

Analisi dei dati in psicologia

Massimiliano Pastore

Premessa: errori già noti da non segnalare

1. Vi sono molti casi in cui la voce bibliografica è riportata erroneamente con il nome dell'autore o degli autori all'interno della parentesi quadra; es. pag. 60, paragrafo dopo la formula (5.2).
2. Tutte i casi con *i*-ma o *j*-ma o simili vanno letti con *i*-esima, *j*-esima e così via.
3. Il data frame `studenti` del pacchetto `ADati` è stato cambiato e quindi gli output che derivano da tale data frame non corrispondono più a quelli dei box.
4. Tutti i “perchè” ed i “poichè” vanno letti come “perché” e “poiché”.
5. Punteggiature mancanti (virgole etc).
6. Le figure del libro sono in bianco e nero, pertanto sono errati tutti i rimandi a colori presenti nelle stesse.

Capitolo 1

1. Pag. 10, paragrafo 1.5, riga 5: ci sono spazi bianchi tra le parole, si legga “**Nel capitolo 4 richiameremo le basi della teoria dei campioni, ...**”
2. Pag. 10, paragrafo 1.5, riga 11: ci sono spazi bianchi tra le parole, si legga “**... fino a considerare i casi multivariati (cap.10) ed i modelli ...**”
3. Pag.10, paragrafo 1.5, riga 12 al posto di concre-ti, si legga “**concreti**”.
4. Pag. 10, paragrafo 1.5, ultima riga: l'indirizzo del pacchetto `ADati` è stato modificato; attualmente è disponibile all'indirizzo <https://github.com/masspastore/ADati>.

Capitolo 2

1. Pag. 16, riga 7: Le colonne del data-frame nel box 2.7 sono 13, si legga “**Si tratta di un data-frame con 58 righe (osservazioni) e 13 colonne ...**”.¹
2. Pag. 16, sezione **Accesso agli elementi di un vettore o di una matrice**, riga 1: nella frase ‘Questo significa che, ciascun elemento ...’ la virgola va tolta, si legga “**Questo significa che ciascun elemento ...**”
3. Pag. 16, penultima riga: in riferimento al box 2.7, per ottenere la colonna `eta`, si deve richiedere la colonna 6, pertanto si legga “**... lo stesso risultato con `studenti[,6]`, cioè richiedendo la sesta colonna.**”
4. Pag. 21, sezione 2.2, riga 1: nella frase iniziale va aggiunta una virgola; si legga “**Nel capitolo 3.1 vedremo come, per analizzare dei dati, sia fondamentale ...**”
5. Pag. 23, riga 9: aggiungere “sezione” tra parentesi, si legga “**... il suffisso `H$` (si veda sezione 2.1).**”

¹ Il dataframe è quello relativo ad una vecchia versione del pacchetto `ADati`; nella versione corrente (5.4.4, del 18/03/2023) tale data-frame ha 14 colonne/variabili.

6. Pag. 23, sezione File Excel, riga 13: manca un punto alla fine della riga, si legga “... nel box 2.12.”

Capitolo 3

1. Pag. 27, sez. 3.1, ultima riga: la parentesi quadra aperta va spostata dopo il nome Fabbris, si legga “... di Fabbris [1997].”
2. Pag. 28, penultima riga: si legga “tutti gli altri tipi di matrici derivano ...”, senza virgola tra “matrici” e “derivano”.
3. Pag. 29, Box 3.2: nella prima riga è sbagliato il numero delle colonne selezionate, il codice corretto è

```
> Y <- stack(studenti[,8:11])
> str(Y)

'data.frame': 232 obs. of  2 variables:
 $ values: num  3 2 4 3 2 2 3 4 3 3 ...
 $ ind   : Factor w/ 4 levels "d1","d2","d3",...: 1 1 1 1 1 1 1 1 1 1 ...

> Y$matricola <- studenti$matricola
> Y$genere <- studenti$genere
> Y$eta <- studenti$eta
> head(Y)

  values ind matricola genere eta
1      3  d1      110      m  23
2      2  d1      111      m  24
3      4  d1      143      f  23
4      3  d1      144      m  24
5      2  d1      159      m  19
6      2  d1      160      m  21
```

in questo modo vengo effettivamente selezionate le colonne d1, d2, d3, d4 coerentemente con quanto descritto nel testo. La stessa correzione nel numero di colonne va fatta nei successivi Box 3.3 e 3.4:

```
> round(cov(studenti[,8:11]),4)

      d1      d2      d3      d4
d1  1.3312 -0.1382 -0.0224 -0.0917
d2 -0.1382  1.2610  0.0048  0.3022
d3 -0.0224  0.0048  0.3799  0.0060
d4 -0.0917  0.3022  0.0060  0.9852

> round(cor(studenti[,8:11]),4)

      d1      d2      d3      d4
d1  1.0000 -0.1067 -0.0315 -0.0800
d2 -0.1067  1.0000  0.0070  0.2711
d3 -0.0315  0.0070  1.0000  0.0099
d4 -0.0800  0.2711  0.0099  1.0000
```

4. Pag. 29, Esempio 3.2, ultime tre righe: se si considera la correzione al punto 3 relativa al box 3.2 si legga coerentemente “... il soggetto con matricola 110, di genere m e di 23 anni,

alla domanda d1 ha risposto 3; nella riga 2 il soggetto con matricola 111, di genere m e di 24 anni, alla stessa domanda d1 ha risposto 2 e così via.”

