Introduzione alla filosofia dei Linear mixed-effects models

enrico.toffalini@unipd.it

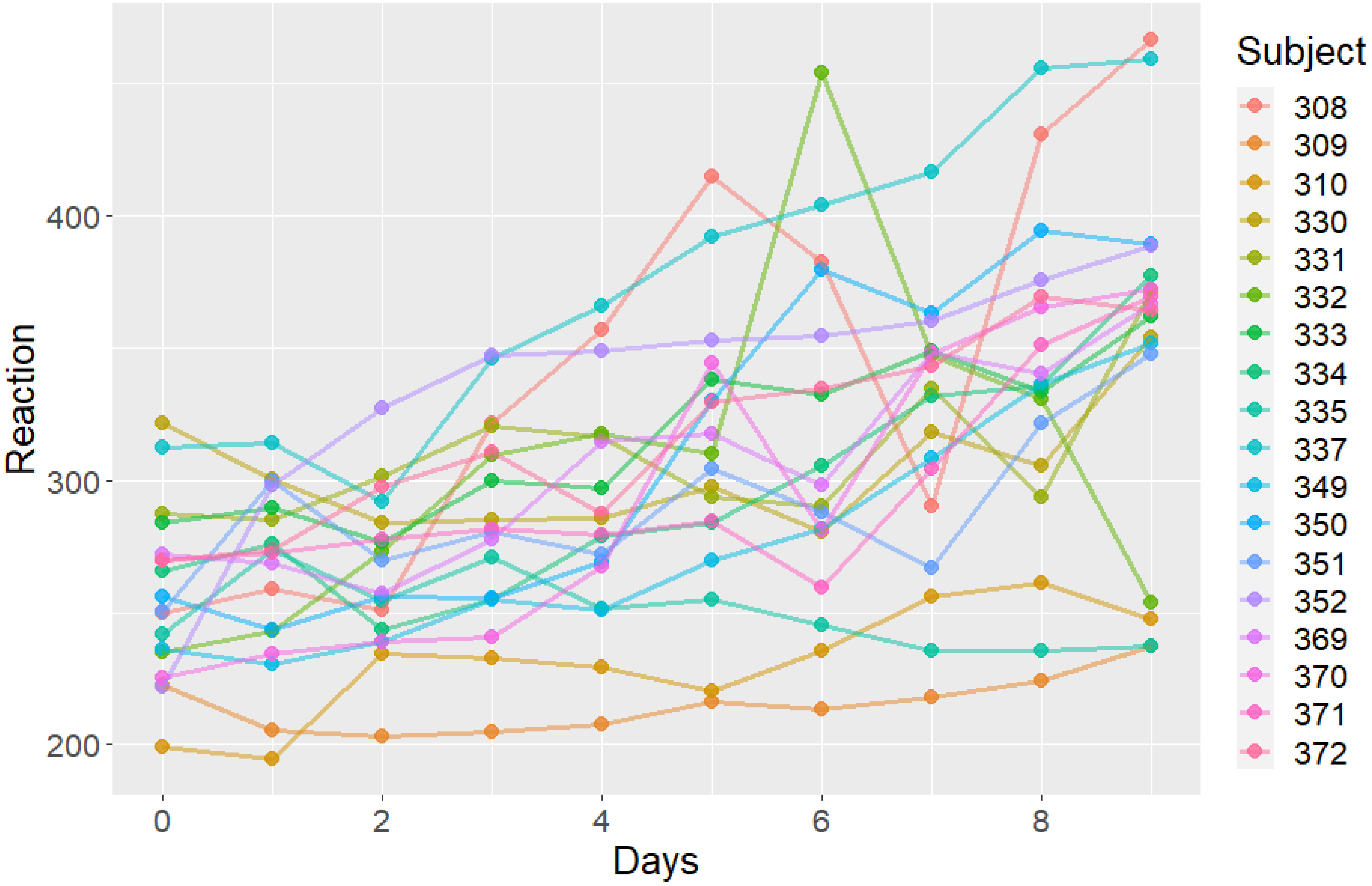
Cosa sono i LMM, in poche parole?

- Sono modelli lineari che includono, oltre agli «effetti fissi» (ovvero effetti che si assumono uguali in tutta la popolazione), degli «effetti random» che sono campionati «casualmente» da una più ampia popolazione
- Gli effetti random corrispondono (di solito) a fattori di raggruppamento delle osservazioni che creano dipendenze locali dei dati
- Esempi di effetti random (ma ci torneremo): partecipanti che forniscono misure ripetute; item di un questionario/test; stimoli di un esperimento; classi, scuole, città in cui sono raggruppati partecipanti
- I casi in cui i partecipanti forniscono numerose risposte (es. sottoposti a numerosi trial o item), o sono raggruppati per cluster locali (es. classi, scuole), è estremamente frequente in psicologia, e solo se applico forzature ai dati (es. lavoro su dati medi per condizione, o fingo che i partecipanti siano campionati in modo indipendente anche se raggruppati) approdo alla situazione della classica ANOVA (sia pure a misure ripetute)

```
> library(lme4)
> d = lme4::sleepstudy
```

```
> ggplot(data=d, aes(x=Days, y=Reaction, group=Subject, color=Subject))+
    geom_point(size=3, alpha=.8) + geom_line(linewidth=1, alpha=.5)+
    scale_x_continuous(breaks=seq(0,10,2)) +  theme(text=element_text(size=18))
```

Vogliamo studiare l'effetto dei giorni di deprivazione di sonno sui tempi di reazione medi. Che analisi possiamo fare?

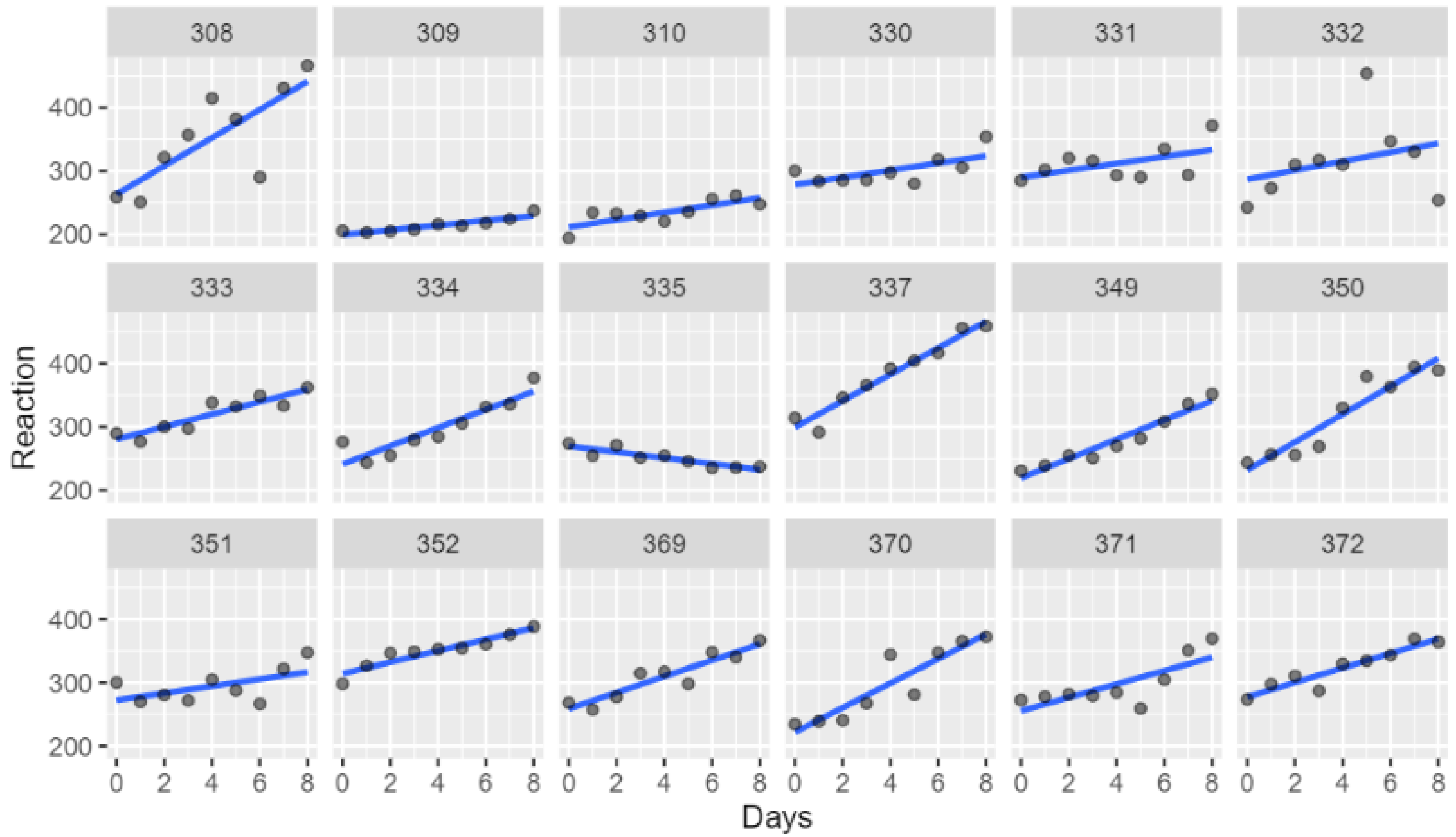


```
> library(lme4)
> d = lme4::sleepstudy
```

da slide del **prof. Altoè (*PsicoStat*)**: https://osf.io/b7tkp/

Vogliamo studiare l'effetto dei giorni di deprivazione di sonno sui tempi di reazione medi.
Che analisi possiamo fare?

```
> ggplot(data=d, aes(x=Days, y=Reaction)) + facet_wrap(~Subject,ncol=6)
+ geom_smooth(method="lm",se=FALSE,formula="y~x") +
+ geom_point(alpha=.5)
```



da slide del **prof. Altoè (*PsicoStat*)**: https://osf.io/b7tkp/

bellissimo corso sui (generalized) linear mixed models con slide dettagliate e un sacco di codice R: assolutamente da vedere e tutto open

```
> library(lme4)
> d = lme4::sleepstudy
```

