
8	2	Cognitive Load	-0.36044158	3	0.492408109
9	2	Cognitive Load	-0.36044158	4	0.483458141
10	2	Cognitive Load	-0.36044158	5	0.432801733
11	3	Cognitive Load	-1.76705741	1	0.612178698
12	3	Cognitive Load	-1.76705741	2	0.431445717
13	3	Cognitive Load	-1.76705741	3	0.446141329
14	3	Cognitive Load	-1.76705741	4	0.509173965
15	3	Cognitive Load	-1.76705741	5	0.572991706
16	4	Emotional Stressor	1.63388771	1	0.290753013
17	4	Emotional Stressor	1.63388771	2	0.702075839

GLM

- Potremmo analizzare questi dati mediante un modell GLM, ma incontreremo dei (gravi) problemi

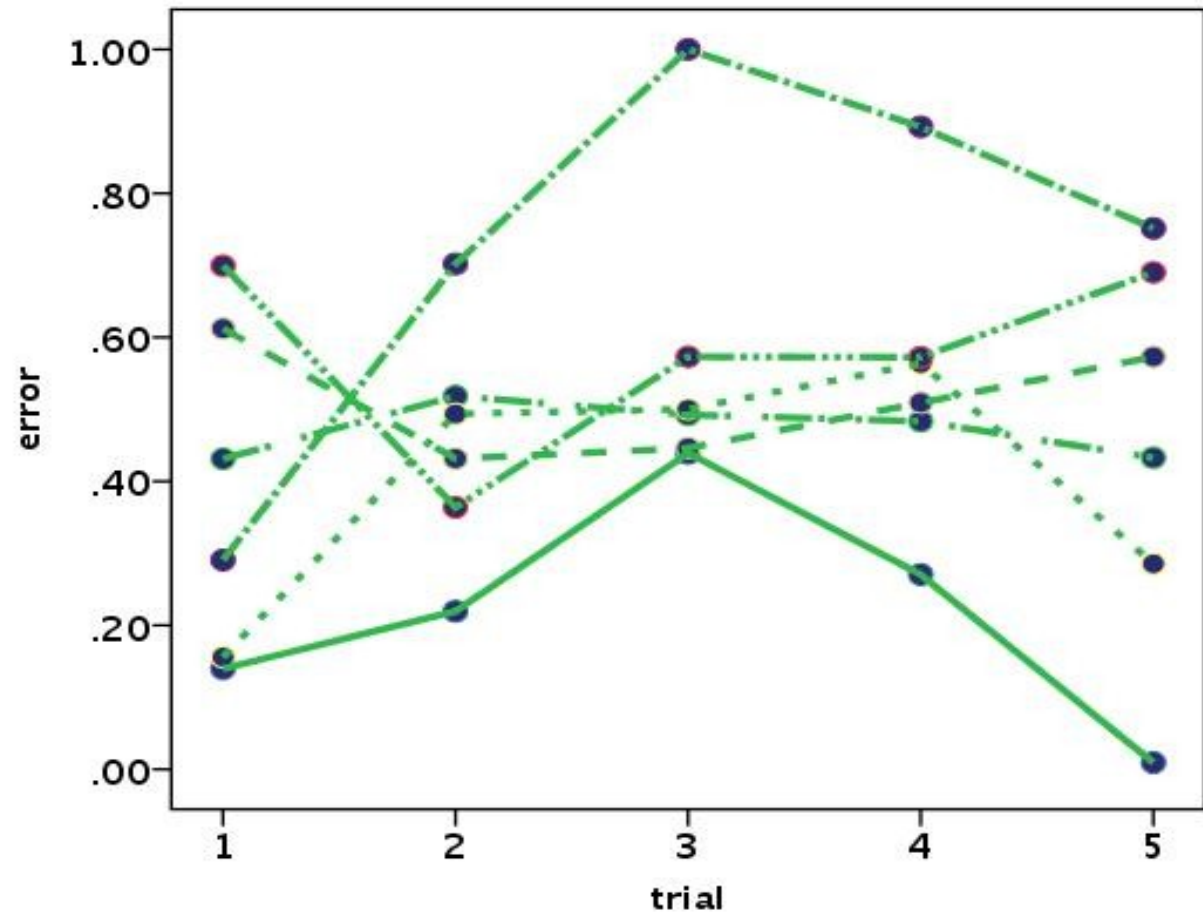
Effetti fissi

Medie dei 5 trials



Problemi

- Il primo problema è ogni soggetto, avendo tutte le misurazioni, esprime il suo proprio effetto di trial



Esempio per 6
soggetti

Soluzione (1)

Dunque per analizzare correttamente il disegno, dobbiamo considerare nel modello un termine che rappresenti la specificità di ogni soggetto. Questo termine sarà lo stesso in ogni soggetto

$$\begin{aligned} Y_{11} &= a + b \cdot T_1 + u_1 + e_{11} \\ Y_{12} &= a + b \cdot T_2 + u_1 + e_{12} \\ Y_{13} &= a + b \cdot T_3 + u_1 + e_{13} \end{aligned}$$

Stesso soggetto,
stesso errore

.....

