

R workshop #3: Modelli a effetti misti

Nicola Romanò

Introduzione

Come abbiamo visto nelle lezioni, sia le misure ripetute che i design annidati pongono alcune sfide all'analisi dei dati. In particolare, questi tipi di design sono problematici in quanto contengono osservazioni correlate.

Ad esempio, nel caso di un design di misure ripetute, in cui misuriamo un determinato parametro nello stesso soggetto in momenti diversi, è probabile che ciascuna delle misurazioni dipenda (ovvero sia correlata) dal valore precedente.

Uno dei modi migliori per gestire questi dati è l'uso di modelli a effetti misti, un'estensione del modello lineare che consente di tenere conto di effetti casuali nel modello.

Per il contesto di questo workshop avremo bisogno di usare il pacchetto `nlme`¹. Come al solito, questo pacchetto può essere installato usando il seguente comando:

```
install.packages("nlme")
```

L'installazione deve essere eseguita solo una volta. Il pacchetto può quindi essere utilizzato dopo il caricamento utilizzando:

```
library(nlme)
```

Nota: i modelli a effetti misti sono uno strumento piuttosto complesso da usare. Lo scopo di questo workshop è di presentarti questi modelli e darti un'idea di come usarli per alcune analisi di base. Questa è ben lungi dall'essere una presentazione completa², e l'analisi di progetti più complessi può non essere così banale!

Obiettivi formativi

Dopo aver completato questo workshop sarai in grado di:

- Creare un modello a effetti misti per esplorare semplici misure ripetute o design annidati.
- Interpretare l'output di un modello a effetti misti

Design a misure ripetute

Consideriamo un semplice design a misure ripetute.

Siamo interessati a valutare l'effetto dell'agonista dopaminergico bromocriptina sulla crescita dei prolattinomi (adenomi pituitari che

¹ L'altro pacchetto comunemente usato è `lme4`, con la funzione `lmer`. La sintassi è leggermente diversa ma si applica un ragionamento simile

² Se sei veramente interessato ad ampliare le tue conoscenze in questo settore, un ottimo libro è "Mixed Effects Models in S and S-PLUS" di Pinheiro e Bates (questo libro usa S, che è il linguaggio da cui è derivato R; il codice S può essere eseguito in R senza troppi problemi)

secernono prolattina). Prendiamo 45 topi e li assegnamo casualmente a uno dei tre gruppi (15 / gruppo).

Il gruppo di controllo riceve un intervento chirurgico simulato, mentre gli altri due gruppi ricevono un impianto sottocutaneo contenente 1 o 10 mg di bromocriptina. Seguiamo quindi la crescita dell'adenoma nel tempo misurando i livelli plasmatici di prolattina (PRL).

Abbiamo due effetti fissi in questo design: tempo e trattamento, e un effetto casuale, il topo. Poiché ogni topo viene misurato più volte, le misure provenienti dallo stesso animale non saranno indipendenti, da cui la necessità di un modello a effetti misti.

Per questo esempio, iniziamo caricando il file `bromocriptine-workshop3.csv`

```
bromocriptine <- read.csv("bromocriptine-workshop3.csv")
```

```
summary(bromocriptine)
```

##	Time	Mouse	PRL	Group
##	Min. : 0	Min. : 1	Min. : 0.4252	Bromo1 :90
##	1st Qu.: 30	1st Qu.:12	1st Qu.: 31.4421	Bromo10:90
##	Median : 75	Median :23	Median : 78.0624	CTRL :90
##	Mean : 75	Mean :23	Mean :148.6180	
##	3rd Qu.:120	3rd Qu.:34	3rd Qu.:222.3295	
##	Max. :150	Max. :45	Max. :775.6684	

Come al solito, esplora i dati, prova a tracciarli e vedere possibili relazioni tra variabili ³.