esempio di informazioni fornite dal mixed-effects model

```
> fit = lmer(Reaction ~ Days + (Days | subject), data = d)
> summary(fit)
Linear mixed model fit by REML. t-tests use Satterthwaite's method
['lmerModLmerTest']
Formula: Reaction ~ Days + (Days | Subject)
   Data: d

REML criterion at convergence: 1743.6

Scaled residuals:
    Min      1Q  Median      3Q     Max
-3.9536 -0.4634  0.0231  0.4634  5.1793

Random effects:
 Groups   Name        Variance Std.Dev. Corr
 Subject  (Intercept) 612.10   24.741
          Days         35.07    5.922   0.07
 Residual             654.94   25.592
Number of obs: 180, groups:  Subject, 18

Fixed effects:
            Estimate Std. Error      df t value Pr(>|t|)
(Intercept)  251.405      6.825  17.000  36.838  < 2e-16 ***
Days          10.467      1.546  17.000   6.771 3.26e-06 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Effetti fissi del modello

Effetti random del modello (con una intercetta random e una slope random di Days nel fattore di raggruppamento Subject)

Deviazione standard dell’intercetta random (cioè SD stimata tra soggetti al giorno 0)

Deviazione standard della slope random (cioè SD stimata tra soggetti dell’effetto dei giorni di deprivazione di sonno sui Reaction time)

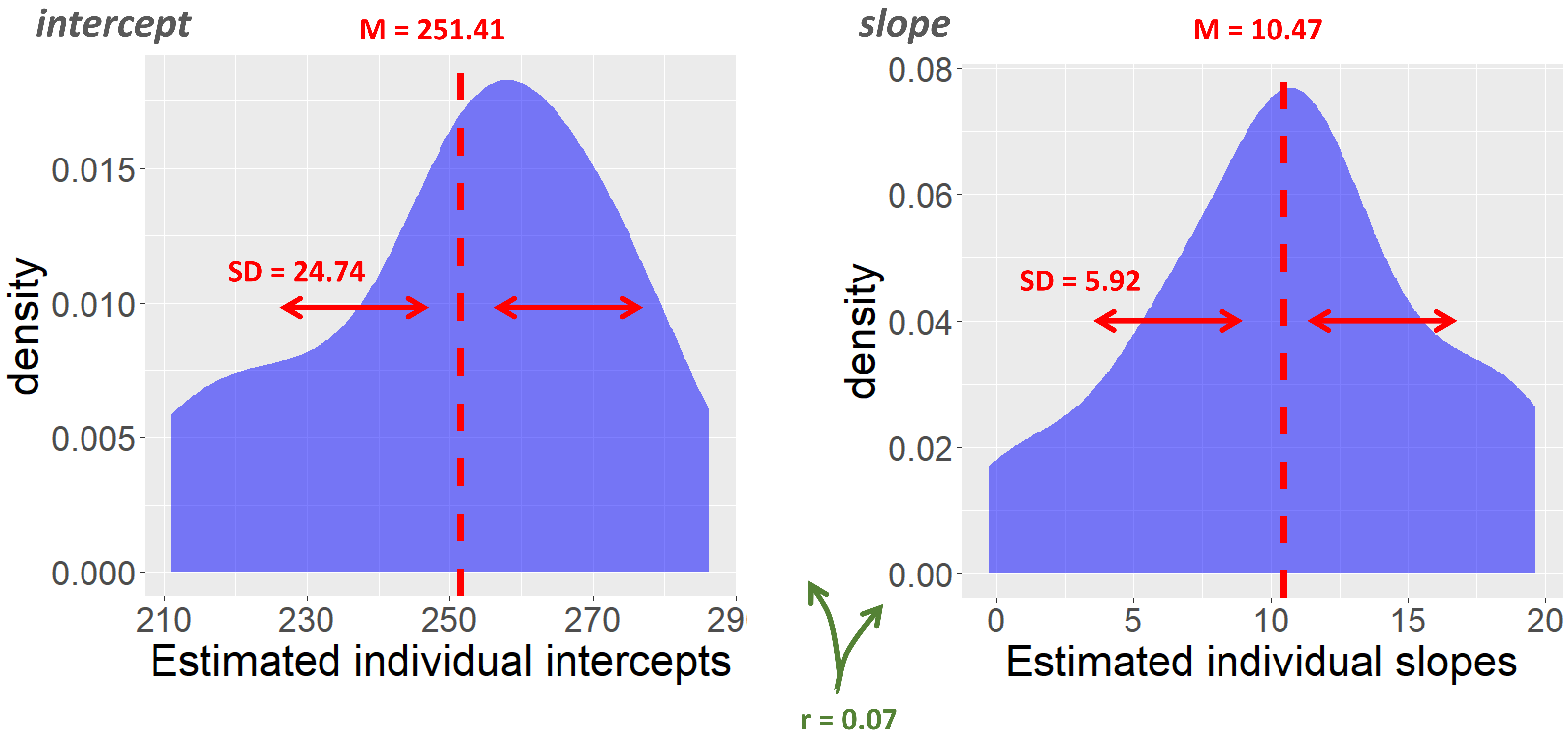
Correlazione tra effetti random: chi ha un’intercetta più alta ha anche una slope più ripida? in positivo? in negativo?

Sigma: SD stimata dei residui (variabilità residua non spiegata dal modello)

Intercetta fissa generale: media stimata generale del campione al giorno 0

Slope fissa generale: effetto medio stimato generale dei giorni di deprivazione di sonno sui Reaction time

Distribuzione degli effetti individuali stimati *(subjects)*



ATTENZIONE: struttura del dataset

da «wide» → a «long»

id	age	Gender	RT.day0	RT.day1	RT.day2
1	19	M	306	315	318
2	22	F	214	235	240
3	25	F	293	307	330
4	20	M	226	230	241

id	age	Gender	Day	RT
1	19	M	0	306
1	19	M	1	315
1	19	M	2	318
2	22	F	0	214
2	22	F	1	235
2	22	F	2	240
3	25	F	0	293
3	25	F	1	307
3	25	F	2	330
4	20	M	0	226
4	20	M	1	230
4	20	M	2	241

- Il formato «wide» è considerato tradizionale da alcuni perché associato alle ANOVA fatte in SPSS o altri software
- Il formato «long» di solito è il default degli output dei software per programmare esperimenti (es. Eprime, Matlab) (ho visto trasformare «long» in «wide»… per poi doverlo ritrasformare in «long» quando si capisce che è meglio fare mixed models!)
- Il formato «long» fa capire immediatamente qual è la variabile dipendente, mettendola in un'unica colonna anziché spezzarla su colonne diverse
- Un altro vantaggio del formato «long» è che possiamo associare le osservazioni ripetute a specifiche proprietà (es. un flag su un evento occorso in un dato giorno a un dato soggetto, o una proprietà di un item), che poi possono molto facilmente diventare delle covariate
- A proposito: un dato mancante NON è un problema nel LMM come lo è nell'ANOVA

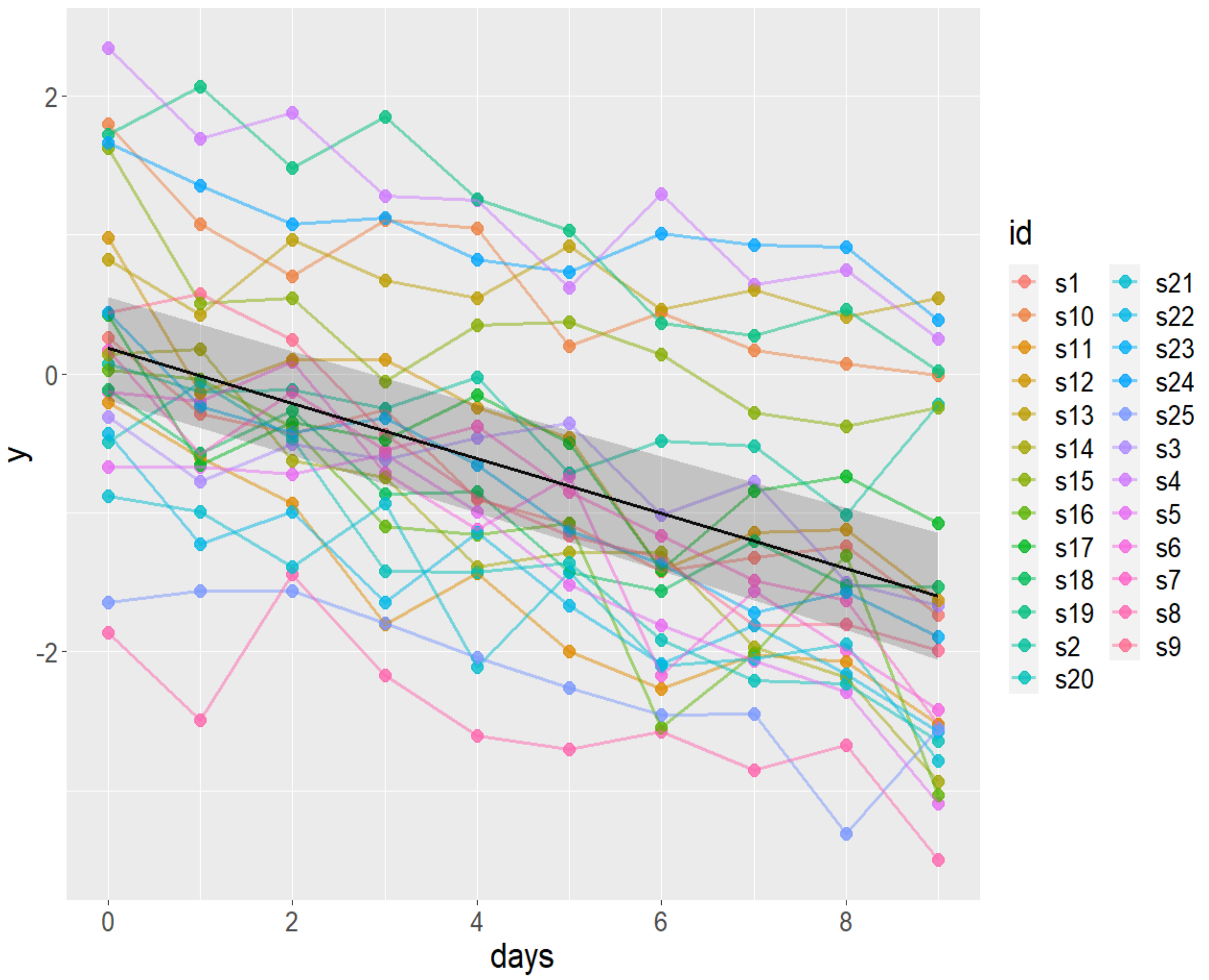
Altro esempio (simulato)

```
> fit4 = lmer(y ~ days + (days|id), data=d)
```

```
> summary(fit4)
Linear mixed model fit by REML. t-tests use Satterthwaite's
method ['lmerModLmerTest']
Formula: y ~ days + (days | id)
   Data: d

REML criterion at convergence: 270.1

Scaled residuals:
     Min       1Q   Median       3Q      Max
-2.88873 -0.50733  0.02217  0.62049  3.08739

Random effects:
 Groups   Name        Variance Std.Dev. Corr
 id       (Intercept) 0.861566 0.92821
          days        0.004585 0.06771  0.07
 Residual             0.090188 0.30031
Number of obs: 250, groups:  id, 25

Fixed effects:
            Estimate Std. Error       df t value Pr(>|t|)
(Intercept)  0.18135    0.18897 23.99947    0.96    0.347
days        -0.19808    0.01507 23.99994  -13.14 1.86e-12 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Cosa significano questi coefficienti?

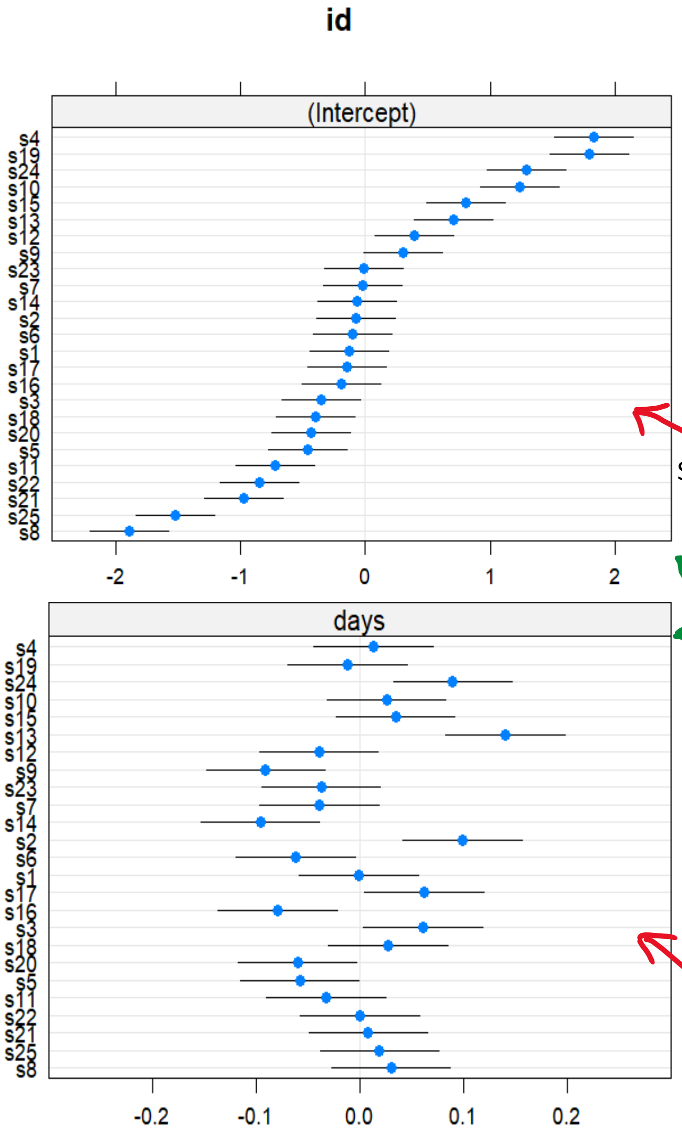
(y è una variabile dipendente di prestazione misurata in punti z)



Altro esempio (simulato)

```
> lattice::dotplot(ranef(fit4))
```

analisi dei random effect dei soggetti

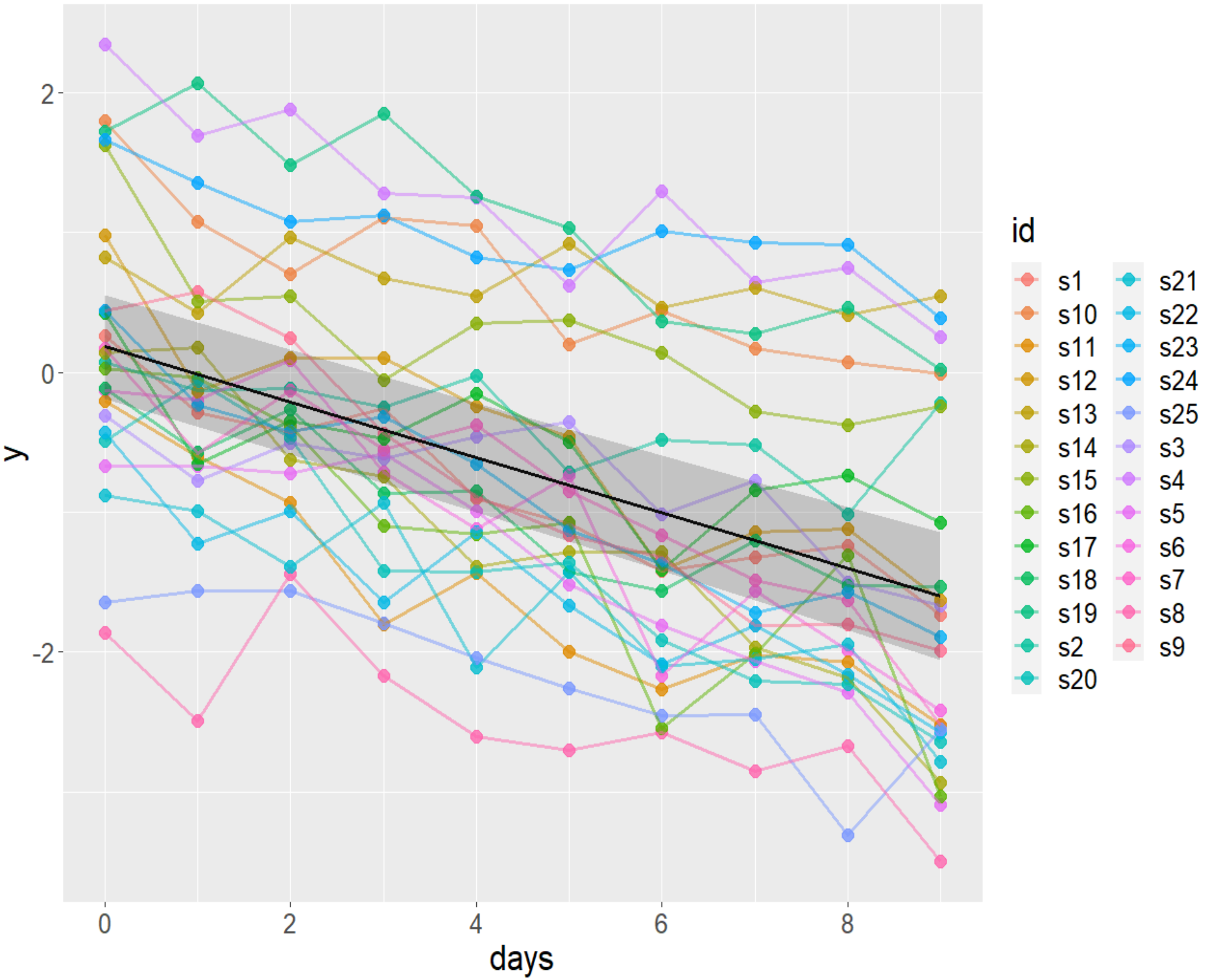


SD = 0.9282

corr = 0.07

SD = 0.0677

(y è una variabile dipendente di prestazione misurata in punti z)



