

***Embrace variation
and accept uncertainty***

Andrew Gelman

BREVE RECAP su modelli misti

- L'interesse non è solo per la parte *fissa* (stima effetto medio a livello di popolazione), ma anche per le variabilità tra fattori *random* («di raggruppamento»)
- Tenere conto dei fattori *random* in alcuni casi è indispensabile per garantire la validità delle stime, dei loro livelli di precisione, e dell'inferenza

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

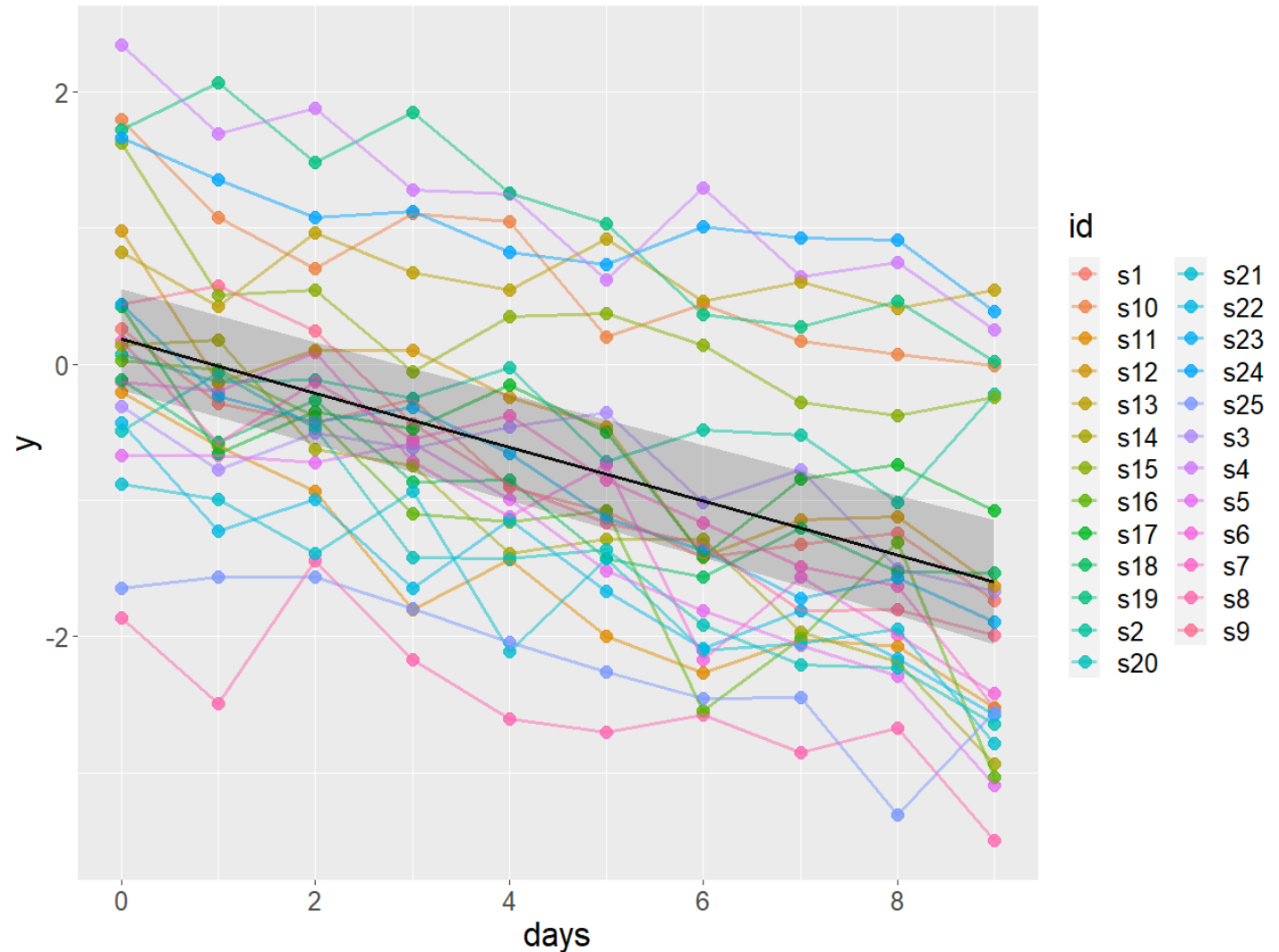
```
> summary(fit4)
Linear mixed model fit by REML. t-tests use Satterthwaite's
method ['lmerModLmerTest']
Formula: y ~ days + (days | id)
Data: d

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 ***
```