$$\begin{aligned} Y_{i1} &= a + b \cdot T_1 + u_i + e_{i1} \\ Y_{i2} &= a + b \cdot T_2 + u_i + e_{i2} \\ Y_{i3} &= a + b \cdot T_2 + u_i + e_{i3} \end{aligned}$$

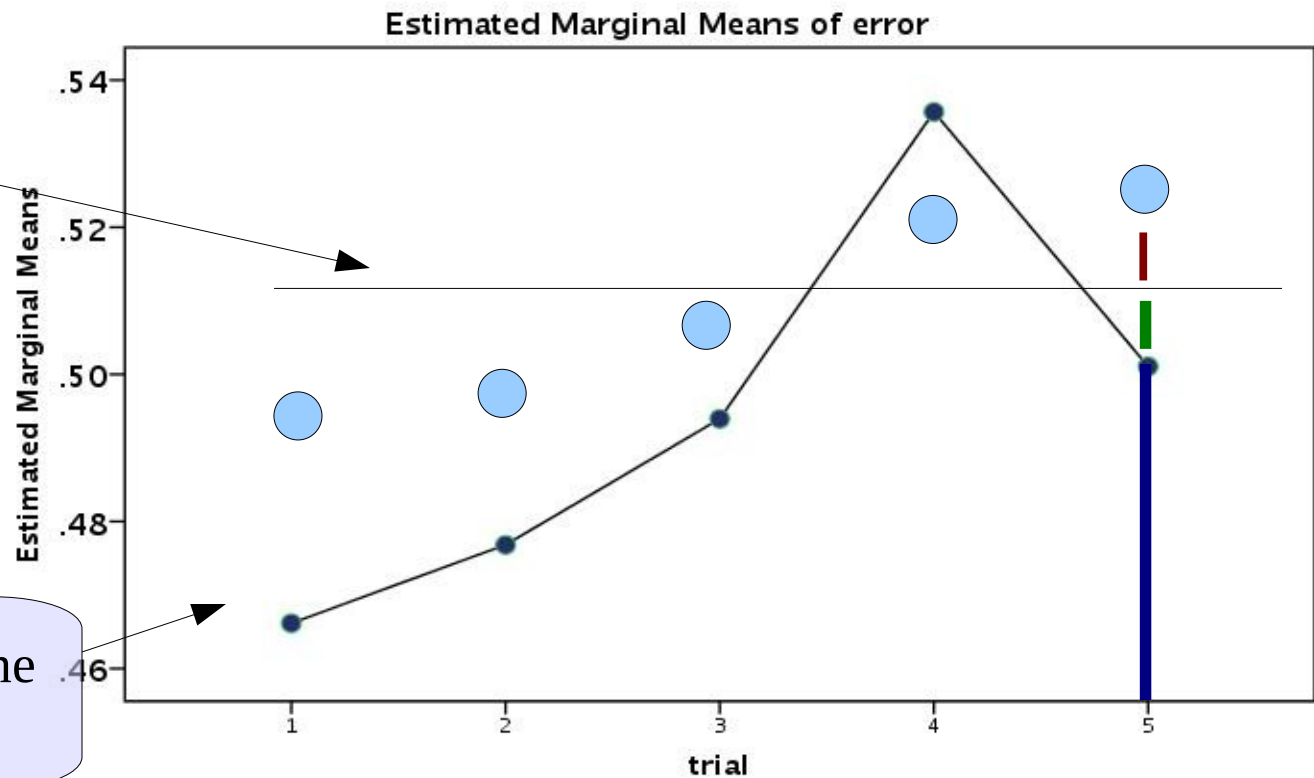
Stesso soggetto,
stesso errore

Componente individuale

Esempio per un
soggetto

Valore del tratto
individuale

Medie del campione
(effetti fissi)