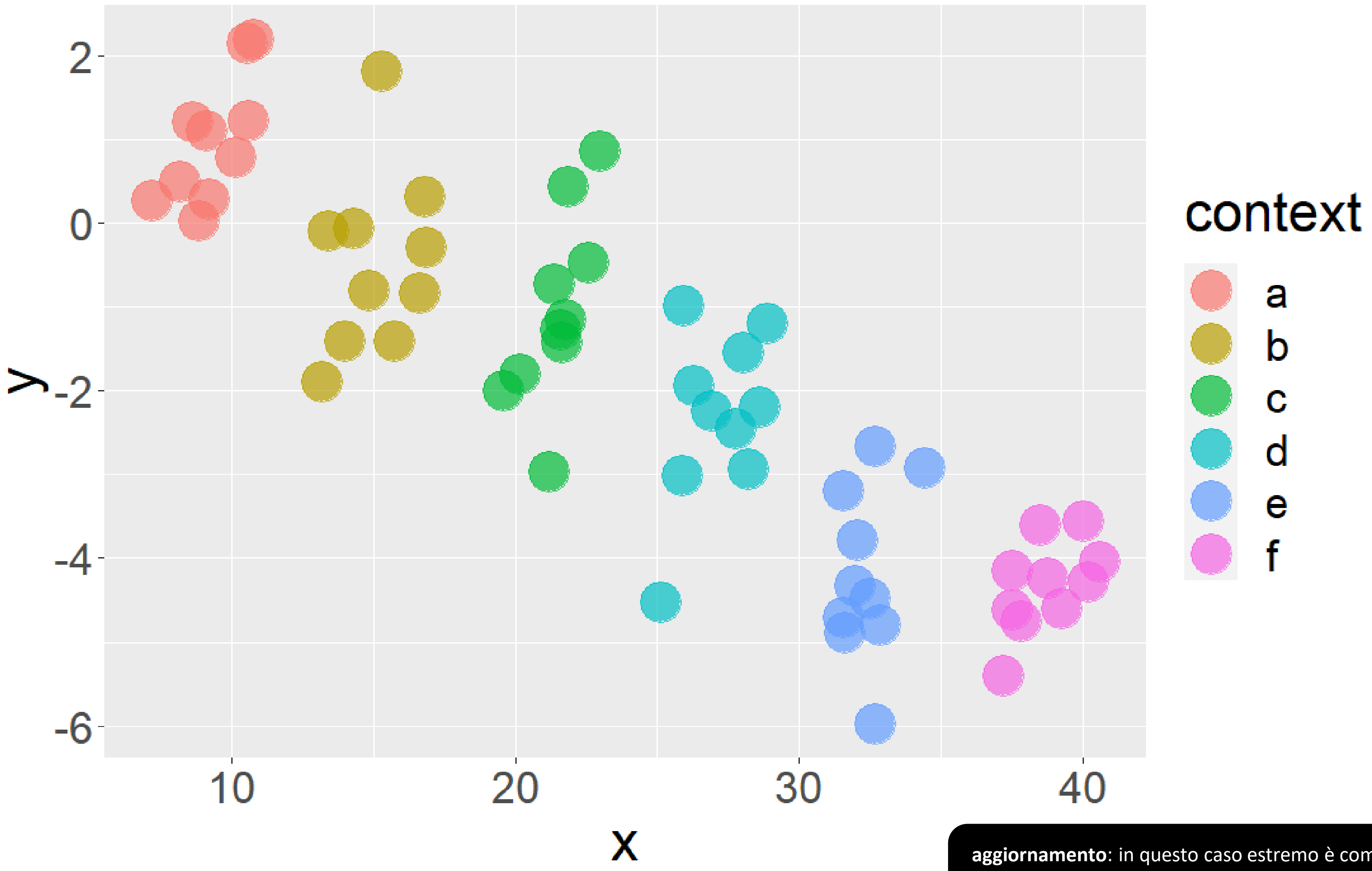
Altro esempio (simulato)

MODEL COMPARISON

```
> # no effects at all
> fit0 = lm(y ~ 1, data=d)
>
> # days has a general fixed effect, which is the same for all subjects, and subjects have no variability in mean scores
> fit1 = lm(y ~ days, data=d)
>
> # days has a general fixed effect, which is the same for all subjects, but subjects still have different mean levels (random intercepts)
> fit2 = lmer(y ~ days + (1|id), data=d)
>
> # days has no general fixed effect, but just has a different unpredictable effect on each subject
> fit3 = lmer(y ~ 1 + (days|id), data=d)
>
> # days has a general fixed effect, and this effect is also heterogeneous across subjects (REAL MODEL)
> fit4 = lmer(y ~ days + (days|id), data=d)
>
> # model comparison
> anova(fit4,fit3,fit2,fit1,fit0)
refitting model(s) with ML (instead of REML)
Data: d
Models:
fit0: y ~ 1
fit1: y ~ days
fit2: y ~ 1 + (days | id)
fit3: y ~ days + (1 | id)
fit4: y ~ days + (days | id)
     npar    AIC    BIC  logLik deviance   Chisq Df Pr(>Chisq)
fit0    2 798.36 805.40 -397.18   794.36
fit1    3 734.85 745.41 -364.43   728.85  65.505  1  5.797e-16 ***
fit2    4 315.29 329.38 -153.64   307.29 421.561  1  < 2.2e-16 ***
fit3    5 324.62 342.23 -157.31   314.62   0.000  1          1      ← fit2 NON è nested in fit3: il likelihood ratio test non si può fare
fit4    6 274.02 295.15 -131.01   262.02  52.596  1  4.096e-13 ***          ma possiamo comunque fare confronti con AIC e BIC
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Paradossi di Simpson

Nel caso precedente, il coefficiente dell'effetto fisso del semplice lm e dell'lmer con random intercept e random slope erano uguali, ma nel secondo caso avevo maggiore precisione (minore Std.Err.). Tuttavia, non è sempre detto che sia così



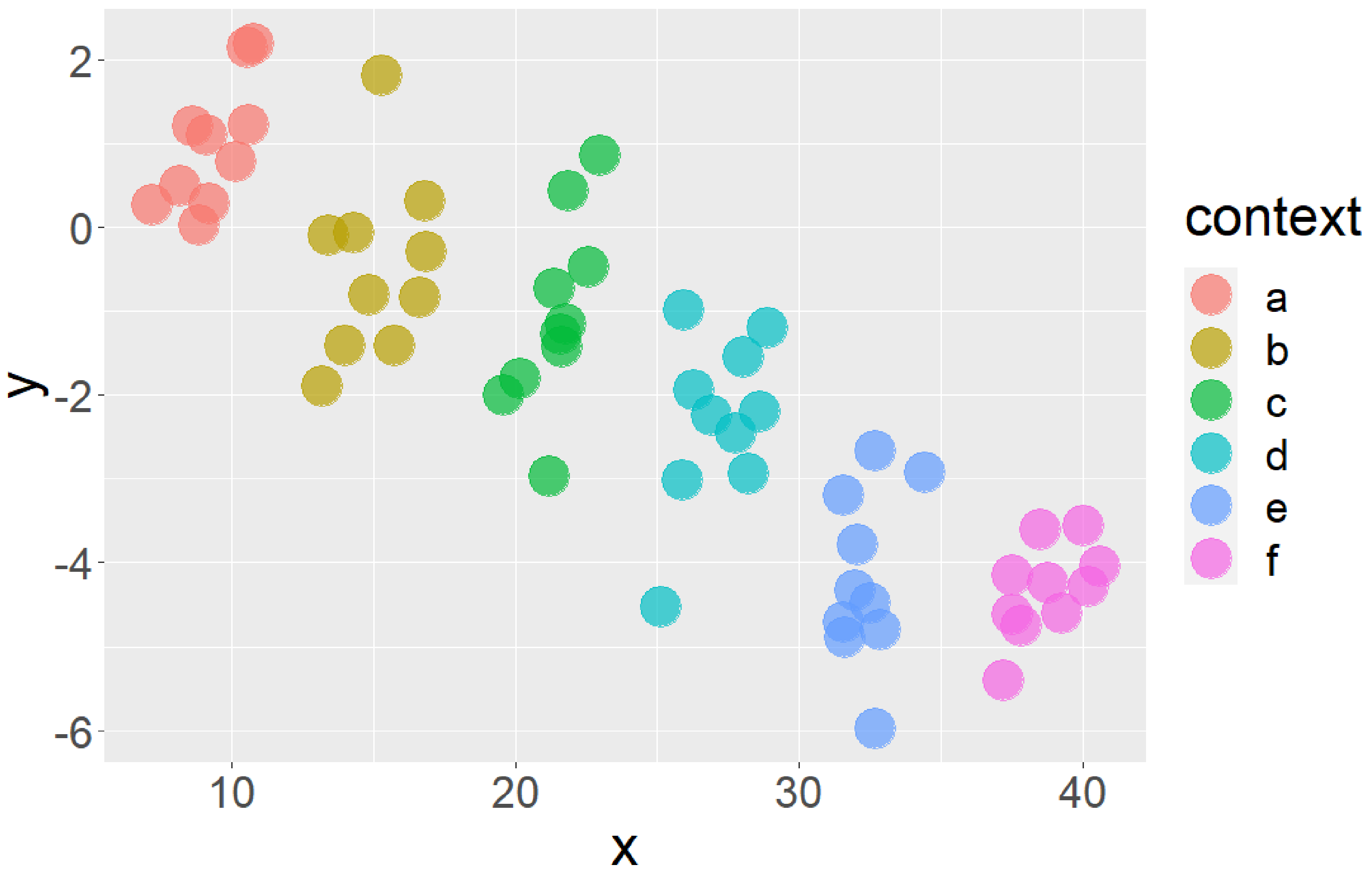
Qui la relazione y ~ x è positiva o negativa?

- <u>Negativa</u> secondo un semplice lm:
  lm(y ~ x, data=d)
  B = –0.18, p < .001
- <u>Positiva</u> se inserisco almeno un'intercetta random per il fattore di raggruppamento
  lmer(y ~ x + (1|context), data=d)
  B = +0.28, p < .001

aggiornamento: in questo caso estremo è comunque violata l'assunzione di ortogonalità tra regressore ed effetti random: le intercette random sono correlate ai valori di x; una soluzione ottimale può essere aggiungere nel modello, come predittore, il valore medio di x del «context»: http://www.stat.columbia.edu/~gelman/research/unpublished/Bafumi_Gelman_Midwest06.pdf

Paradossi di Simpson

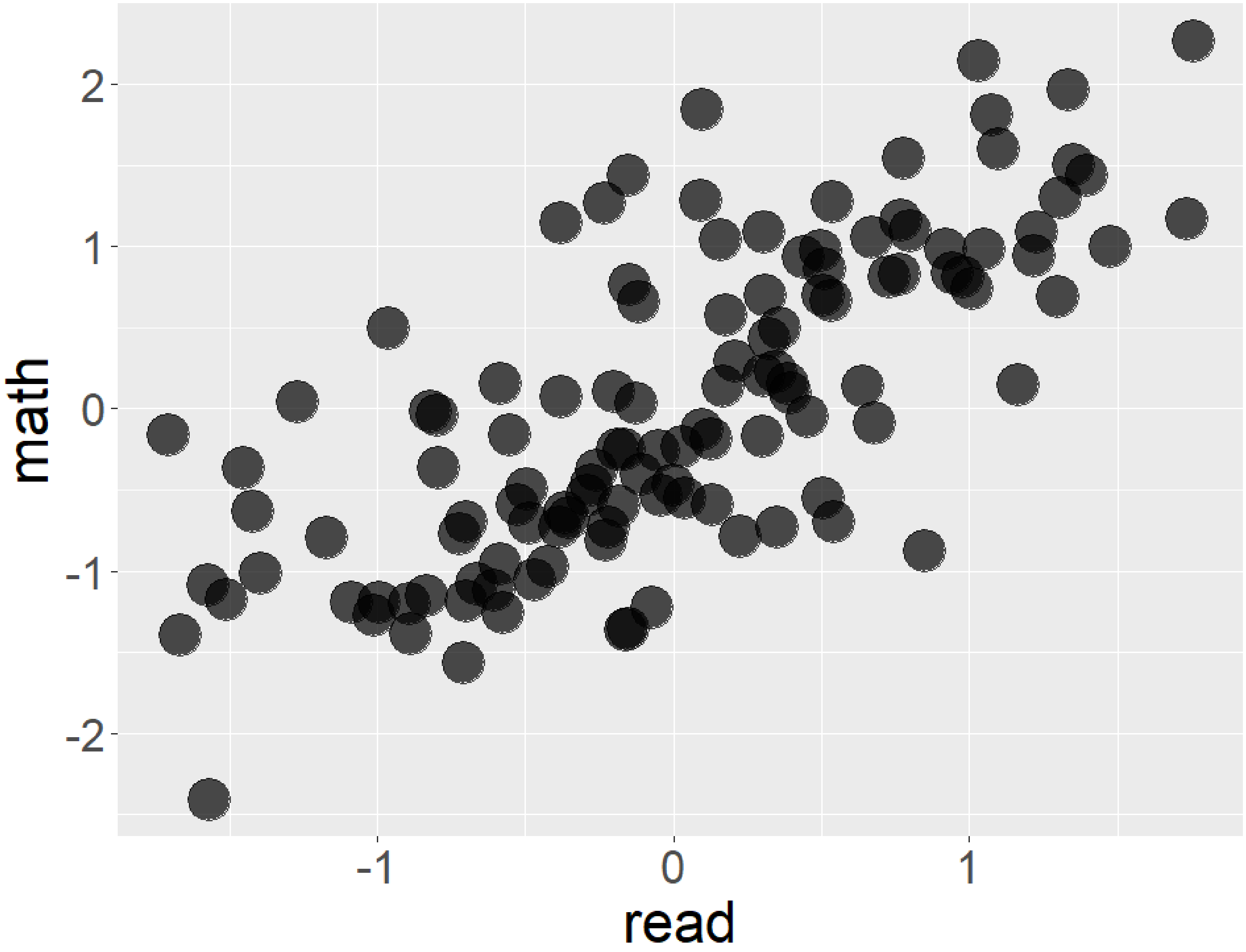
Ma qui una soluzione vincente per analizzare entrambi i livelli potrebbe essere un SEM multilivello fittato in lavaan:




```
> library(lavaan)
> m  = "
      Level: 1
      y ~ x
      Level: 2
      y ~ x
"
> fit = sem(m, data=d, cluster="context")
> summary(fit)

[...]

Level 1 [within]:

Regressions:
                   Estimate  Std.Err  z-value  P(>|z|)
  y ~
    x                 0.377    0.097    3.880    0.000

[...]