RECAP su struttura del dataset

*i dati simulati su cui lavoreremo ora
sono già in forma long*

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

- La variabile dipendente sta in UNA colonna
- Per ogni riga c'è UNA osservazione della variabile dipendente
- Ciascun predittore (categoriale o quantitativo) sta in un UNA colonna
- Ciascun fattore di raggruppamento/effetto random (spesso il soggetto, ma non sempre e non necessariamente) è indicato in una colonna

PACCHETTI CONSIGLIATI

- «*lme4*» → per la stima dei LMM
- «*lmerTest*» → se vogliamo p-value nei summary dei LMM
- «*effects*» → utile per estrarre e visualizzare gli effetti stimati
- «*ggplot2*» → per un tocco di classe nella visualizzazione degli effetti
- «*performance*» → aggiunge una serie di funzioni per valutare la qualità dei modelli e alcuni indici (es. stime degli R^2 nei LMM, variance inflation factor per collinearità, check di omoschedasticità, e moltissime altre cose)
- «*influence.ME*» → aggiunge funzioni per il test dei casi influenti (es. calcolo delle distanze di Cook per gli effetti random)

ESERCIZIO 1A (warming up...): Rigeneratività e memoria di lavoro

RESEARCH QUESTION:

- Essere nel gruppo «Restorative» fa aumentare la prestazione al digitSpan più che essere nel gruppo «Control» quando si passa dal tempo «Pre» al tempo «Post»?

Qui abbiamo un'unica misura della variabile dipendente al «Pre» e un'unica misura al «Post»: possiamo (e dobbiamo) modellare un'intercetta random per soggetto, ma non possiamo modellare una slope random

id	group	time	digitSpan
Co2	Control	0.Pre	7
Co2	Control	1.Post	8
Co3	Control	0.Pre	6
Co3	Control	1.Post	6
Co4	Control	0.Pre	8
Co4	Control	1.Post	7
Co5	Control	0.Pre	7
Co5	Control	1.Post	8
Co6	Control	0.Pre	7
Co6	Control	1.Post	7
Co7	Control	0.Pre	6
Co7	Control	1.Post	7
Co8	Control	0.Pre	9
Co8	Control	1.Post	8
Co9	Control	0.Pre	6

ESERCIZIO 1B: Efficacia del trattamento per la dislessia

RESEARCH QUESTION:

- Il trattamento è stato efficace nel migliorare l'abilità di lettura in punti z nel gruppo trattato rispetto al controllo?
- C'è variabilità tra soggetti nell'efficacia del trattamento?

Questo esercizio assomiglia al precedente, ma qui abbiamo ottenuto diverse misure di lettura in punti z al «Pre» e al «Post» per ciascun soggetto; questo ci permette di stimare eventualmente la variabilità nella slope tra soggetti

id	time	group	read_z
Co1	0.Pre	Control	-3.052652
Co1	0.Pre	Control	-2.572984
Co1	0.Pre	Control	-2.590617
Co1	0.Pre	Control	-2.892973
Co1	0.Pre	Control	-2.933491
Co1	0.Pre	Control	-3.412163
Co2	0.Pre	Control	-2.238359
Co2	0.Pre	Control	-2.179174
Co2	0.Pre	Control	-2.648526
Co2	0.Pre	Control	-2.383780
Co2	0.Pre	Control	-2.256374
Co2	0.Pre	Control	-2.514444
Co3	0.Pre	Control	-1.910107
Co3	0.Pre	Control	-2.287200
Co3	0.Pre	Control	-1.514228
Co3	0.Pre	Control	-1.886045

ESERCIZIO 2: Chi fa meglio in geometria? Maschi o femmine?

RESEARCH QUESTION: fin da giovanissimi maschi e femmine mostrano piccole differenze in compiti di matematica e comprensione del testo. Cosa possiamo dire della geometria in quinta elementare?