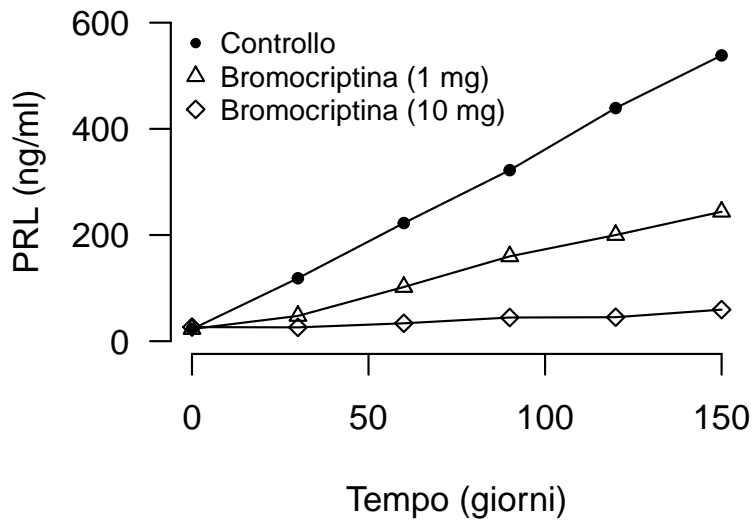
Inoltre, riorganizziamo l'ordine dei livelli nel fattore `Group` in modo che il gruppo di controllo venga utilizzato come livello di riferimento.

```
bromocriptine$Group <- factor(bromocriptine$Group,
  levels = c("CTRL", "Bromo1", "Bromo10"))
```

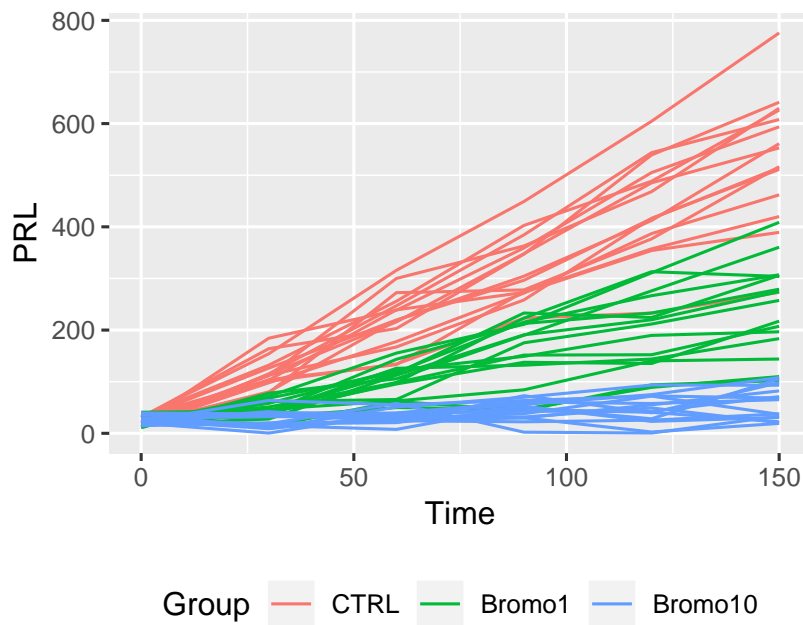
Qui sotto ho tracciato i valori medi di PRL nel tempo nei tre gruppi ⁴

³ Fai riferimento ai workshop precedenti se non ricordi come

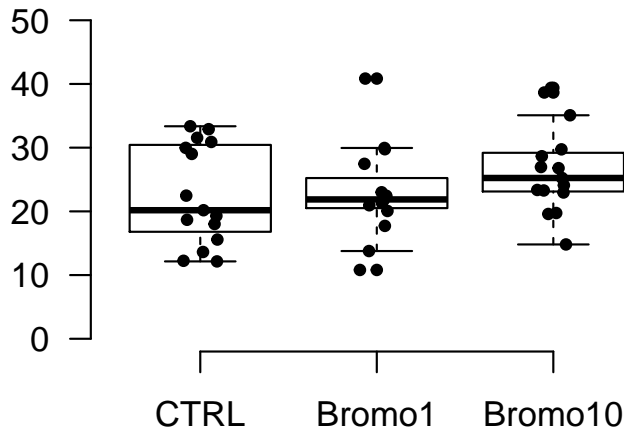
⁴ Come sfida, prova a riprodurre questo grafico!



È chiaro che esiste un effetto dose-dipendente del farmaco e che, nel complesso, il cambiamento dei livelli di PRL nel tempo può essere studiato utilizzando un modello lineare. Se osserviamo le risposte dei singoli topi possiamo vedere che iniziano più o meno allo stesso livello, ma poi aumentano in modi diversi (con pendenze diverse) a seconda del trattamento, ma anche all'interno del trattamento.