Level 2 [context]:

Regressions:
                   Estimate  Std.Err  z-value  P(>|z|)
  y ~
    x                -0.190    0.015  -12.322    0.000

[...]
```

Un caso verosimile : Bambini nelle classi

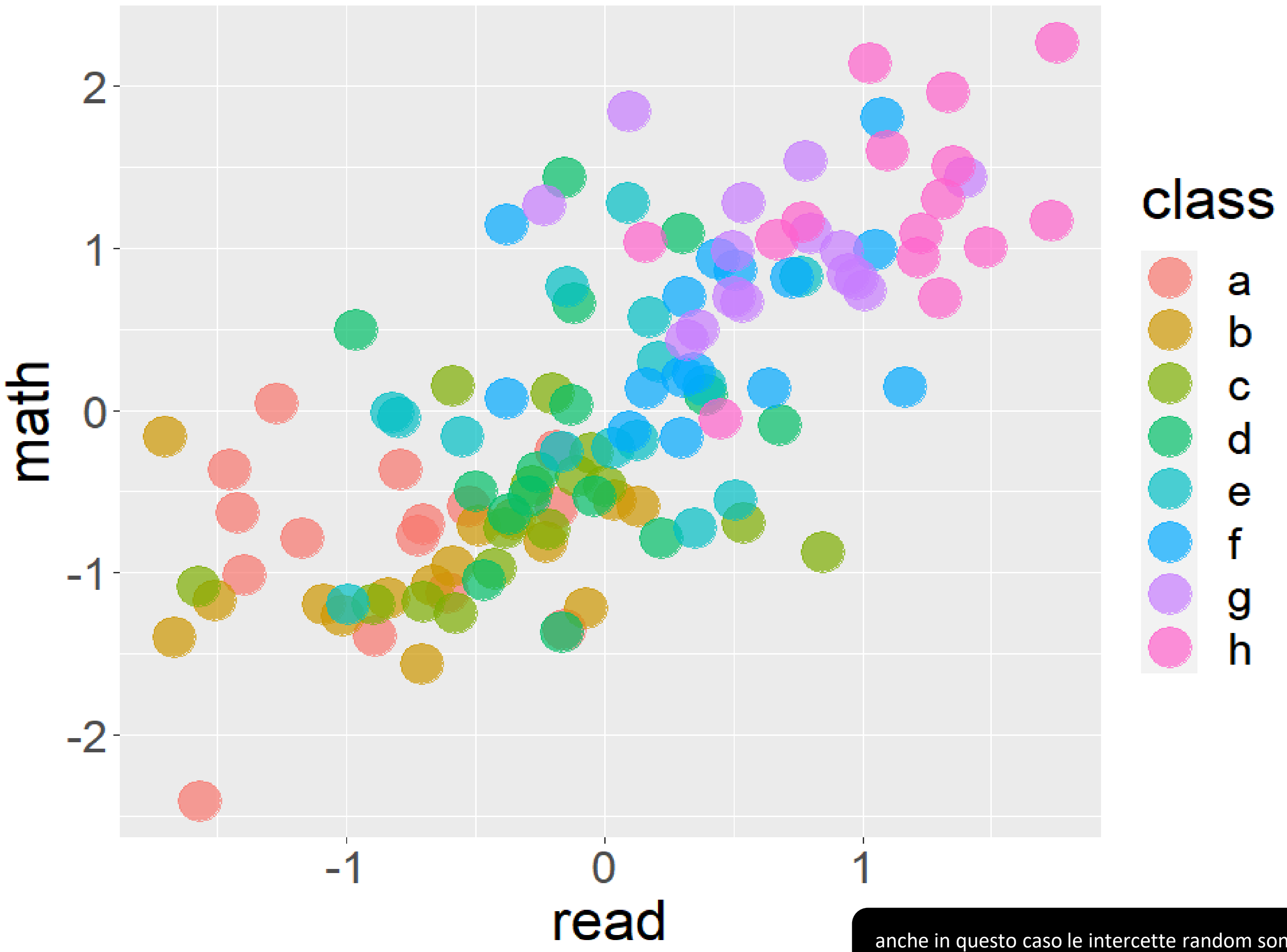
Supponiamo di voler studiare l'effetto dell'abilità di lettura sulla capacità dei bambini di risolvere problemi matematici scritti (*math* e *read* sono parzializzati per il fattore g). Raccogliamo i nostri N = 120 bambini e questa è la relazione.




```
> fit0 = lm(math ~ read, data=d)
> summary(fit0)

Call:
lm(formula = math ~ read, data = d)

Residuals:
     Min       1Q   Median       3Q      Max 
-1.63367 -0.44389 -0.08438  0.40325  1.74312 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)  0.01982    0.06067   0.327    0.745    
read         0.87312    0.07650  11.413   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 
‘ ’ 1

Residual standard error: 0.6646 on 118 degrees of freedom
Multiple R-squared:  0.5247,	Adjusted R-squared:  0.5206 
F-statistic: 130.3 on 1 and 118 DF,  p-value: < 2.2e-16
```

Un caso verosimile : Bambini nelle classi

Poi il revisore chiede di controllare il raggruppamento dei bambini per classe (15 bambini x 8 classi) … e l'effetto cambia!
Nelle classi dove *read* è insegnato meglio, anche *math* è insegnato meglio, ma noi vogliamo vedere l'effetto individuale



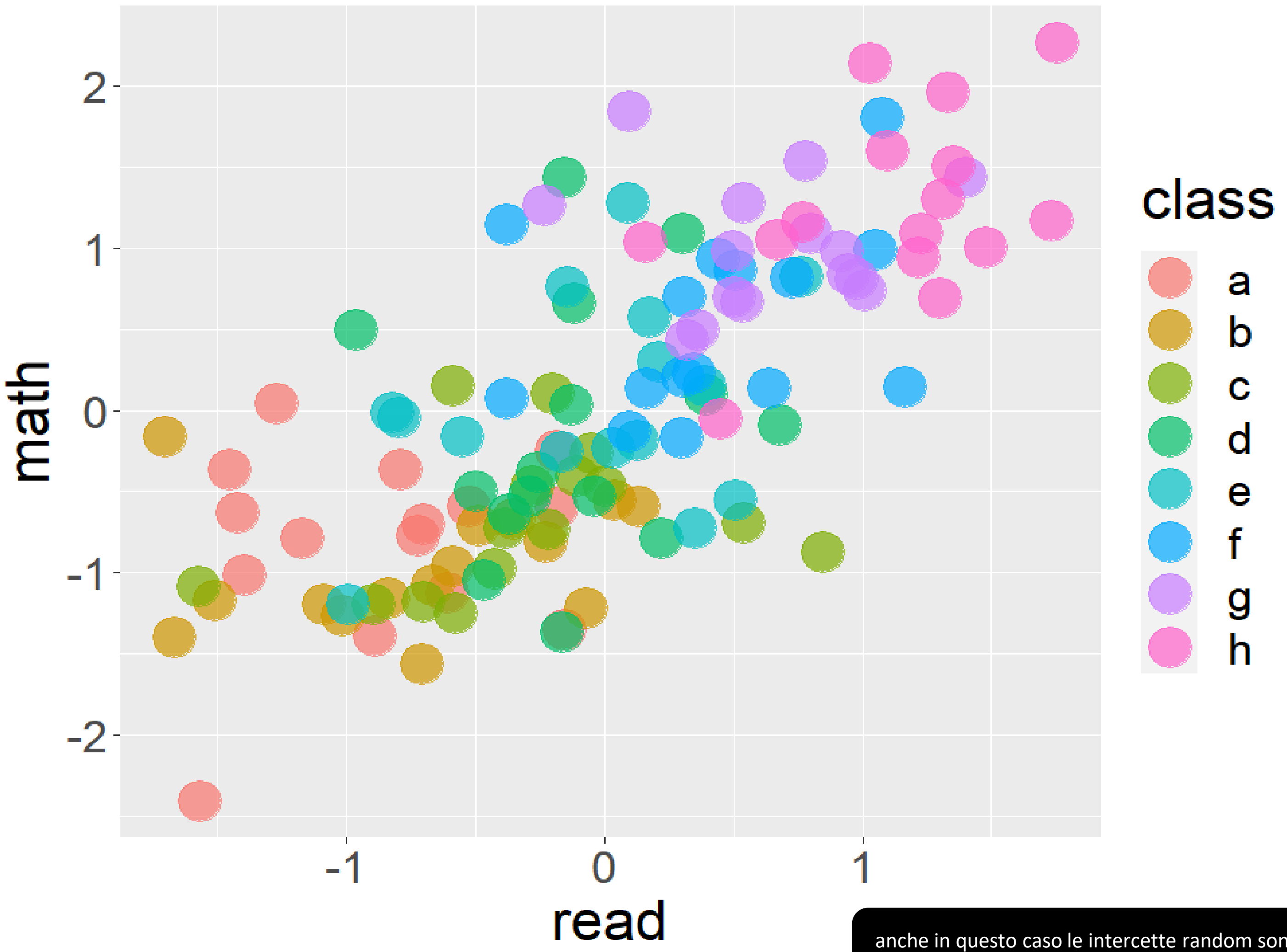

```
> fit1 = lmer(math ~ read + (1|class), data=d)
> summary(fit1)

Linear mixed model fit by REML. t-tests use Satterthwaite's
method [lmerModLmerTest]
Formula: math ~ read + (1 | class)
   Data: d

REML criterion at convergence: 222.5

Scaled residuals:
    Min      1Q  Median      3Q     Max
-2.4809 -0.5700 -0.0681  0.4344  2.8667

Random effects:
 Groups   Name        Variance Std.Dev.
 class    (Intercept) 0.3516   0.5929
 Residual             0.3030   0.5504
Number of obs: 120, groups:  class, 8

Fixed effects:
            Estimate Std. Error        df t value Pr(>|t|)
(Intercept)  0.02614    0.21557   5.87126   0.121 0.907541
read         0.35720    0.10098 115.77060   3.537 0.000583 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

e dov'è l'R^2?!

anche in questo caso le intercette random sono correlate ai valori di «read» e una soluzione ottimale può essere aggiungere nel modello come predittore il valore medio di «read» della classe: http://www.stat.columbia.edu/~gelman/research/unpublished/Bafumi_Gelman_Midwest06.pdf

Un caso verosimile : Bambini nelle classi

e dov'è l'R^2?




```
> library(MuMIn)
> MuMIn::r.squaredGLMM(fit1)

           R2m       R2c
[1,] 0.1100071 0.5880173
```

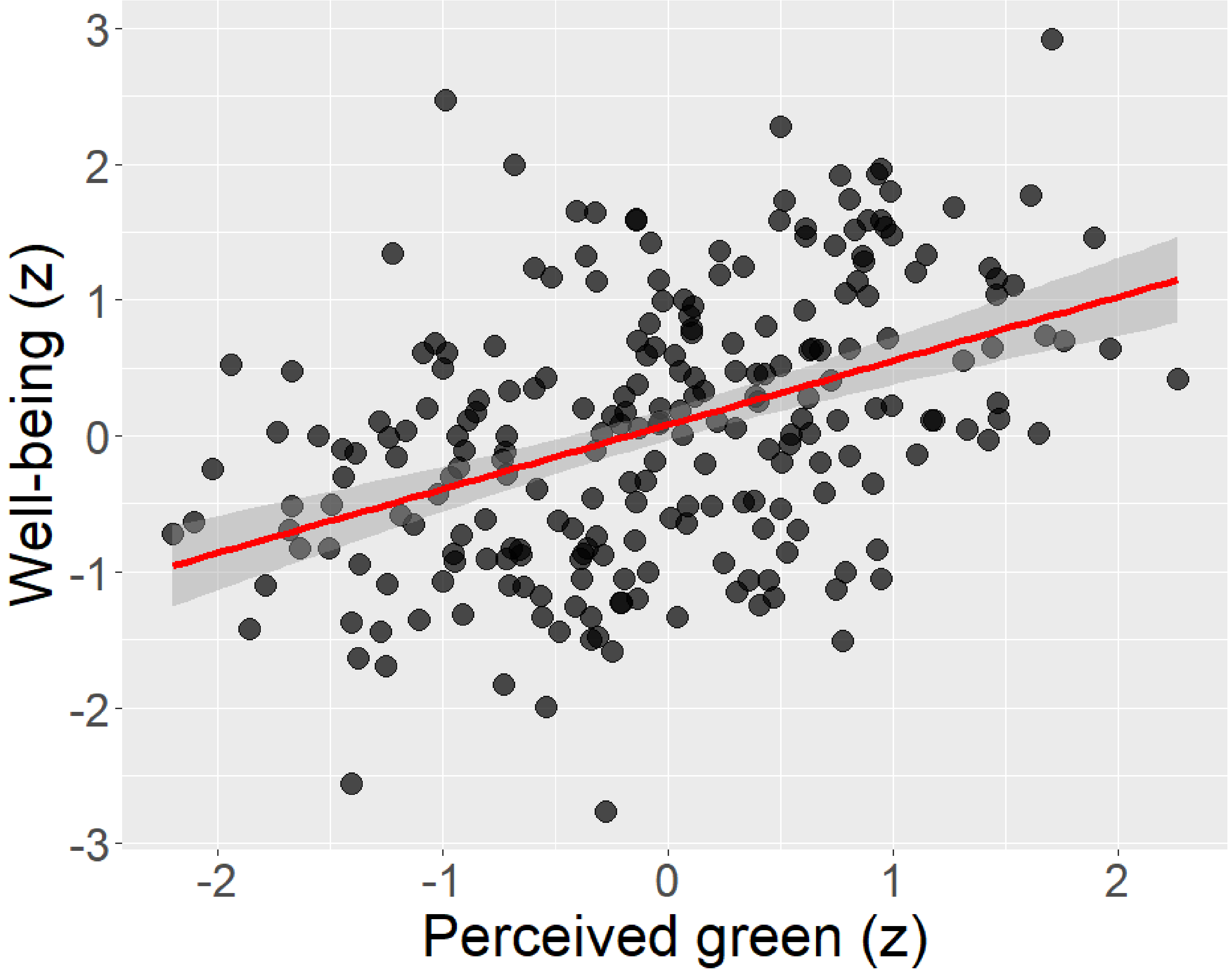
gli R^2 stimati sono due:

- R^2 marginale (R2m) è la proporzione di varianza spiegata dagli *effetti fissi* rispetto alla *varianza totale*
- R^2 condizionato (R2c) è la proporzione di varianza spiegata dagli *effetti fissi + random* (cioè tutta la varianza che il modello riesce a spiegare) rispetto alla *varianza totale*

anche in questo caso le intercette random sono correlate ai valori di «read» e una soluzione ottimale può essere aggiungere nel modello come predittore il valore medio di «read» della classe: http://www.stat.columbia.edu/~gelman/research/unpublished/Bafumi_Gelman_Midwest06.pdf

Un altro caso verosimile : Persone nelle città

La percezione del verde nel proprio ambiente correla con un maggior benessere riportato?