METODO: per scoprirlo sono stati raccolti dati da vari istituti scolastici per un po' di anni (stile invalsi). Ora ci chiedono di analizzarli...

Nel dataset troviamo molte informazioni, tra cui la classe, la scuola, l'anno di somministrazione, il numero di classi per scuola e di studenti per classe...quale sarà il modello generativo dei dati?

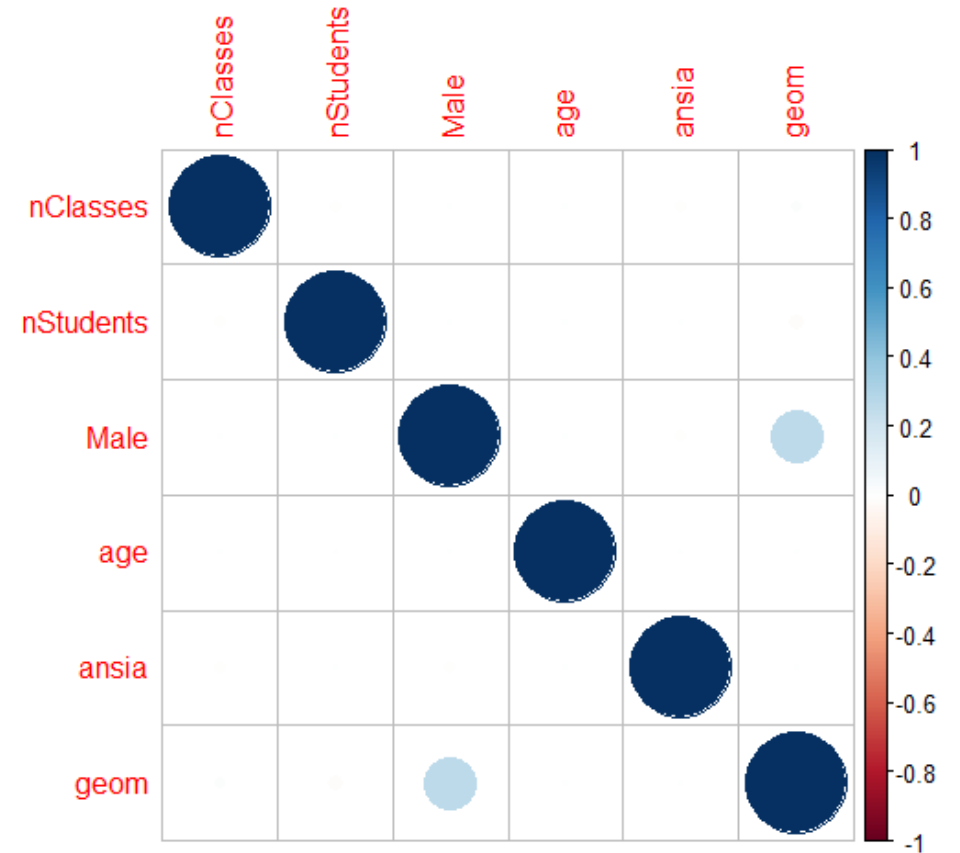
Year	idSchool	idClass	idChild	Gender	age	ansia	geom
2009	School1	Class1	Child1	F	10.19	0.920	0.424
2009	School1	Class1	Child2	M	10.28	-0.529	0.059
2009	School1	Class1	Child3	F	9.71	-0.576	-0.590
2009	School1	Class1	Child4	M	10.08	0.916	0.564
2009	School1	Class1	Child5	F	10.01	0.162	0.269
2009	School1	Class1	Child6	F	10.34	-0.220	-0.329
2009	School1	Class1	Child7	M	9.70	0.375	0.951
2009	School1	Class1	Child8	F	9.80	0.516	0.552
2009	School1	Class1	Child9	M	9.50	-0.730	1.961
2009	School1	Class1	Child10	M	9.63	-1.206	0.810
2009	School1	Class1	Child11	M	10.23	1.093	0.779
2009	School1	Class1	Child12	M	9.78	1.199	-0.008
2009	School1	Class1	Child13	M	9.52	-0.526	0.913
2009	School1	Class1	Child14	M	10.11	-0.532	1.188
2009	School1	Class1	Child15	M	10.27	0.188	0.893
2009	School1	Class1	Child16	M	10.50	1.103	-0.655
2009	School1	Class1	Child17	F	9.53	-1.126	0.230
2009	School1	Class1	Child18	M	10.09	-0.839	1.675
2009	School1	Class1	Child19	M	10.13	0.686	-0.586
2009	School1	Class1	Child20	F	9.67	2.237	0.433
2009	School1	Class1	Child21	F	10.32	-2.245	-0.567
2009	School1	Class2	Child22	F	10.09	-0.356	0.069
2009	School1	Class2	Child23	M	9.69	0.062	1.233
2009	School1	Class2	Child24	F	9.59	-0.122	0.901