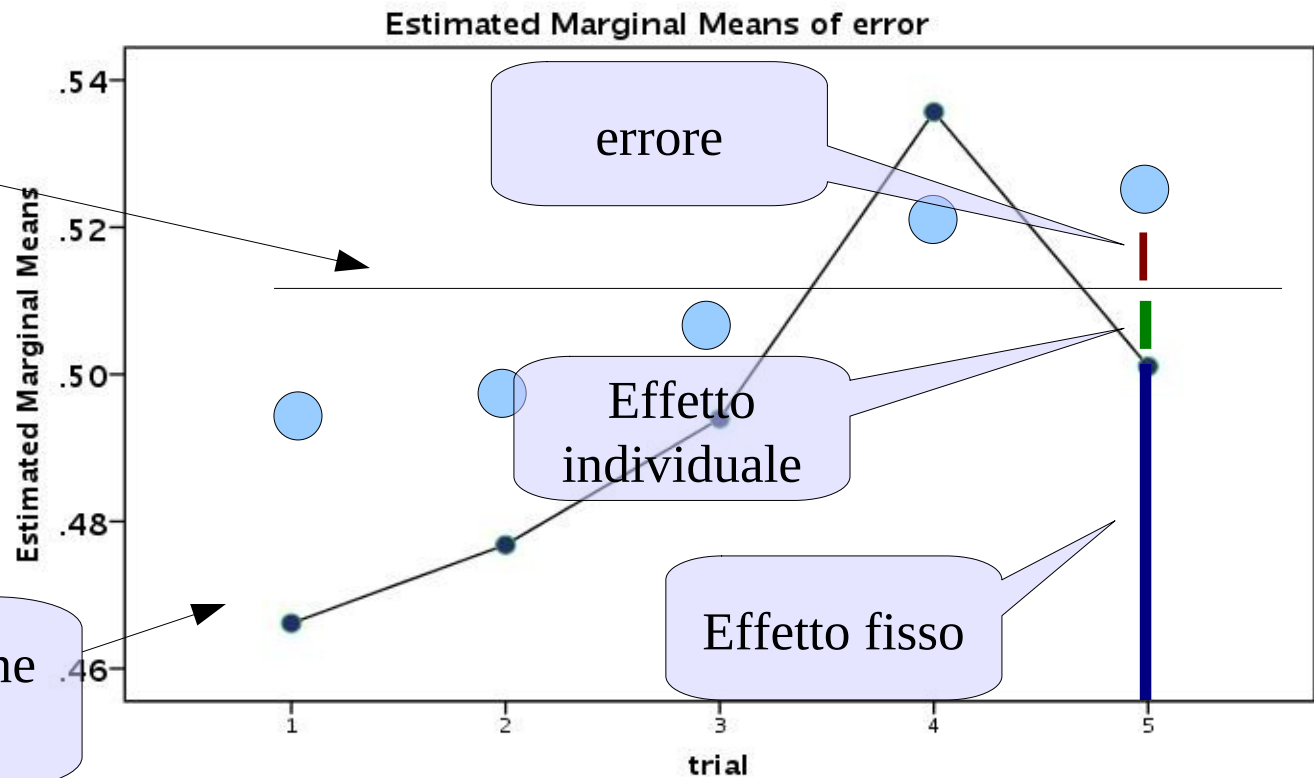


Componente individuale

Esempio per un
soggetto

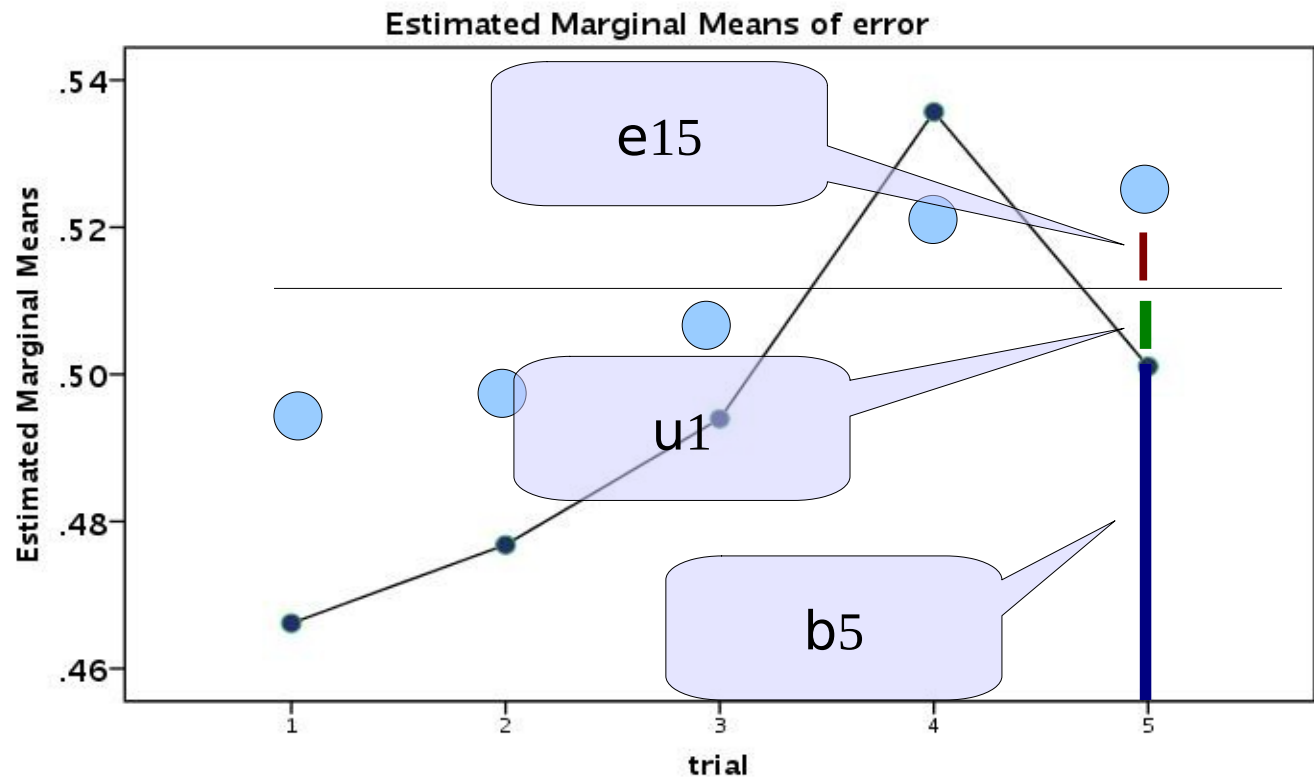
Valore del tratto
individuale

Medie del campione
(effetti fissi)



Componente individuale

$$Y_{11} = a + b \cdot T_1 + u_1 + e_{11}$$



Soluzione (1)