5. Pag. 30, Esempio 3.3, terza riga: al posto di “... nel seguente modo:”, si legga “**come illustrato nel box 3.3**”.
6. Pagg. 31, box 3.3, e 32, box 3.4: l’arrotondamento è alla terza cifra decimale pertanto gli output corretti sono quelli riportati nella nota precedente.
7. Pagg. 30 (sez. **Matrice di covarianza**, riga 3) e 31 (sez. **Matrice di correlazione**, riga 3 e successive): tutte le “i-ma” e le “j-ma” vanno sostituite con “**i-esima**” e “**j-esima**”.
8. Pag. 37, comandi R: il secondo comando

```
> hist(donne$attegg,breaks=as.numeric(Q),col="gray(.8)")
```

non funziona, bisogna togliere gli apici alla funzione `gray()`

```
> hist(donne$attegg,breaks=as.numeric(Q),col=gray(.8))
```

9. Pag. 37, nona riga dal basso: al posto di “(chiamati baffi)“, si legga “**(chiamate baffi)**”
10. Pag. 37, nona riga dal basso: alla definizione “Le righe orizzontali esterne alla scatola (chiamate baffi) sono poste ad una distanza da Q1 e Q3 pari a 1.5 volte la distanza tra Q1 e Q3.” si aggiunga la seguente nota: In alcuni casi questi valori possono essere aggiustati con una formula leggermente diversa.

Capitolo 4

1. Pag. 50, va aggiunta una frase dopo la formula del campionamento senza reinserimento, si legga:

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} \sqrt{\left(\frac{N-n}{N-1}\right)}$$

in cui N indica la numerosità della popolazione.

Capitolo 5

1. Colori errati. Le figure sono ovviamente in bianco e nero, per cui sono considerate errate tutte le volte che si fa riferimento a linee o aree colorate (es. pag. 70 riga 5 “... **linea verticale rossa.**”)
2. Pag. 60, paragrafo dopo la formula (5.2): la parentesi quadra aperta va spostata dopo il nome Cohen, si legga “... **ripreso da Cohen [1994], ...**”
3. Pag. 63, riga 1: nella citazione il nome Oakes va fuori parentesi, si legga “... **Oakes [1986] ...**”.
4. Pag. 64, riga 3 al basso: al posto di “... x il numero di volte che esce il sei, ...”, si legga: “... **x il numero di volte in cui esce il sei, ...**”
5. Pag. 67, riga 8 dal basso: al posto di “... mentre la standarizzazione è ...”, si legga: “... **mentre la standardizzazione è ...**”
6. Pag. 68, box 5.2: negli ultimi tre comandi bisogna togliere gli apici alla funzione `gray()`:

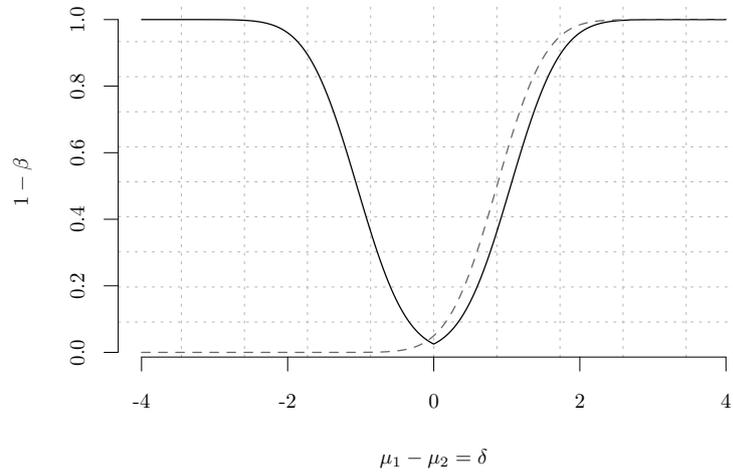
```
[...]
```

```

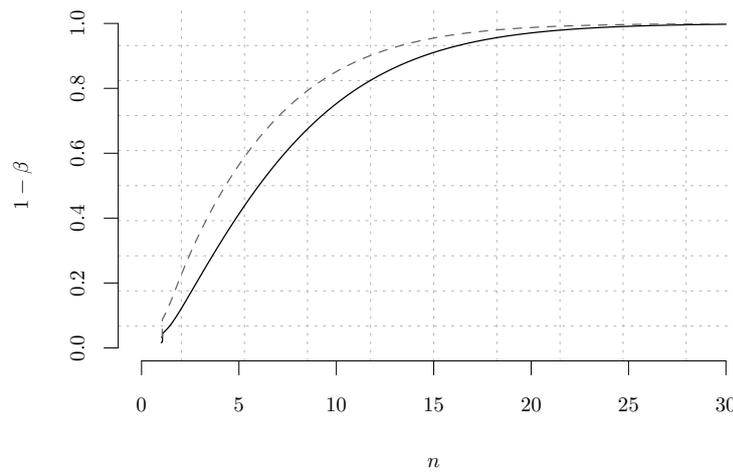
> hist(x,nclass=28,col=gray(.7),xlab="x",ylab="",main="[A]")
> hist(z,nclass=28,col=gray(.5),xlab="z",ylab="",main="[B]")
> hist(nx,nclass=28,col=gray(.3),xlab="n",ylab="",main="[C]")