ESERCIZIO 2: soluzioni?

CORRELAZIONI:

Un veloce sguardo alle correlazioni ci mostra una soluzione molto semplice: solo il genere ha un effetto sulla geometria

```
> (ctab <- round(cor(subset(db, select = c
+ (nClasses, nStudents, Male, age, ansia, geom))),2))
      nClasses nStudents  Male age  ansia  geom
nClasses    1.00   -0.01   0.00  0 -0.01   0.01
nStudents   -0.01    1.00   0.00  0  0.00  -0.02
Male         0.00    0.00   1.00  0 -0.01   0.26
age          0.00    0.00   0.00  1  0.00   0.00
ansia       -0.01    0.00  -0.01  0  1.00   0.00
geom         0.01   -0.02   0.26  0  0.00   1.00
```

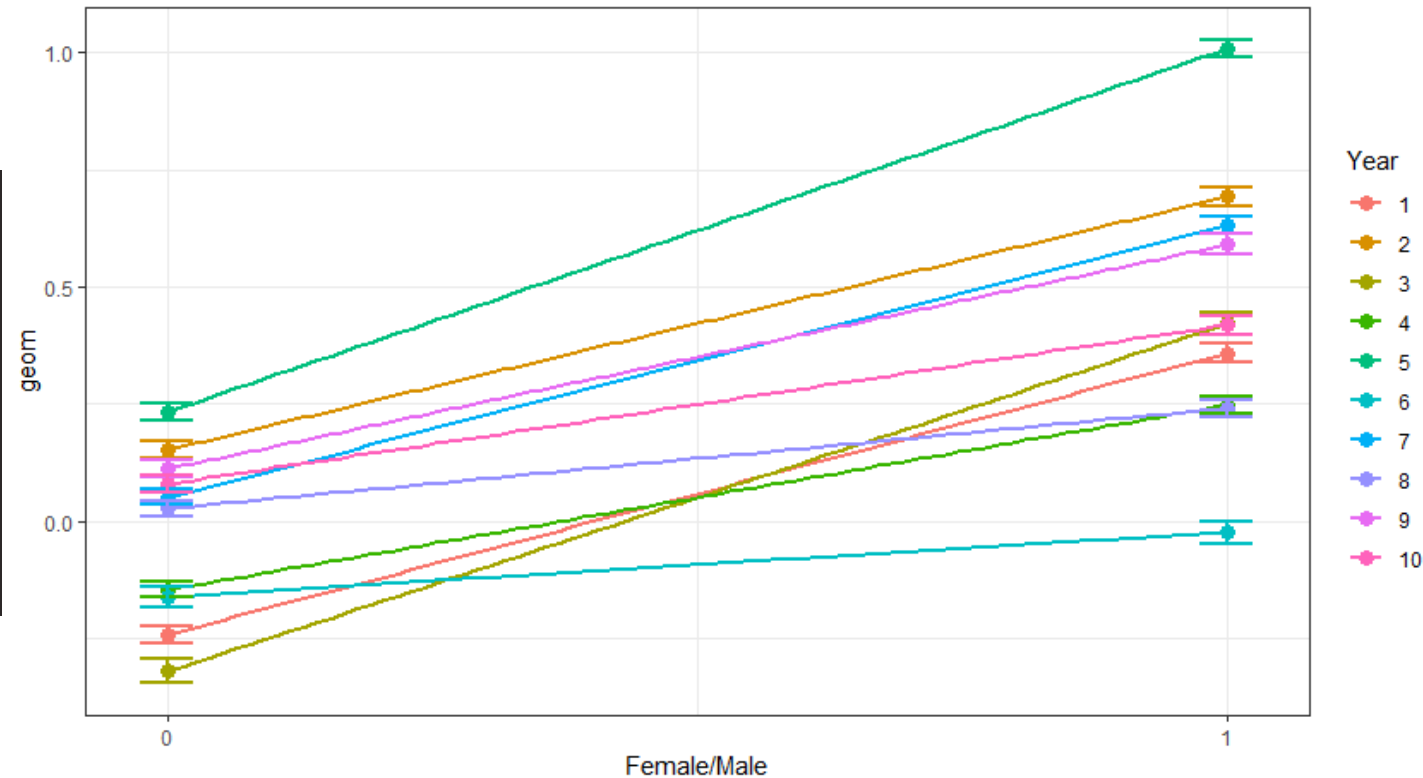


ESERCIZIO 2: soluzioni?

Ma uno sguardo un po' più approfondito ci complica un po' le cose...

1: Per colpa degli anni di somministrazione

```
#Plotto
ggplot(db, aes(y = geom, x = Male, group = as.factor(idYear),
               color = as.factor(idYear))) +
  geom_smooth(method = "lm", se = F) +
  stat_summary(fun=mean, geom=c("point"),
              aes(group = idYear, color = as.factor(idYear)),
              size = 3) +
  stat_summary(fun.data=mean_se, geom=c("errorbar"),
              aes(group = idYear, color = as.factor(idYear)),
              width = .05, size = 1) +
  scale_color_discrete(name = "Year") +
  scale_x_continuous(breaks = c(0,1)) +
  xlab("Female/Male") +
  theme_bw()
```