Possiamo vedere come le differenze individuali tra i topi non siano molto pronunciate al tempo 0.



Costruiamo ora il nostro modello ⁵. Modelliamo il gruppo e il tempo come effetti fissi, così come la loro interazione, e modelliamo *Mouse* come fattore casuale. Creiamo un modello di pendenza casuale, dal momento che ciò che sembra essere variabile tra i soggetti è la pendenza. Specifichiamo l'effetto casuale come `Time - 1 | Mouse`, nel senso che vogliamo usare il mouse come un effetto casuale, e vogliamo avere pendenze casuali sul Tempo ma non intercette casuali (da cui il -1).

⁵ Ricorda di caricare prima il pacchetto nlme!

```
model <- lme(PRL ~ Group * Time, data = bromocriptine, random = ~Time - 1 | Mouse)
summary(model)
```

```
## Linear mixed-effects model fit by REML
## Data: bromocriptine
##      AIC      BIC    logLik
## 2564.473 2593.08 -1274.236
##
## Random effects:
## Formula: ~Time - 1 | Mouse
##           Time Residual
## StdDev: 0.54899 20.62831
##
## Fixed effects: PRL ~ Group * Time
##              Value Std.Error DF   t-value p-value
## (Intercept)  17.239483  3.854825 222    4.472183  0.0000
## GroupBromo1  -3.546567  5.451545  42   -0.650562  0.5189
```

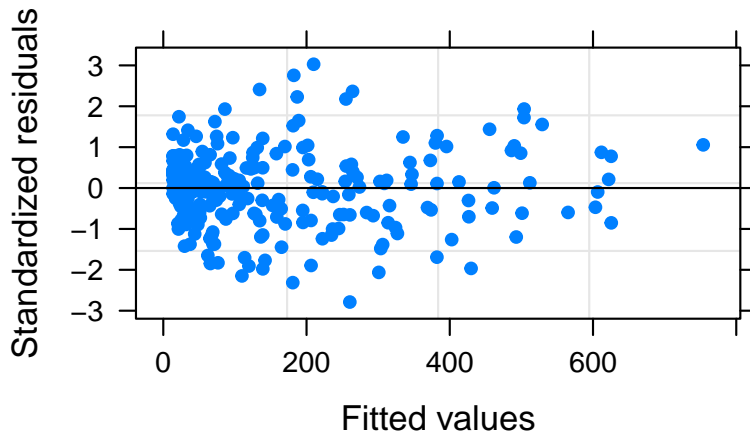
```
## GroupBromo10      5.390355  5.451545  42   0.988776  0.3284
## Time              3.466547  0.147966 222  23.428049  0.0000
## GroupBromo1:Time -1.924222  0.209255 222  -9.195581  0.0000
## GroupBromo10:Time -3.244864  0.209255 222 -15.506739  0.0000
## Correlation:
##              (Intr) GrpBr1 GrpB10 Time   GrB1:T
## GroupBromo1    -0.707
## GroupBromo10   -0.707  0.500
## Time           -0.237  0.167  0.167
## GroupBromo1:Time 0.167 -0.237 -0.118 -0.707
## GroupBromo10:Time 0.167 -0.118 -0.237 -0.707  0.500
##
## Standardized Within-Group Residuals:
##           Min           Q1           Med           Q3           Max
## -2.78846661 -0.59600115  0.02844046  0.52908281  3.02785483
##
## Number of Observations: 270
## Number of Groups: 45
```

Questo è un modello abbastanza complesso, le parti importanti che vogliamo vedere nel sommario sono:

- AIC, BIC e logLik: sono misure di bontà del fit. Sia per AIC che per BIC più piccolo il valore, migliore è il fit; per la `_log likelihood`, più alta è meglio.
- La parte dell'effetto casuale ci dà le deviazioni standard per gli effetti casuali e per i residui.
- La parte a effetti fissi ci fornisce le stime per i coefficienti del modello. In questo caso, l'intercetta 17.23 è il livello PRL per un topo nel gruppo di controllo (poiché questo è il livello di riferimento della nostra variabile di gruppo), al tempo 0. Gli altri coefficienti sono interpretati come abbiamo visto nella lezione #2. Si noti che esiste, come previsto, una forte interazione tra tempo e trattamento, cioè PRL varia in modo diverso nel tempo per trattamenti diversi.
- Una tabella di correlazione. Per la maggior parte, se non tutte, delle situazioni con cui ti occuperai, puoi tranquillamente ignorarla.
- Un riassunto della distribuzione dei residui
- Il numero di osservazioni e gruppi. Puoi utilizzare questi numeri per verificare di aver correttamente specificato la struttura dell'esperimento nel modello. Abbiamo un totale di 270 osservazioni (puoi verificare eseguendo `ncol(bromocriptina)`) e 45 gruppi (cioè unità sperimentali), corrispondenti ai 45 topi. Questo ci dice che R ha capito che quelle 270 osservazioni non sono indipendenti, ma provengono da 45 topi, quindi più osservazioni sono associate allo stesso topo.

Possiamo controllare come di consueto la distribuzione dei residui rispetto ai valori fittati (sebbene l'output abbia un formato leggermente diverso rispetto a `lm`)

```
plot(model, pch = 20)
```

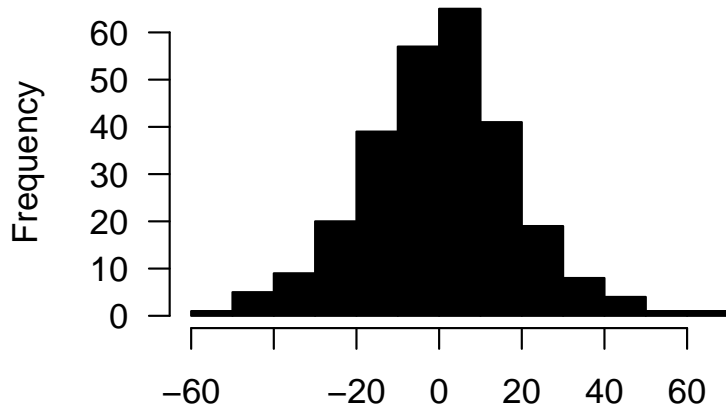
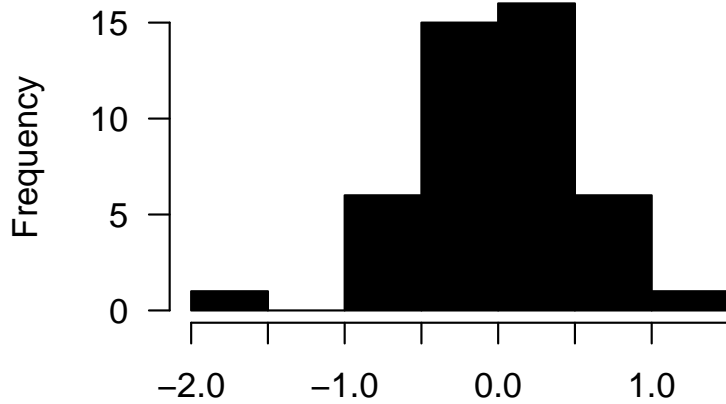


Analizziamo ora la distribuzione degli effetti casuali e confrontiamola con quella dei residui. La funzione `random.effects` restituisce gli effetti casuali. Possiamo combinarla con `hist` e possiamo vedere che gli effetti casuali provengono da una distribuzione normale, come previsto, con una varianza diversa da quella dei residui (notare la diversa scala x).

```
head(random.effects(model))
```

```
##           Time
## 1  0.49678070
## 2  0.44133005
## 3 -0.05455666
## 4 -0.31447629
## 5  0.58834643
## 6 -1.55698489
```

```
par(mfrow = c(2, 1))
hist(random.effects(model)$Time, main = "", col = "black",
     xlab = "Effetti casuali", las = 1)
hist(resid(model), main = "", col = "black", xlab = "Residui",
     las = 1)
```



Possiamo anche verificare che R abbia fittato correttamente un modello di pendenza casuale.