```

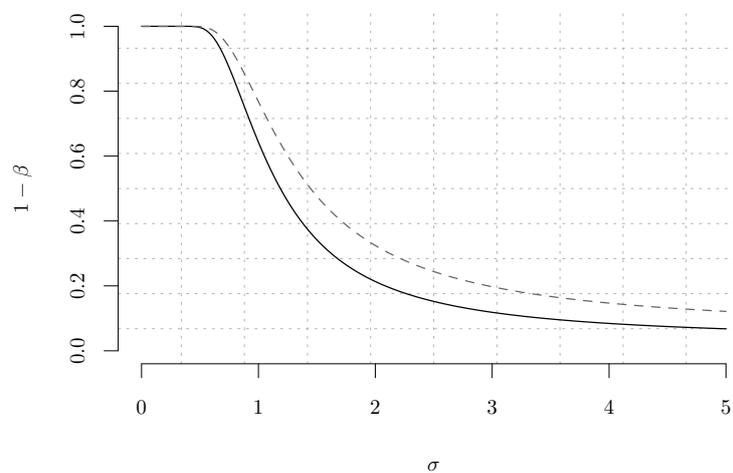
7. Pag. 72, la figura 5.5 è la seguente:



8. Pag. 73, la figura 5.6 è la seguente:



9. Pag. 74, la figura 5.7 è la seguente:



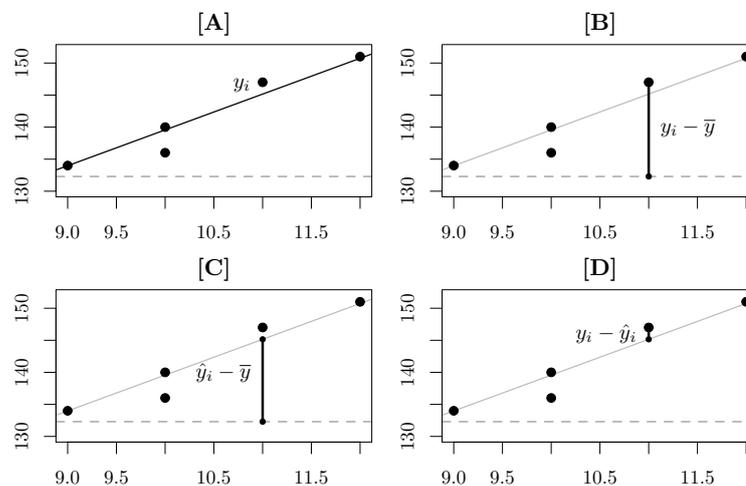
10. Pag. 80, legenda della figura 5.9: il pannello della distribuzione a posteriori è a destra, si legga pertanto: “Il pannello centrale rappresenta la funzione di verosimiglianza del risultato, il pannello a destra la distribuzione a posteriori.”

Capitolo 6

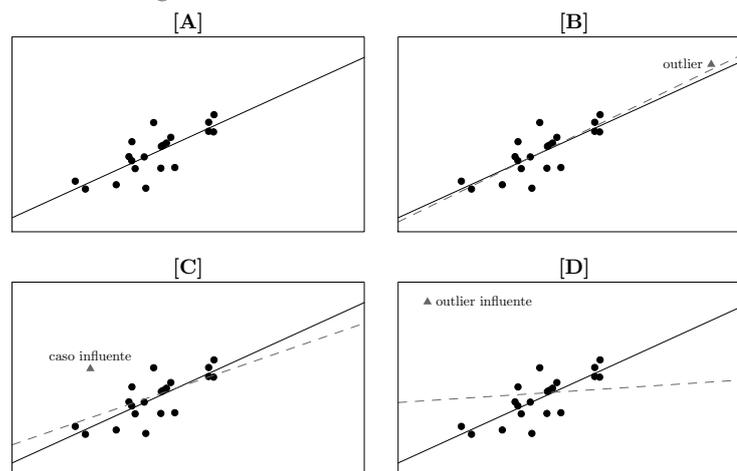
- Pag. 89, Esempio 6.2, sesta riga: si legga “... (la linea nera in fig. 6.2B) ...”
- Pag. 96, undicesima riga dal basso: si legga “... da triangoli neri.”
- Pag. 101, quarta riga dal fondo e box 6.6, il coefficiente di correlazione, coerentemente con quanto calcolato nel box 6.2, è negativo: $r = -0.12$. Non cambia comunque l'interpretazione della potenza associata (power = 0.12).

