AIP

UNIVERSITÀ DEGLI STUDI DI MILANO BICOCCA

Quarta giornata

Il modello lineare misto

Marcello Gallucci
Università Milano-Bicocca

GLM

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

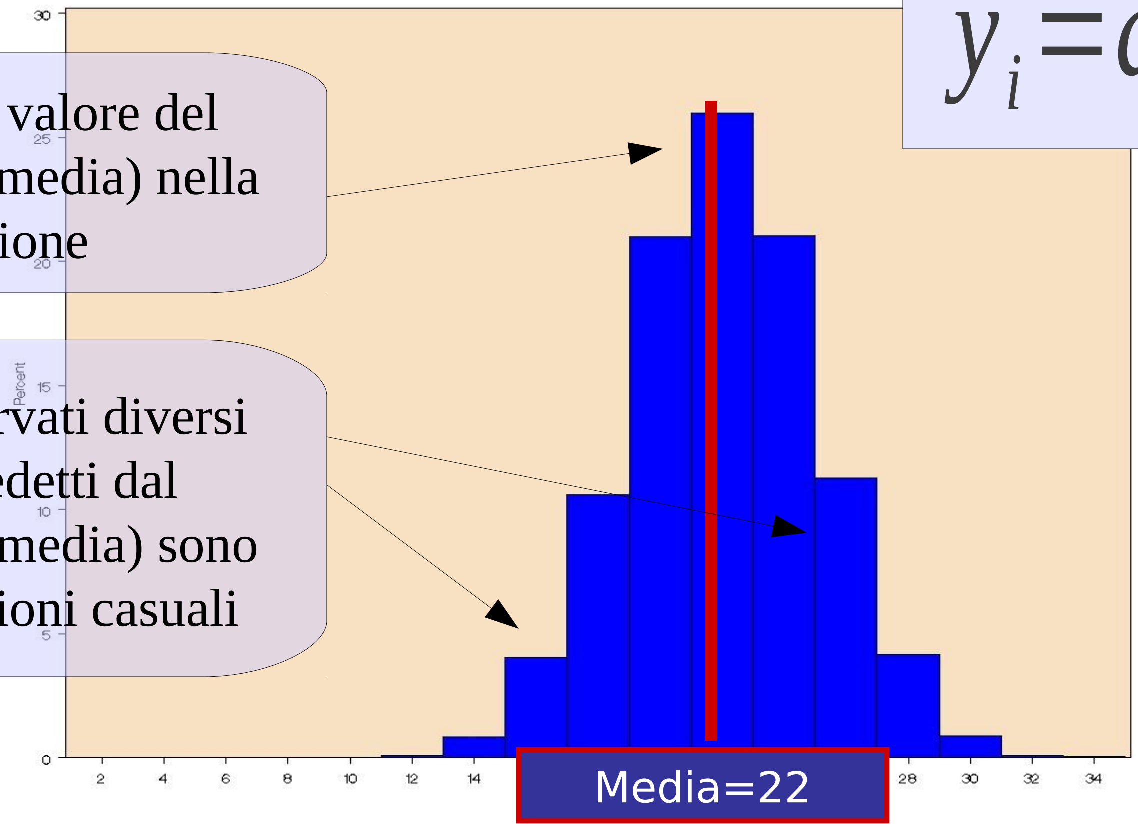
Assunzioni GLM

Modello Lineare Generale

$$y_i = a + e_i$$

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

I punteggi osservati diversi da quelli predetti dal parametro (e.g. media) sono dovuti a variazioni casuali



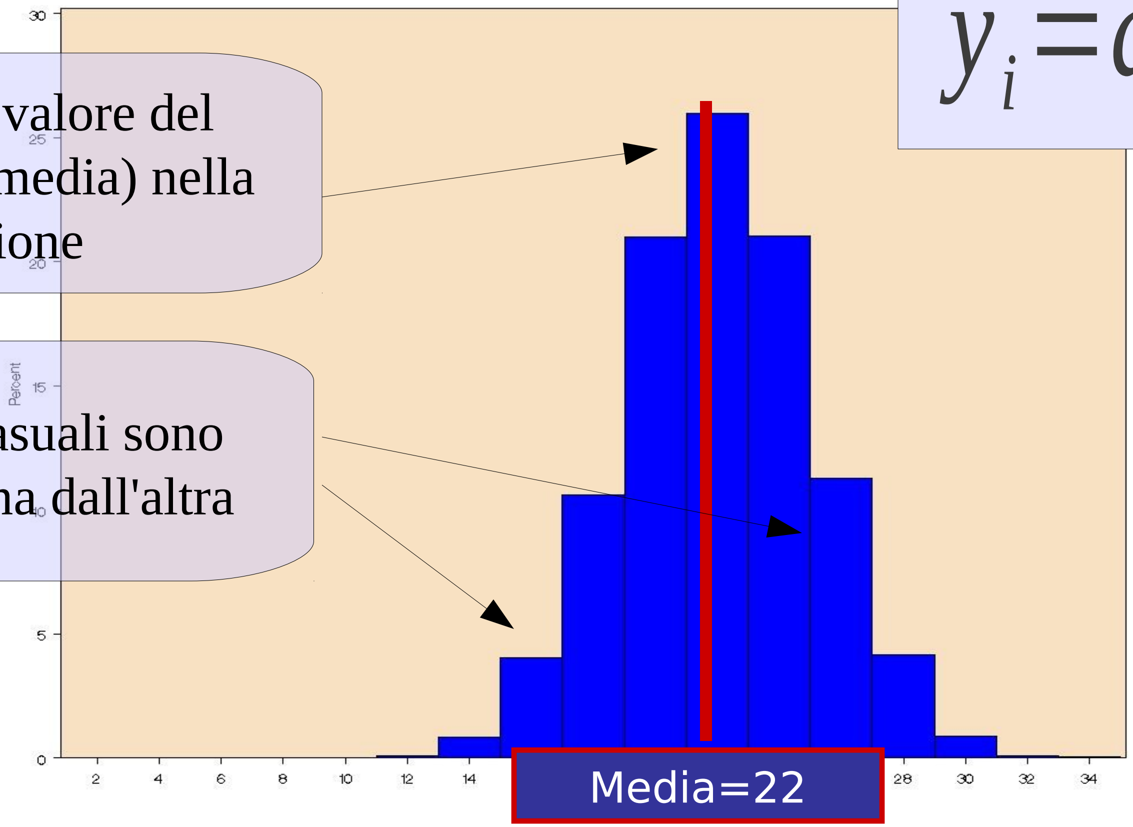
Assunzioni GLM

Modello Lineare Generale

$$y_i = a + e_i$$

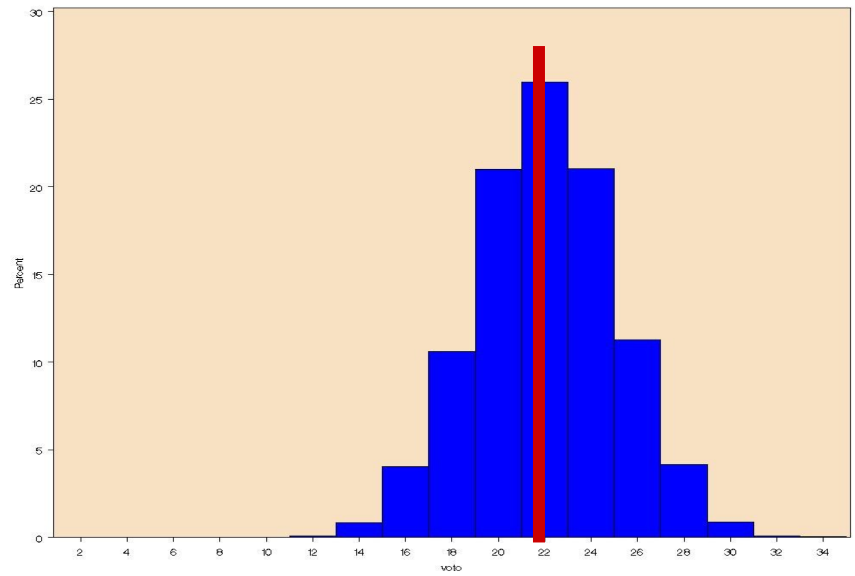
Esiste un solo valore del parametro (e.g. media) nella popolazione

Le variazioni casuali sono indipendenti l'una dall'altra



Assunzioni GLM

Modello Lineare Generale



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

Le variazioni casuali sono indipendenti l'una dall'altra

$$y_i = a + e_i$$

$$corr(e_i, e_j) = 0$$

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 ripeture
- 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

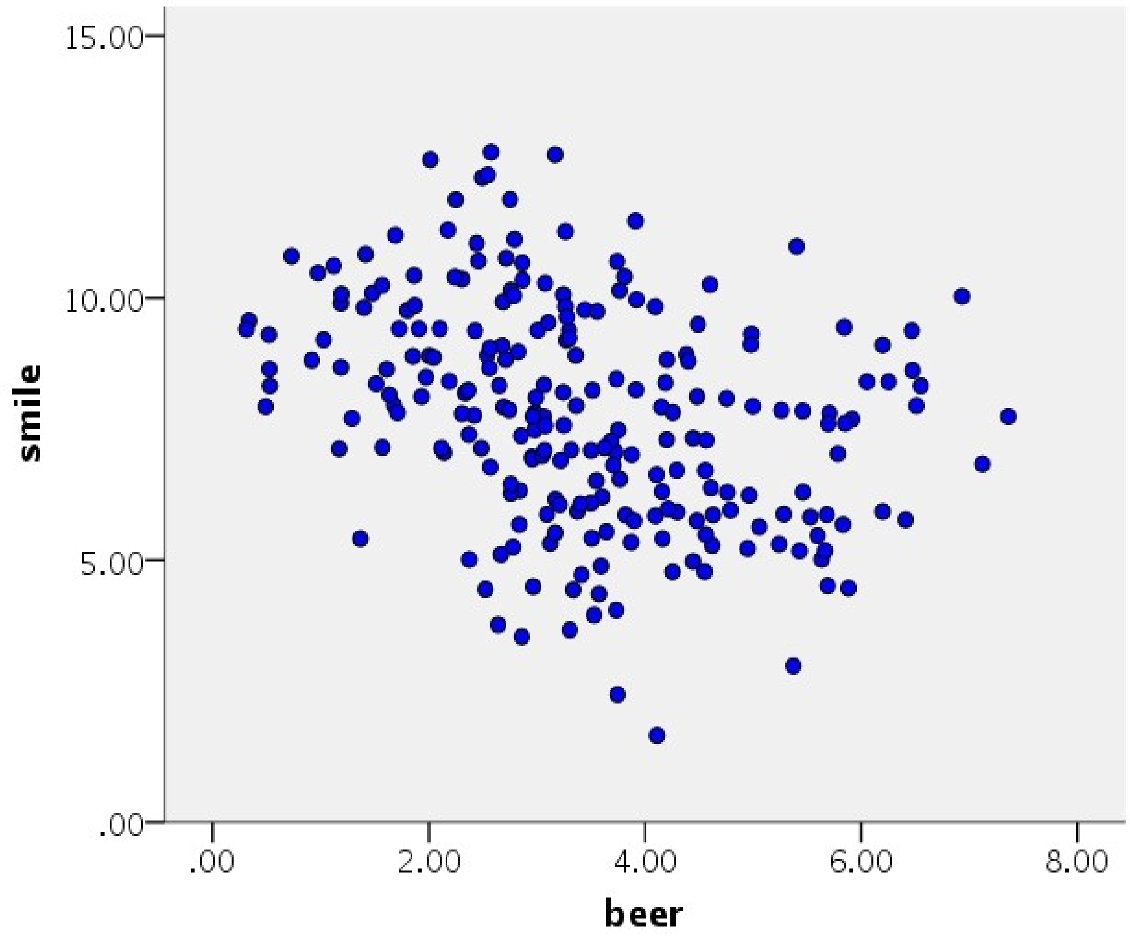
bar

		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

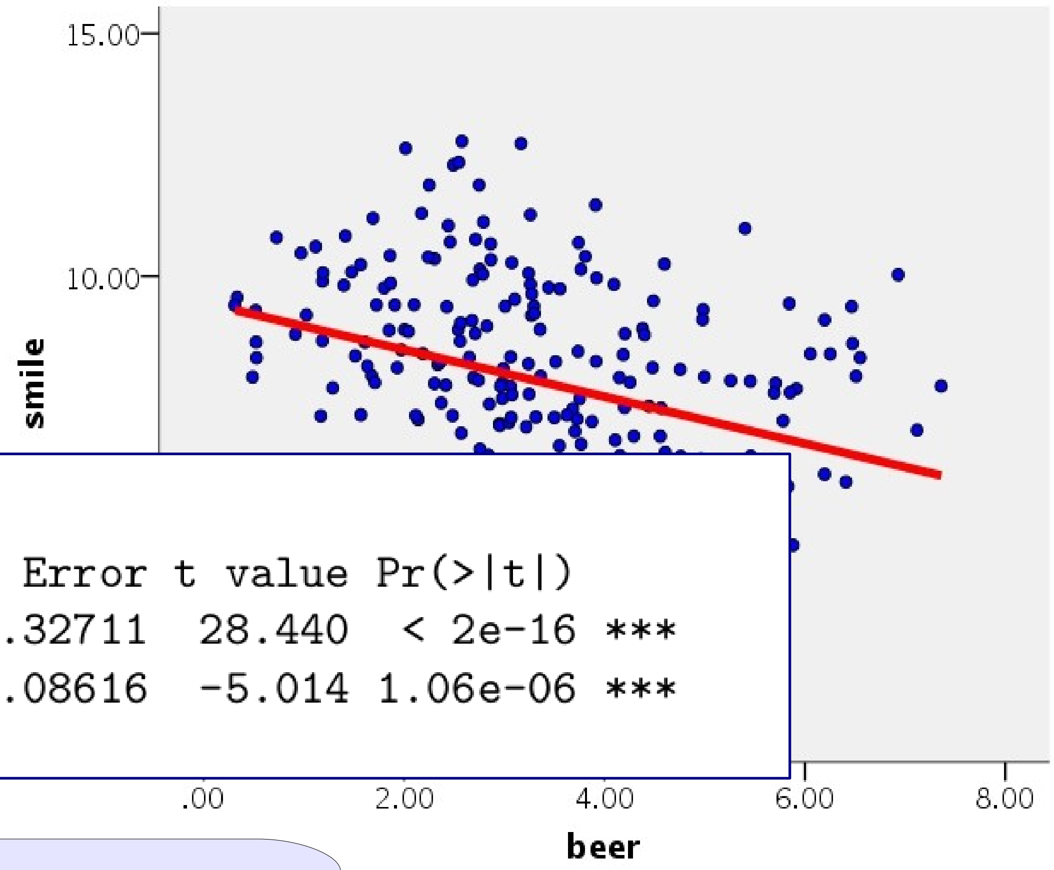
Esempio “birre” 2

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

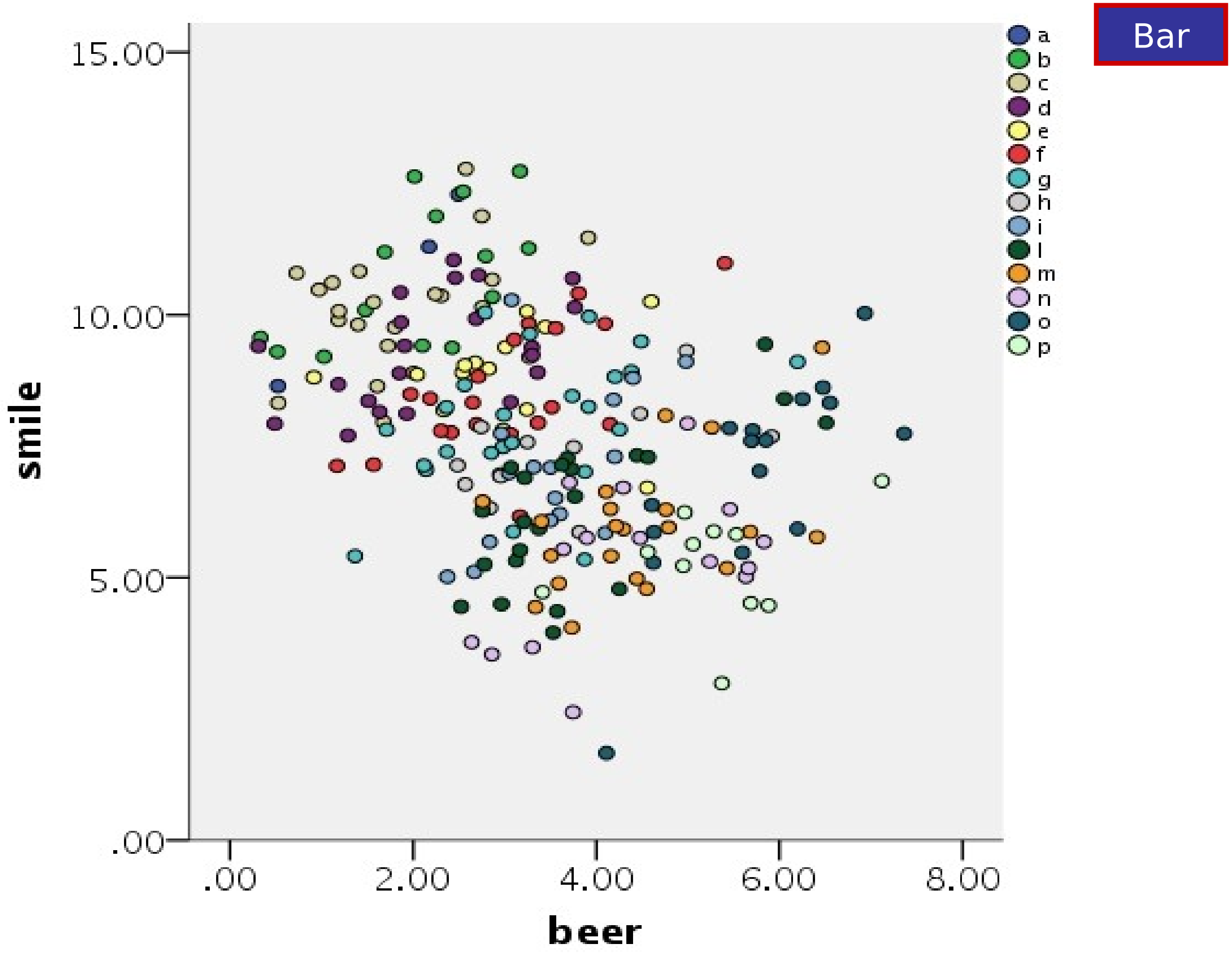
Possibili spiegazioni

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 caretteristiche 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