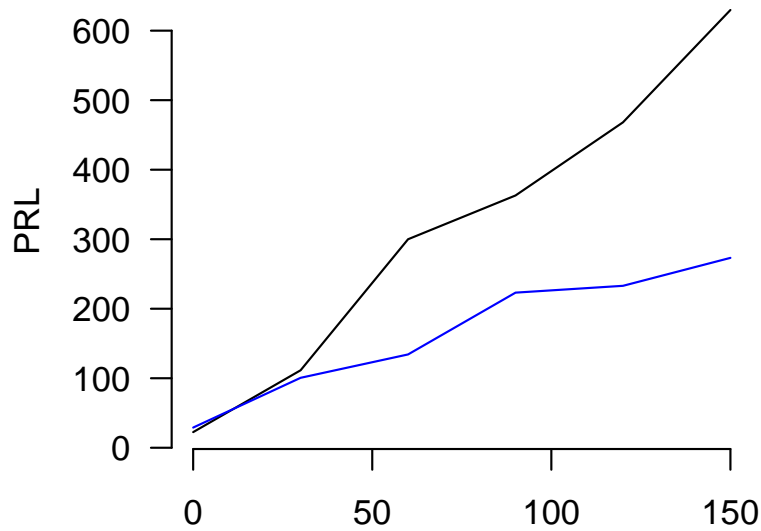
```
head(coef(model))
```

```
##      (Intercept) GroupBromo1 GroupBromo10      Time GroupBromo1:Time
## 1      17.23948   -3.546567    5.390355  3.963328   -1.924222
## 2      17.23948   -3.546567    5.390355  3.907877   -1.924222
## 3      17.23948   -3.546567    5.390355  3.411991   -1.924222
## 4      17.23948   -3.546567    5.390355  3.152071   -1.924222
## 5      17.23948   -3.546567    5.390355  4.054894   -1.924222
## 6      17.23948   -3.546567    5.390355  1.909562   -1.924222
##      GroupBromo10:Time
## 1          -3.244864
## 2          -3.244864
## 3          -3.244864
## 4          -3.244864
## 5          -3.244864
## 6          -3.244864
```

Nota come tutti i coefficienti sono uguali per tutti gli animali, ma

la pendenza per il tempo è stata cambiata per ciascun animale. Ad esempio, tracciamo i dati per i topi 1 e 6. Questi sono entrambi topi di controllo, ma hanno profili abbastanza diversi.

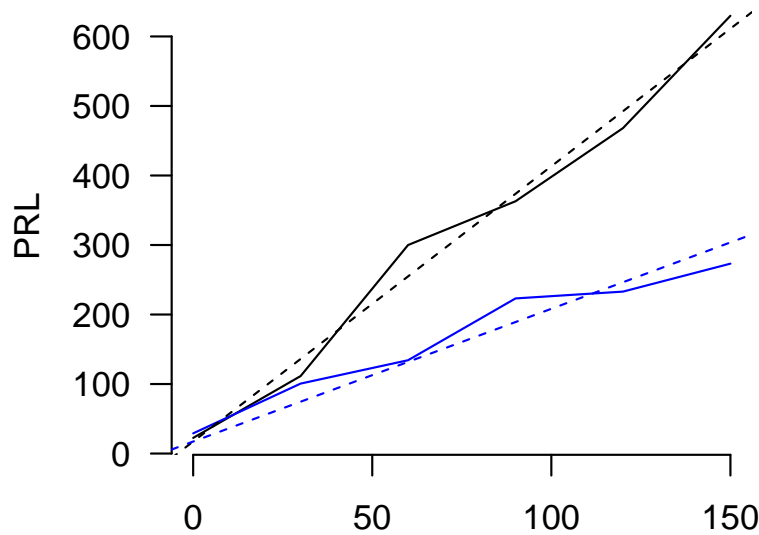
```
plot(PRL ~ Time, bromocriptine, subset = bromocriptine$Mouse ==
     1, t = "l", las = 1, bty = "n")
lines(PRL ~ Time, bromocriptine, subset = bromocriptine$Mouse ==
     6, col = "blue")
```



Infatti, guardando le pendenze fittate, sono 3.96 e 1.91, mentre l'intercetta è 17.24 per entrambi.