Dato che \mathbf{u} è la stesso dentro ogni soggetto, le componenti delle misure ripetute che non sono legate agli effetti fissi saranno correlate

$$\begin{aligned} Y_{11} &= a + b \cdot T_1 + u_1 + e_{11} \\ Y_{12} &= a + b \cdot T_2 + u_1 + e_{12} \\ Y_{13} &= a + b \cdot T_3 + u_1 + e_{13} \end{aligned}$$

Stesso soggetto,
stesso errore

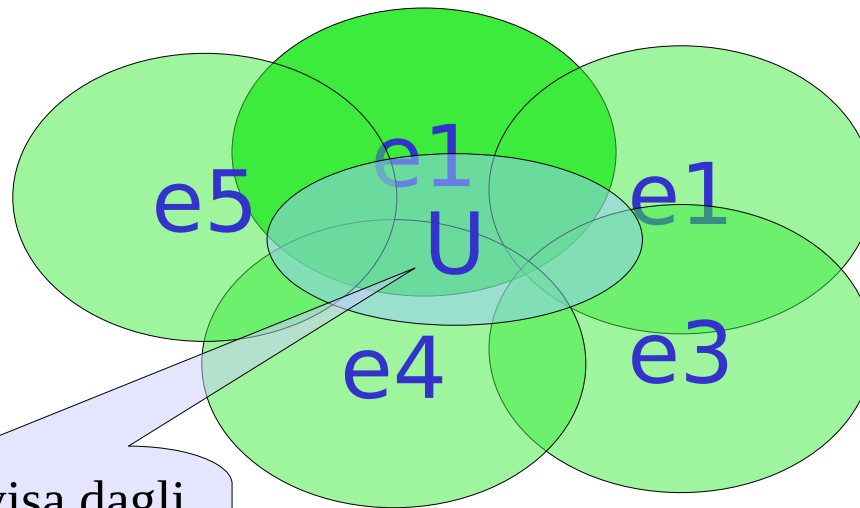
.....

$$\begin{aligned} Y_{i1} &= a + b \cdot T_1 + u_i + e_{i1} \\ Y_{i2} &= a + b \cdot T_2 + u_i + e_{i2} \\ Y_{i3} &= a + b \cdot T_2 + u_i + e_{i3} \end{aligned}$$

Stesso soggetto,
stesso errore

Soluzione (1)

Dato che \mathbf{u} è la stesso dentro ogni soggetto, le componenti delle misure ripetute che non sono legate agli effetti fissi saranno correlate



Varianza condivisa dagli
errori dovuta alla
componente individuale

Modello misto

- Si può specificare il modello in maniera alternativa, più in linea con la teoria dei modelli misti vista fin ora
- Possiamo modellare la componente individuale come un parametro random

$$Y_{ij} = \bar{a} + a_i + b \cdot T_j + e_{ij}$$

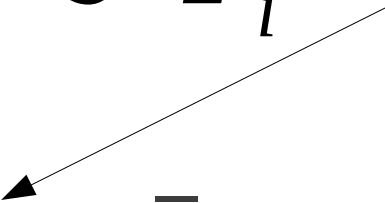
Intercetta media
(Effetto fisso)

Intercetta random,
diversa per ogni
soggetto

Definizione del modello

Riportiamo il modello con la terminologia del MM

$$Y_{ij} = a + b \cdot T_i + u_j + e_{ij}$$


$$y_{ij} = \bar{a} + a_j + \bar{b} \cdot x_{ij} + e_{ij}$$

Definizione del modello

- Modello completo

$$Y_{ij} = \bar{a} + a_i + b \cdot T_j + e_{ij}$$

Intercetta
(Effetto fisso)

Intercetta
random

Effetto fisso di
Trial

Errore
Random
(IID)

i =soggetti, j =trials

Definizione del modello


$$y_{ij} = \bar{a} + a_j + \bar{b} \cdot x_{ij} + e_{ij}$$



- Quali sono gli effetti fissi? Intercetta e effetto di trial
- Quali sono gli effetti random? Intercetta
- Quali sono i cluster su cui variano gli effetti random? Soggetto (id)

Input

Variabile categoriale


Variabile cluster

Mixed Model 

 group
 x


→

Dependent Variable

 error

→

Factors



 trial

→

Covariates

→

Cluster variables

 id 

Estimation **Confidence Intervals**

☒ REML ☒ Confidence intervals Interval %

Coefficienti random

Intercette random

Tutti i possibili
coefficienti

The screenshot shows a software interface for defining random effects. At the top, there is a dropdown menu labeled "Random Effects" with a downward arrow. Below this, the interface is divided into two main panels. The left panel, titled "Components", contains a list box with the text "trial | id" highlighted in blue. A right-pointing arrow button is located between the two panels. The right panel, titled "Random Coefficients", contains the text "Intercept | id". At the bottom of the interface, there is a checkbox labeled "Correlated Effects" which is currently checked.