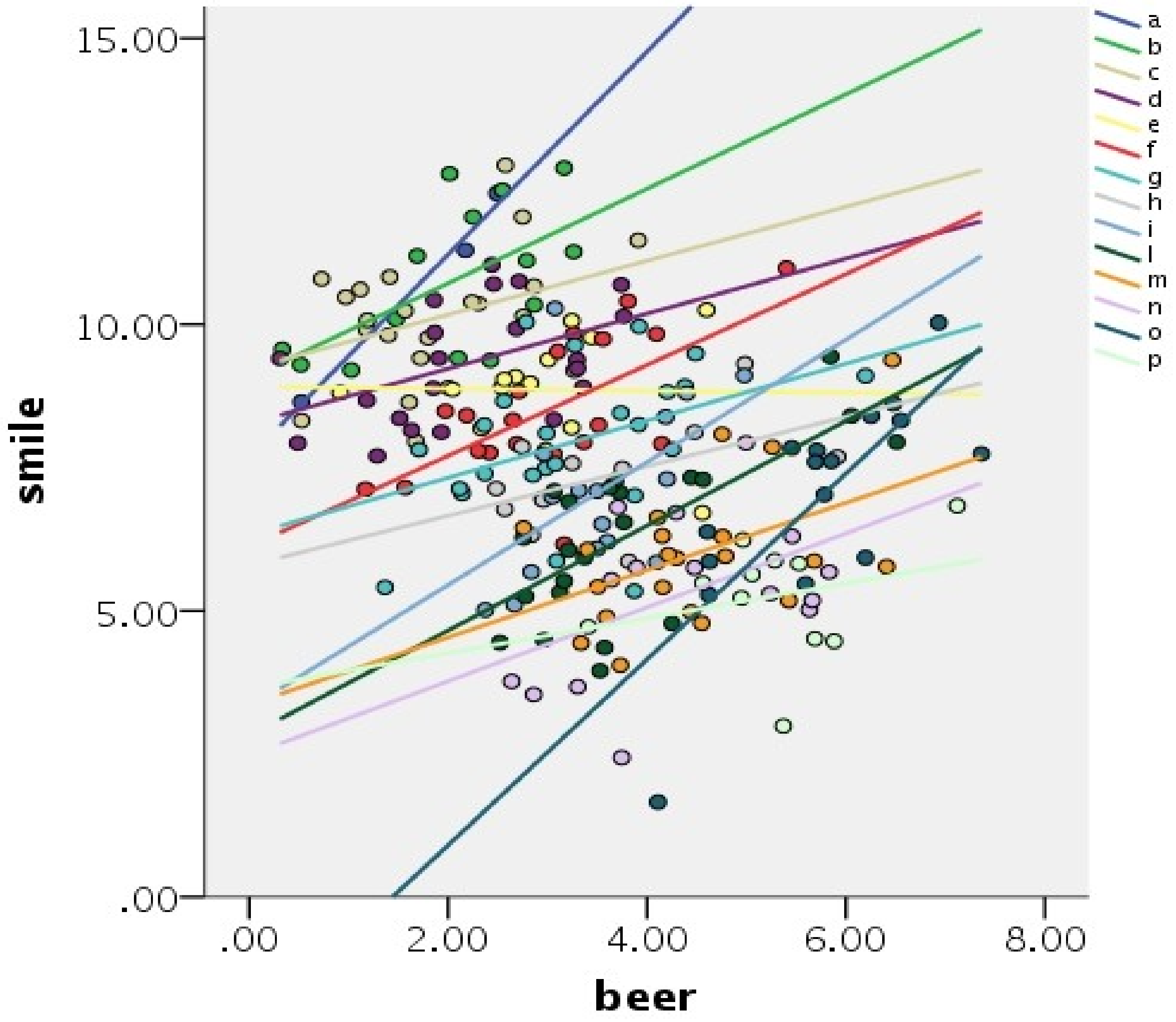
Scatterplot per Bar

Lo scatterplot, distinguendo per bar, offre indicazioni



Scatterplot per Bar

Lo scatterplot, distinguendo per bar, offre indicazioni

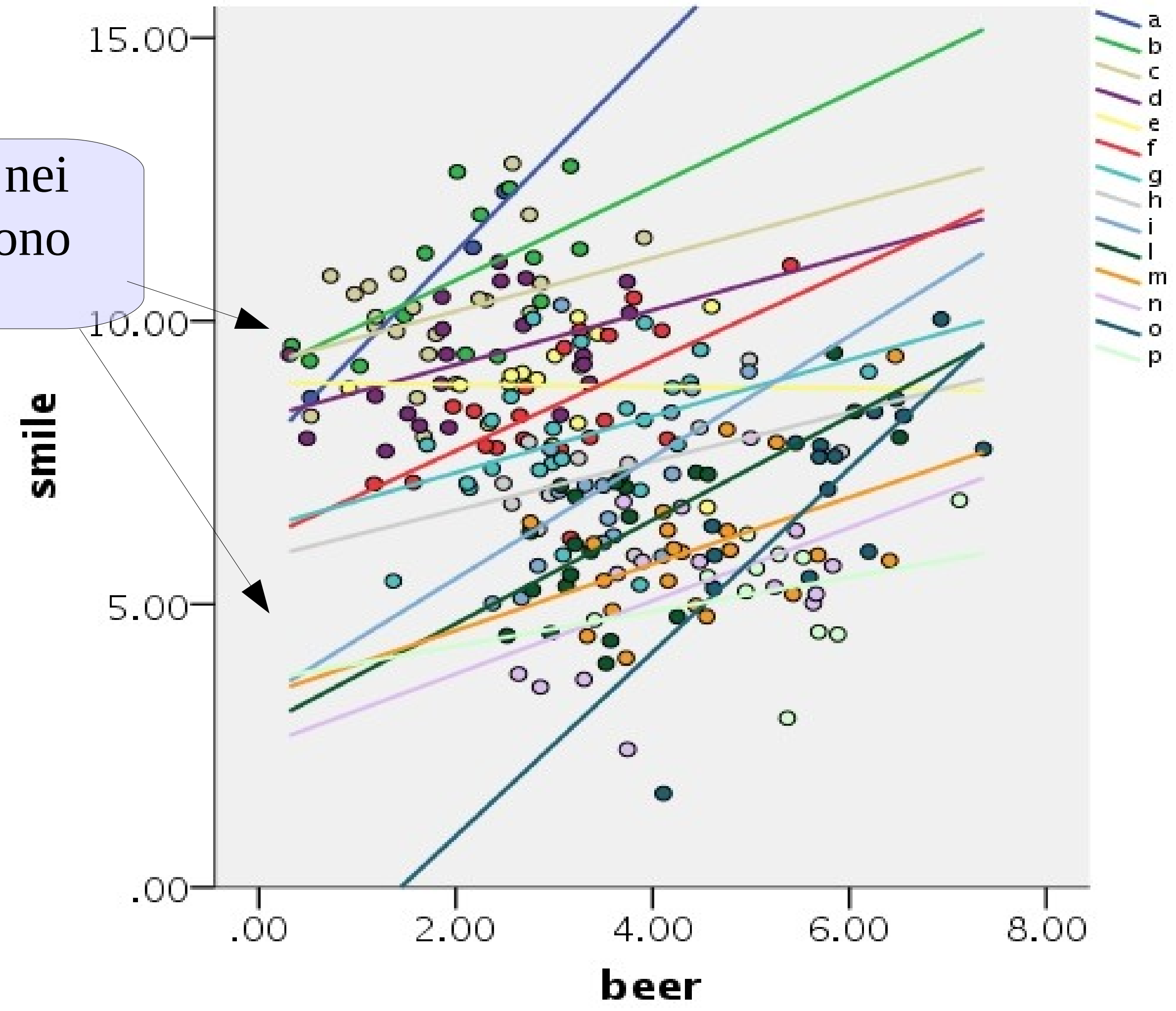


Bar

Scatterplot per Bar

Lo scatterplot, distinguendo per bar, offre indicazioni

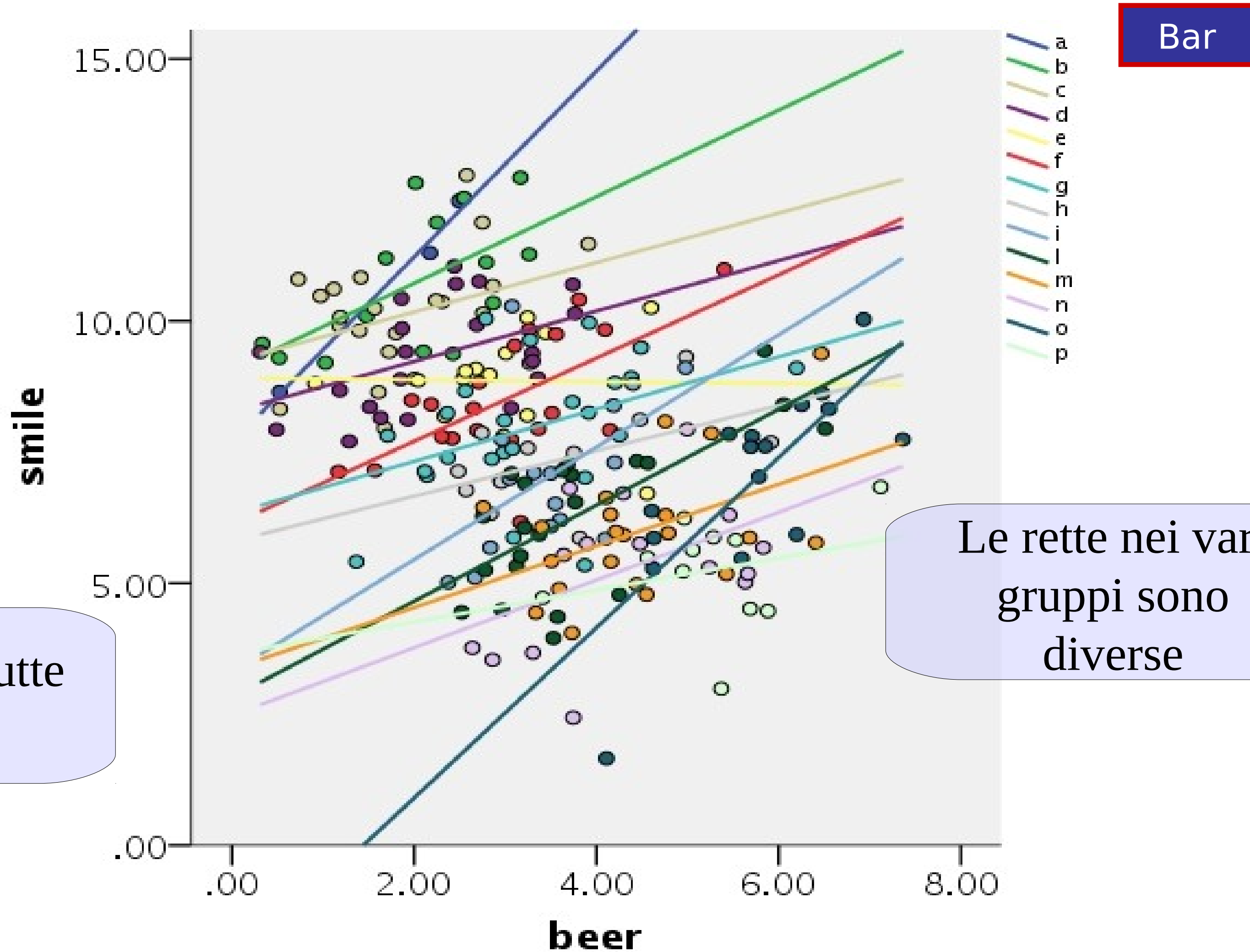
Bar



Le intercette nei vari gruppi sono diverse

Scatterplot per Bar

Lo scatterplot, distinguendo per bar, offre indicazioni



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

$$\hat{y}_{ia} = a_a + b_a \cdot x_{ia}$$
$$\hat{y}_{ib} = a_b + b_b \cdot x_{ib}$$
$$\hat{y}_{ic} = a_c + b_c \cdot x_{ic}$$

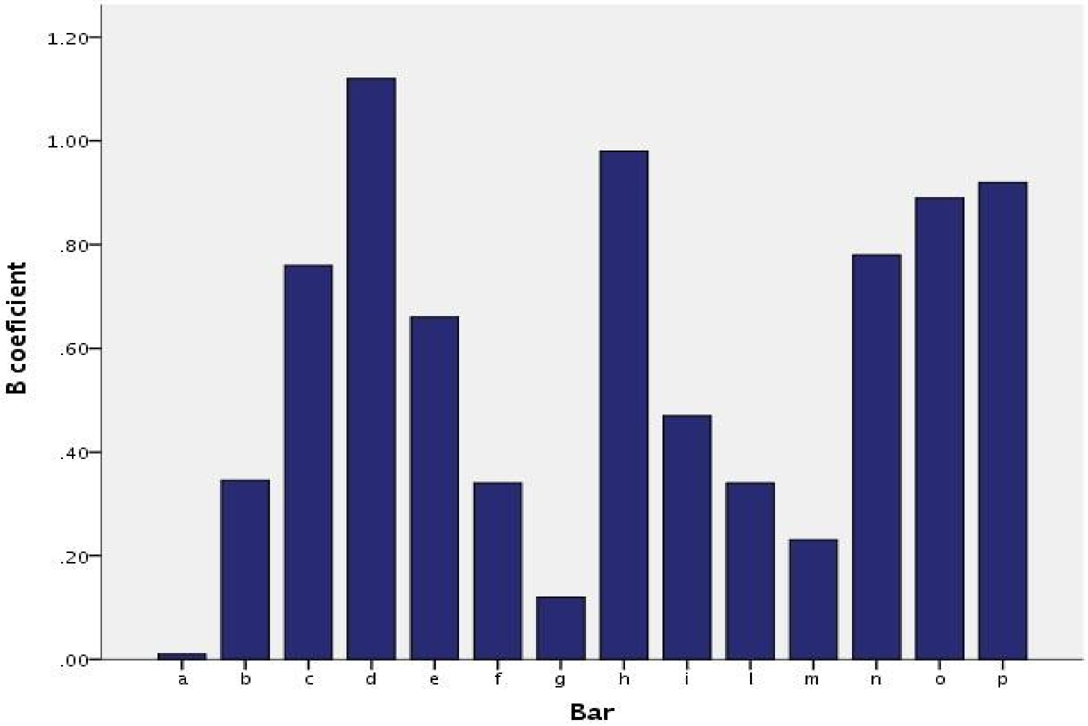
$$\hat{y}_{ij} = a_j + b_j \cdot x_{ij}$$