Possiamo usare `abline` per aggiungere le due rette fittate al grafico.

```
abline(17.24, 3.96, lty = "dashed") # Mouse 1
abline(17.24, 1.91, lty = "dashed") # Mouse 6
```

Infine, usiamo `emmeans` per confrontare i tre diversi trattamenti. Qui usiamo un argomento extra (`cov.reduce = range`) per dire a `emmeans` di guardare agli estremi `[0, 150]` dell'intervallo di tempo. Se non usiamo questo argomento, `emmeans` `epairs` effettueranno confronti solo nel tempo medio (cioè, 75).

```
library(emmeans)

marginals <- emmeans(model, ~Group * Time, cov.reduce = range)
pairs(marginals, by = "Time")

## Time = 0:
## contrast      estimate      SE df t.ratio p.value
## CTRL - Bromo1    3.5466  5.451545 42   0.651  0.7931
## CTRL - Bromo10  -5.3904  5.451545 42  -0.989  0.5879
## Bromo1 - Bromo10 -8.9369  5.451545 42  -1.639  0.2407
##
## Time = 150:
## contrast      estimate      SE df t.ratio p.value
## CTRL - Bromo1   292.1799 30.559605 42   9.561 <.0001
## CTRL - Bromo10  481.3392 30.559605 42  15.751 <.0001
## Bromo1 - Bromo10 189.1593 30.559605 42   6.190 <.0001
##
## P value adjustment: tukey method for comparing a family of 3 estimates
```

La nostra analisi post-hoc ha rivelato che c'è una differenza significativa tra tutti i gruppi al tempo 150, ma non al momento 0. Se vuoi testare su tutti i tempi, puoi usare `cov.reduce = unique`.

Design annidata: un esperimento “split plot”

Analizzeremo ora un esperimento *split plot* come esempio di design annidato ⁶.

Per questo esempio utilizzeremo il set di dati `Oats` fornito con il pacchetto `nlme` (questo viene automaticamente caricato quando si carica `nlme`).

Questi dati provengono da un esempio di split plot design di Yates (1935).

La struttura di trattamento utilizzata nell’esperimento era un fattoriale completo 3 x 4, con tre varietà di avena e quattro concentrazioni di azoto. Le unità sperimentali sono state disposte in sei blocchi, ciascuno con tre interi lotti suddivisi in quattro sotto-lotti. Le varietà di avena sono state assegnate casualmente ai lotti interi e alle concentrazioni di azoto nei sotto-lotti. Tutte e quattro le concentrazioni di azoto sono state utilizzate su ogni lotto.

In questo esperimento abbiamo sei appezzamenti di terra, ognuno dei quali è diviso in 3 lotti, ciascuno dei quali è ulteriormente suddiviso in 4 sotto-lotti. In ogni lotto coltiviamo una varietà di avena (scelta a caso), e in ogni sotto-lotto usiamo una diversa concentrazione di azoto (l’azoto è usato come fertilizzante). Poiché abbiamo 4 sotto-lotti, utilizziamo 4 diverse concentrazioni di azoto. Le varietà di avena sono chiamate Golden Rain, Marvelous e Victory. La nostra variabile di output è la resa del lotto specifica. Questo è un disegno fattoriale completo 3 x 4, il che significa che ciascuna delle 3 varietà è stata trattata con ciascuna delle 4 combinazioni di azoto (quindi è “completo”, poiché tutte le possibili combinazioni sono prese in considerazione).

Ci chiediamo se 1) l’azoto ha un effetto sulla resa e 2) se le diverse varietà danno rese diverse 3) se c’è un’interazione tra i due.

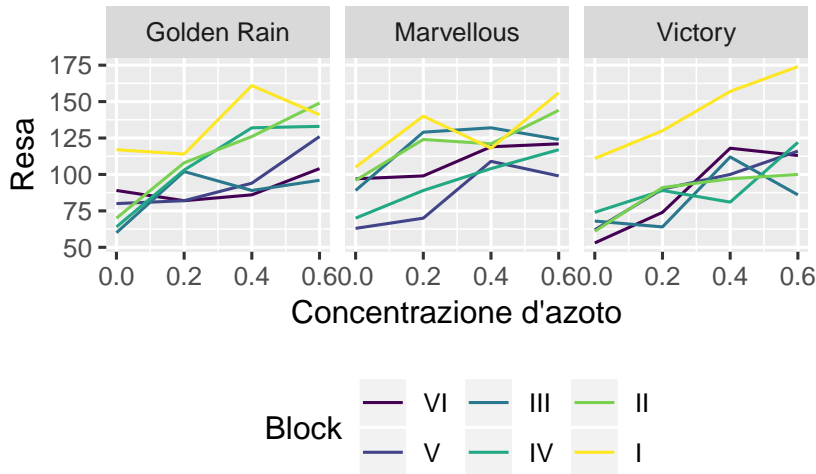
Iniziamo come al solito esplorando e tracciando il set di dati

```
summary(Oats)
```