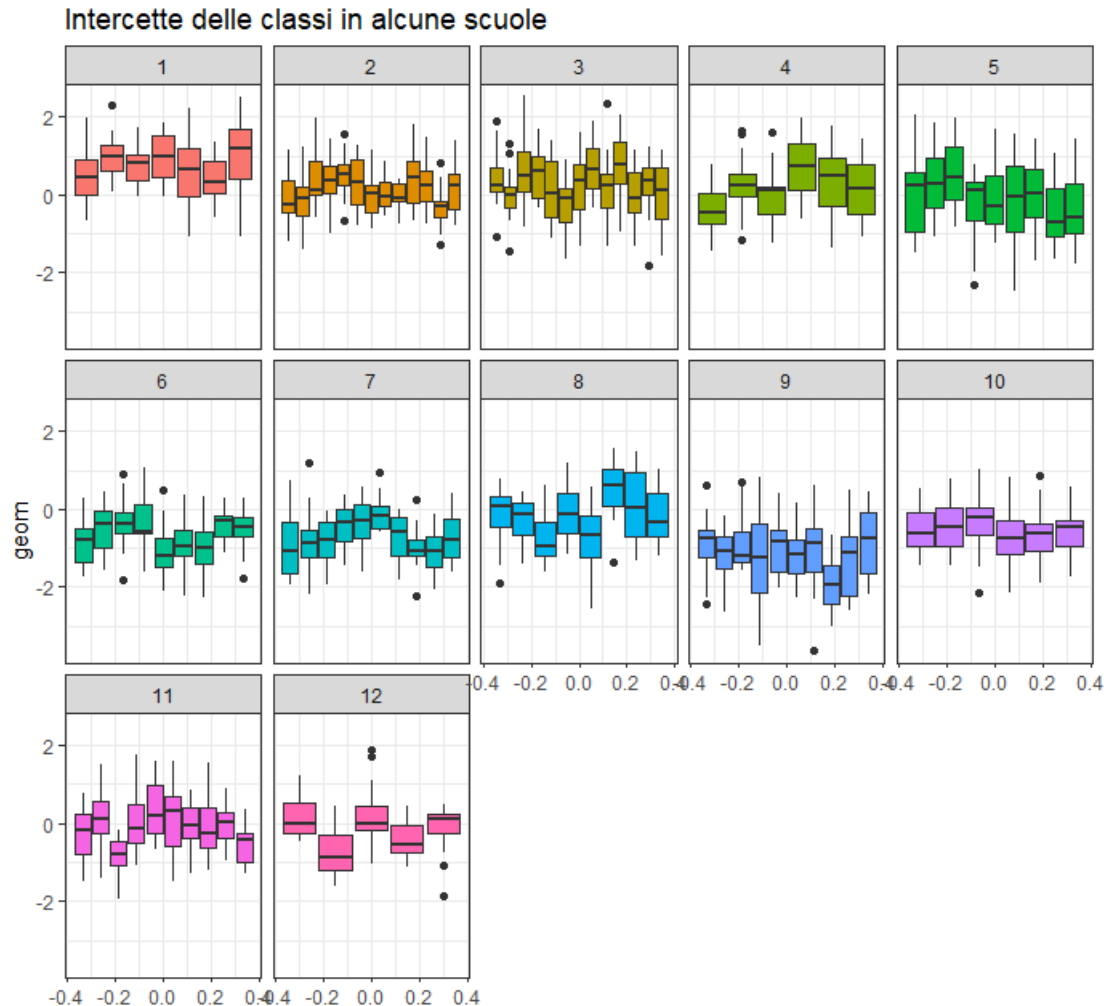


ESERCIZIO 2: soluzioni?

Ma uno sguardo un po' più approfondito ci complica un po' le cose...

2: Perché nessuna scuola e nessuna classe è uguale!

```
ggplot(db[1:2100,], aes(y = geom,  
                        group = as.factor(idclass),  
                        fill = as.factor(idschool))) +  
  geom_boxplot() +  
  labs(title = "Intercette delle classi in alcune scuole") +  
  facet_wrap(~as.factor(idschool), ncol = 5) +  
  theme_bw()
```