Capitolo 7

- Pag. 103, ultima riga: si legga “... osservazione i -esima, ...”.
- Pag. 106, la figura 7.2 è la seguente:

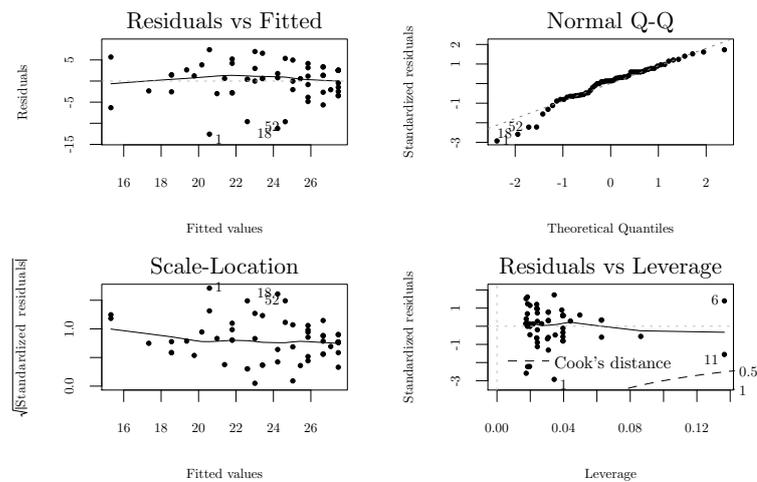


- Pag. 109, la figura 7.5 è la seguente:



- Pag. 109, riga 9, in coerenza con la figura, la linea del pannello [A] è continua; si legga “Il pannello [A] rappresenta ..., la linea continua rappresenta ...”
- Pag. 110, descrizione della figura 7.6: la linea continua nel pannello [B] è nera e non grigia.

6. Pag. 113, la figura 7.7 è la seguente:



- Pag. 116, ultima riga: al posto di "... la stessa rappresentata in rosso ..." si legga "... la stessa rappresentata in nero ...".
- Pag. 117, righe 1-2: al posto di "... in funzione delle ore di studio." si legga "... in funzione delle ore di frequenza."
- Pag. 119, riga 10 dal basso: al posto di "... ovvero che le ore di studio hanno ..." si legga "... ovvero che le ore di frequenza hanno ...".
- Pag. 119/120, l'espressione $\log(x)$ si intende come logaritmo naturale.

Capitolo 8

- Pag. 125, riga 4 dopo il box 8.4: " $1/0.33 = 3.03$ ".
- Pag. 126, sezione **Analisi con R**, terza riga: si veda nota 1 del Capitolo 2.
- Pag. 130, **Valutazione dell'evidenza**, riga 4: " $1/0.21 = 4.76$ ".
- Pag. 131, riga 1: si legga "... i valori associati ai parametri (i quadrati neri) ...".
- Pag. 131, penultima riga: le barre del grafico sono nere, si tenga presente che producendo lo stesso grafico in R sono rosse.

Capitolo 9

- Pag. 137, quarta riga dal basso: al posto di "figura 9.3" si legga "figura 9.2".
- Pag. 139, riga 5: il coefficiente di regressione β_1 corretto è 2.14 e non 15.56, che invece è l'intercetta del modello.
- Pag. 143, box 9.7: manca un numero nel p -value dell'ultima riga, si legga:

```
F-statistic: 46.12 on 3 and 254 DF, p-value: < 2.2e-16
```

- Pag. 143, elenco puntato alla fine: al posto di

- ...
- Modello 4: i due predittori dei modelli 1 e 2 senza effetto di interazione.
- Modello 5: i due predittori dei modelli 1 e 2 con effetto di interazione.