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

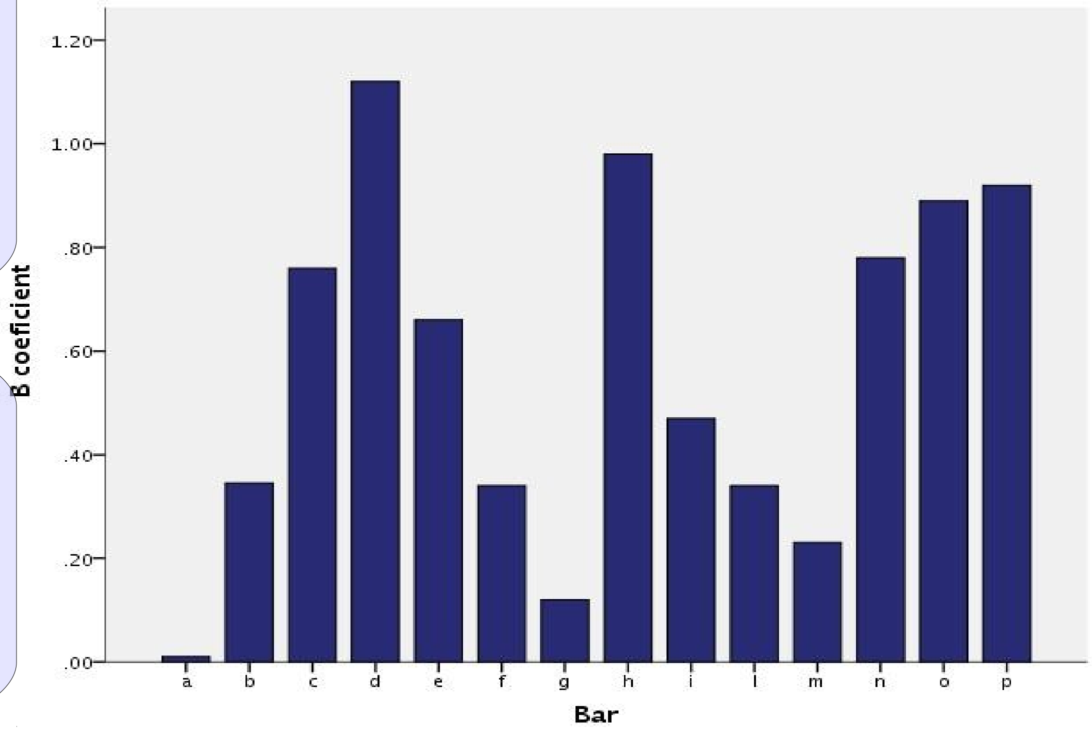


Coefficienti random

- 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

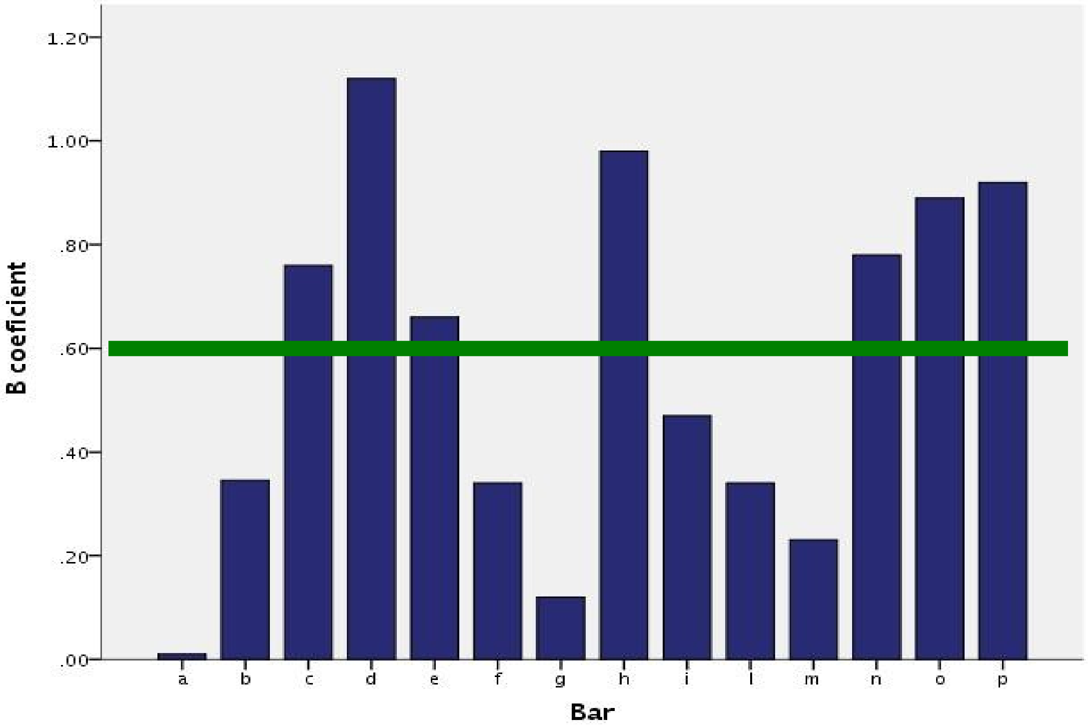


Media dei Coefficiente

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

MEDIA

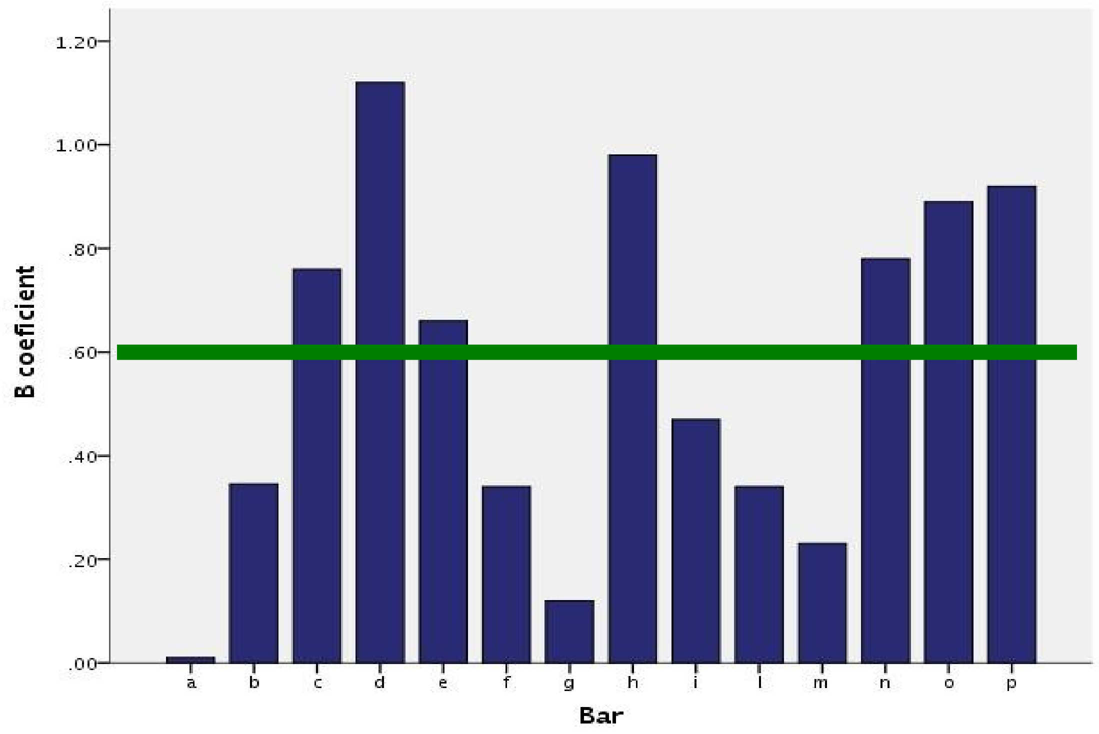
$$\bar{b} = \frac{\sum_j b_j}{k}$$



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

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

Modello Misto

- 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

Modello Misto

- 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**

Modelli Misti

- 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**

Modelli Misti

- 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

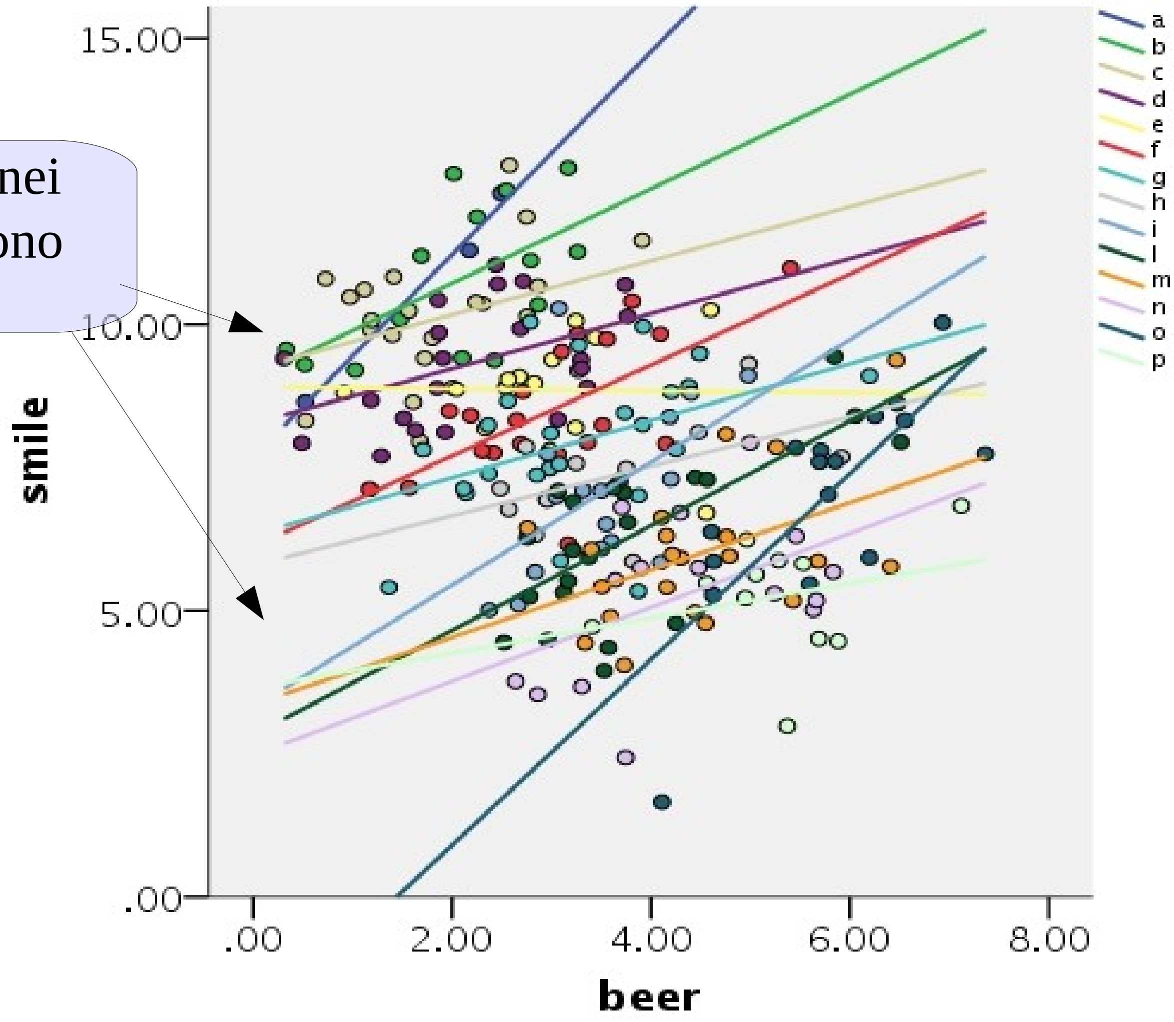
- La definizione corretta del modello, consente di ottenere stime e errori standard (e dunque test inferenziali) corretti

Birre al Bar

Definiamo il modello, iniziando dal più semplice

Bar

Le intercette nei vari gruppi sono diverse



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

R syntax

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

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

```
mod<-lmer(smile~(1|bar)+beer,data=bars)
summary(mod)
```

Effetti fissi (intercetta è inclusa di default)

Effetti random
1=intercetta

Variabile cluster

Output

```
## Linear mixed model fit by REML t-tests use Satterthwaite approximations
##   to degrees of freedom [lmerMod]
## Formula: smile ~ (1 | bar) + beer
```

```
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  bar      (Intercept) 6.532    2.556
##  Residual             1.285    1.133
## Number of obs: 234, groups:  bar, 15
##
```

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

Output

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

```
## Fixed effects:
##               Estimate Std. Error        df t value Pr(>|t|)
## (Intercept)    5.55107    0.72419  18.29000   7.665 4.03e-07 ***
## beer           0.63870    0.07769 227.90000   8.221 1.53e-14 ***
## ---
```

Intercetta: In media, per zero birre ci attendiamo 5.5 sorrisi

Output

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

```
## Fixed effects:
##             Estimate Std. Error        df t value Pr(>|t|)
## (Intercept)  5.55107    0.72419  18.29000   7.665 4.03e-07 ***
## beer         0.63870    0.07769 227.90000   8.221 1.53e-14 ***
## ---
```

Coefficiente **b**: In media, per ogni birra in più i sorrisi aumentano di .638

R Output

E' possibile testare la significatività delle varianze degli effetti random

```
dat<-read.spss('data/regression_beers_bars.sav',to.data.frame = T)
head(dat)
mm1<-lmer(smile~1+beer+(1|bar),data=dat)

rand(mm1)
```

```
> rand(mm1)
Analysis of Random effects Table:
    Chi.sq Chi.DF p.value
bar    201      1  <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

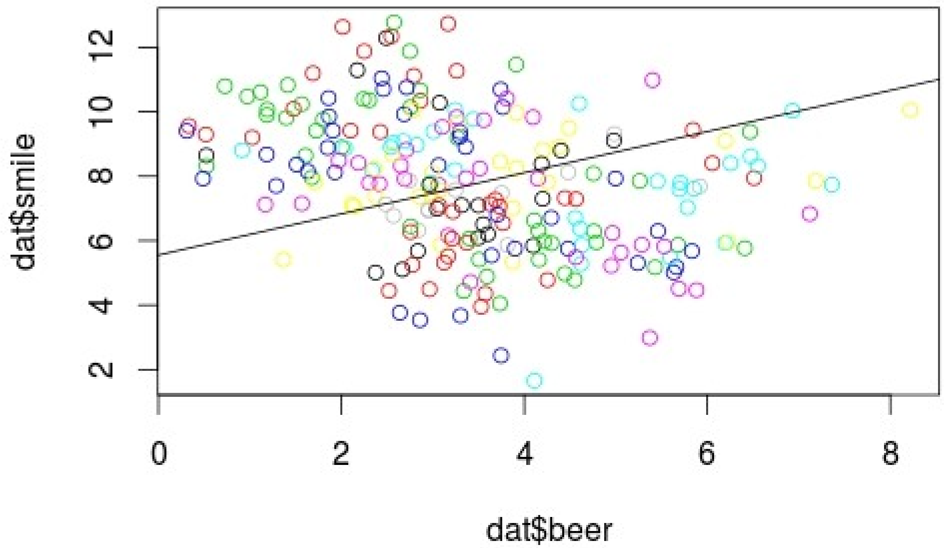
OK!

R Plot

Possiamo plottare gli effetti fissi

```
summary(mm1)
plot(dat$smile~dat$beer,col=dat$bar)
abline(fixef(mm1))
```

fixef tira fuori dal modello gli effetti fissi

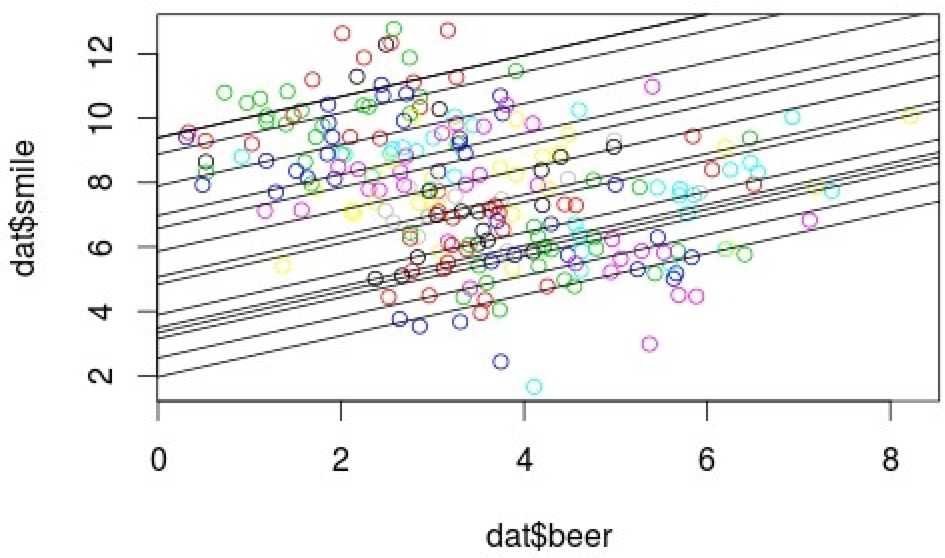


R Plot

Possiamo anche plottare gli effetti random

```
plot(dat$smile~dat$beer,col=dat$bar)
apply(coef(mm1)[[1]],1,abline)
```

coef tira fuori dal modello gli effetti random



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

R syntax

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

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

```
mod<-lmer(smile~(1+beer|bar)+beer,data=bars)
summary(mod)
```

Effetti fissi (intercetta è inclusa di default)

Effetti random

Variabile cluster

Output

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

```
## Random effects:
##  Groups   Name        Variance Std.Dev. Corr
##  bar      (Intercept) 9.33420  3.0552
##           beer        0.03452  0.1858   -0.79
##  Residual             1.25881  1.1220
## Number of obs: 234, groups:  bar, 15
##
```

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

Output

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

```
## Fixed effects:
##             Estimate Std. Error      df t value Pr(>|t|)
## (Intercept)   5.3733     0.8509 11.5970   6.315 4.51e-05 ***
## beer          0.6417     0.0924  9.3360   6.945 5.61e-05 ***
## ---
```

Effetti non diversi da prima

Varianze

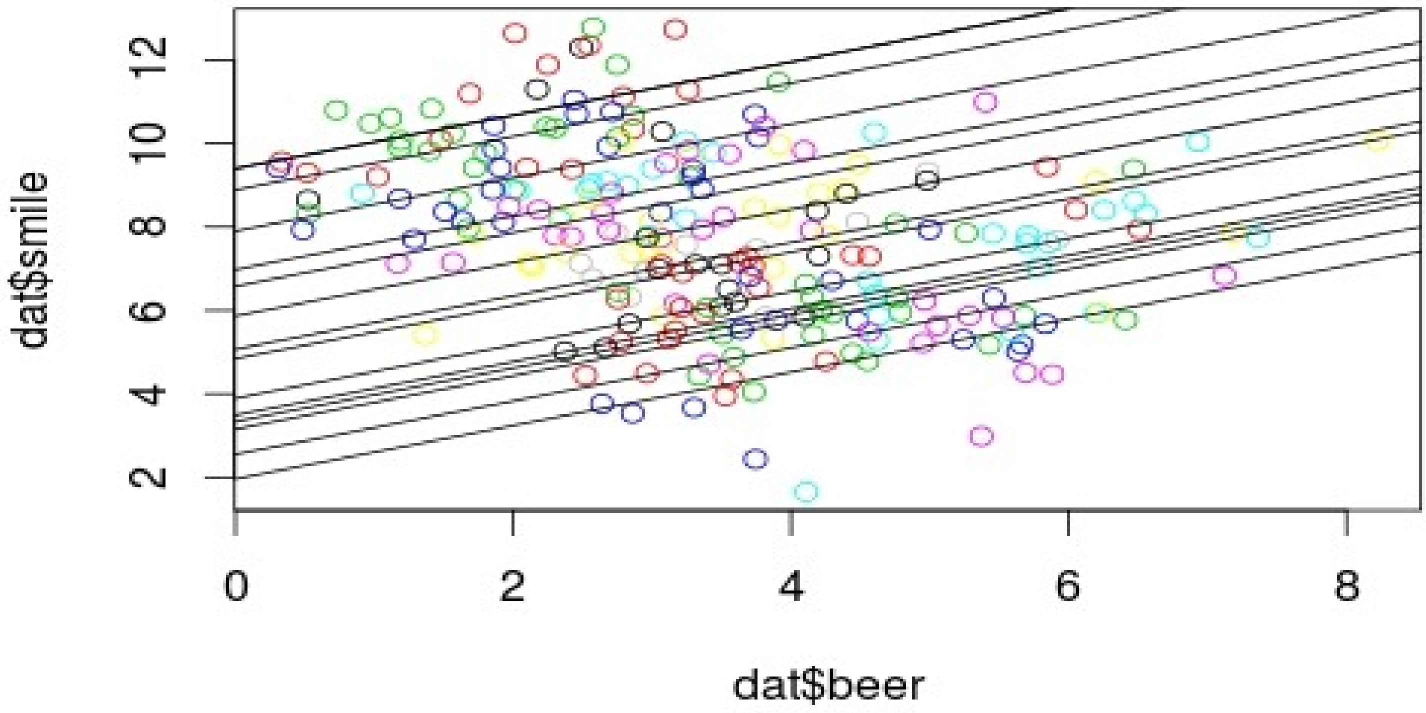
La varianza degli effetti random ci indica quanta variabilità c'è tra i cluster nell'effetto

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