Effetti fissi

Effetti fissi

The screenshot shows a software interface for defining fixed effects. At the top, there is a tab labeled "Fixed Effects" with a dropdown arrow. Below this, there are two main panels: "Components" on the left and "Model Terms" on the right. The "Components" panel contains the text "trial". The "Model Terms" panel also contains the text "trial". Between these two panels, there are two buttons: a right-pointing arrow (→) and a right-pointing arrow followed by a dropdown arrow (→ ▾). A callout line points from the "trial" text in the "Components" panel to the "trial" text in the "Model Terms" panel. At the bottom of the interface, there is a checkbox labeled "Fixed Intercept" which is checked.

Tutti i possibili effetti

Risultati: modello

Model Info

Info

R-quadro

Estimate	Linear mixed model fit by REML
Call	error ~ 1 + (1 id) + trial
AIC	-463.8270
R-squared Marginal	0.0148
R-squared Conditional	0.2171

**R-squared Marginal: Quanta
varianza è spiegata dai fixed
effects da soli**

**R-squared Conditional: quanta
varianza è spiegata dai fixed e
dai random effects tutti insieme**

Risultati: variance

Varianza dell'intercette

Random Components

Groups	Name	SD	Variance
id	(Intercept)	0.0883	0.00780
	Residual	0.1738	0.03020

Note. Numer of Obs: 1000 , groups: id , 200

Se è maggiore di 0, ok

GAMLj: Results: fixed

F-tests

Fixed Effect ANOVA

	F	Num df	Den df	p
trial	4.72	4	796	< .001

Note. Satterthwaite method for degrees of freedom

Coefficienti

Fixed Effects Parameter Estimates

Effect	Contrast	Estimate	SE	95% Confidence Interval		df	t	p
				Lower	Upper			
(Intercept)	Intercept	0.49474	0.00832	0.4784	0.51104	199	59.4620	< .001
trial1	2 - (1, 2, 3, 4, 5)	-0.01791	0.01099	-0.0395	0.00363	796	-1.6296	0.104
trial2	3 - (1, 2, 3, 4, 5)	-7.92e-4	0.01099	-0.0223	0.02075	796	-0.0720	0.943
trial3	4 - (1, 2, 3, 4, 5)	0.04094	0.01099	0.0194	0.06248	796	3.7246	< .001
trial4	5 - (1, 2, 3, 4, 5)	0.00634	0.01099	-0.0152	0.02788	796	0.5764	0.564