si legga

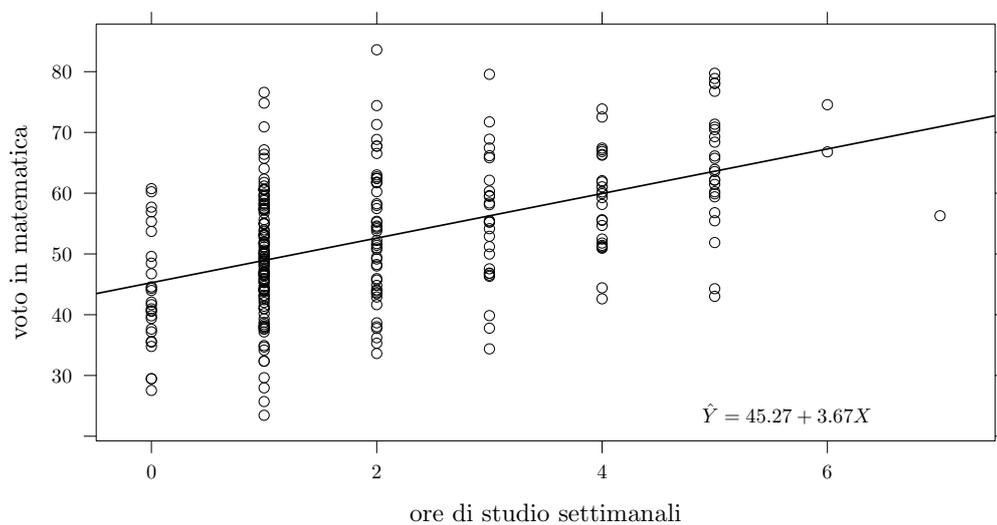
- ...
 - Modello 4: i due predittori dei modelli 2 e 3 senza effetto di interazione.
 - Modello 5: i due predittori dei modelli 2 e 3 con effetto di interazione.
5. Pag. 147, terzultima riga: al posto di “Nella sezione 8.1 ...” si legga “Nella sezione 7.4 ...”.

Capitolo 10

1. Pag. 156, terzo elemento dell’elenco puntato: la dimensione corretta della porzione residua è $p \times (p + q)$ che in questo caso equivale infatti a $1 \times (1 + 3) = 1 \times 4$.
2. Pag. 165: i nomi delle variabili `Organ.commit` e `Unit.perf.stand` vanno letti senza gli spazi: `Organa.commit` e `Unit.perf.stand`.
3. Pag. 171, penultima riga prima del box 10.7: si legga “... varianza riprodotta della i -esima variabile, ...”.
4. Pag. 172, terza riga dopo il box 10.8: si legga “... prendendo il modello analizzato nel box 10.4.” al posto di “... prendendo il modello 10.4’..”.
5. Pag. 173, sezione **Confronto tra modelli**, riga 4: errore di battitura, “risucire”, si legga “... utile riuscire ...”.
6. Pag. 173, sezione **Confronto tra modelli**, righe 6–7: si legga “... relativamente al modello analizzato nel box 10.4” al posto di “... relativamente al modello dell’esempio 10.4’.

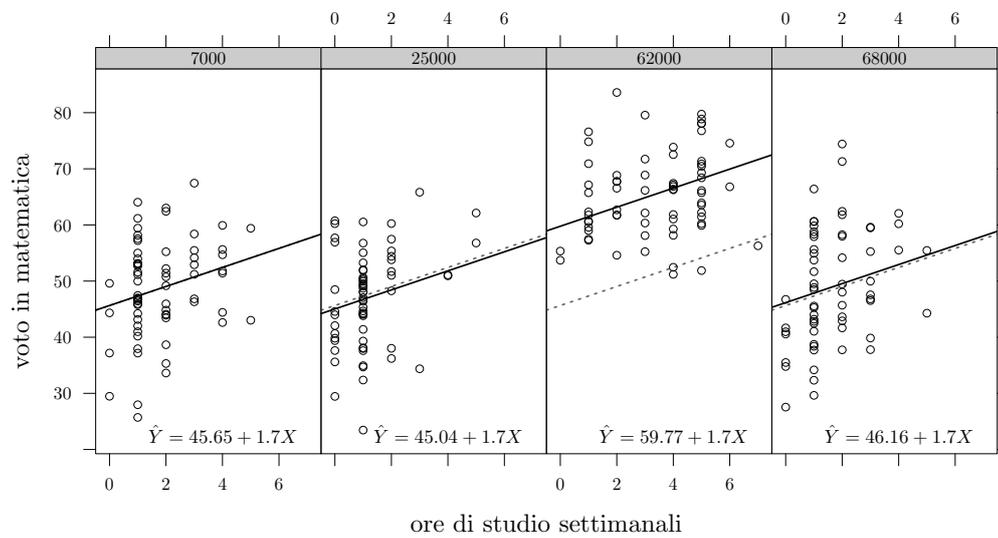
Capitolo 11

1. Pag. 183, figura 11.1, ultima riga della didascalia: al posto di “... intercetta e pendenza variabile” si legga “... intercetta e pendenza variabili”.
2. Pag. 185, figura 11.2; coerentemente con il risultato del box 11.3, la figura corretta è la seguente:

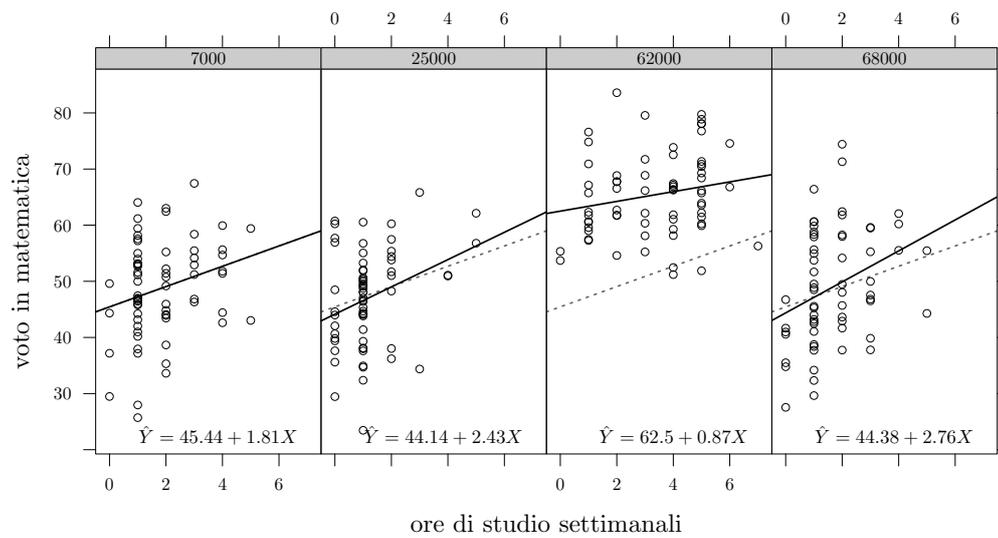


3. Pag. 185, riga 6 dal basso: manca una “a” nella frase “... relativa alla scuola ...”.

4. Pag. 186, figura 11.3; coerentemente con il risultato del box 11.4, la figura corretta è la seguente:



5. Pag. 187, figura 11.4; la figura corretta è la seguente:



6. Pag. 189, riga 5 dopo il sistema di equazioni, il primo decimale di 45.72 è sbagliato, si legga "... in cui il valore 45.62 non è altro che ..."

7. Pag. 189, box 11.6; il secondo valore riportato nel box manca del punto che separa il decimale, l'output corretto è il seguente:

```
> ranef(fit1)

$schid
      (Intercept)
7000          -3.420
25000         -3.980
62000          10.307
68000         -2.907

with conditional variances for "schid"
```

per cui il valore è **-3.98**, coerentemente con il valore corrispondente nel sistema di equazioni posto nella stessa pagina poco sopra.

8. Pag. 189, box 11.7; i valori riportati sono errati e non coerenti con il modello 11.3 cui fanno riferimento, ovvero con solo intercette variabili. L'output corretto si presenta nel modo seguente:

```
> coef(fit1)

$schid
      (Intercept) homework
7000           45.62    1.754
25000          45.06    1.754
62000           59.35    1.754
68000           46.13    1.754

attr(,"class")
[1] "coef.mer"
```

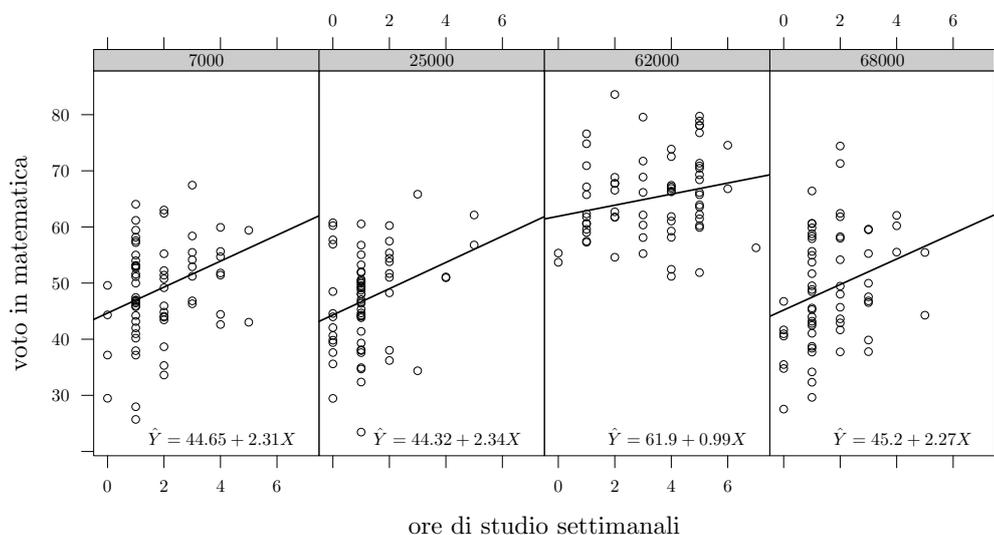
9. Pag. 190, riga 3 dal basso, il valore di $\hat{\sigma}_\epsilon = 8.648$ è diverso da quello riportato nel box 11.8, si legga: "... e $\hat{\sigma}_\epsilon = 8.468$ (variabilità residua) ..."
10. Pag. 191, riga 5: manca uno zero nel codice della scuola, si legga "... per la scuola 25000 ..."
11. Pag. 192, box 11.10; i valori riportati sono riferiti ad una vecchia versione del dataset `school`. L'output che si ottiene con la versione attuale si presenta nel modo seguente:

```
> coef(fit2)

$schid
      (Intercept) homework
7000           44.65    2.3140
25000          44.32    2.3398
62000           61.90    0.9856
68000           45.20    2.2720

attr(,"class")
[1] "coef.mer"
```