```
## Random effects:
##  Groups   Name        Variance Std.Dev. Corr
##  bar      (Intercept) 9.33420  3.0552
##           beer        0.03452  0.1858   -0.79
##  Residual             1.25881  1.1220
## Number of obs: 234, groups:  bar, 15
##
```

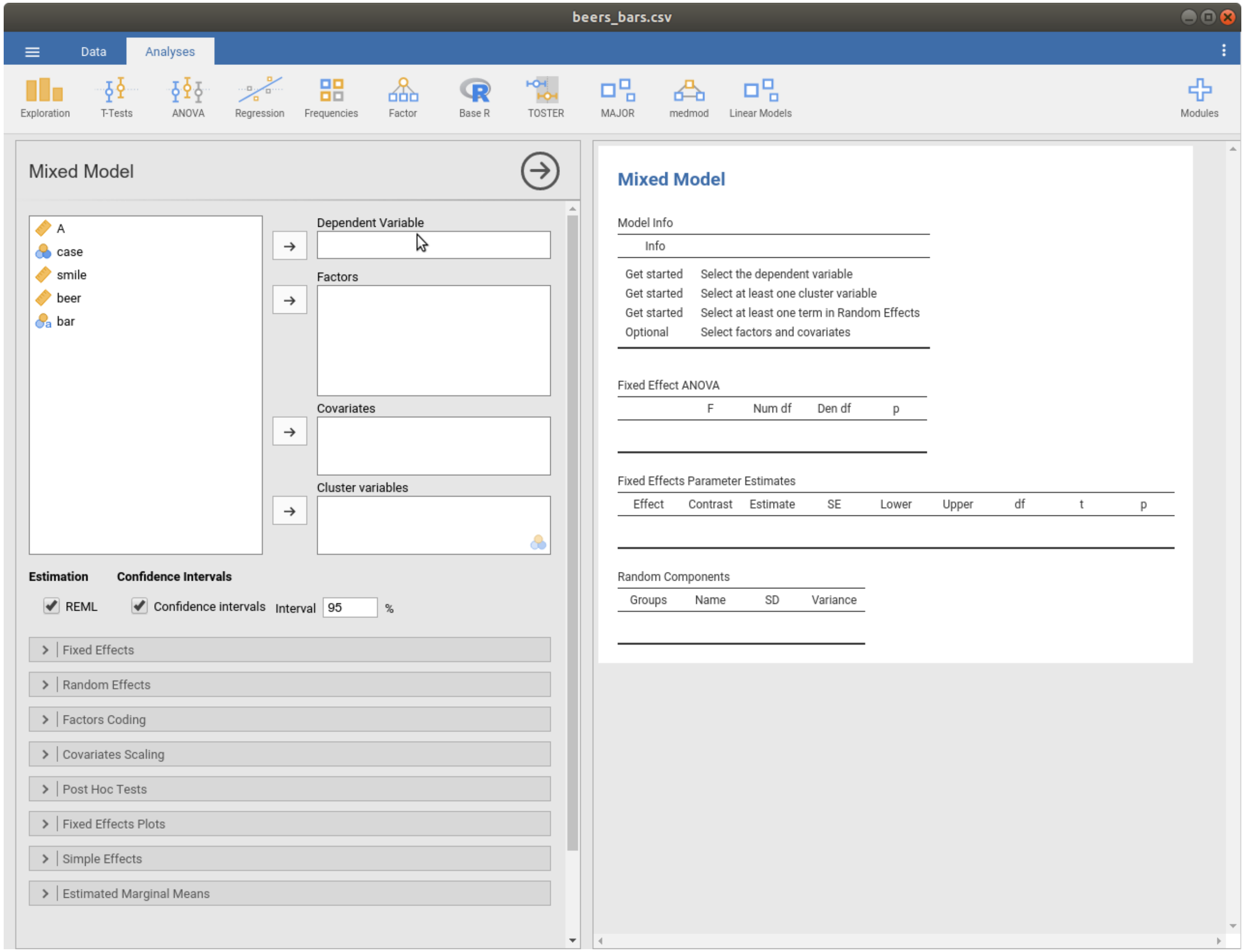
Plots

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



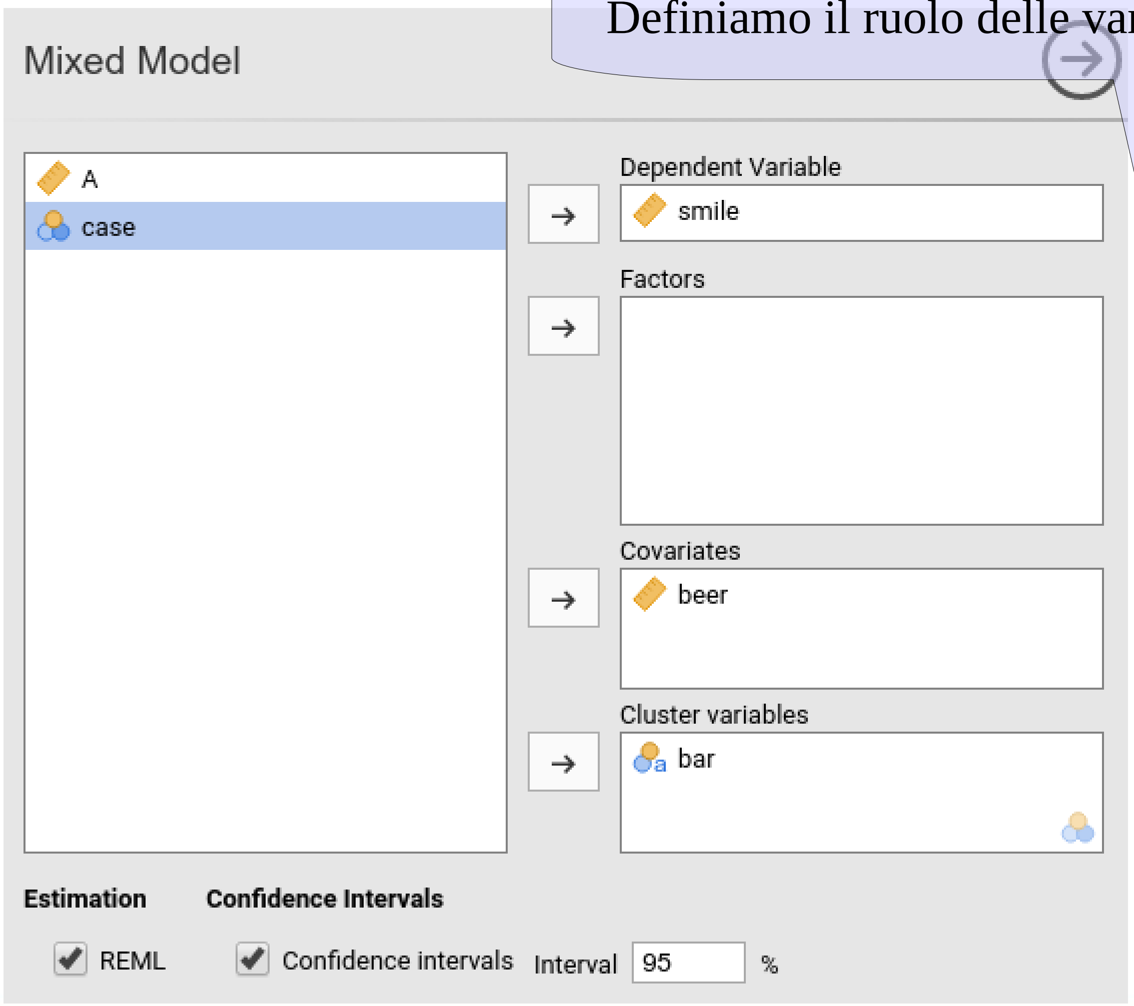
Jamovi

- In jamovi "mixed model" è nel modulo GAMLj



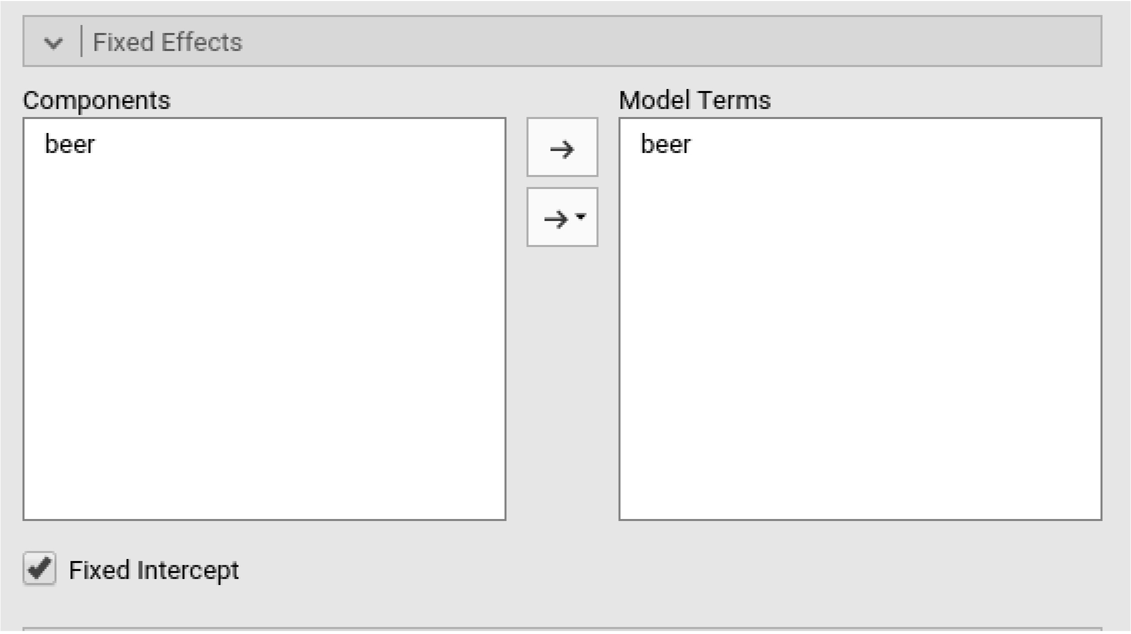
Jamovi

Definiamo il ruolo delle variabili



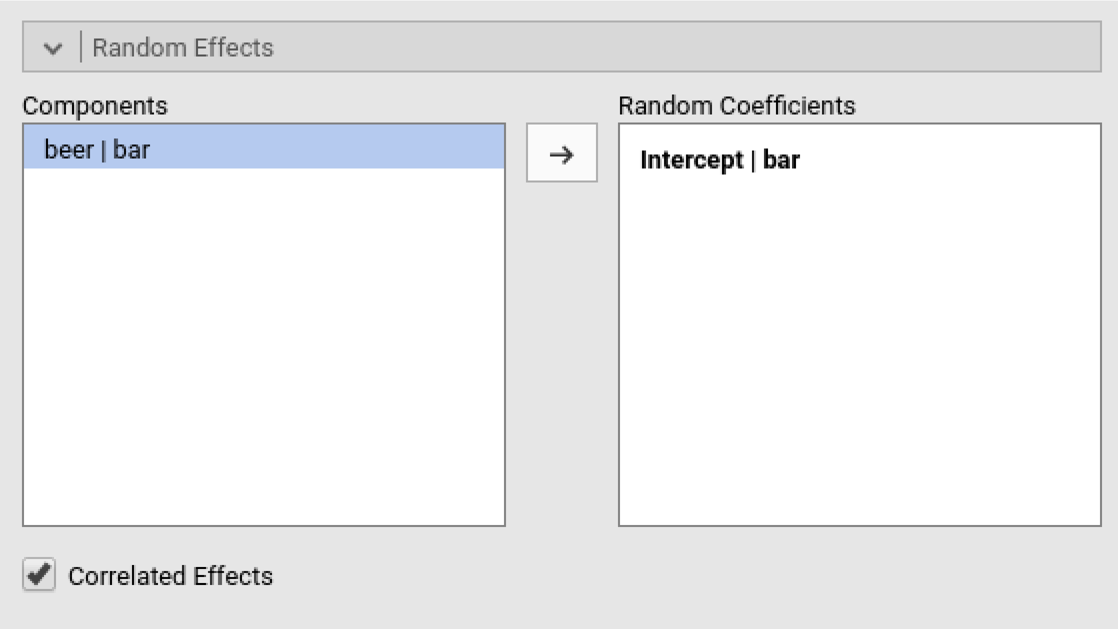
Jamovi

Definiamo gli effetti fissi



Jamovi

Definiamo la componente random



Jamovi

- 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

Jamovi

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

Jamovi

Coefficienti per l'effetto fisso di beer

Fixed Effects Parameter Estimates

Effect	Contrast	Estimate	SE	95% Confidence Interval Lower	95% Confidence Interval Upper	df	t	p
(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

Jamovi

Componente Random

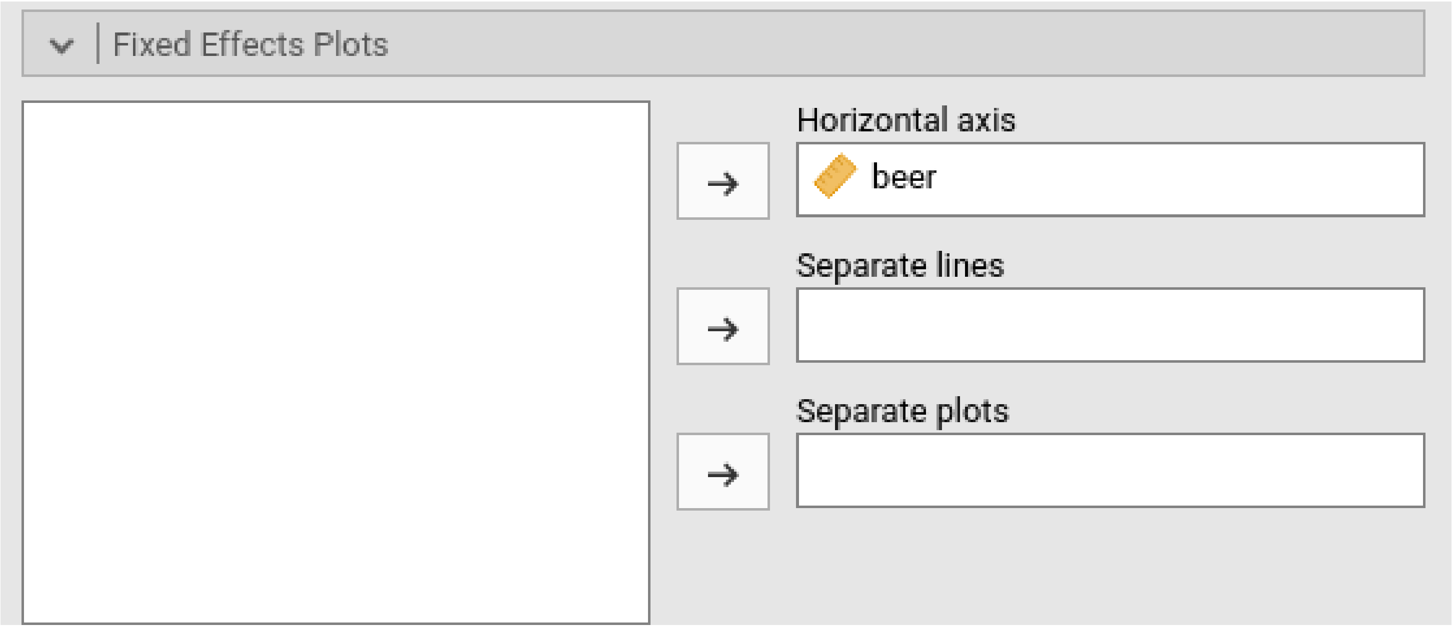
Random Components

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

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

Jamovi

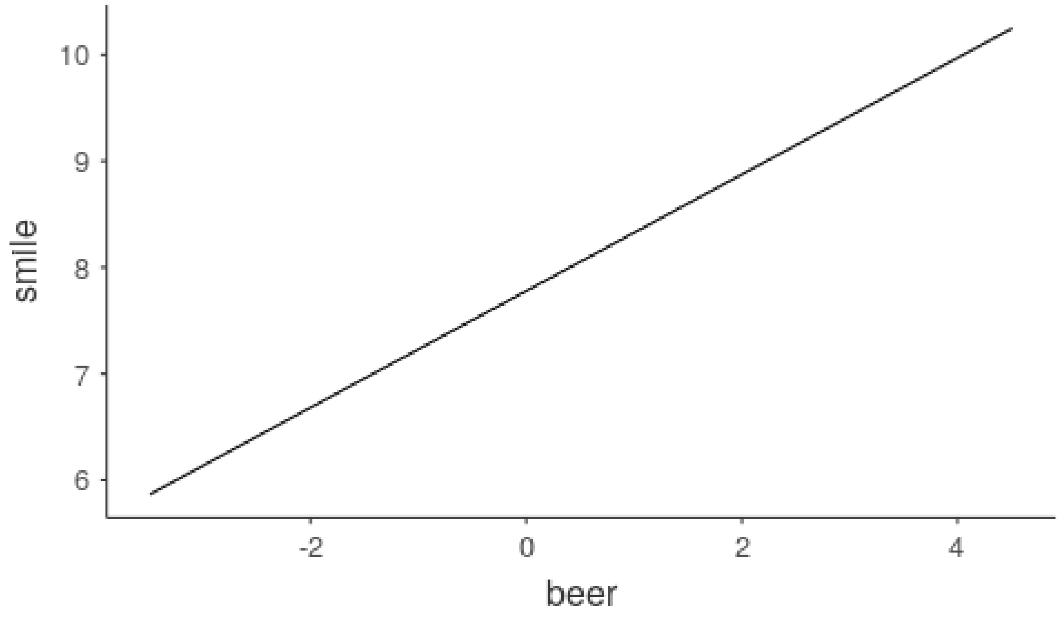
- Plots



Jamovi

Plot dell'effetto fisso

Fixed Effects Plots



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

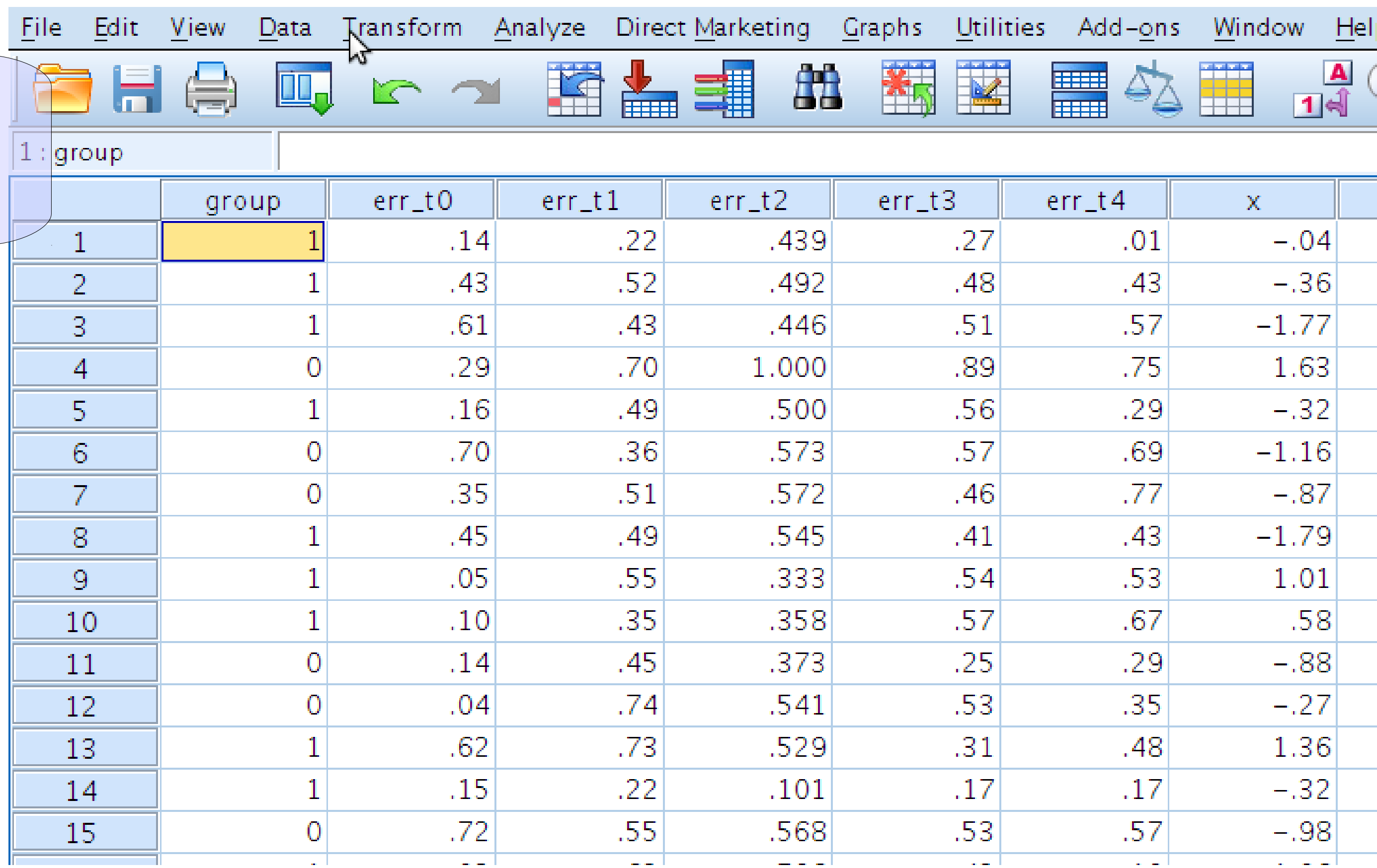
trial

Soggetti	1	2	3	4	5
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

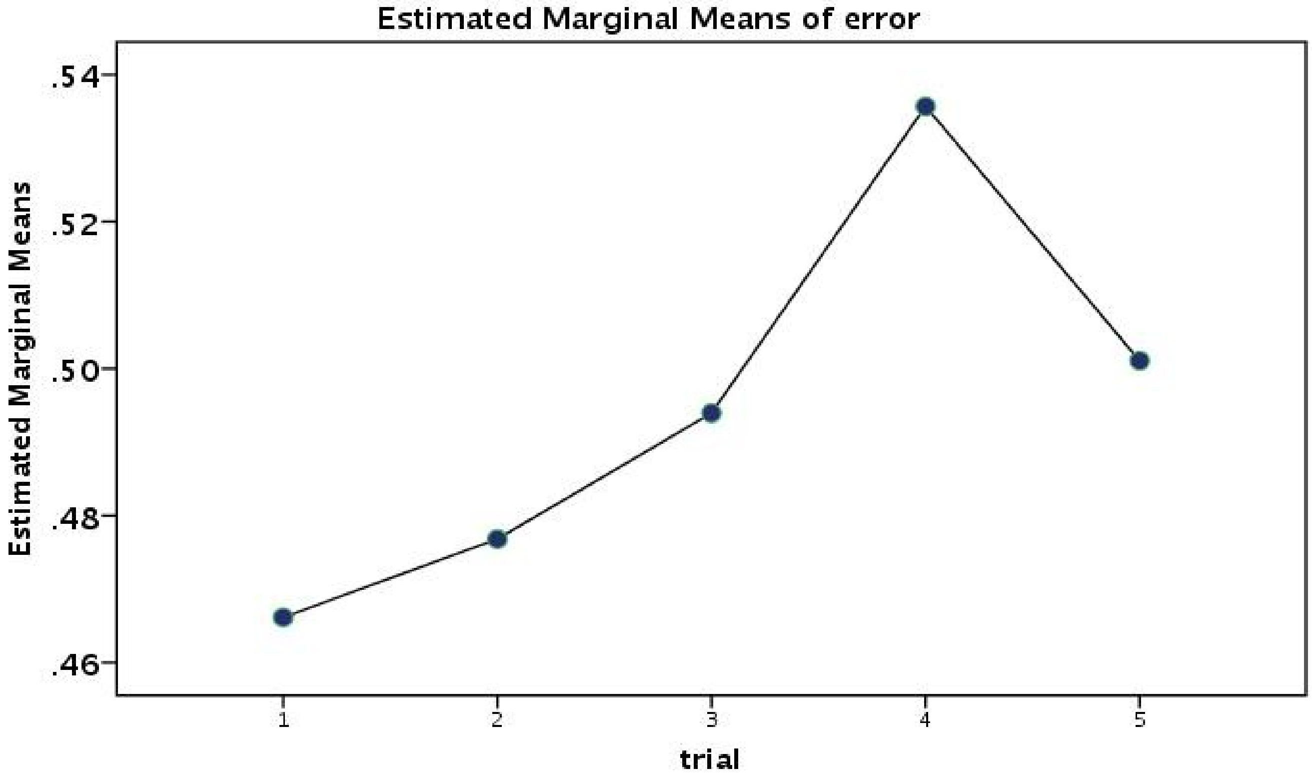
	id	group	x	trial	error
1	1	Cognitive Load	-0.03721233	1	0.139644148
2	1	Cognitive Load	-0.03721233	2	0.219611404
3	1	Cognitive Load	-0.03721233	3	0.439030338