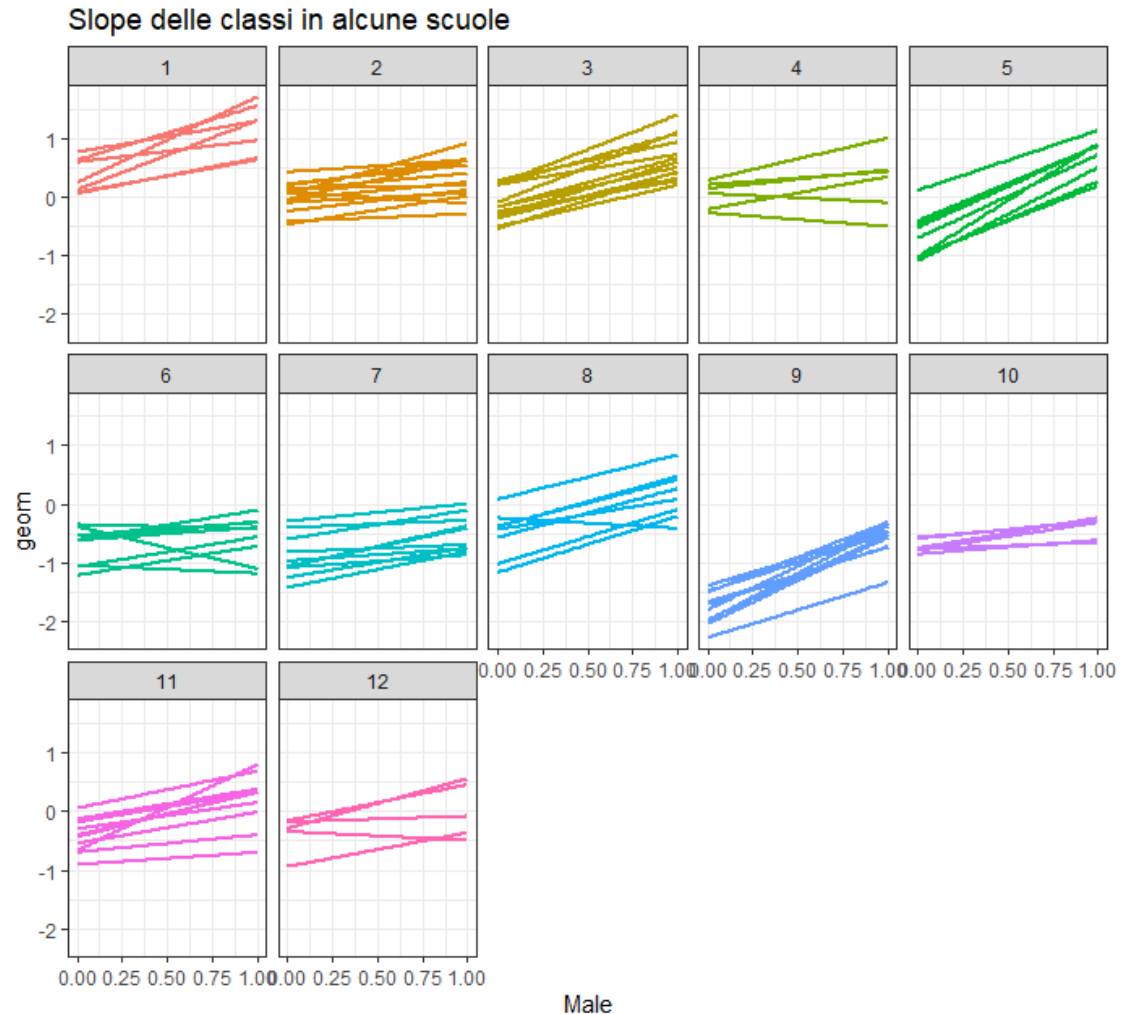


ESERCIZIO 2: soluzioni?

Ma uno sguardo un po' più approfondito ci complica un po' le cose...

2: Perché nessuna scuola e nessuna classe è uguale!

```
ggplot(db[1:2100,], aes(y = geom, x = Male,  
                        group = as.factor(idclass),  
                        color = as.factor(idschool))) +  
  geom_smooth(method = "lm", se = F) +  
  labs(title = "slope delle classi in alcune scuole") +  
  facet_wrap(~as.factor(idschool), ncol = 5) +  
  theme_bw()
```



ESERCIZIO 2: soluzioni?

E allora...

```
R 4.2.2 · C:/Users/feraco/OneDrive - Università degli Studi di Padova/LabMeeting/ToffaLab/Slides-e-codice-2023_03_08/ ↗
> fit <- lmer(geom ~ Male + (Male|idYear/idschool/idclass), data = db,
+           control=lmerControl(optimizer="bobyqa",optCtrl = list(maxfun=2e4)))
> summary(fit)
Linear mixed model fit by REML. t-tests use Satterthwaite's method ['lmerModLmerTest']
Formula: geom ~ Male + (Male | idYear/idschool/idclass)
Data: db
Control: lmerControl(optimizer = "bobyqa", optCtrl = list(maxfun = 20000))

REML criterion at convergence: 78391.5

Scaled residuals:
   Min       1Q   Median       3Q      Max