12. Pag. 192, figura 11.5; la figura è coerente con i valori stimati nel box 11.10 ma se si fa riferimento ai nuovi valori della correzione precedente la figura diventa la seguente:



13. Pag. 193, il box 11.11 presenta un output relativo a dati di una vecchia versione del pacchetto AData, con la versione corrente (5.4.4, del 18/03/2023) lo stesso box si presenta così:

```
> summary(fit1)
```

```
[...]
Random effects:
Groups   Name      Variance Std.Dev.
schid    (Intercept) 48.6     6.97
Residual                72.6     8.52
Number of obs: 260, groups: schid, 4

Fixed effects:
              Estimate Std. Error t value
(Intercept)   49.038     3.615    13.57
homework       1.754     0.395     4.44
[...]
```

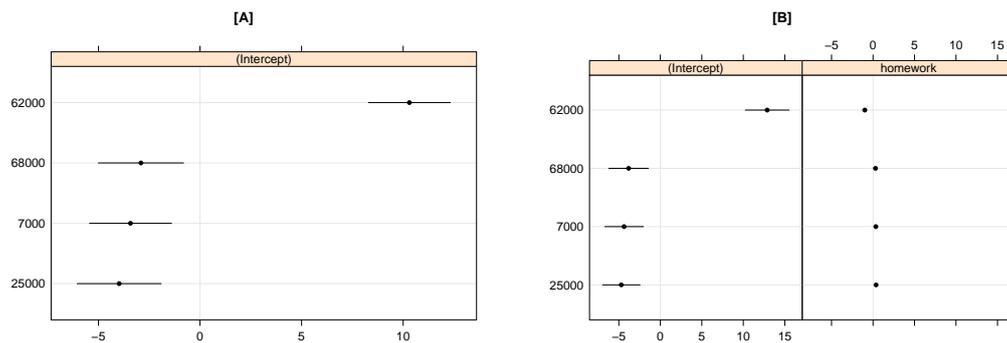
14. Pag. 193, la didascalia del box 11.11 deve fare riferimento al box 11.5 e non, come scritto, al box 11.3.
15. Pag. 193, **Significatività dei parametri**, righe 3-4, i valori dei $\hat{\beta}_0$ e $\hat{\beta}_1$ non corrispondono ai valori riportati nel box 11.11 in quanto riferiti ad una vecchia versione del data-frame `school`; si legga “... il valore di $\hat{\beta}_0 = 54.677$ rappresenta il punteggio atteso al test di matematica per gli studenti con 0 ore di studio, il valore di $\hat{\beta}_1 = 1.879$ indica ...”. Se però si considera la correzione precedente, riferita alla versione aggiornata, i valori sono corretti.
16. Pag. 194, terzultima riga: al posto di “... intercetta e pendenza variabile” si legga “... intercetta e pendenza variabili”.
17. Pag. 194, il box 11.12 deve avere le stime dei parametri coerenti con quelle del box 11.5 quindi:

```
> library(lmerTest)
> summary(lmer(math~homework+(1|schid),data=school))
```

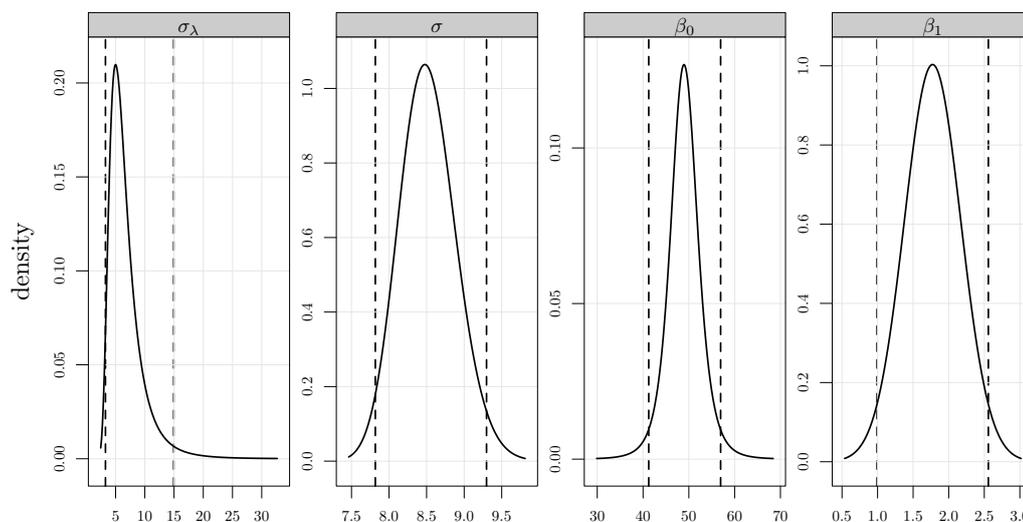
```
[...]
Random effects:
Groups   Name             Variance Std.Dev.
schid    (Intercept)  48.6    6.97
Residual                    72.6    8.52
Number of obs: 260, groups:  schid, 4

Fixed effects:
              Estimate Std. Error      df t value Pr(>|t|)
(Intercept)  49.038      3.615    3.265  13.57  0.00055 ***
homework      1.754      0.395  257.640   4.44  1.3e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
[...]
```