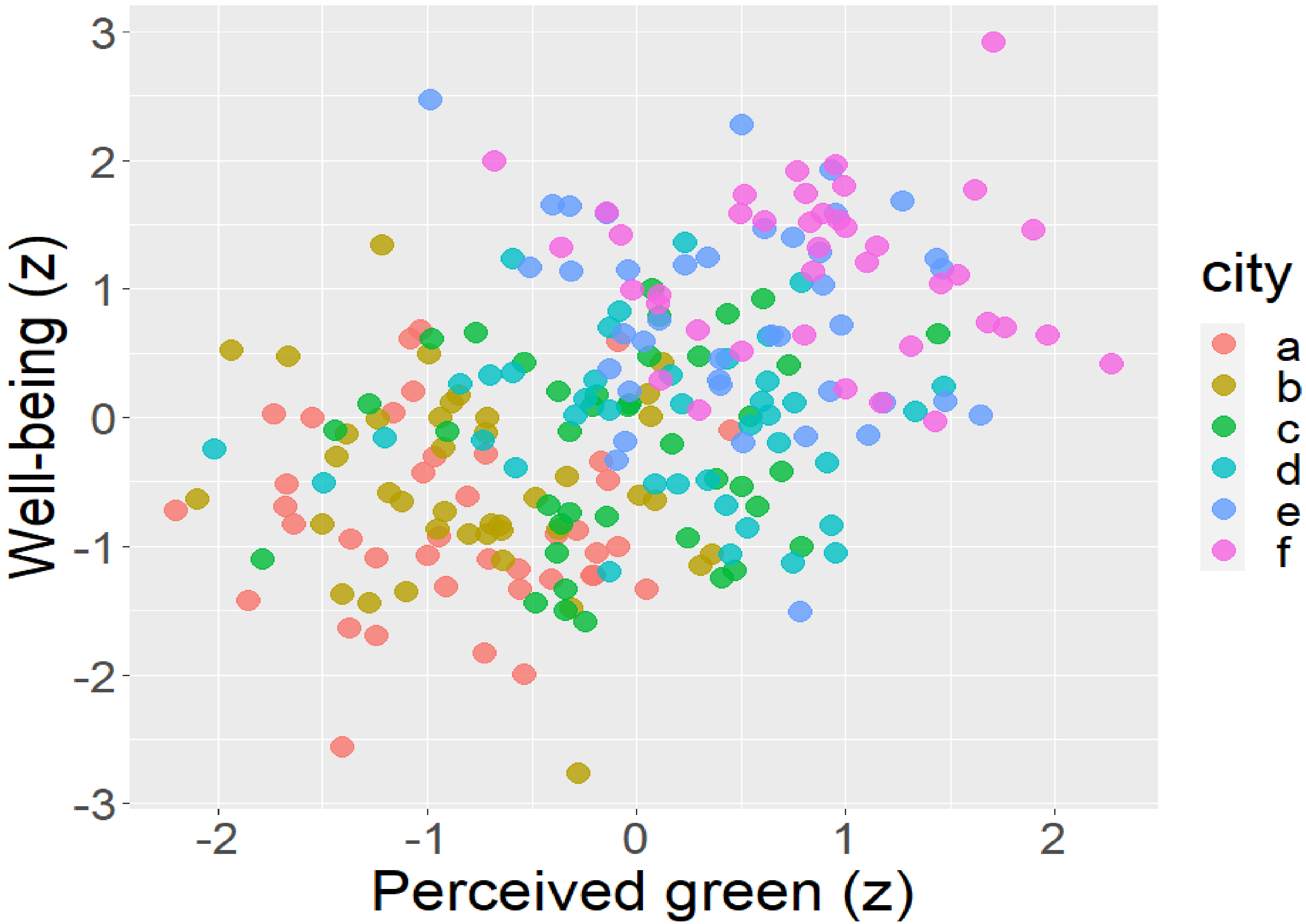



```
> summary(lm(wb~pg,data=d))
 Call:
lm(formula = wb ~ pg, data = d)

Residuals:
     Min       1Q   Median       3Q      Max 
-2.71471 -0.66516  0.03043  0.60281  2.85218 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)  0.08106    0.05827   1.391    0.165    
pg           0.47139    0.06477   7.278 4.91e-12 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 
‘ ’ 1

Residual standard error: 0.9016 on 238 degrees of freedom
Multiple R-squared:  0.182,	Adjusted R-squared:  0.1786 
F-statistic: 52.96 on 1 and 238 DF,  p-value: 4.911e-12
```

Sembra di sì, MA…

Un altro caso verosimile : Persone nelle città

La percezione del verde nel proprio ambiente correla con un maggior benessere riportato?

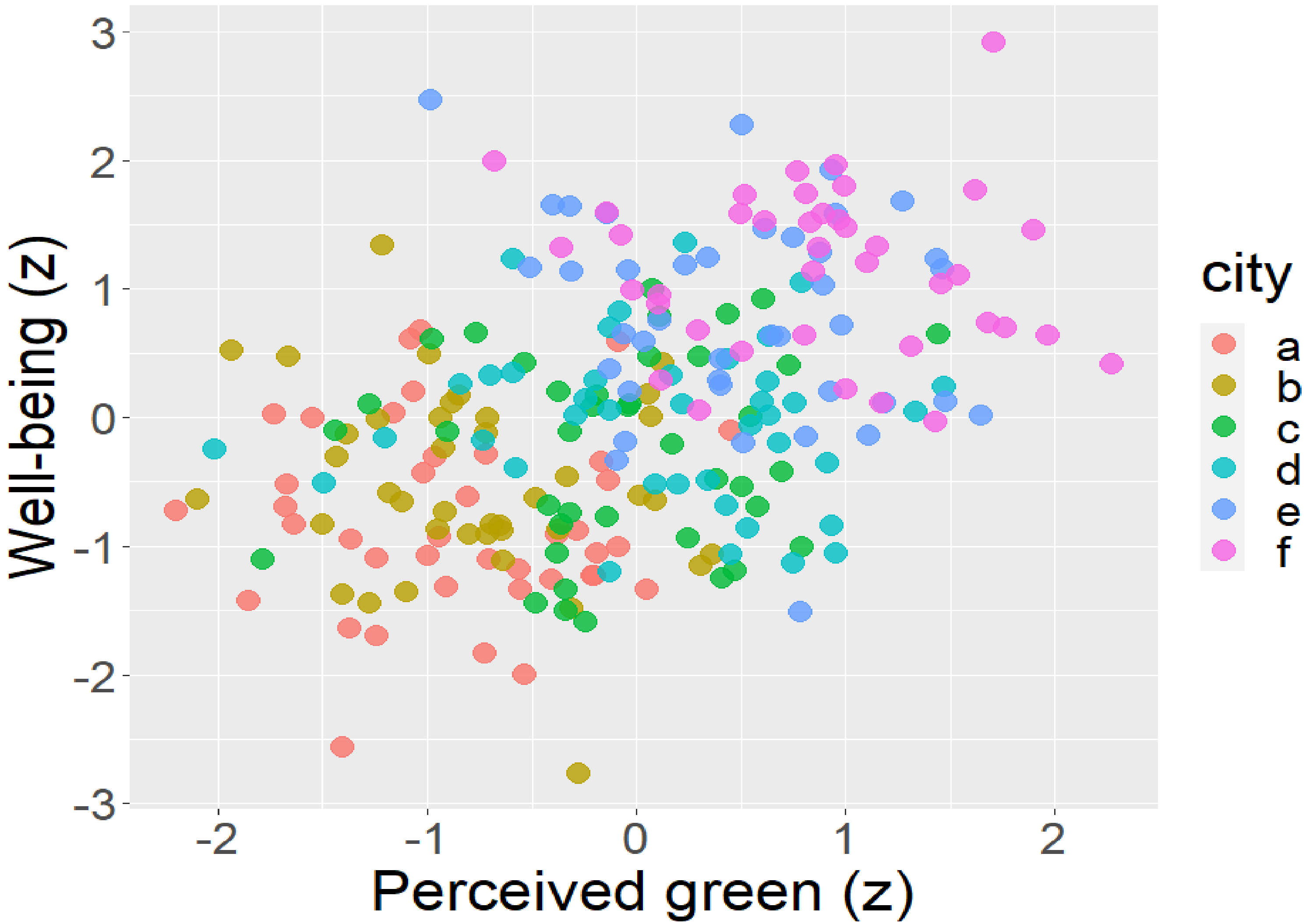


Abbiamo raccolto i dati in sole 8 città. Dopo aver controllato per le differenze medie tra città in *perceived green*, l'effetto non c'è più

```
> summary(lmer(wb~pg+(1|city),data=d))
Linear mixed model fit by REML. t-tests use
Satterthwaite's method [lmerModLmerTest]
Formula: wb ~ pg + (1 | city)
   Data: d

REML criterion at convergence: 544.3

Scaled residuals:
    Min      1Q  Median      3Q     Max
-3.1861 -0.6810 -0.0193  0.6537  2.5842

Random effects:
 Groups   Name        Variance Std.Dev.
 city     (Intercept) 0.6493   0.8058
 Residual             0.5084   0.7130
Number of obs: 240, groups:  city, 6

Fixed effects:
            Estimate Std. Error        df t value Pr(>|t|)
(Intercept)  0.05658    0.33217   4.83720    0.17    0.872
pg          -0.07064    0.07066 237.87250   -1.00    0.318
```

Un altro caso verosimile : Persone nelle città

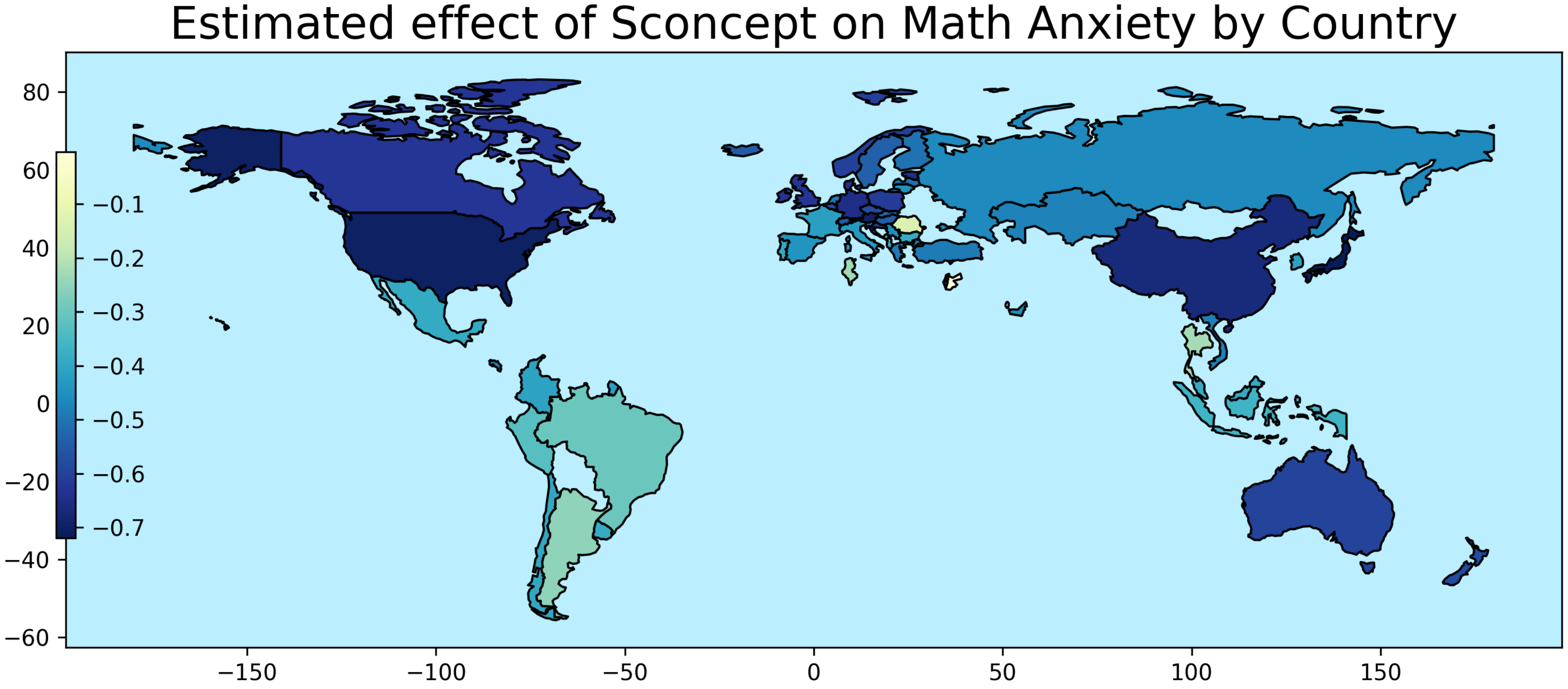
Anche qui, per avere entrambi i livelli potrebbe essere vincente un multilevel-SEM :




```
> library(lavaan)
> m  = "
      Level: 1
      wb ~ pg
      Level: 2
      wb ~ pg
"
> fit = sem(m, data=d, cluster="city")
> summary(fit)

[...]

Level 1 [within]:

Regressions:
          Estimate  Std.Err  z-value  P(>|z|)
  wb ~
    pg      -0.091    0.071   -1.282    0.200

[...]

Level 2 [city]:

Regressions:
          Estimate  Std.Err  z-value  P(>|z|)
  wb ~
    pg       1.105    0.134    8.226    0.000

[...]
```

Un caso reale: Eterogeneità dell'effetto tra nazioni

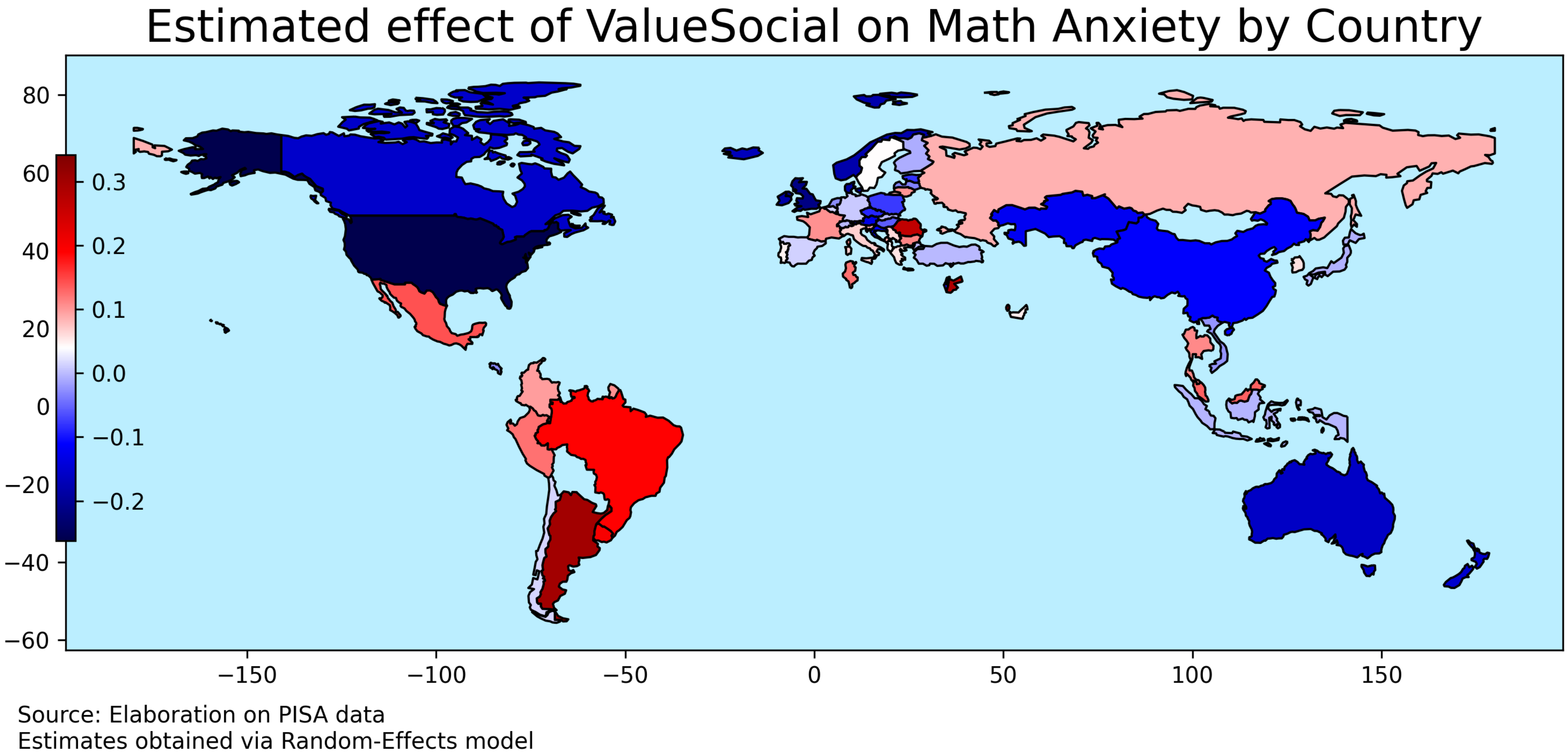
(Senza aggiungere covariate) c'è un forte effetto fisso significativo del Self concept sulla propria Ansia Matematica ($B = -0.480$ $p < 0.001$), con N = 151.745 osservazioni in k = 63 nazioni, tuttavia…



Source: Elaboration on PISA data
Estimates obtained via Random-Effects model

Un caso reale: Eterogeneità dell’effetto tra nazioni

(Senza aggiungere covariate) qui invece NON troviamo un effetto fisso significativo del Valore Sociale percepito della matematica sull’Ansia Matematica ($B = 0.009$, $p = 0.609$), nonostante $N = 151.745$ osservazioni in $k = 63$ nazioni, tuttavia…



Ma cosa vuol dire «effetti random»?