4	1	Cognitive Load	-0.03721233	4	0.270522729
5	1	Cognitive Load	-0.03721233	5	0.009309587
6	2	Cognitive Load	-0.36044158	1	0.431302228
7	2	Cognitive 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 incontremo dei (gravi) problemi

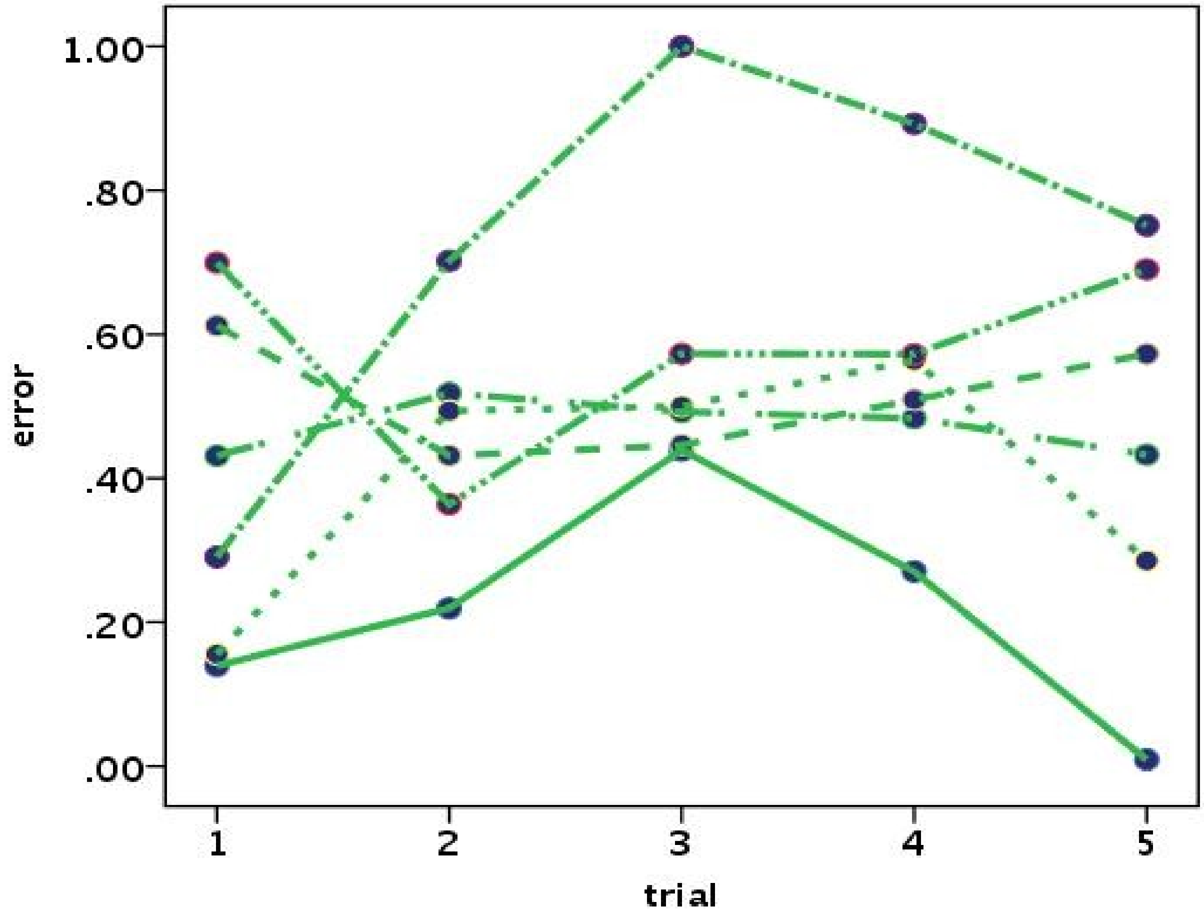
Effetti fissi

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

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

Stesso soggetto, stesso errore

$$\ldots\ldots$$

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

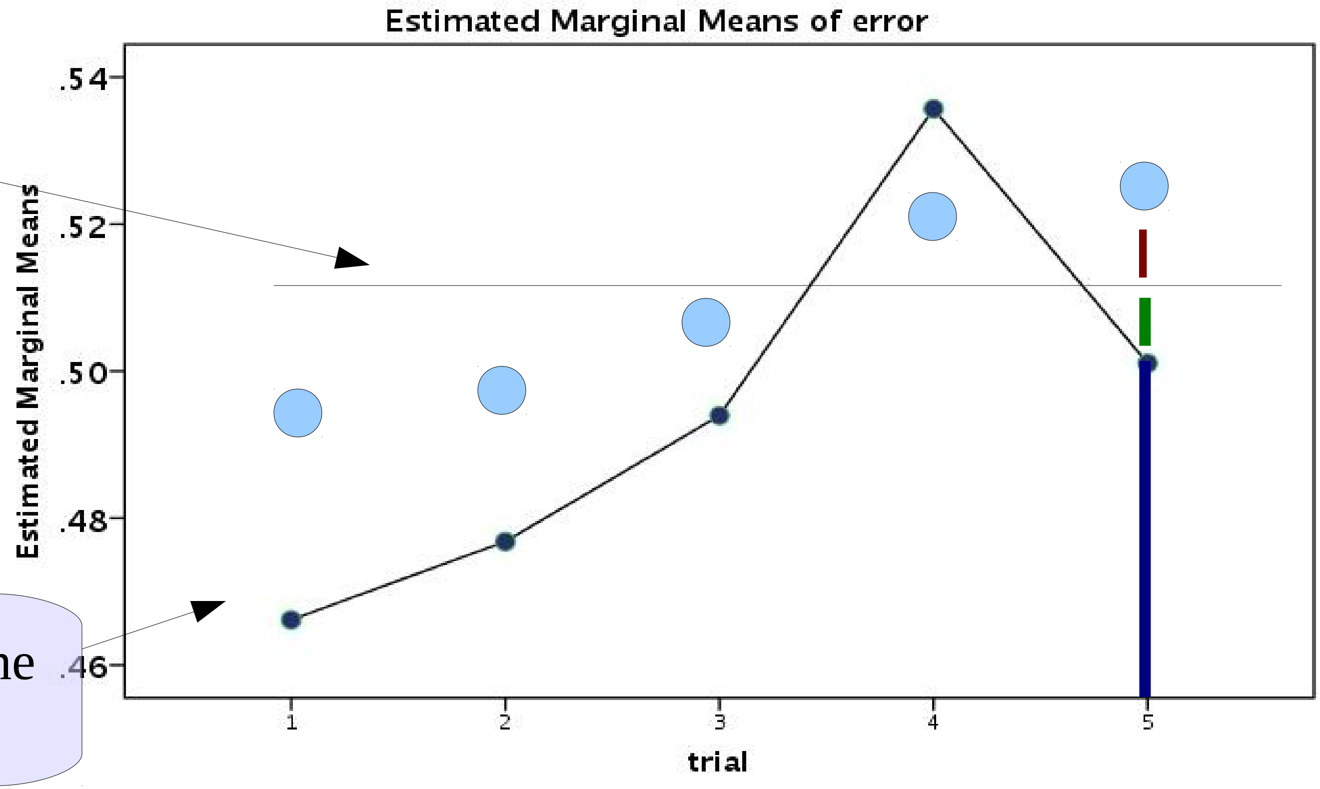
Stesso soggetto, stesso errore

Componente individuale

Esempio per un soggetto

Valore del tratto individuale

Medie del campione (effetti fissi)

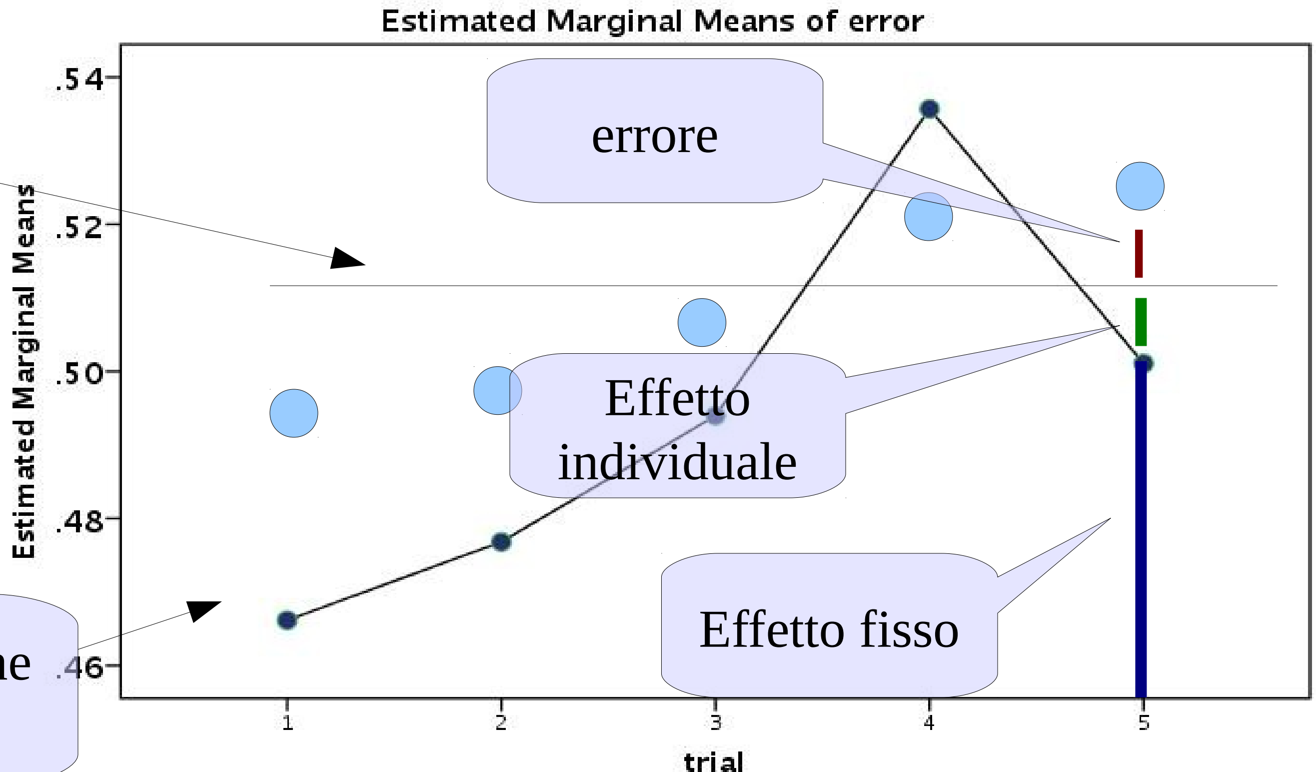


Componente individuale

Esempio per un soggetto

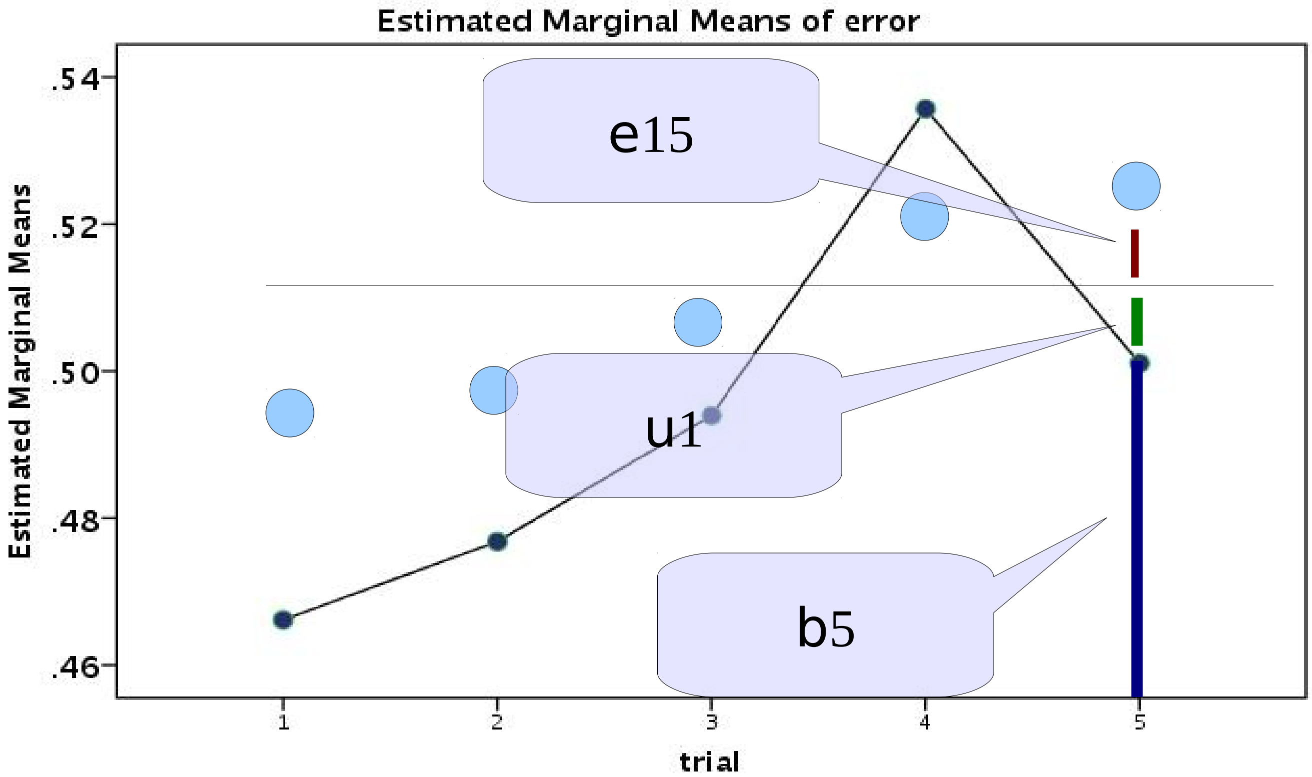
Valore del tratto individuale

Medie del campione (effetti fissi)



Componente individuale

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



Soluzione (1)

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

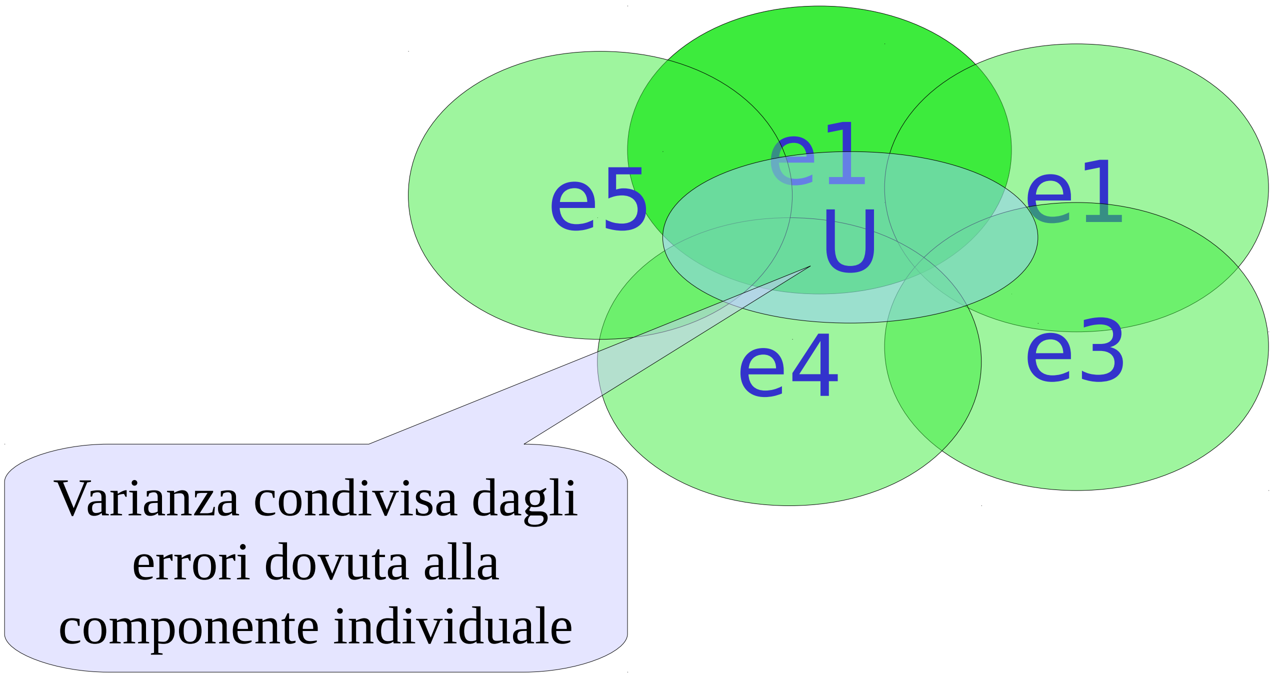
$$\left.\begin{aligned} Y_{11} &= a + b \cdot T_1 + u_1 + e_{11} \\ Y_{12} &= a + b \cdot T_2 + u_1 + e_{12} \\ Y_{13} &= a + b \cdot T_3 + u_1 + e_{13} \end{aligned}\right\} \text{Stesso soggetto, stesso errore}$$

$$\ldots\ldots$$

$$\left.\begin{aligned} Y_{i1} &= a + b \cdot T_1 + u_i + e_{i1} \\ Y_{i2} &= a + b \cdot T_2 + u_i + e_{i2} \\ Y_{i3} &= a + b \cdot T_2 + u_i + e_{i3} \end{aligned}\right\} \text{Stesso soggetto, stesso errore}$$

Soluzione (1)

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



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)

R syntax

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