18. Pag. 196, **Analisi degli effetti random**, riga 15; manca la parentesi chiusa a fine riga, si legga “(62000)”.
19. Pag. 197, coerentemente con i risultati dei box 11.6 e 11.9, la figura 11.6 è la seguente, pertanto il testo a pagina 196 va inteso come corretto in quanto fa riferimento a questa:



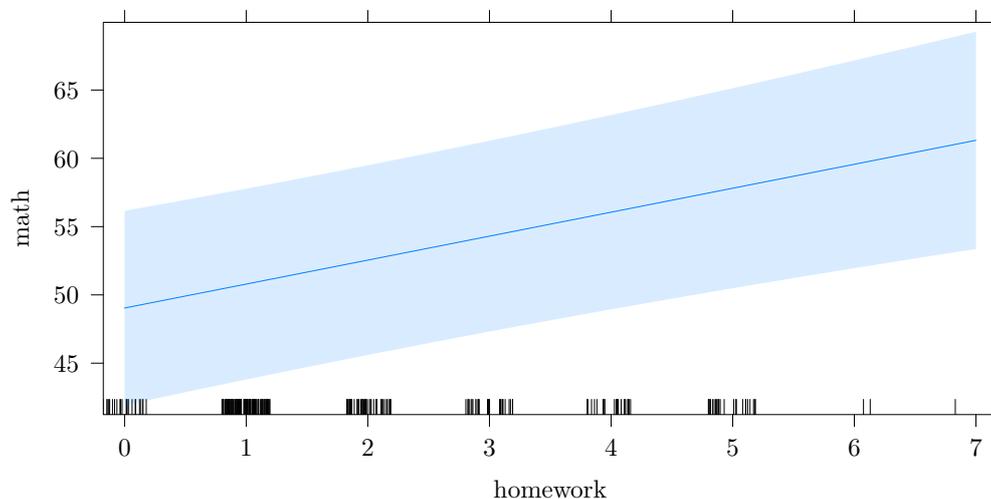
20. Pag. 199, la figura 11.8 corretta e coerente con i valori riportati nel box 11.15 è la seguente:



21. Pag. 199, box 11.15: il primo valore superiore riportato 14.985 è errato, si legga “14.945”.

22. Pag. 200, la figura 11.9 è corretta ma si riferisce ad una versione vecchia del pacchetto ADati; con la versione 5.4.4 del pacchetto ADati la figura è leggermente diversa ed appare come segue:

homework effect plot



23. Pag. 200, riga 1; l’arrotondamento del valore superiore dell’intervallo di confidenza è sbagliato rispetto al valore riportato nel box 11.15, ma corretto se si tiene conto della correzione precedente.

24. Pag. 202, seconda riga, il valore del Bayes Factor “relativo al fattore random schid (5.197e+33)” non coincide con quello riportato nel box 11.17; si legga “... relativo al fattore random schid (2.437e+27) ...”.

Ultimo aggiornamento: 26/06/2023.

Le segnalazioni di errori presenti nel manuale vanno inviate a: massimiliano.pastore@unipd.it

Per le correzioni si ringraziano i seguenti studenti: Sonia Ali, Alberto Arletti, Nicola Bertozzi, Valeria Bianco, Miriana Bressan, Salvador Casara Bruno Gabriel, Debora Brusini, Rosaria Capasso, Anna Giorgia Carollo, Valentina Cinelli, Giada Ciresola, Stefano Dalla Bona, Chiara Dalla Costa, Nicola Furlanis, Giulia Gringiani, Francesca Eugenia Ialuna, Gianluca Lorenzon, Alberto Lucchetta, Veronica Mafrika, Eleonora Mannino, Sofia Mariani, Giorgia Martini, Arianna Mercurio, Luca Muhlbauer, Enrica Murgia, Arturo Palma, Federica Paolini, Elisa Pesenti, Anna Piazza, Riccardo Poggioli, Giulia Quintarelli, Emma Rotolo, Michele Saggio, Vito Fabio Sblano, Marcella Stelluto, Francesca Tacchella, Laura Tonin, Maria Uggeri, Eleonora Zamuner, Giulia Maria Zoratti.