-3.8689 -0.6559  0.0083  0.6526  4.1620

Random effects:
Groups                Name      Variance Std.Dev. Corr
idclass:(idschool:idYear) (Intercept) 0.06004  0.2450
                        Male      0.02456  0.1567  0.04
idschool:idYear      (Intercept) 0.28866  0.5373
                        Male      0.10050  0.3170  0.08
idYear                (Intercept) 0.01622  0.1274
                        Male      0.04014  0.2003  0.11
Residual              0.35952  0.5996

Number of obs: 40508, groups: idclass:(idschool:idYear), 2018; idschool:idYear, 205; idYear, 10

Fixed effects:
              Estimate Std. Error    df t value Pr(>|t|)
(Intercept) -0.01136    0.05577  8.40407  -0.204    0.843
Male         0.48550    0.06764  8.81598   7.177 5.79e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Correlation of Fixed Effects:
      (Intr)
Male 0.090
>
```

ESERCIZIO 3: RATING DI FOTOGRAFIE

RESEARCH QUESTION:

- Bambini e ragazzi dai 6 ai 16 anni hanno dato rating (0-100) a diverse fotografie di animali vs paesaggi. Vogliamo indagare da cosa dipendano i rating.
- Vogliamo anche valutare la variabilità random non solo tra soggetti rispondenti, ma anche tra fotografie.