```
mod<-lmer(error~(1|id)+trial,data=stress)
```

Effetti fissi (intercetta è inclusa di default)

Effetti random

Variabile cluster

Check delle dummies

```
stress$trial<-factor(stress$trial)
contrasts(stress$trial)
```

```
##   2 3 4 5
## 1 0 0 0 0
## 2 1 0 0 0
## 3 0 1 0 0
## 4 0 0 1 0
## 5 0 0 0 1
```

Trial 1 è il *reference category*

Output: effetti random

```
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  id       (Intercept) 0.007804 0.08834
##  Residual             0.030204 0.17379
## Number of obs: 1000, groups:  id, 200
##
```

C'è variabilità delle intercette dunque OK

Output: effetti fissi

```
summary(mod)
```

```
## Fixed effects:
##              Estimate Std. Error        df t value Pr(>|t|)
## (Intercept)   0.46616    0.01379 851.40000  33.815  < 2e-16 ***
## trial2        0.01066    0.01738 796.00000   0.613   0.5398
## trial3        0.02778    0.01738 796.00000   1.598   0.1103
## trial4        0.06951    0.01738 796.00000   4.000 6.94e-05 ***
## trial5        0.03491    0.01738 796.00000   2.009   0.0449 *
## ---
```

Comparazione media di ogni trial con il trial 1

Output: effetti fissi

```
anova(mod)
```

```
## Analysis of Variance Table of type III  with  Satterthwaite
## approximation for degrees of freedom
##          Sum Sq Mean Sq NumDF DenDF F.value Pr(>F)
## trial 0.57079  0.1427     4   796  4.7244  9e-04 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

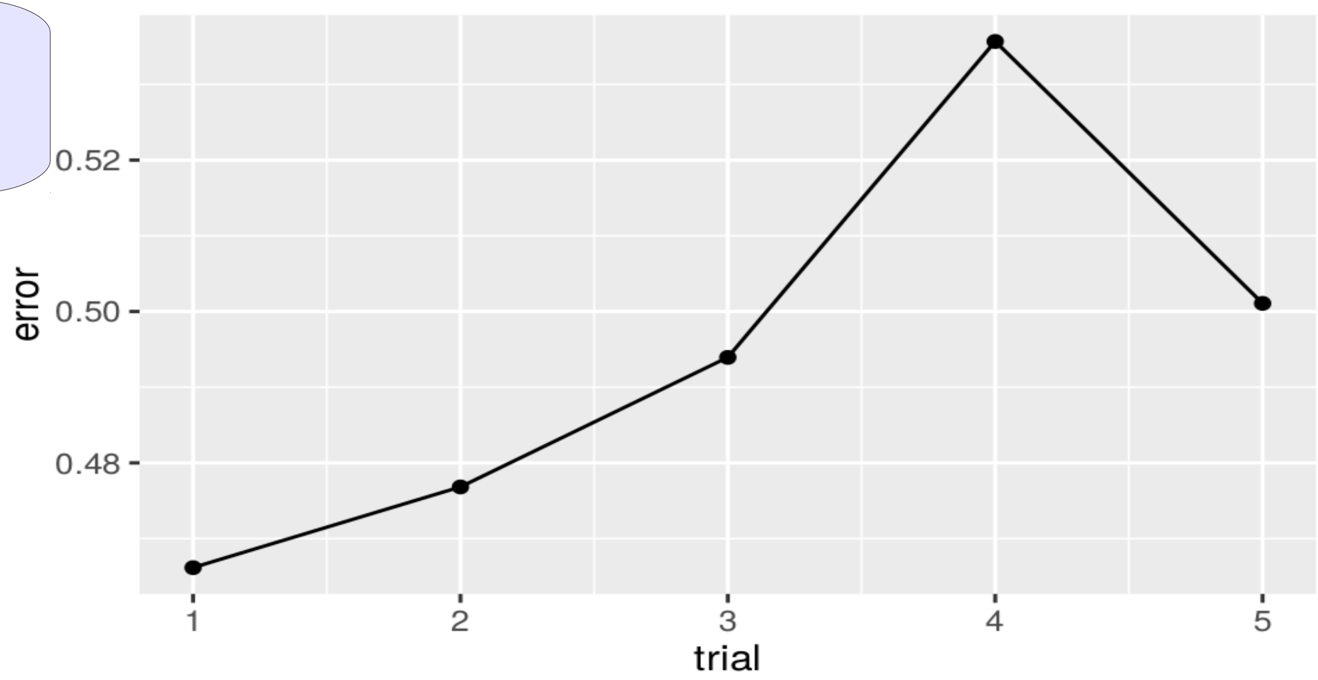
Test F per l'effetto principale della variabile categorica

Interpretazione

- L'interpretazione sarà la medesima che per una ANOVA qualunque, cioè guarderemo le medie

```
library(ggplot2)
ggplot(mm, aes(x=trial, y=error,group=1))+
  geom_line()+
  geom_point()
```

Effetti fissi



Dipendenza tra i punteggi

Quanta variabilità dei punteggi è spiegata dalla componente individuale?

Quantifichiamo cioè quanto è stato necessario considerare i soggetti come cluster

```
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  id       (Intercept) 0.007804 0.08834
##  Residual             0.030204 0.17379
## Number of obs: 1000, groups:  id, 200
##
```

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**

Coefficiente di dipendenza

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

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

```
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  id       (Intercept) 0.007804 0.08834
##  Residual             0.030204 0.17379
## Number of obs: 1000, groups:  id, 200
##
```

σ_a

σ

Coefficiente di dipendenza

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

$$ICR = \frac{.0078}{.0078 + .0302} = .205$$

Il 20% della variabilità della dipendente è attribuibile alla differenza tra soggetti

```
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  id       (Intercept) 0.007804 0.08834
##  Residual             0.030204 0.17379
## Number of obs: 1000, groups:  id, 200
##
```

σ_a

σ

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 efficente dei valori mancanti
- Possibilità di modellare variabili continue come variabili ripetute nel tempo
- Possibilità di combinare il disegno a misure ripetute con disegni gerachici 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

Definizione variabile indipendente

- Iniziamo a considerare il tempo come variabile categorica

```
genderage$age<-factor(genderage$age)
contrasts(genderage$age)
```

```
##   2 3 4
## 1 0 0 0
## 2 1 0 0
## 3 0 1 0
## 4 0 0 1
```

Definisco una categorica
e centro le dummies

```
contrasts(genderage$age)<-contr.sum(4)/2
contrasts(genderage$age)
```

```
##    [,1] [,2] [,3]
## 1   0.5  0.0  0.0
## 2   0.0  0.5  0.0
## 3   0.0  0.0  0.5
## 4  -0.5 -0.5 -0.5
```

Definizione variabile indipendente

- Gender è categorica

```
genderage$gender<-factor(genderage$gender)
contrasts(genderage$gender)
```

```
##       Girls
## Boys      0
## Girls     1
```

```
contrasts(genderage$gender)<-contr.sum(2)/2
contrasts(genderage$gender)
```

Definisco una categorica
e centro le dummies

```
##        [,1]
## Boys    0.5
## Girls  -0.5
```

Il modello

- Il modello ora comprende due fattori, la loro interazione, ed una intercetta random per catturare le differenze individuali

$$Y_{ik} = \bar{a} + a_i + b_a Age_k + b_g Gender + b_x Age * Gender + e_{ij}$$

Effetto random

Effetto Age (within)

Effetto Between-subject

Interazione

Definizione del modello

$$Y_{ik} = \bar{a} + a_i + b_a Age_k + b_g Gender + b_x Age * Gender + e_{ij}$$

- Quali sono gli effetti fissi? Intercetta, Age, Gender, Age*Gender
- Quali sono gli effetti random? Intercetta
- Quali sono i cluster su cui variano gli effetti random? Soggetto (id)

```
mod<-lmer(y~(1|id)+age*gender,data=genderage)
```

Risultati

- Varianza delle intercette

```
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  id       (Intercept)  78.02    8.833
##  Residual             127.38   11.286
## Number of obs: 400, groups:  id, 50
##
```

$$ICR = \frac{78.02}{78.02 + 127.38} = .379$$

Risultati ANOVA

- I risultati ci daranno la F e la significatività per gli effetti richiesti

```
anova(mod)
```

```
## Analysis of Variance Table of type III  with  Satterthwaite
## approximation for degrees of freedom
##             Sum Sq Mean Sq NumDF DenDF F.value   Pr(>F)
## age        1542.23  514.08     3   343  4.0357 0.007678 **
## gender      596.32  596.32     1   343  4.6820 0.031180 *
## age:gender 1070.43  356.81     3   343  2.8011 0.039949 *
## ---
```

Risultati ANOVA

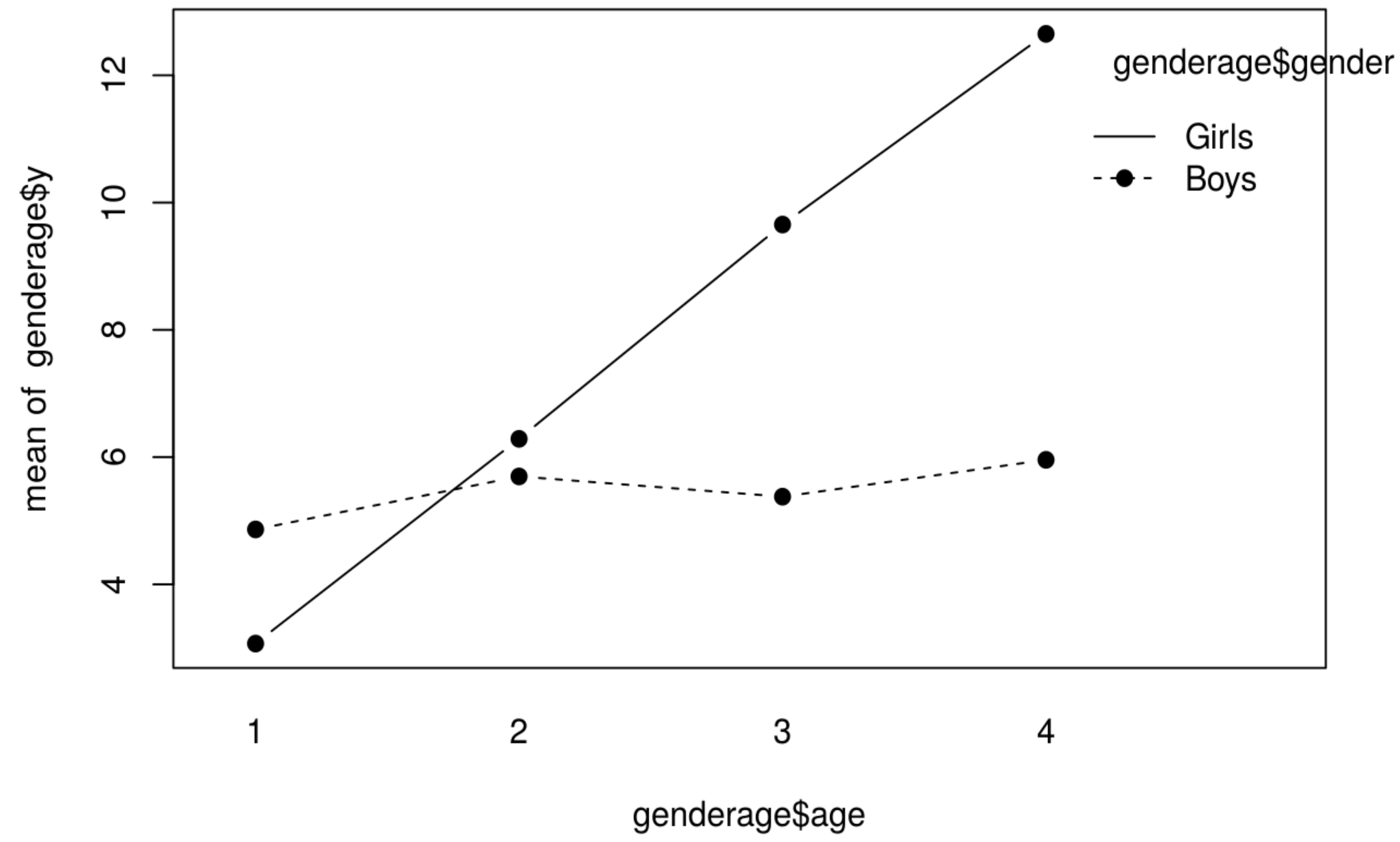
- I risultati dei coefficienti possiamo ignorarli

```
## Fixed effects:
##                 Estimate Std. Error      df t value Pr(>|t|)
## (Intercept)        6.695      1.371  49.000   4.884 1.15e-05 ***
## age1              -5.456      1.955 343.000  -2.791  0.00554 **
## age2              -1.406      1.955 343.000  -0.719  0.47248
## age3               1.644      1.955 343.000   0.841  0.40108
## gender1           -2.442      1.129 343.000  -2.164  0.03118 *
## age1:gender1       8.472      3.910 343.000   2.167  0.03094 *
## age2:gender1       3.702      3.910 343.000   0.947  0.34435
## age3:gender1      -3.669      3.910 343.000  -0.938  0.34869
## ---
```

Effetti medi

- Le medie della combinazione dei livelli dei fattori per l'interpretazione

```
interaction.plot(genderage$age,genderage$gender,genderage$y,type="b",pch=19)
```



Disegno a misure ripetute

- Abbiamo 2 gruppi - Control vs Treatment, measurati 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

*) https://www.uvm.edu/~dhowell/StatPages/More_Stuff/Mixed-Models-Repeated/Mixed-Models-for-Repeated-Measures1.html

Disegno a misure ripetute

- Abbiamo 2 gruppi - Control vs Treatment, measurati 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

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

*) https://www.uvm.edu/~dhowell/StatPages/More_Stuff/Mixed-Models-Repeated/Mixed-Models-for-Repeated-Measures1.html

Disegno a misure ripetute: dati

- Data sono in “long format”

Ogni soggetto ha 4 righe

Data | Analyses

Exploration | T-Tests | ANOVA | Regression | Frequencies | Factor | Linear Models