Nella nostra ricerca possiamo vedere gli effetti random come qualsiasi fattore di raggruppamento delle osservazioni che è stato campionato «casualmente» da una più ampia popolazione, esempi:

- i singoli **partecipanti**, laddove forniscano più di una risposta/osservazione (misure ripetute)
- gli **item** di un test, laddove per ciascun item ci sono risposte di molti partecipanti
- fattori di raggruppamento di partecipanti, ad esempio **classi**, **scuole**, **province**, **nazioni**, ecc., se campionati casualmente

NON sono effetti random:

- fattori di raggruppamenti di partecipanti quali **classi, scuole, province, nazioni** ecc. se selezionati a priori per specifici motivi, generalmente in numero limitato, per essere confrontati tra loro
- **raggruppamenti di item** se fissati a priori per determinati motivi e per essere confrontati tra loro (es. item verbali vs item non-verbali)

La linea di confine tra effetto «fisso» e «random» in alcuni (rari) casi è sfumata…

---

https://statmodeling.stat.columbia.edu/2005/01/25/why_i_dont_use/

Why I don't use the term "fixed and random effects"

Posted on January 25, 2005 6:41 PM by Andrew

se siete molto nerd potreste lasciar perdere le definizioni «random» e «fixed» e parlare sempre solo di multilevel modeling

Crossed vs Nested random effects

Quando ho più effetti random, posso averli «crossed» o «nested», o anche entrambe le situazioni assieme (per ora non approfondiamo)

- Se ciascun caso di un effetto random si associa a diversi casi dell’altro effetto random e viceversa, allora gli effetti random sono **incrociati** (*crossed*). Ad esempio **partecipanti** e **item**: ciascun partecipante risponde a molti item, e ciascun item viene affrontato da molti partecipanti

- Se ciascun caso di un effetto random si può presentare solo all’interno di un particolare caso di un altro effetto random sovraordinato, allora gli effetti random sono annidati (*nested*). Ad esempio, nei dati INVALSI ogni partecipante appartiene a una **classe**, che appartiene a una **scuola**, che appartiene a una **provincia**, che appartiene a una **regione**, e tutti questi livelli rappresentano fattori di raggruppamento con una variabilità random

Random effect e potenza statistica… abbiamo già un'idea intuitiva

Supponiamo di voler vedere se persone nate in giorni dispari siano più belle di persone nate in giorni pari. Portiamo partecipanti in Lab e chiediamo di valutare fotografie di persone nate in giorni pari vs dispari.

- Abbiamo N = 10.000 partecipanti (sample size): bene ✓
- **MA** ci serve anche un ampio numero k di stimoli (es. foto volti) di uomini nati in giorni pari vs dispari

→ Cosa succede se il *campione di stimoli* è troppo ridotto?

Nati giorni pari

Nati giorni dispari

Random effect e potenza statistica… abbiamo già un’idea intuitiva

Ciascun partecipante (N = 10.000) valuta *tutti* gli stimoli/volti (condizione «pari vs dispari» è *within-participant*). **Che analisi faccio?**

Potrei calcolare la media di ciascun rispondente nei 2 livelli (*pari* vs *dispari*) e fare l’ANOVA a misure ripetute? Certo! Però perdo del tutto:

- **L’incertezza legata al limitato campione di stimoli: <u>questo DISTRUGGE l’inferenza</u>**
- Stime e variabilità dei singoli volti (alcuni stimoli/volti sono mediamente più belli, altri meno, indipendentemente dalla condizione «pari vs dispari»)
- Stima delle tendenze individuali verso le condizioni (a qualcuno potrebbero piacere di più i nati nei giorni pari, a qualcun altro i nati nei giorni dispari, per qualcun altro è lo stesso)
- *Risposte idiosincratiche dei soggetti agli stimoli (qualche rispondente potrebbe trovare Steve Buscemi più bello di Cristiano Ronaldo → questo però non è comunque formalmente stimabile da questo eventuale mixed-model, perché ciascun soggetto risponderebbe una singola volta a ogni stimolo)*

Come scrivere il modello

parte fissa

fit = lmer (scoreBellezza ~ natiPariDispari +

(1 + natiPariDispari | IDsubj) + (1 | item) , data = d)

parte random

Parte fissa

- natiPariDispari → Tendenza media generale dei nati nei giorni pari (vs dispari) a essere più belli

Parte random

- 1 | IDsubj → tendenza generale del soggetto a dare complessivamente score medi più alti o più bassi rispetto alla media dei soggetti
- natiPariDispari | IDsubj → tendenza del soggetto a trovare più belli i nati pari o dispari, *nella misura in cui ciò devia dalla tendenza generale dei soggetti*
- 1 | item → tendenza dei singoli volti a ricevere score di bellezza più alti, *nella misura in cui si discosta dalla media della loro categoria (pari vs dispari)*

Come scrivere il modello

Le intercette random potrebbero anche essere stimate con una certa precisione

L'effetto di interesse (sia nella parte fissa che random) è stimato in modo del tutto impreciso se non ho un numero adeguato di item/volti diversi

Parte fissa

- natiPariDispari → Tendenza media generale dei nati nei giorni pari (vs dispari) a essere più belli

Parte random

- 1 | IDsubj → tendenza generale del soggetto a dare complessivamente score medi più alti o più bassi rispetto alla media dei soggetti
- natiPariDispari | IDsubj → tendenza del soggetto a trovare più belli i nati pari o dispari, *nella misura in cui ciò devia dalla tendenza generale dei soggetti*
- 1 | item → tendenza dei singoli volti a ricevere score di bellezza più alti, *nella misura in cui si discosta dalla media della loro categoria (pari vs dispari)*

Un caso verosimile : Elettroconduttanza cutanea e valenza delle foto

→ Caso TOTALMENTE analogo al primo esempio sui volti

Posso campionare 100 partecipanti, ma il power dipende anche da quanti stimoli ho campionato

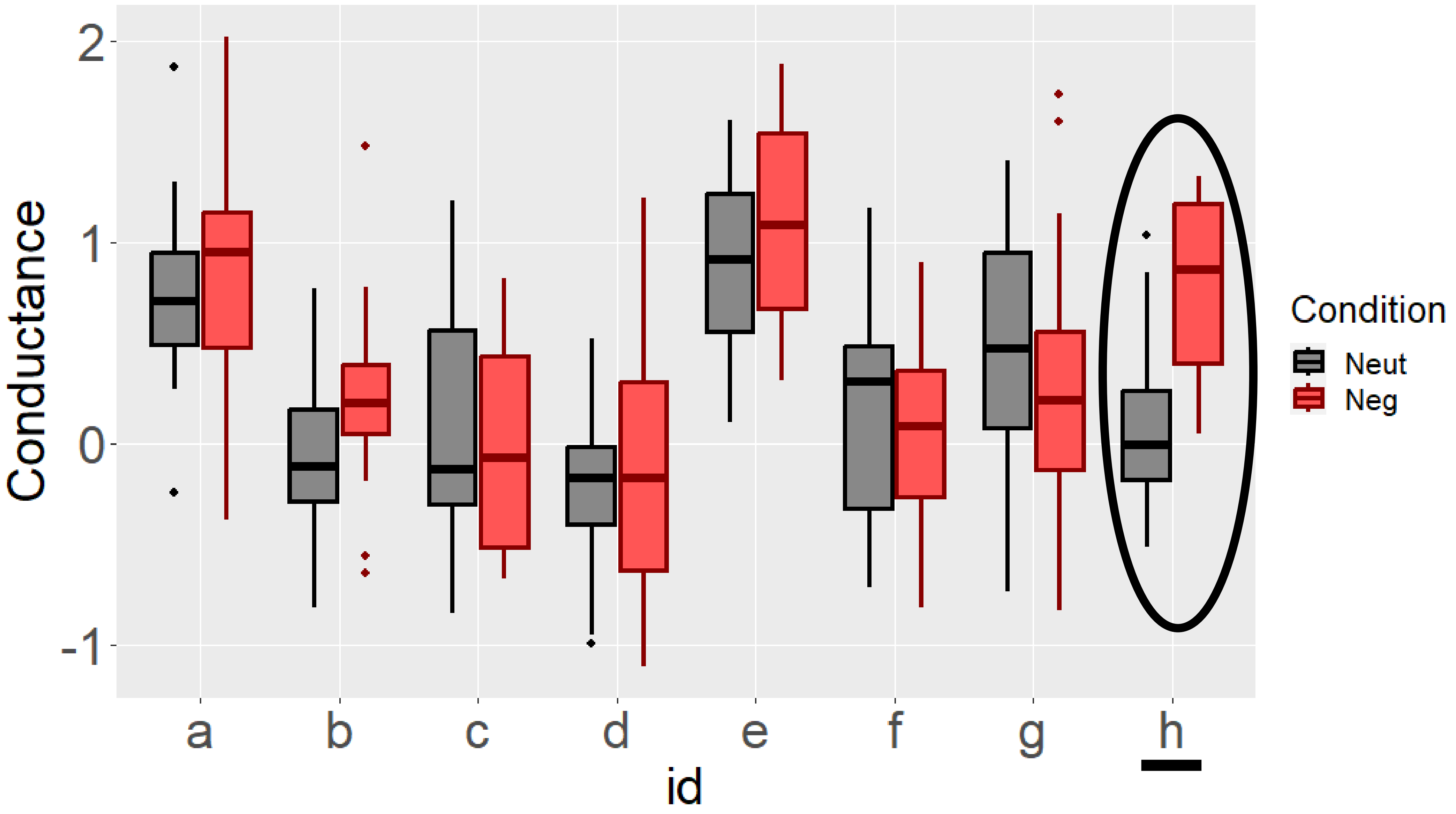
Diciamo che utilizzo 20 IAPS negative vs 20 IAPS neutre

→ Se ho buona varietà di stimoli (es. > decine di item per livello), potrei anche fare il «medione» per partecipante e usarlo in un'ANOVA a misure ripetute (*safe option?*)

→ Oppure posso usare i LMM per estrarre tutta l'informazione, ma devo stare attento:

- Un'imprecisione rischiosa è considerare solo l'*intercetta random* senza valutare l'eventualità che serva anche una *slope random* (ERRORE FREQUENTE)
- Cosa succede se la maggioranza dei soggetti NON mostra l'effetto, ma alcuni sì in modo decisamente più marcato (es. perché più ansiosi)?
- È sempre opportuno considerare i ***casi influenti***

Un caso verosimile : Elettroconduttanza cutanea e valenza delle foto



Un caso verosimile : Elettroconduttanza cutanea e valenza delle foto

gestisco solo la random intercept: tengo conto che l'elettroconduttanza media differisce tra un soggetto e l'altro