Constrasti

GAMLj: plot

Plot degli effetti

Fixed Effects Plots

Horizontal axis → trial

Separate lines →

Separate plots →

Display

☒ None

☐ Confidence intervals

Interval %

☐ Standard Error

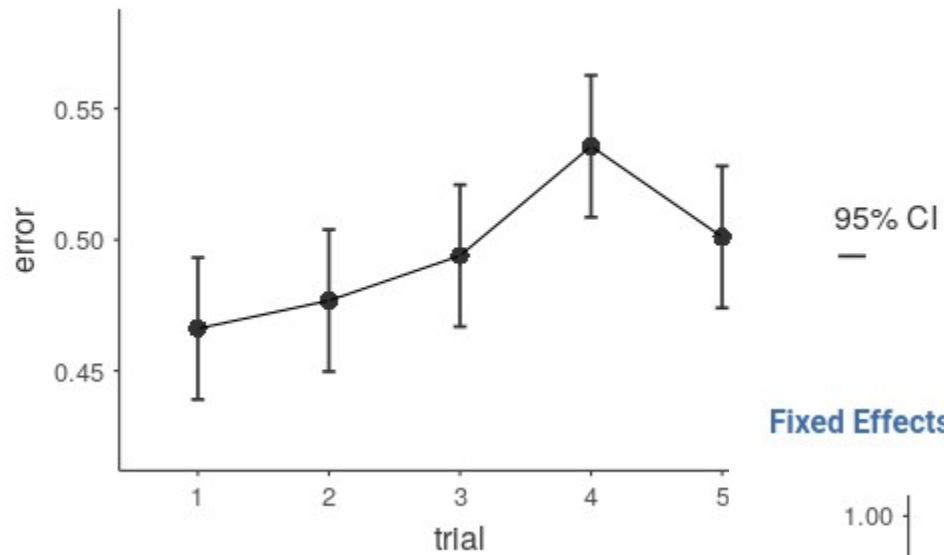
Plot

☐ Observed scores

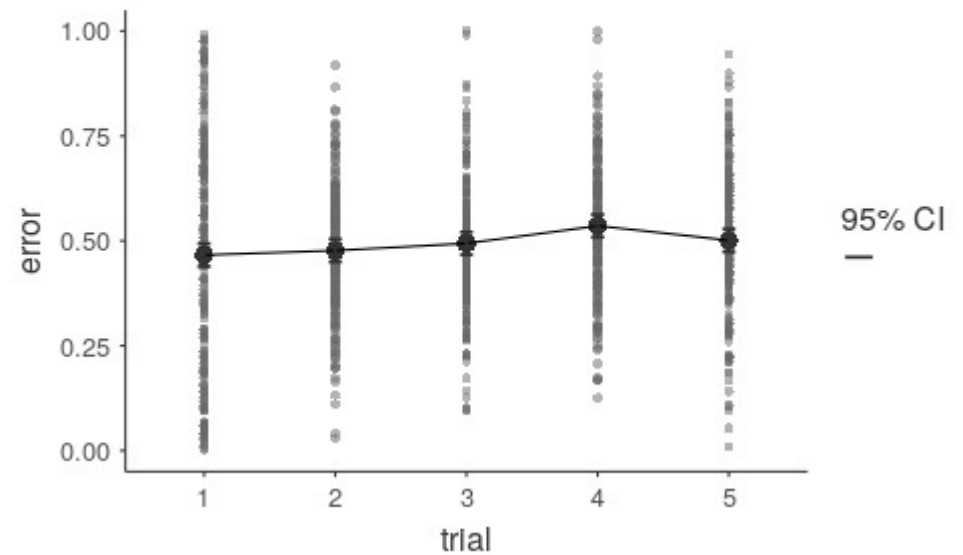
☐ Y-axis observed range

Risultati: plot

Fixed Effects Plots



Fixed Effects Plots



MM nelle misure ripetute

- Il modello misto permette di analizzare le misure ripetute con una vasta gamma di opzioni
- Applicazione delle varie tecniche come in regressione/Anova
- Gestione efficiente dei valori mancanti
- Possibilità di modellare variabili continue come variabili ripetute nel tempo
- Possibilità di combinare il disegno a misure ripetute con disegni gerarchici o clusterizzati

Moderazione





Interazione → Moderazione

- Come per il modello lineare generale la moderazione si stima mediante l'interazione
- L'interazione nel modello misto funziona **esattamente** come nel modello lineare generale

Variabili continue

- Abbiamo Una serie di scuole. In ogni scuola abbiamo misurato il QI degli studenti (*intel*). Di ogni scuola abbiamo anche la dotazione economica annuale (*funds*).
- La variabile dipendente è uno score di performance
- Vogliamo sapere se il QI è associato alla performance, e se tale relazione è moderata dai fondi a disposizione della scuola

Dati

Empire		variables			rows
	 school	 funds	 perform	 intel	
5	1	372.327	10.029	102.145	
6	1	372.327	9.196	90.859	
7	1	372.327	10.187	103.771	
8	1	372.327	9.124	91.466	
9	1	372.327	10.752	108.944	
0	1	372.327	9.493	96.176	
1	1	372.327	9.954	101.083	
2	1	372.327	9.752	99.053	
3	1	372.327	9.459	95.260	
4	1	372.327	10.367	106.097	
5	1	372.327	9.788	98.727	
6	2	368.685	12.081	109.591	
7	2	368.685	6.328	80.896	
8	2	368.685	10.551	101.516	
9	2	368.685	8.189	90.094	
0	2	368.685	9.145	94.737	
1	2	368.685	7.815	87.620	
2	2	368.685	10.799	103.318	

Input

Mixed Model

case

→

Dependent Variable

perform

→

Factors

→

Covariates

intel
funds

→

Cluster variables

school

Estimation

Confidence Intervals

☒ REML

☒ Confidence intervals

Interval

95

%

Effetti fissi

▼ | Fixed Effects

Components

intel

funds

→

→ ▼

Model Terms

intel

funds

intel * funds

☒ Fixed Intercept

Effetti random

▼ | Random Effects

Components

funds | school

intel : funds | school

→

Random Coefficients

Intercept | school

intel | school

Effects correlation

Tests

Risultati

Spieghiamo tanto (dati didattici!)

Model Info

Info	
Estimate	Linear mixed model fit by REML
Call	perform ~ 1 + intel + funds + intel:funds+(1 + intel school)
AIC	-1252.950
BIC	-1177.120
LogLikel.	613.418
R-squared Marginal	0.599
R-squared Conditional	0.999
Converged	yes
Optimizer	bobyqa

Risultati: variance

**Le varianze non sono
0: ok**

Random Components

Groups	Name	SD	Variance	ICC
school	(Intercept)	0.1026	0.01053	0.792
	intel	0.0885	0.00783	
Residual		0.0526	0.00277	

Note. Number of Obs: 500 , groups: school 20

Random Parameters correlations

Groups	Param.1	Param.2	Corr.
school	(Intercept)	intel	0.487

Risultati: effetti fissi

Test F

Fixed Effect Omnibus tests

	F	Num df	Den df	p
intel	24.1264	1	18.0	< .001
funds	0.0471	1	18.0	0.831
intel * funds	6.4748	1	18.0	0.020