	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

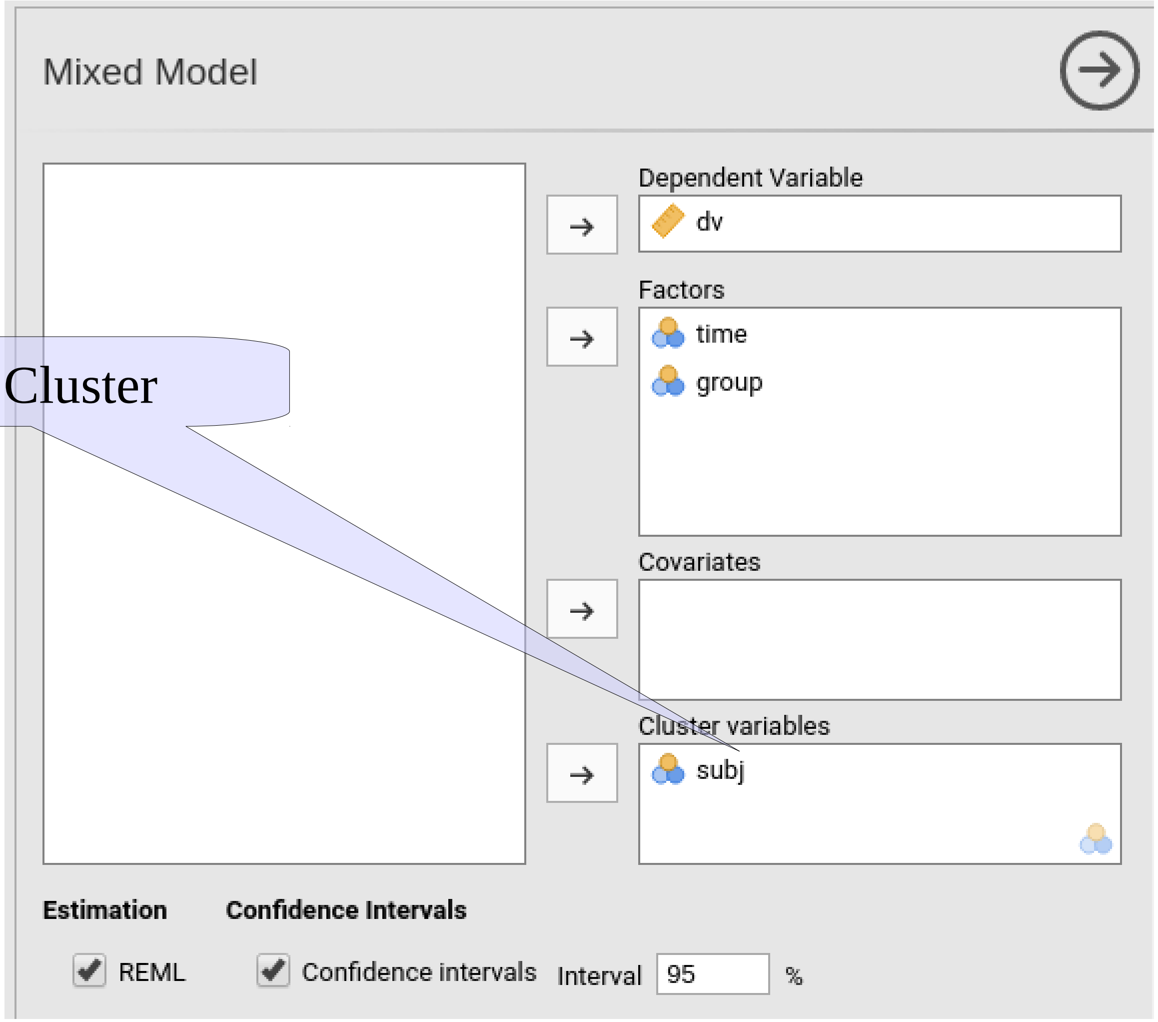
Mixed model

Traduciamo il disegno in un modello misto

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

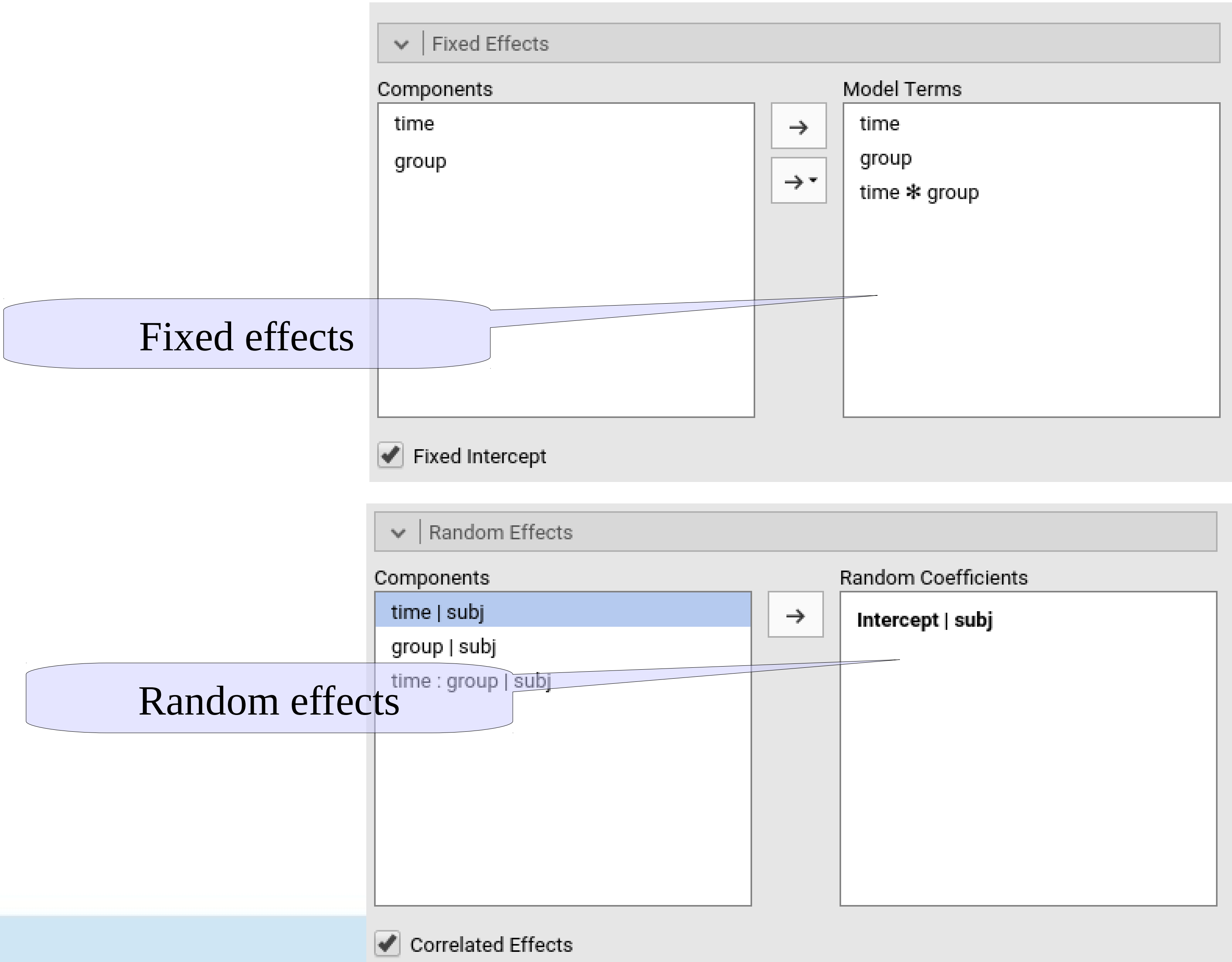
Variables

- Definisco le variabili



Variabile Cluster

Modello


Fixed Effects
Components
time
group
Model Terms
time
group
time ✻ group
Fixed effects
Fixed Intercept
Random Effects
Components
time | subj
group | subj
time : group | subj
Random Coefficients
Intercept | subj
Random effects
Correlated Effects

Risultati

- Interpretazione dei risultati

Modello

Effetti random

Mixed Model

Model Info

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

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

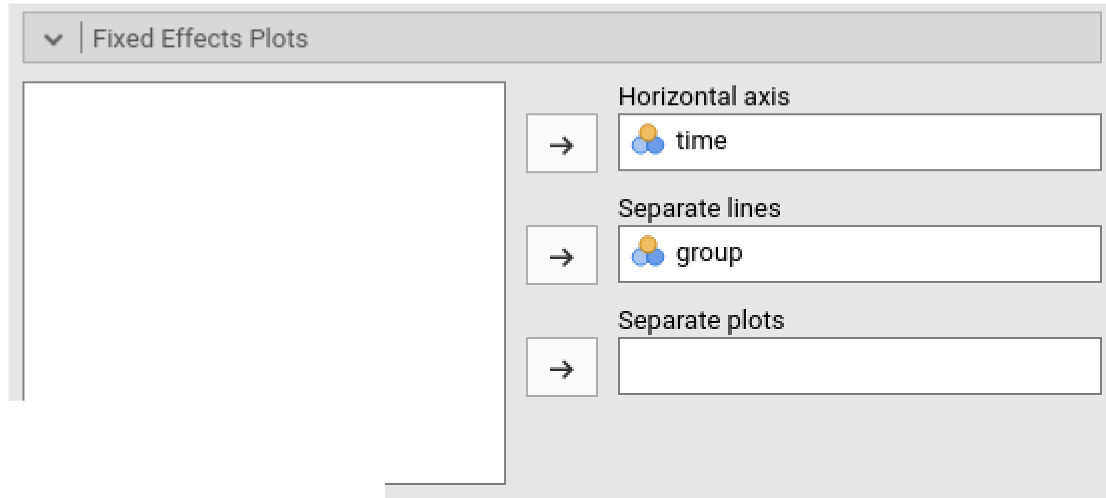
Fixed Effect ANOVA

	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

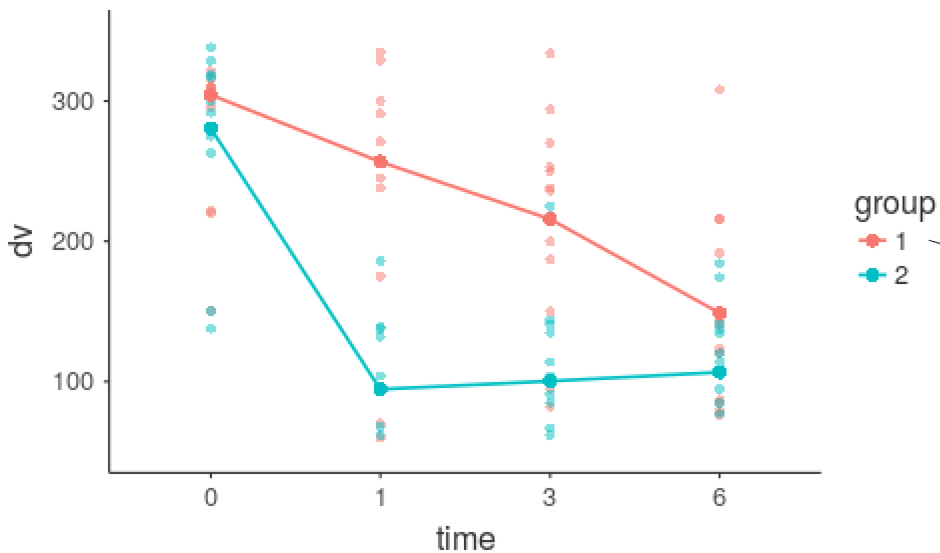
Note. Satterthwaite method for degrees of freedom

- Per il momento ignoriamo la stima dei coefficienti

Results: plot



Fixed Effects Plots



Rosso è il gruppo di controllo

Analisi sulla interazione

- Per analizzare ulteriormente l'interazione possiamo fare (come nel GLM):

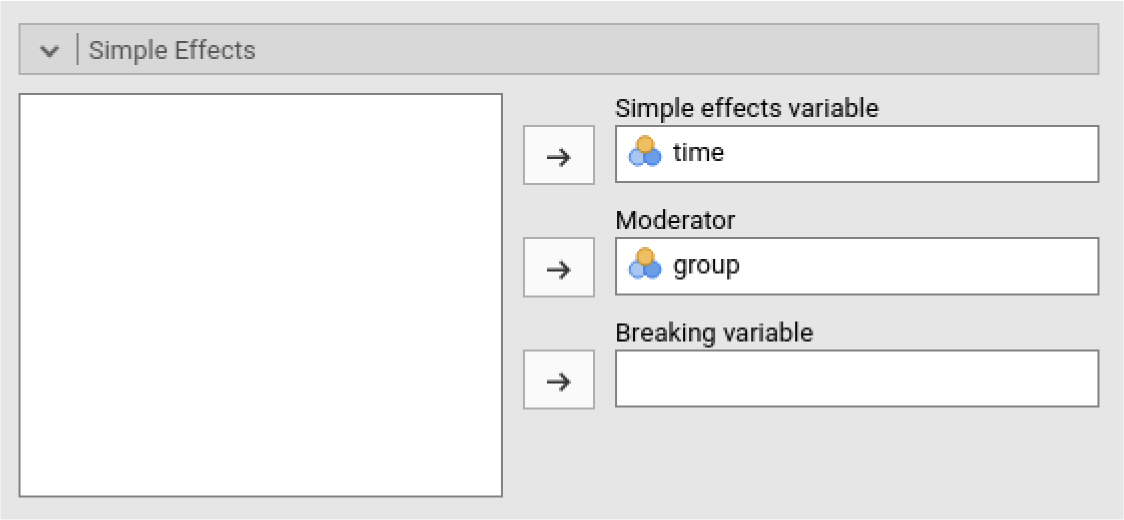
- Simple effects: Testare se l'effetto di tempo è presente in ognuno dei gruppi

- 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

- Chiediamo di stimare gli effetti ti tempo in ogni gruppo

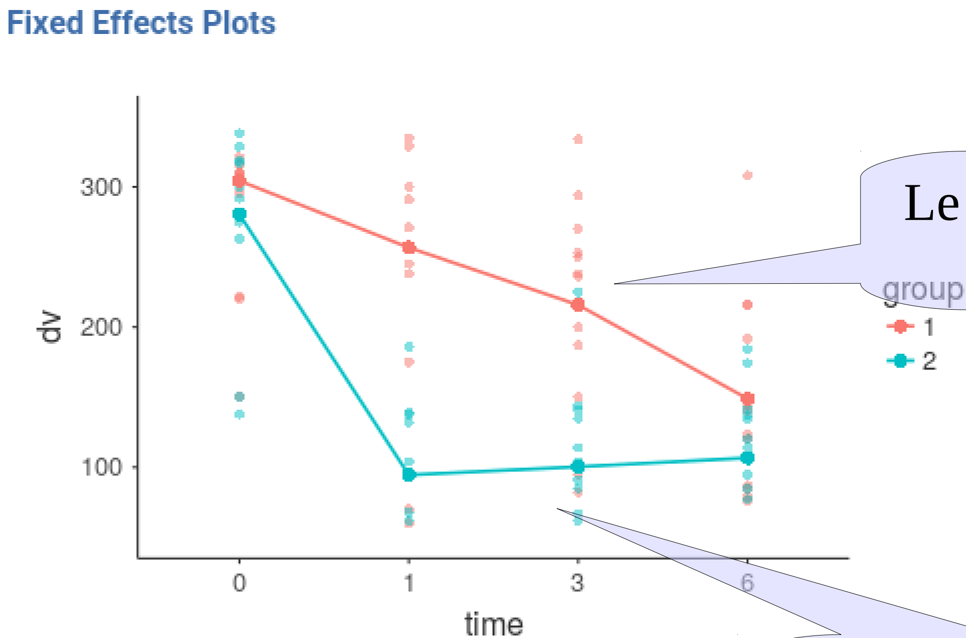
Simple Effects ANOVA

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



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

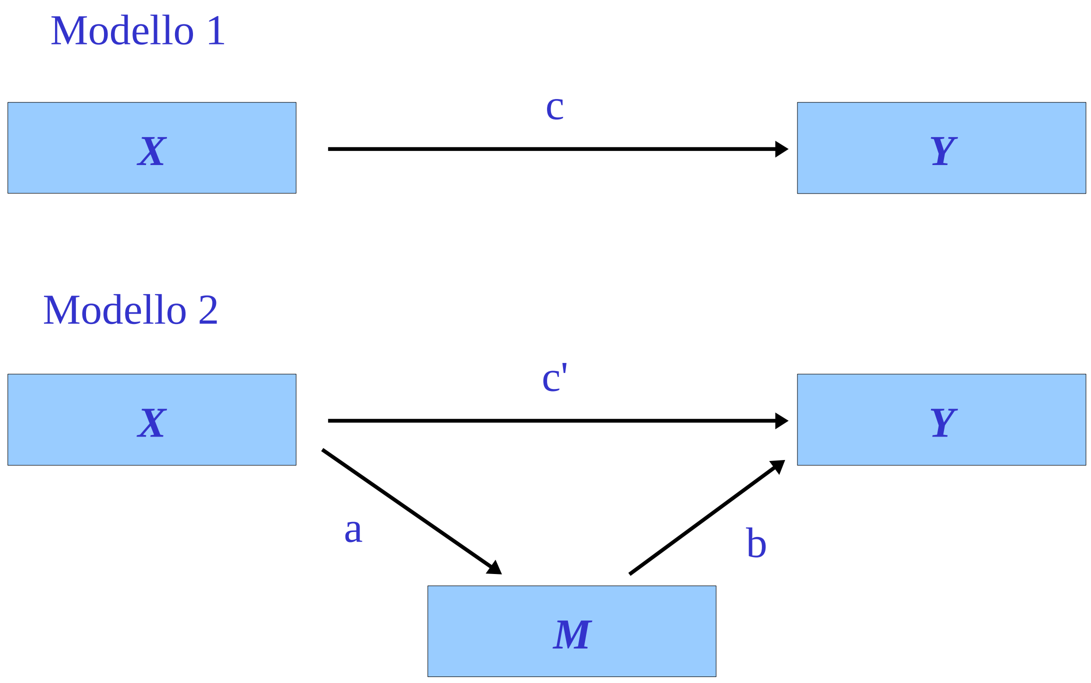
- Ciò consente di applicare tutte le conoscenze dell'ANOVA/Regressione al caso delle misure ripetute

- I modelli misti consentono dunque di stimare modelli a misure ripetute combinandoli con altre strutture complesse dei dati

La mediazione nel modello misto

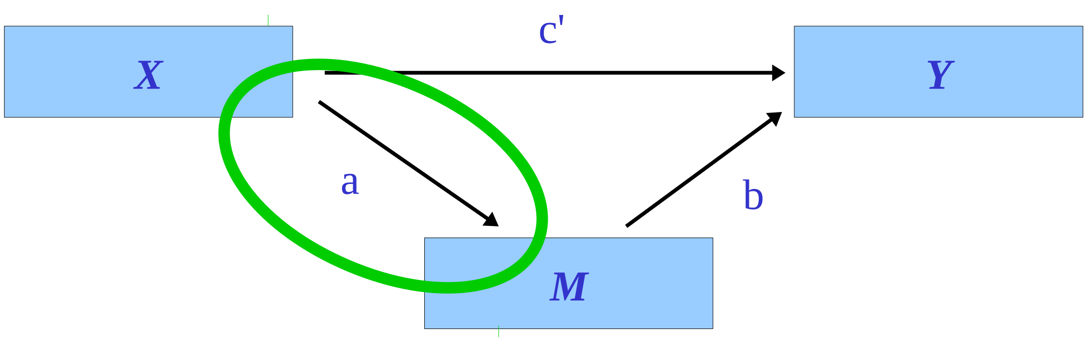
La mediazione

- Logicamente, la mediazione nel modello misto è identica al modello lineare generale



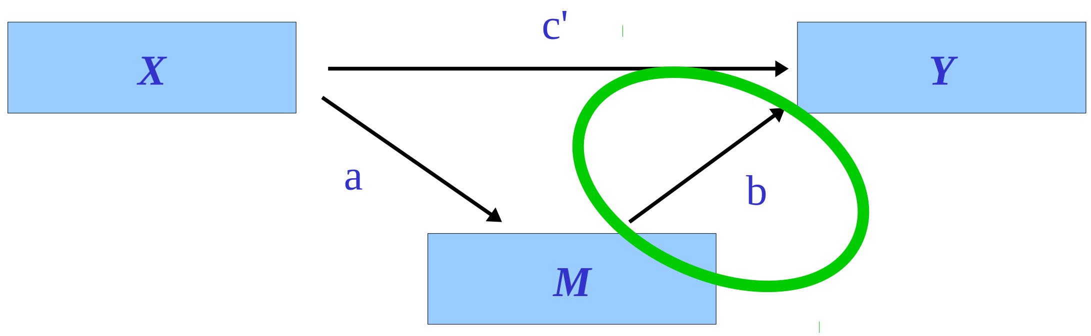
Condizioni statistiche

- Nel modello misto i coefficienti della mediazione si ottengono stimando due modelli
  - **X sul mediatore**
    - L'effetto si ottiene con un regressione semplice con X come IV e Y come DV
    - E si prende **l'effetto fisso**



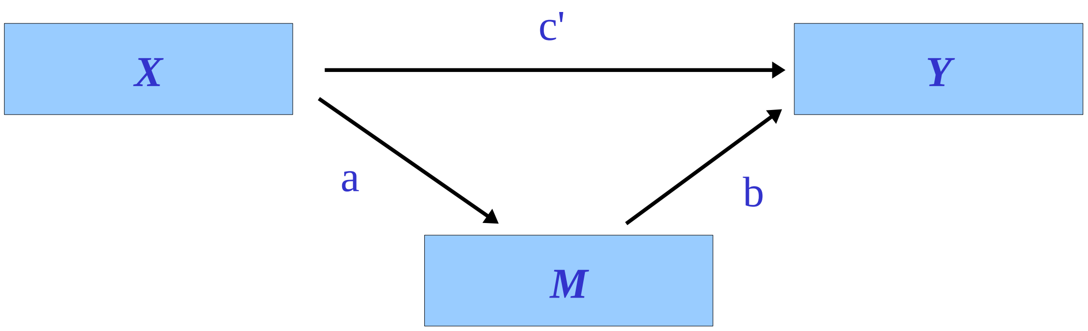
Condizioni statistiche

- Nel modello misto i coefficienti della mediazione si ottengono stimando due modelli
  - **Mediatore sulla dipendente al netto della X**
    - L'effetto si ottiene con un regressione multiple con X e M come IV e Y come DV
    - E si prende **l'effetto fisso**



L'effetto mediato

- Il problema è che l'effetto mediato non corrisponde al prodotto dei coefficienti fissi