idSubj	age	gender	condition	idPicture	rating
subj1	15.04017	F	animal	anim1	75
subj1	15.04017	F	animal	anim2	91
subj1	15.04017	F	animal	anim3	76
subj1	15.04017	F	animal	anim4	68
subj1	15.04017	F	animal	anim5	65
subj1	15.04017	F	animal	anim6	60
subj1	15.04017	F	animal	anim7	70
subj1	15.04017	F	animal	anim8	86
subj1	15.04017	F	animal	anim9	68
subj1	15.04017	F	animal	anim10	90
subj1	15.04017	F	animal	anim11	82
subj1	15.04017	F	animal	anim12	80
subj1	15.04017	F	animal	anim13	76
subj1	15.04017	F	animal	anim14	83
subj1	15.04017	F	animal	anim15	69
subj1	15.04017	F	animal	anim16	57

ESERCIZIO 4: BAMBINI E OPERAZIONI ARITMETICHE

RESEARCH QUESTION:

- Bambini di 4° primaria di età omogenea affrontano operazioni matematiche, e vogliamo modellare la probabilità di risposta in funzione del tipo di operazione (addizione vs sottrazione), della condizione (baseline vs time pressure), dell'ansia per la matematica, ed eventualmente del genere.
- Vogliamo tenere conto della variabilità random a vari livelli: soggetto, ma anche classe in cui si trova, e item.

Qui la situazione è complicata dal fatto che la «risposta» registrata è la correttezza dell'operazione svolta, ed è dicotomica: 0 (sbagliato) vs 1 (giusto). Dobbiamo usare una regressione binomiale/logistica

idClass	idSubj	Gender	Age	MathAnx	Condition	idItem	Type	Correct
cl5	aes110	F	9.8	-0.9968873	Baseline	item48	Subtraction	1
cl5	aes110	F	9.8	-0.9968873	Baseline	item46	Subtraction	1
cl5	aes110	F	9.8	-0.9968873	Baseline	item44	Subtraction	1
cl5	aes110	F	9.8	-0.9968873	Baseline	item42	Subtraction	0
cl5	aes110	F	9.8	-0.9968873	Baseline	item40	Subtraction	0
cl5	aes110	F	9.8	-0.9968873	Baseline	item38	Subtraction	1
cl5	aes110	F	9.8	-0.9968873	Baseline	item36	Subtraction	0
cl5	aes110	F	9.8	-0.9968873	Baseline	item34	Subtraction	1
cl5	aes110	F	9.8	-0.9968873	Baseline	item32	Subtraction	0
cl5	aes110	F	9.8	-0.9968873	Baseline	item30	Subtraction	1
cl5	aes110	F	9.8	-0.9968873	Baseline	item28	Subtraction	1
cl5	aes110	F	9.8	-0.9968873	Baseline	item26	Subtraction	1
cl5	aes110	F	9.8	-0.9968873	TimePressure	item24	Subtraction	0
cl5	aes110	F	9.8	-0.9968873	TimePressure	item22	Subtraction	1
cl5	aes110	F	9.8	-0.9968873	TimePressure	item20	Subtraction	0
cl5	aes110	F	9.8	-0.9968873	TimePressure	item18	Subtraction	1
cl5	aes110	F	9.8	-0.9968873	TimePressure	item16	Subtraction	1
cl5	aes110	F	9.8	-0.9968873	TimePressure	item14	Subtraction	0
cl5	aes110	F	9.8	-0.9968873	TimePressure	item12	Subtraction	1
cl5	aes110	F	9.8	-0.9968873	TimePressure	item10	Subtraction	0
cl5	aes110	F	9.8	-0.9968873	TimePressure	item8	Subtraction	0
cl5	aes110	F	9.8	-0.9968873	TimePressure	item6	Subtraction	0
cl5	aes110	F	9.8	-0.9968873	TimePressure	item4	Subtraction	0
cl5	aes110	F	9.8	-0.9968873	TimePressure	item2	Subtraction	0
cl1	afy13	F	9.7	-0.2964118	Baseline	item2	Subtraction	1
cl1	afy13	F	9.7	-0.2964118	Baseline	item4	Subtraction	0
cl1	afy13	F	9.7	-0.2964118	Baseline	item6	Subtraction	0
cl1	afy13	F	9.7	-0.2964118	Baseline	item8	Subtraction	1

***All models are wrong,
some are useful***

George Box