 # Intercetta random

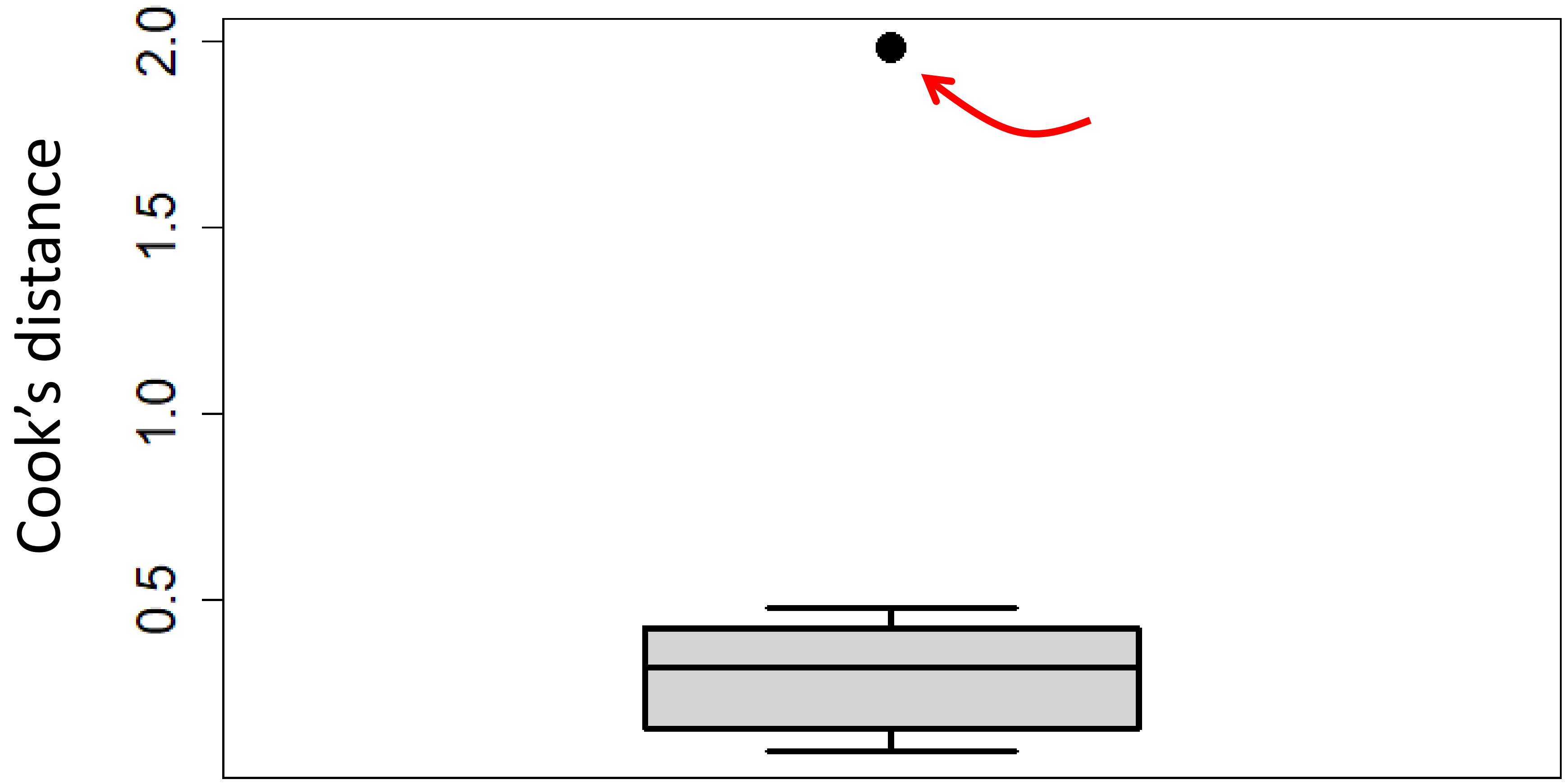
```
fit <- lmer(Conductance ~ Condition + (1|id) + (1|item) data=d)
summary(fit)
(…)
Fixed effects:
              Estimate Std. Error        df t value Pr(>|t|)
(Intercept)    0.27260    0.14078   7.42388   1.936  0.09166 .
ConditionNeg   0.11312    0.04017 612.00001   2.816  0.00502 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Problema analogo nelle META-ANALISI senza effetti random (cioè, a effetti fissi) → un singolo caso/studio con effetto forte per far risultare significativo l'effetto per l'intera meta-analisi (ma si tratterebbe comunque di un **CASO INFLUENTE**)

Un caso verosimile : Elettroconduttanza cutanea e valenza delle foto

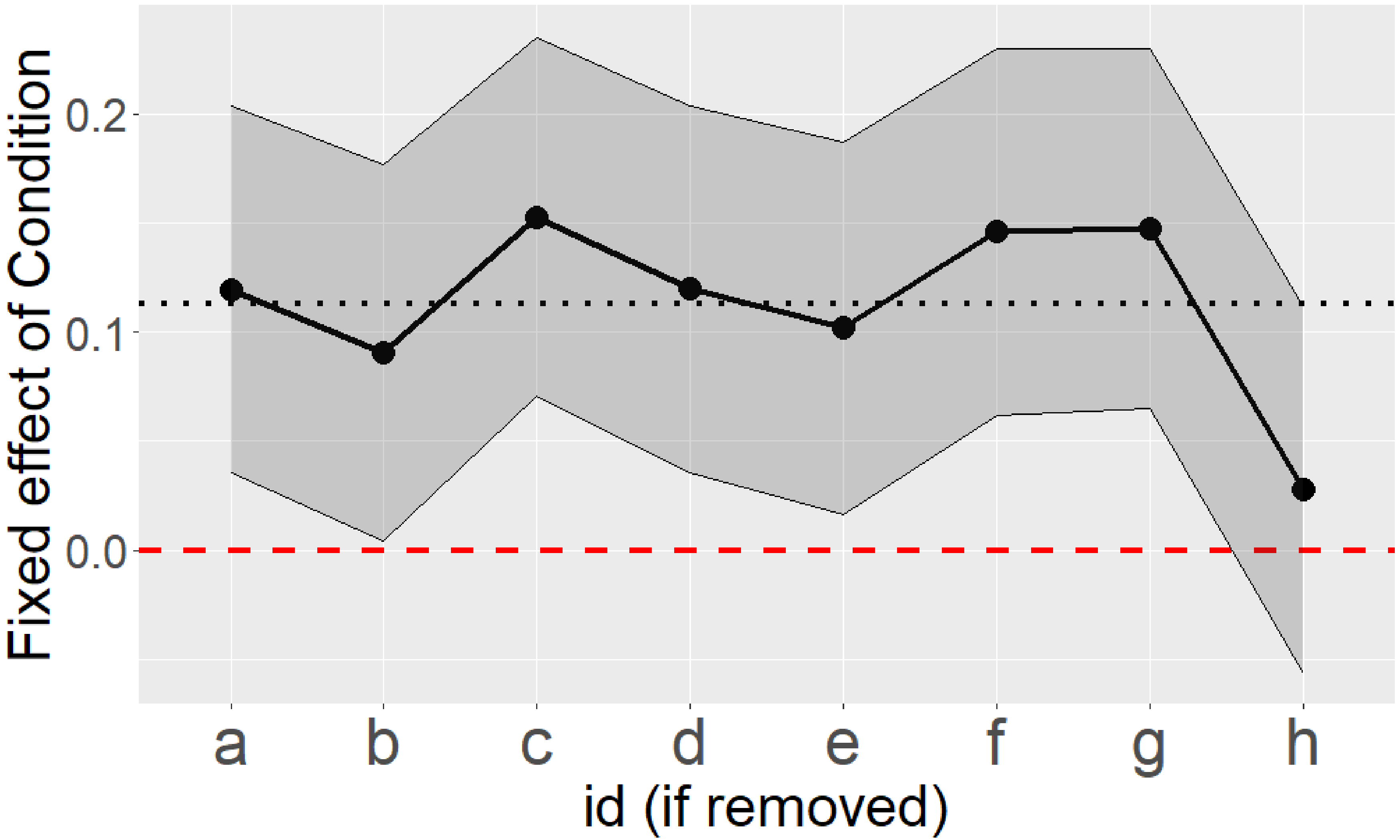
Analisi dei casi influenti: Cook's distance

```
cookdID = cooks.distance(influence.ME::influence(fit0,group="id"))
boxplot(cookdID,ylab="Cook's distance",pch=19,pars=list(boxlwd=2,whisklwd=2,staplelwd=2,outcex=1.5))
```



Un caso verosimile : Elettroconduttanza cutanea e valenza delle foto

Analisi dei casi influenti: leave-one-out range of variation



Un caso verosimile : Elettroconduttanza cutanea e valenza delle foto

gestisco sia la random intercept che la random slope: non solo l'elettroconduttanza cutanea media differisce tra partecipanti, ma anche l'effetto della condizione è eterogeneo

 # Intercetta e slope random

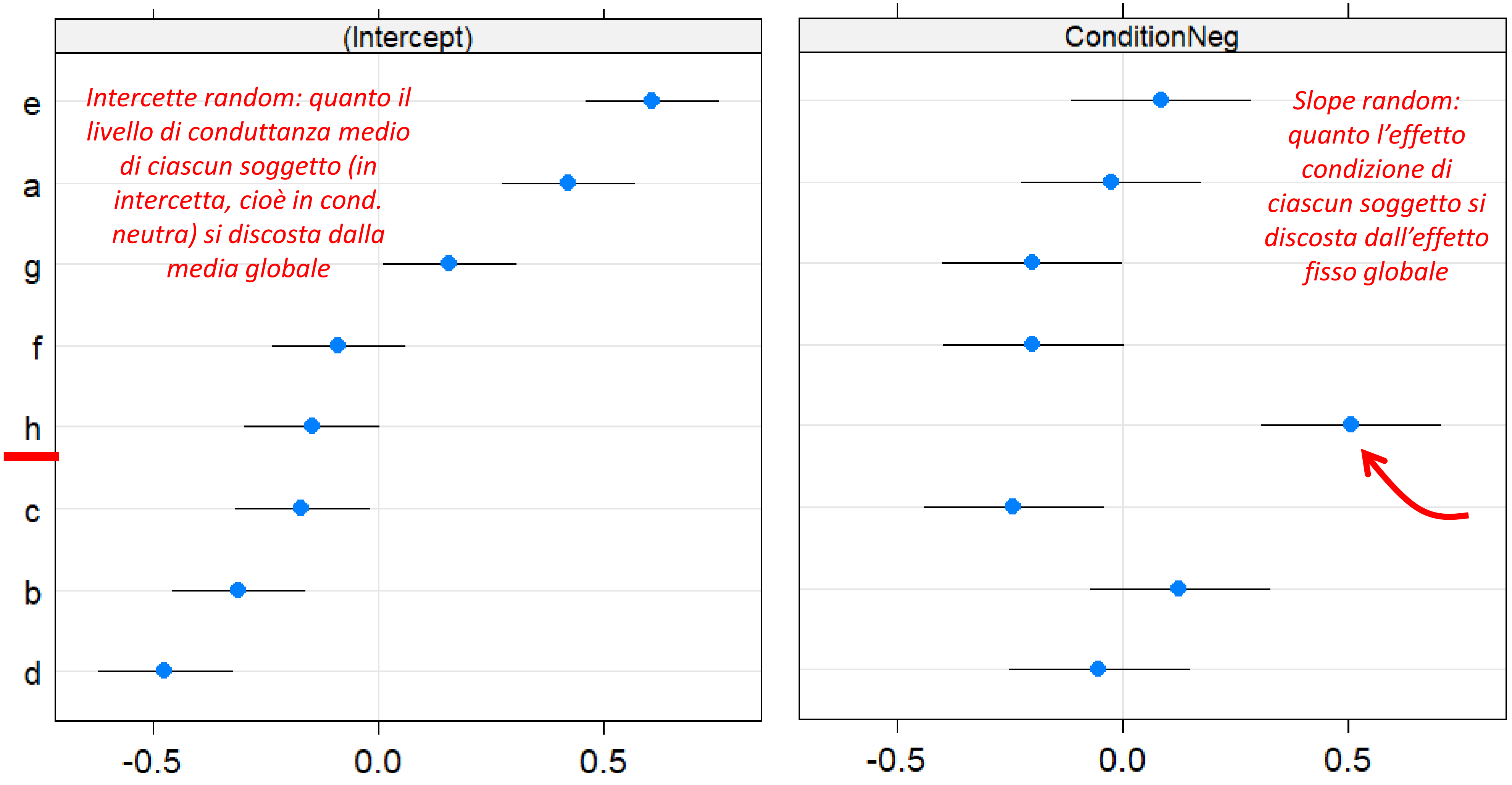
```
fit <- lmer(Conductance ~ Condition + (Condition|id) + (1|item), data=d)
summary(fit)
(…)
Fixed effects:
              Estimate Std. Error      df t value Pr(>|t|)
(Intercept)     0.2726     0.1365  7.1486   1.997   0.0851 .
ConditionNeg    0.1131     0.1015  7.0001   1.115   0.3018
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

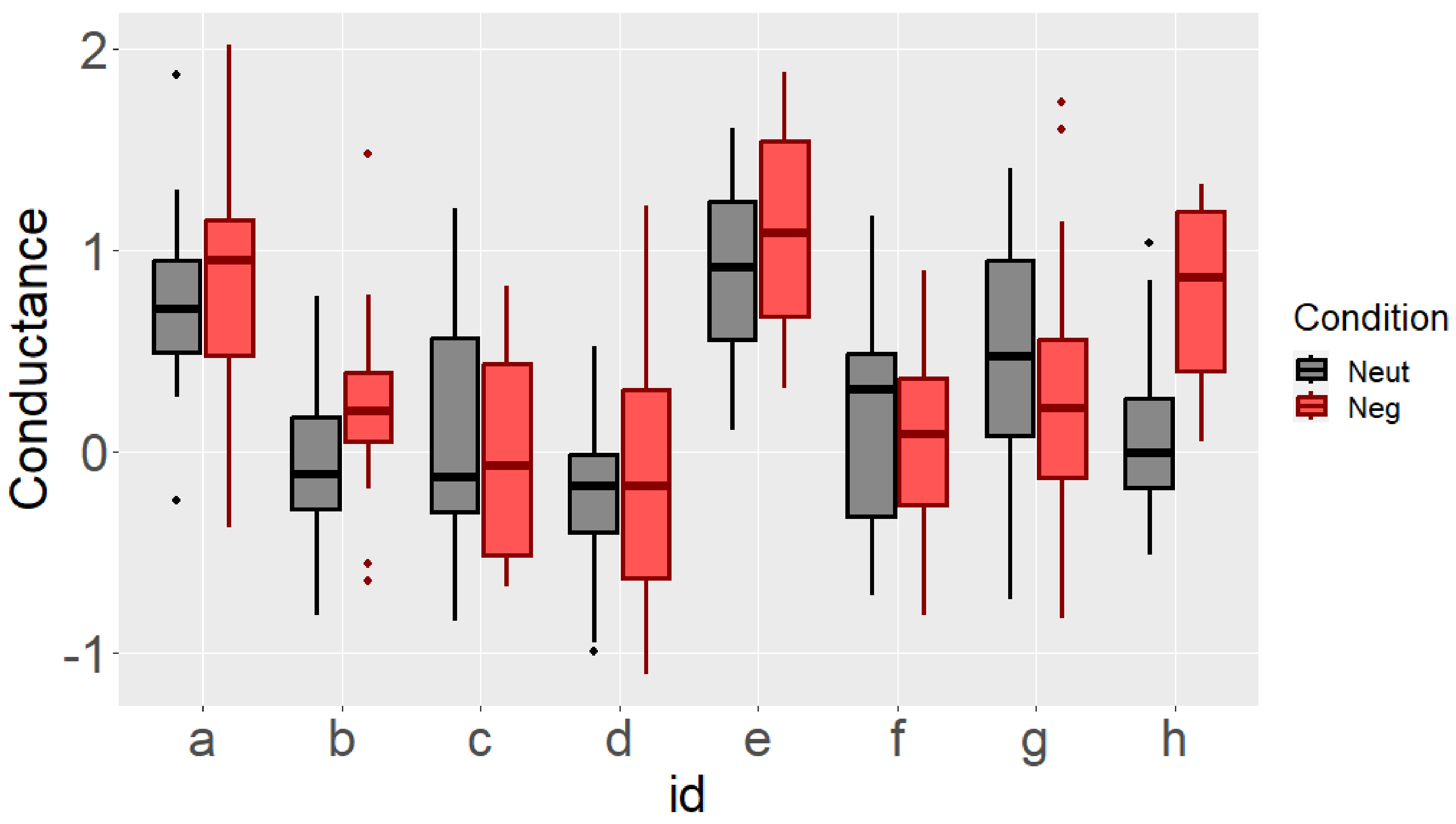
Ho comunque scampato il pericolo gestendo correttamente le *slope random* … per quanto il caso influente sulle stime rimanga

È interessante notare come le stime dei coefficienti siano rimaste le stesse, ma l'incertezza sia aumentata enormemente: ora l'effetto fisso della Conduttanza NON è più significativo (ed è giusto che non lo sia, perché era un effetto particolare di un singolo partecipante)