Note. Satterthwaite method for degrees of freedom

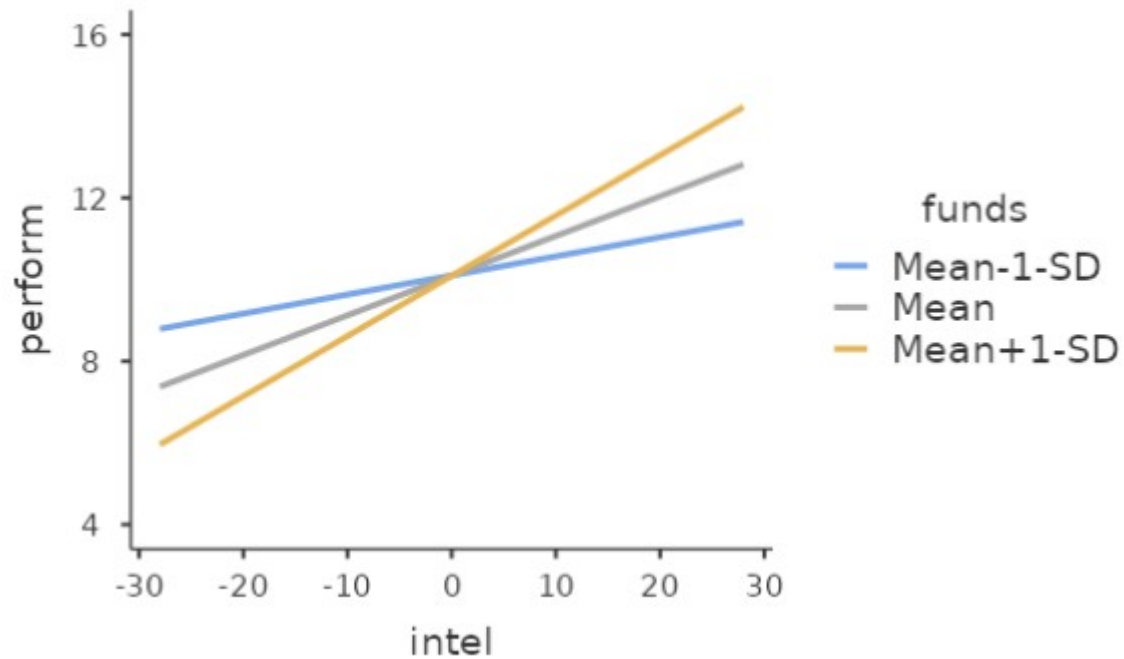
Coefficienti

Fixed Effects Parameter Estimates

Names	Estimate	SE	95% Confidence Interval		df	t	p
			Lower	Upper			
(Intercept)	10.0981	0.0231	10.0529	10.14331	18.0	437.597	< .001
intel	0.0972	0.0198	0.0584	0.13602	18.0	4.912	< .001
funds	-5.63e-5	2.60e-4	-5.65e-4	4.52e-4	18.0	-0.217	0.831
intel * funds	5.67e-4	2.23e-4	1.30e-4	0.00100	18.0	2.545	0.020

Risultati: plot

Effects Plots



Plot effetti fissi

Disegno a misure ripetute

- Abbiamo 2 gruppi - Control vs Treatment, misurati in 4 tempi diversi. I tempi sono: 1 (pretest), 2 (one month posttest), 3 (3 months follow-up), and 4 (6 months follow-up).
- La variabile dipendente è uno score di depressione (e.g. Beck Depression Inventory) e il trattamento è l'utilizzo di un farmaco versus nessun farmaco. Ci aspettiamo un miglioramento in tutte e due gruppi, ma vogliamo testare che il gruppo in treatment migliori più rapidamente

Disegno a misure ripetute

- Abbiamo 2 gruppi - Control vs Treatment, misurati in 4 tempi diversi. I tempi sono: 1 (pretest), 2 (one month posttest), 3 (3 months follow-up), and 4 (6 months follow-up).

96 osservazioni
24 soggetti

Contingency Tables

Contingency Tables

time	group		Total
	1	2	
0	12	12	24
1	12	12	24
3	12	12	24
6	12	12	24
Total	48	48	96

Disegno a misure ripetute: dati

- Data sono in “long format”

Ogni soggetto ha 4 righe

	subj	time	group	dv
1	1	0	1	296
2	1	1	1	175
3	1	3	1	187
4	1	6	1	192
5	2	0	1	376
6	2	1	1	329
7	2	3	1	236
8	2	6	1	76
9	3	0	1	309
10	3	1	1	238
11	3	3	1	150
12	3	6	1	123
13	4	0	1	222
14	4	1	1	60
15	4	3	1	82
16	4	6	1	85
17	5	0	1	150
18	5	1	1	271
19	5	3	1	250

Mixed model

Traduciamo il disegno in un modello misto

- Effetti Fissi? Intercetta group,time, la loro interazione
- Effetti Random? Intercetta
- Clusters? Soggetto (subj)

Variables

- Definisco le variabili

Mixed Model

Dependent Variable

→ dv

Factors

→ time
group

Covariates

→

Cluster variables

→ subj

Estimation

☒ REML

Confidence Intervals

☒ Confidence intervals Interval 95 %

Variable Cluster

Modello

Fixed Effects

Components		Model Terms
time	→	time
group	→ ▾	group
		time * group

☒ Fixed Intercept

Fixed effects

Random Effects

Components		Random Coefficients
time subj	→	Intercept subj
group subj		
time : group subj		

☒ Correlated Effects

Random effects

Risultati

- Interpretazione dei risultati

Mixed Model

Modello

Model Info

Info

Estimate	Linear mixed model fit by REML
Call	<code>dv ~ 1 + (1 subj) + time + group + time:group</code>
AIC	1011.895
R-squared Marginal	0.554
R-squared Conditional	0.768

Effetti random

Random Components

Groups	Name	SD	Variance
subj	(Intercept)	50.4	2539
	Residual	52.5	2761

Note. Numer of Obs: 96 , groups: subj , 24

Risultati

- Interpretazione dei risultati

F-tests per gli effetti fissi

Fixed Effect ANOVA

	F	Num df	Den df	p
time	45.14	3	66.0	< .001
group	13.71	1	22.0	0.001
time:group	9.01	3	66.0	< .001