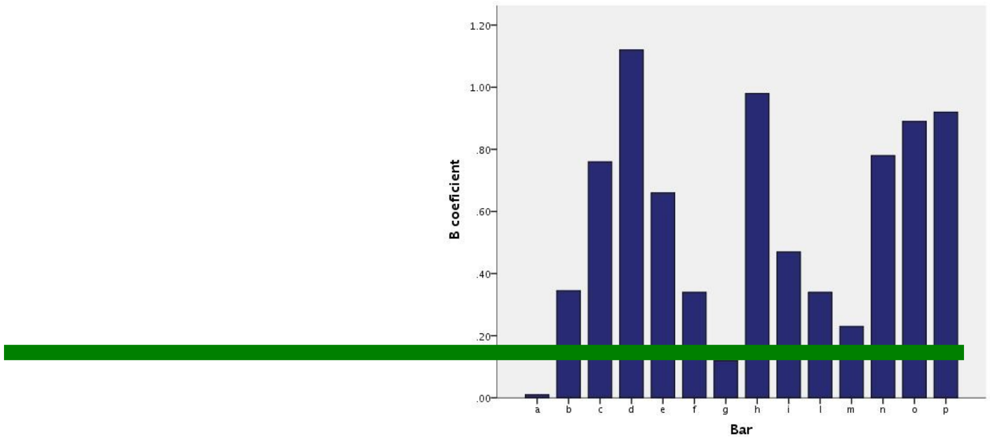
$$EM \neq a \cdot b$$

Media dei Coefficiente

- Ricordiamo che il coefficiente fisso in un modello misto è la media dei coefficienti random

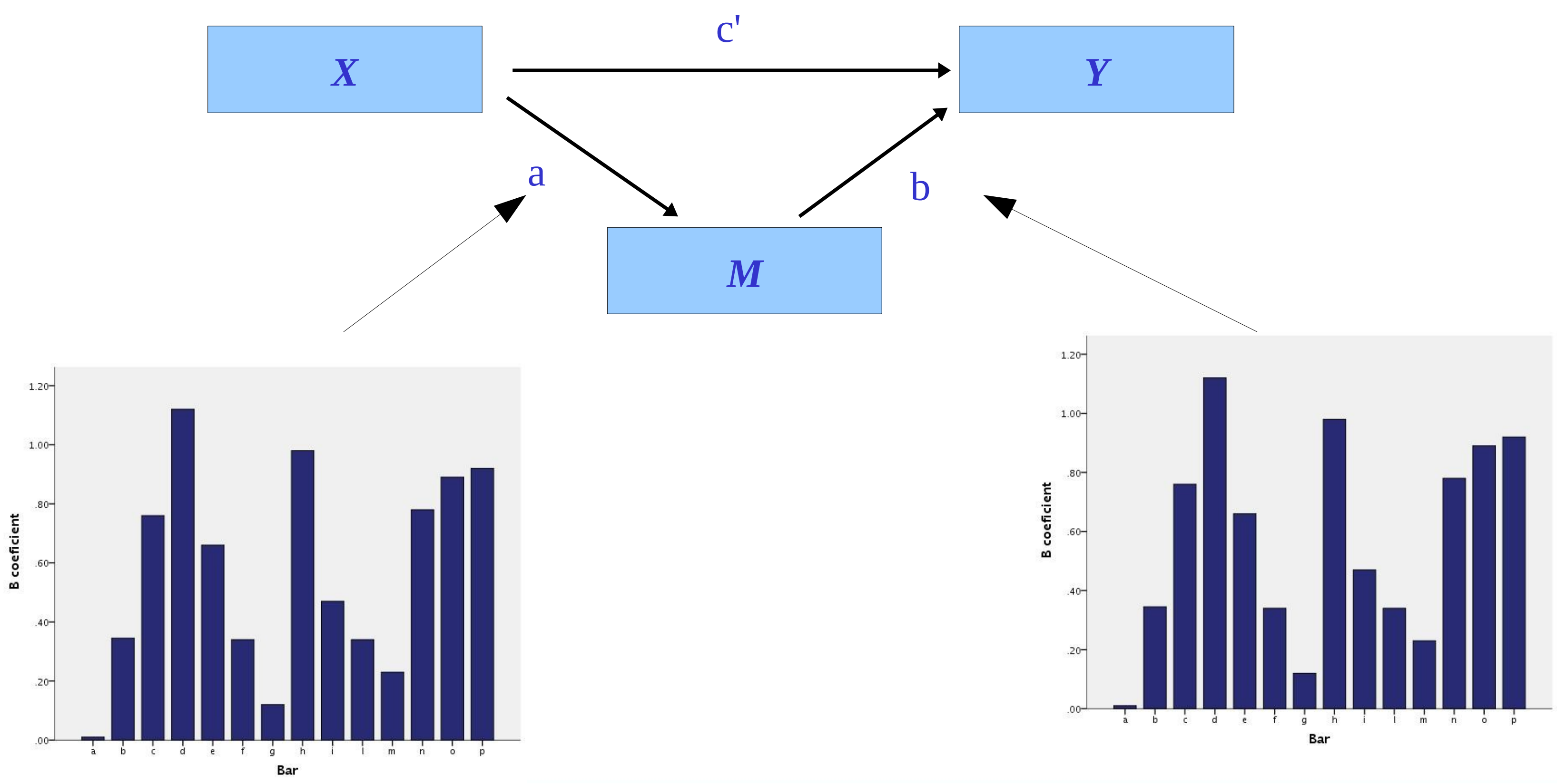
MEDIA

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



L'effetto mediato

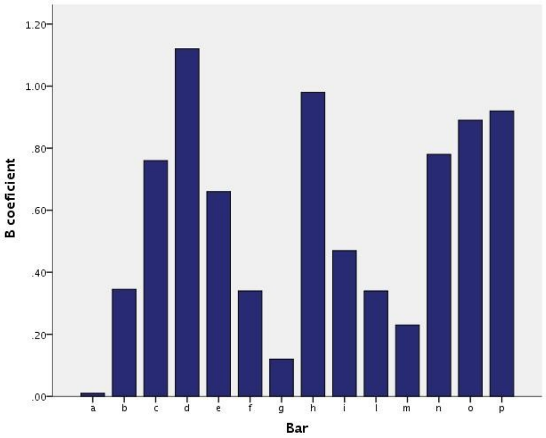
- Dunque per ogni coefficiente della mediazione (a e b) avremo una distribuzione di coefficienti.



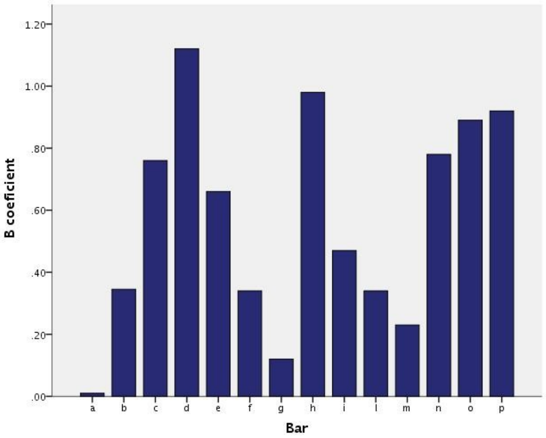
L'effetto mediato

- Dunque per ogni coefficiente della mediazione (a e b) avremo una distribuzione di coefficienti.

Distribuzione di **a**



Distribuzione di **b**

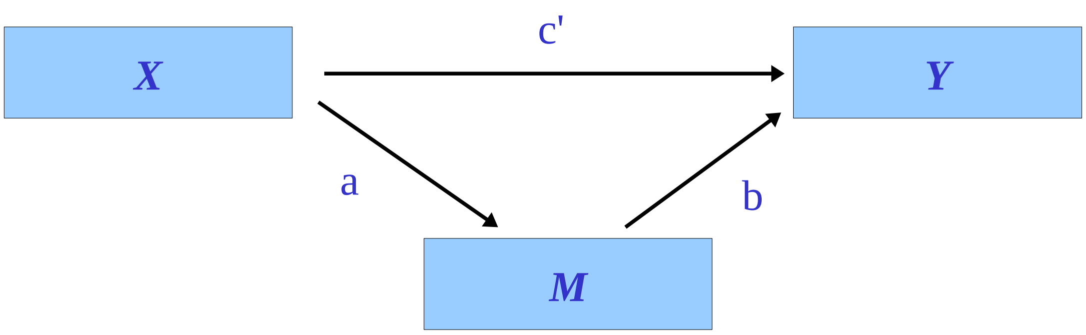


Queste coefficienti possono essere correlati

L'effetto mediato

- E l'effetto mediato è dato dal prodotto **a** e **b** più la covarianza tra le distribuzioni di coefficienti

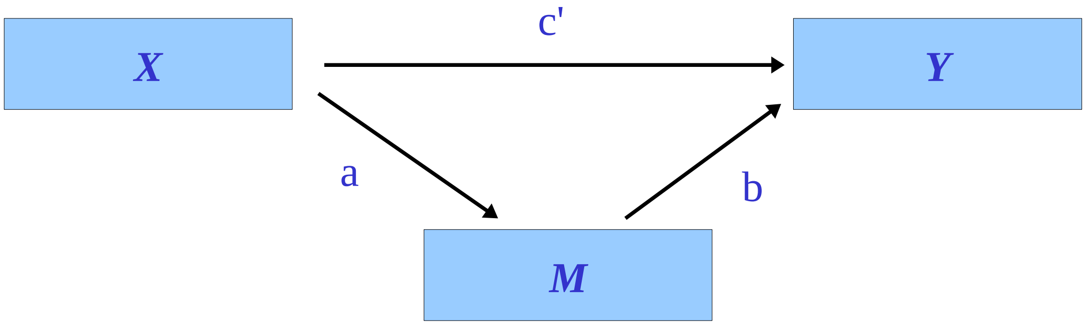
- Stimare questa covarianza è un incubo!!



$$EM = a \cdot b + cov(a, b)$$

Intervalli di confidenza

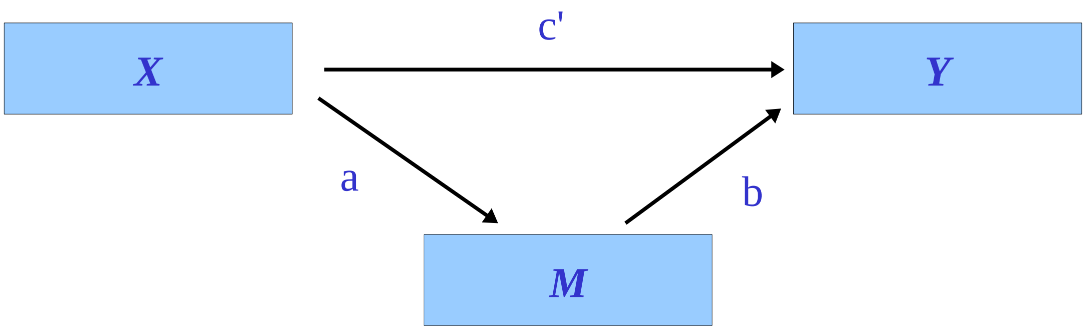
- Inoltre, il metodo bootstrap non è chiaro se funzioni bene con il modello misto
- Si preferisce un altro metodo detto “Monte Carlo Method”
- Stimare questi intervalli è un incubo!!



$$EM = a \cdot b + cov(a, b)$$

bert17

- Noi useremo bert17 (su queste funzioni è ancora un po' instabile)
- Dunque opereremo come la mediazione normale



$$EM = a \cdot b + cov(a, b)$$

Esempio

- Una ricerca sperimentale è volta a studiare l'effetto della densità di stimoli in un lasso di tempo sulla percezione del passare del tempo

- I soggetti eseguivano un compito con un numero variabile di stimoli da osservare

- Il numero di stimoli è stato manipolato sperimentalmente (0,30,60) in tre condizioni a **misure ripetute**

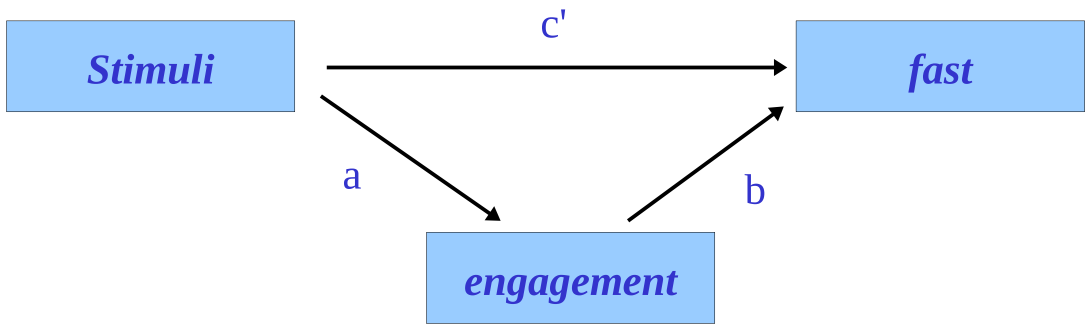
- Per ogni intervallo è stato misurato l'*engagement* (*engagment*) del soggetto nel compito sperimentale e quanto rapidamente avessero percepito il tempo passato durante il compito (*fast*)

```
table(timedata$stimuli)
```

```
##
##   0  30  60
## 924 732 924
```

Esempio

- L'ipotesi è la seguente



- Essendo il disegno a misure ripetute, dobbiamo usare il modello misto

Effetto totale

- Operiamo le solite tre stime, ma usando il modello misto, stimando gli effetti random e fissi tra soggetti (variabile cluster sono i soggetti)

```
mod1<-lmer(fast~(1+cstimuli|id)+cstimuli,data=timedata)
summary(mod1)
```

Effetti random

- Operiamo le solite tre stime, ma usando il modello misto, stimando gli effetti random e fissi tra soggetti (variabile cluster sono i soggetti)

```
mod1<-lmer(fast~(1+cstimuli|id)+cstimuli,data=timedata)
summary(mod1)
```

```
##
## Random effects:
##  Groups   Name        Variance Std.Dev. Corr
##  id       (Intercept) 0.496    0.7043
##           cstimuli    0.254    0.5040   1.00
##  Residual             4.863    2.2053
## Number of obs: 2580, groups:  id, 308
##
```

Notiamo semplicemente che abbiamo della variabilità nei coefficienti

Effetti fissi

- Operiamo le solite tre stime, ma usando il modello misto, stimando gli effetti random e fissi tra soggetti (variabile cluster sono i soggetti)

```
## Fixed effects:
##              Estimate Std. Error        df t value Pr(>|t|)
## (Intercept) 4.610e+00  5.929e-02 3.100e+02  77.746   <2e-16 ***
## cstimuli    1.129e+00  1.734e-01 1.549e+03   6.512    1e-10 ***
## ---
```

Effetto totale fisso

1.129

Effetto sul mediatore

- Operiamo le solite tre stime, ma usando il modello misto, stimando gli effetti random e fissi tra soggetti (variabile cluster sono i soggetti)

```
mod2<-lmer(engagment~(1+cstimuli|id)+cstimuli,data=timedata)
summary(mod1)
```

Effetti random

- Operiamo le solite tre stime, ma usando il modello misto, stimando gli effetti random e fissi tra soggetti (variabile cluster sono i soggetti)

```
## Random effects:
##  Groups   Name        Variance Std.Dev. Corr
##  id       (Intercept) 2.038552 1.4278
##           cstimuli    0.001253 0.0354   1.00
##  Residual             2.547846 1.5962
## Number of obs: 2580, groups:  id, 308
##
```

Notiamo semplicemente che abbiamo della variabilità nei coefficienti

Effetti fissi

- Operiamo le solite tre stime, ma usando il modello misto, stimando gli effetti random e fissi tra soggetti (variabile cluster sono i soggetti)

```
##
## Fixed effects:
##               Estimate Std. Error        df t value Pr(>|t|)
## (Intercept) 5.353e+00  8.734e-02 3.078e+02  61.289  < 2e-16 ***
## cstimuli    7.233e-01  1.238e-01 2.263e+03   5.843 5.86e-09 ***
## ---
```

Effetto sul mediatore

.723

Effetto del mediatore

- Operiamo le solite tre stime, ma usando il modello misto, stimando gli effetti random e fissi tra soggetti (variabile cluster sono i soggetti)

```
mod3<-lmer(fast~+(1+cstimuli|id)+(1+engagment|id)+engagment+cstimuli,data=timedata)
summary(mod3)
```

Effetti random

- Operiamo le solite tre stime, ma usando il modello misto, stimando gli effetti random e fissi tra soggetti (variabile cluster sono i soggetti)

```
.. ..
## Random effects:
##  Groups   Name        Variance Std.Dev. Corr
##  id       (Intercept) 0.257273 0.50722
##           cstimuli    0.007445 0.08628  -1.00
##  id.1     (Intercept) 1.544368 1.24273
##           engagment   0.079323 0.28164  -1.00
##  Residual             3.367156 1.83498
## Number of obs: 2580, groups:  id, 308
##
```

Notiamo semplicemente che abbiamo della variabilità nei coefficienti

Effetti fissi

- Operiamo le solite tre stime, ma usando il modello misto, stimando gli effetti random e fissi tra soggetti (variabile cluster sono i soggetti)

```
##
## Fixed effects:
##                Estimate Std. Error        df t value Pr(>|t|)
## (Intercept) 1.468e+00  1.392e-01 2.067e+02  10.551  < 2e-16 ***
## engagment   5.971e-01  2.599e-02 2.456e+02  22.977  < 2e-16 ***
## cstimuli    6.223e-01  1.451e-01 2.157e+03   4.288 1.88e-05 ***
## ---
```

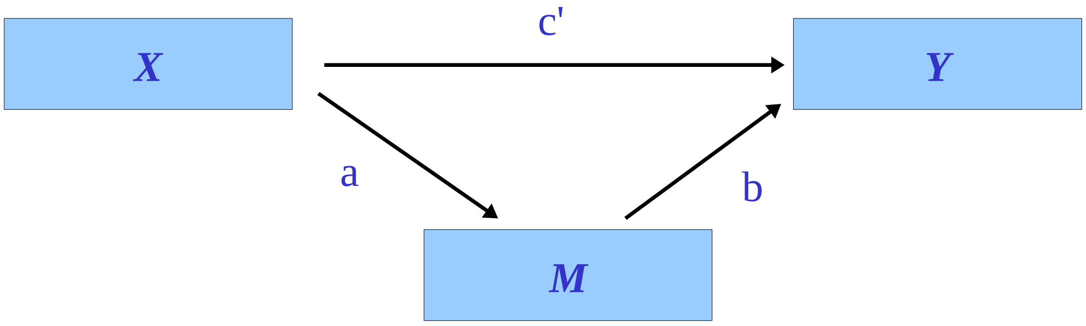
Effetto del mediatore

.597

Effetto mediato

- Dato che abbiamo osservato che gli effetti a e b sono random (cioè variano da soggetto a soggetto) potrebbero covariare (correlare)

- Per stimare la covarianza usiamo bert17, **che usa il medoto boostrap**



$$EM = a \cdot b + cov(a, b)$$

bert17

- Bert17 (per ora) funziona sul modello misto solo per un mediatore, una indipendente ed una indipendente.

Cluster (soggetti)

```
med.mixed("fast","engagment","cstimuli","id",timedata)
```

dipendente

mediatore

indipendente

```
med.mixed(y="fast",mediator = "engagment",x="cstimuli","id",timedata,test = T,BR=500)
```

Queste opzioni chiedono il test bootstrap (può essere molto lento)

Effetto mediato e CI

```
Mediated effect
0.431244
Mediated fixed effect
0.4325846
covariance
-0.00134056

BOOTSTRAP CONFIDENCE INTERVAL CALCULATIONS
Based on 500 bootstrap replicates

CALL :
boot::boot.ci(boot.out = br, type = "perc")

Intervals :
Level     Percentile
95%   ( 0.3561,  0.6961 )
Calculations and Intervals on Original Scale
```

Effetto del mediatore

Intervalli di confidenza