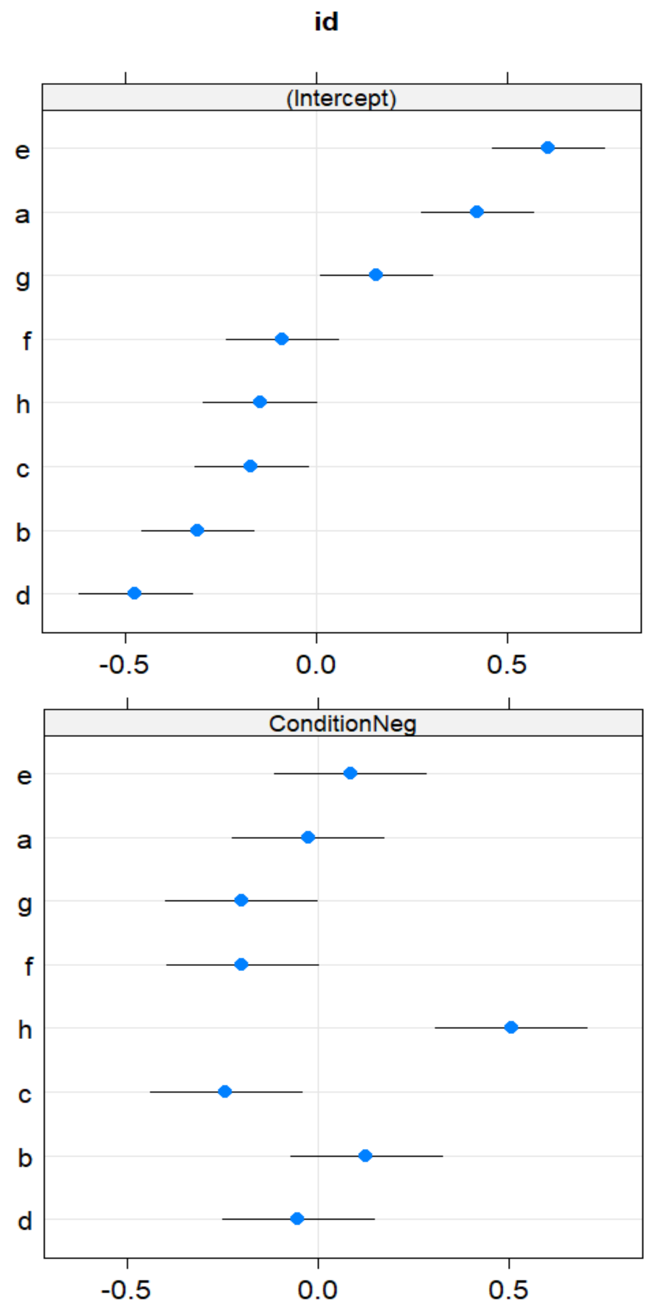
Un caso verosimile : Elettroconduttanza cutanea e valenza delle foto

Intercetta e slope random (del fattore random "id")





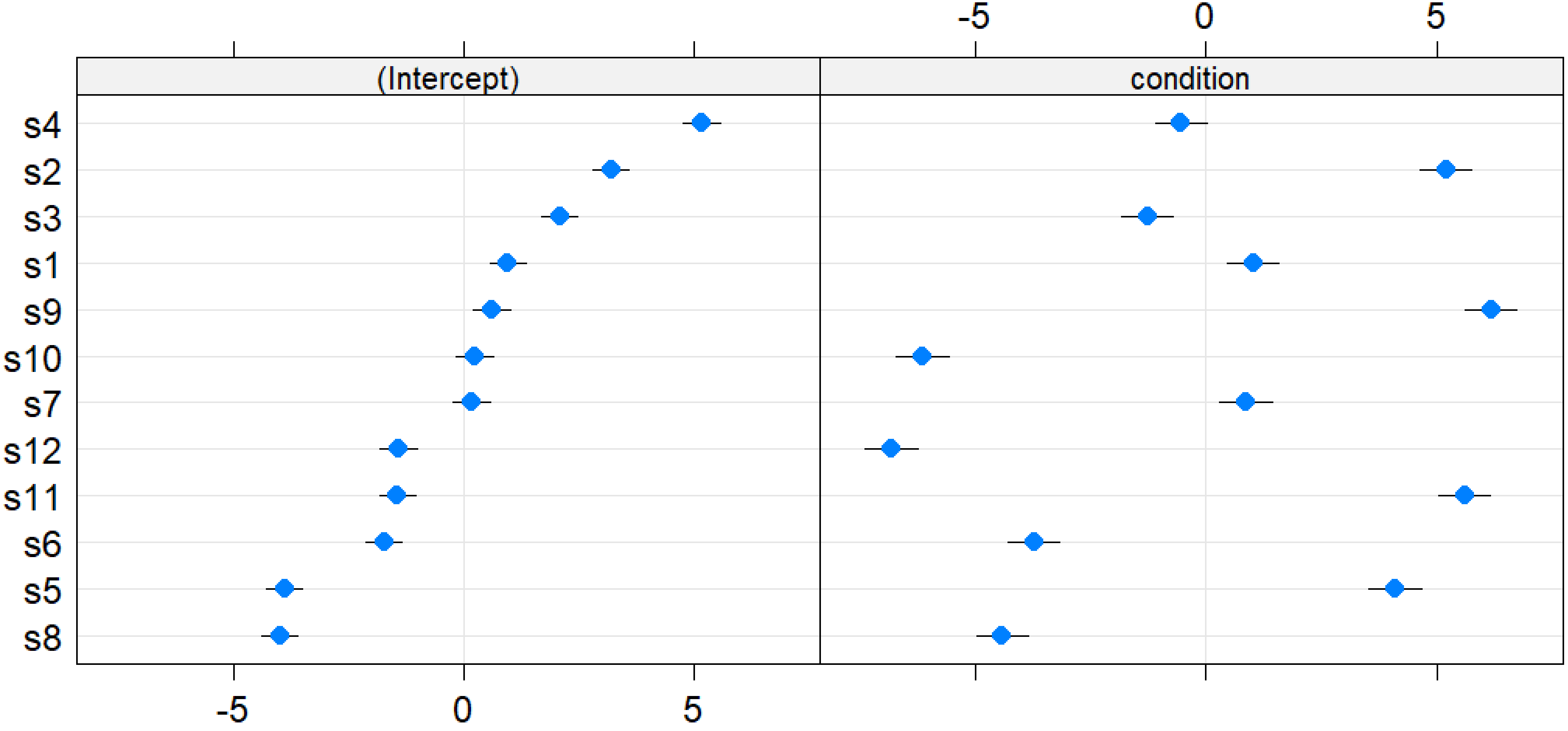
Conductance
2
1
0
-1
a
b
c
d
e
f
g
h
id
Condition
Neut
Neg



id
(Intercept)
e
a
g
f
h
c
b
d
-0.5
0.0
0.5
ConditionNeg
e
a
g
f
h
c
b
d
-0.5
0.0
0.5

SE DIMENTICO LE RANDOM SLOPE QUANDO SERVONO, POTREI ANCHE AVERE TUTTI «CASI INFLUENTI» (SENZA OUTLIERS)

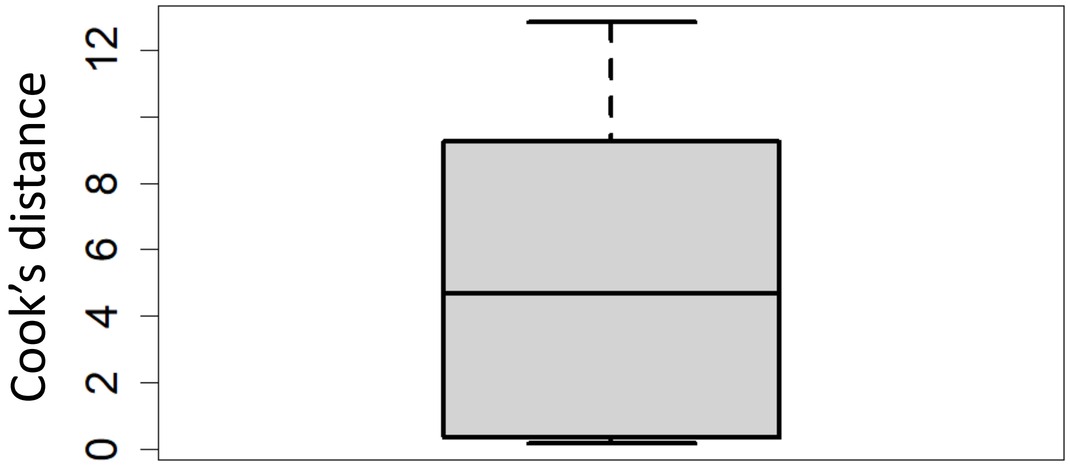
Ad esempio, in questo contesto (del tutto analogo alle IAPS di prima) ho semplicemente tanta variabilità random sia in intercette che in slope:



SE DIMENTICO LE RANDOM SLOPE QUANDO SERVONO, POTREI ANCHE AVERE TUTTI «CASI INFLUENTI» (SENZA OUTLIERS)

Solo random intercept (modello «sbagliato»)

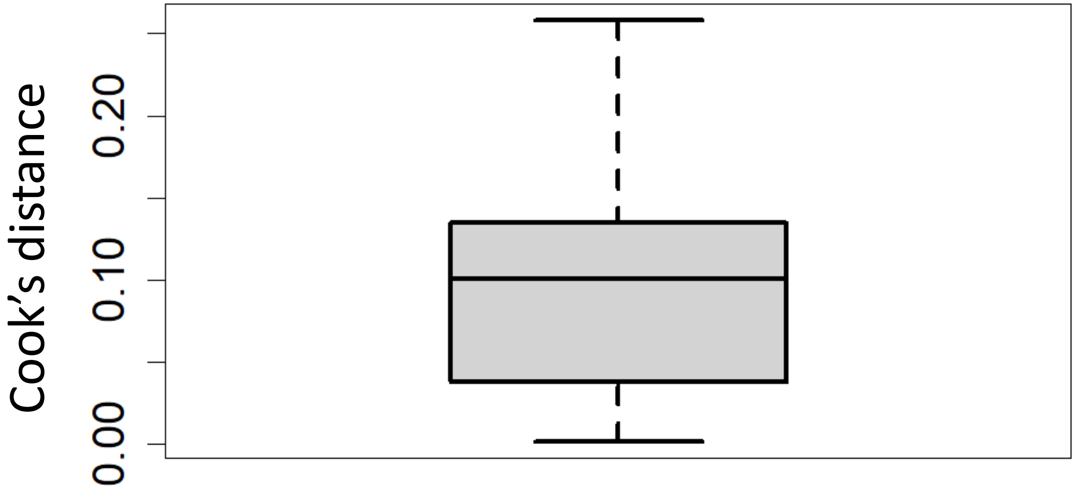
```
> summary(lmer(resp ~ condition + (1|id), data=d))
(...)
Fixed effects:
            Estimate Std. Error        df t value Pr(>|t|)    
(Intercept)   9.8486     1.1185   11.0669   8.806 2.49e-06 ***
condition     1.4003     0.1231 2387.0000  11.377  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```



Indicativo che c'è qualcosa che non va

Random intercept e random slope (modello «vero»)

```
> summary(lmer(resp ~ condition + (condition|id), data=d))
(...)
Fixed effects:
            Estimate Std. Error      df t value Pr(>|t|)    
(Intercept)   9.8486     0.7872 10.9970  12.511 7.59e-08 ***
condition     1.4003     1.3414 11.0005   1.044    0.319    
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```



Un caso verosimile : Risposta al trattamento

Randomized Controlled Trials: Raccogliere misure ripetute sia al pre- che al post-test conferisce non solo più potenza, ma aiuta anche a valutare effetti entro-soggetto

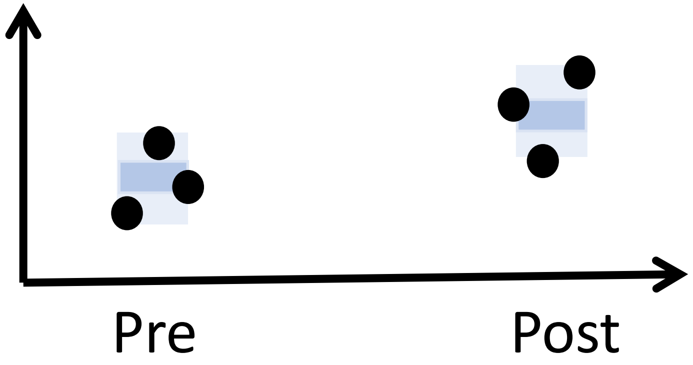

Behavior Research Methods
https://doi.org/10.3758/s13428-021-01549-x

Dyslexia treatment studies: A systematic review and suggestions on testing treatment efficacy with small effects and small samples

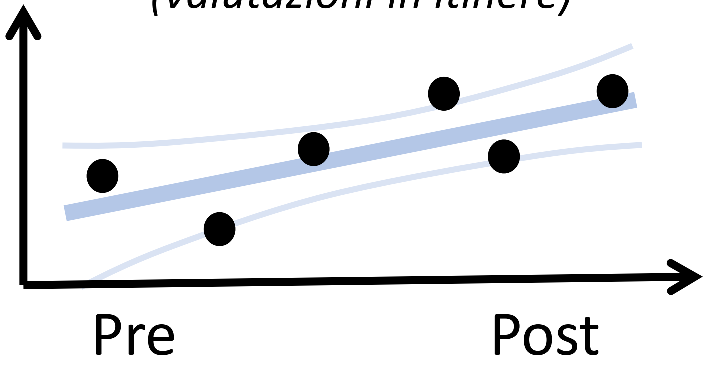
Enrico Toffalini[1] · David Giofrè[2] · Massimiliano Pastore[3] · Barbara Carretti[1] · Federica Fraccadori[1] · Denes Szűcs[4]

Accepted: 20 January 2021

Dati per singolo soggetto trattato (2 tempi)



Dati per singolo soggetto trattato (valutazioni in itinere)



Un caso verosimile : Risposta al trattamento

Randomized Controlled Trials: Raccogliere misure ripetute sia al pre- che al post-test conferisce non solo più potenza, ma aiuta anche a valutare effetti entro-soggetto

- Ovviamente posso migliorare il power usando (molte) misure ripetute solo fino a un certo punto (solo finché ho margine per ridurre ancora in modo rilevante l'errore di misura)
- Un vantaggio cruciale però è che posso stimare i parametri di risposta al trattamento a livello individuale usando un LMM
- *Rischio?* Se dimentico di valutare la *slope random*, uno o pochi soggetti che beneficino molto dal trattamento (o che siano casualmente migliorati per i fatti loro) potrebbero da soli rendere significativo l'effetto fisso, facendomi generalizzare scorrettamente l'inferenza sull'efficacia *(questo è testabile spesso guardando i **casi influenti**)*