Note. Satterthwaite method for degrees of freedom

- Per il momento ignoriamo la stima dei coefficienti

Results: plot

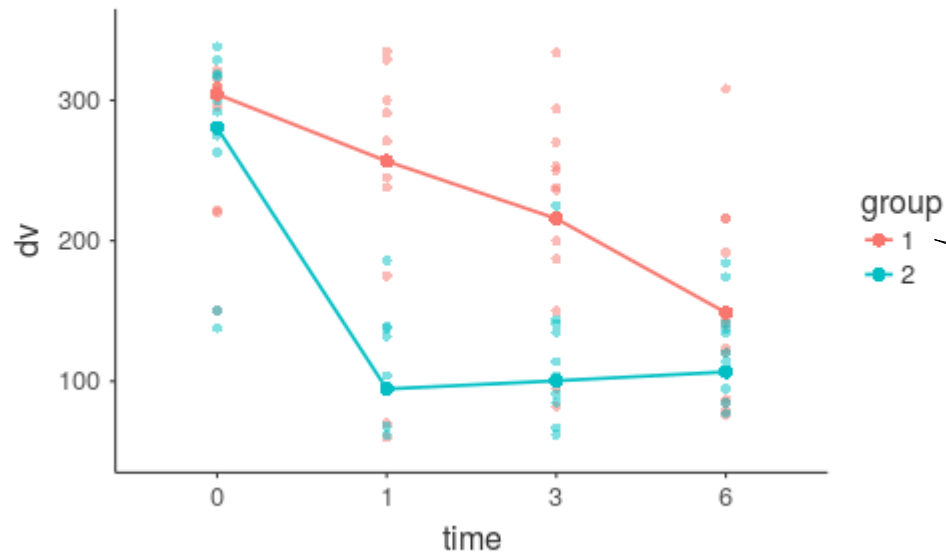
Fixed Effects Plots

Horizontal axis
→ time

Separate lines
→ group

Separate plots
→

Fixed Effects Plots



Rosso è il gruppo di controllo

Analisi sulla interazione

- Per analizzare ulteriormente l'interazione possiamo fare (come nel GLM):
- Simple effects: Testare se l'effetto di tempo è presente in ognuno dei gruppi
- Trend analysis: Testare dei trend specifici nelle nostre medie
- Post-hoc test: confronto delle medie tutto contro tutto


Simple effects

- Chiediamo di stimare gli effetti ti tempo in ogni gruppo

▼ | Simple Effects


→

Simple effects variable

 time

→

Moderator

 group

→

Breaking variable

Simple effects

- Chiediamo di stimare gli effetti ti tempo in ogni gruppo

Simple Effects ANOVA

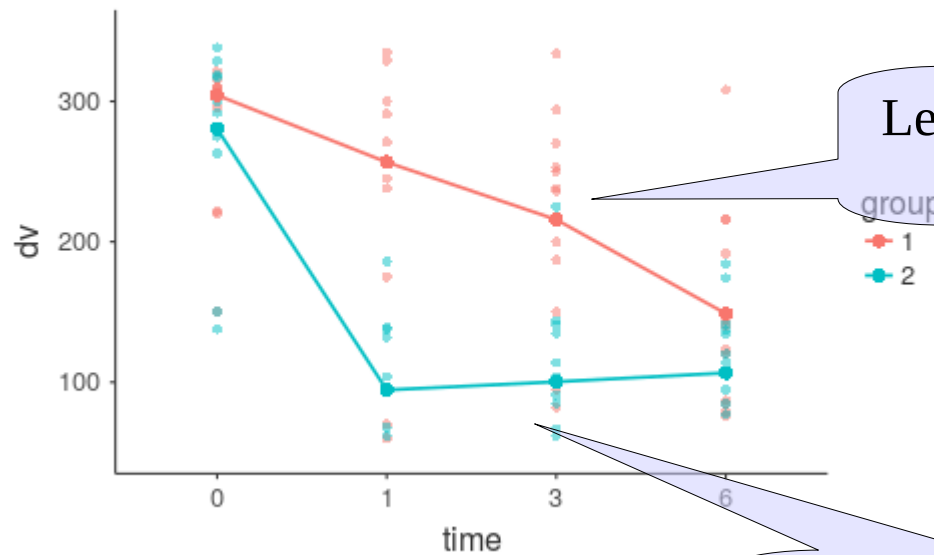
Simple effects of time

Effect	Moderator Levels	df Num	df Den	F	p
time	group at 1	3.00	66.0	18.9	< .001
time	group at 2	3.00	66.0	35.3	< .001

In entrambi i gruppi le medie cambiano nel tempo

Results: plot

Fixed Effects Plots



Le medie rosse mostrano
un effetto di tempo

Le medie verdi mostrano
un effetto di tempo

Ricapitolando

- Dunque, i disegni a misure ripetute possono essere analizzati come qualunque altro disegno “Anova”, ma deve essere modellata la componente individuale che cambia da soggetto a soggetto
- Ciò consente di applicare tutte le conoscenze dell'ANOVA/Regressione al caso delle misure ripetute
- I modelli misti consentono dunque di stimare modelli a misure ripetute combinandoli con altre strutture complesse dei dati