##	Block	Variety	nitro	yield
##	VI :12	Golden Rain:24	Min. :0.00	Min. : 53.0
##	V :12	Marvellous :24	1st Qu.:0.15	1st Qu.: 86.0
##	III:12	Victory :24	Median :0.30	Median :102.5
##	IV :12		Mean :0.30	Mean :104.0
##	II :12		3rd Qu.:0.45	3rd Qu.:121.2
##	I :12		Max. :0.60	Max. :174.0

⁶ Questo esempio è tratto da Pinheiro e Bates, 2000

Puoi produrre diversi grafici interessanti con questi dati e ti invito a farlo. Questo è un grafico che riassume la relazione tra le diverse variabili.



Possiamo vedere chiare differenze tra lotto e lotto, e sembra che il rendimento aumenti approssimativamente linearmente con la concentrazione di azoto.

Ora dobbiamo pensare a quali sono gli effetti casuali e come sono nidificati. Abbiamo 6 appezzamenti, ciascuno diviso in 3 lotti (uno per varietà) e ciascuno diviso in 4 sotto-lotti (uno per ogni concentrazione di azoto). Poiché otteniamo solo una misura di rendimento da ciascun sotto-lotto, non considereremo questo come un effetto casuale, ma considereremo i lotti annidati all'interno degli appezzamenti come i nostri effetti casuali.

Creiamo il nostro modello!

```
model.oats <- lme(yield ~ nitro * Variety, data = Oats,
  random = ~1 | Block/Variety)
```

Nota che abbiamo creato un modello di intercettazione casuale (guarda il grafico qui sopra e pensa perché l'abbiamo fatto!). Inoltre, stiamo usando appezzamenti (**Block**) e lotti (usiamo **Variety**, poiché ogni lotto è associato ad una varietà) come i nostri effetti casuali, annidati l'uno nell'altro. Nota che elenchiamo prima i fattori più grandi, quindi quelli contenuti all'interno. Se avessimo più livelli di nidificazione, potremmo indicarli utilizzando più / (ad esempio Città / Scuola / Classe).

Vediamo il sommario ⁷

```
summary(model.oats)
```

```
## Linear mixed-effects model fit by REML
## Data: Oats
```

⁷ Per un output più conciso puoi anche usare `anova(model.oats)`

```

##           AIC           BIC      logLik
##    581.2372 600.9441 -281.6186
##
## Random effects:
## Formula: ~1 | Block
##           (Intercept)
## StdDev:      14.64487
##
## Formula: ~1 | Variety %in% Block
##           (Intercept) Residual
## StdDev:      10.39931 12.99039
##
## Fixed effects: yield ~ nitro * Variety
##
##              Value Std.Error DF   t-value p-value
## (Intercept)    81.90000  8.570714 51  9.555796  0.0000
## nitro          75.33333 11.858548 51  6.352661  0.0000
## VarietyMarvellous  8.51667  8.684675 10  0.980655  0.3499
## VarietyVictory    -8.60000  8.684675 10 -0.990250  0.3454
## nitro:VarietyMarvellous -10.75000 16.770519 51 -0.641006  0.5244
## nitro:VarietyVictory    5.75000 16.770519 51  0.342864  0.7331
## Correlation:
##              (Intr) nitro  VrtyMr VrtyVc ntr:VM
## nitro          -0.415
## VarietyMarvellous -0.507  0.410
## VarietyVictory    -0.507  0.410  0.500
## nitro:VarietyMarvellous  0.294 -0.707 -0.579 -0.290
## nitro:VarietyVictory    0.294 -0.707 -0.290 -0.579  0.500
##
## Standardized Within-Group Residuals:
##              Min           Q1           Med           Q3           Max
## -1.78878623 -0.64954429 -0.06301223  0.57818033  1.63463810
##
## Number of Observations: 72
## Number of Groups:
##              Block Variety %in% Block
##              6           18

```

Possiamo vedere un chiaro effetto dell'azoto, ma non della varietà e nessuna interazione significativa. Inoltre, si noti che l'output è leggermente diverso dall'esempio precedente in quanto: - Ci dice che ci sono 72 osservazioni, da 18 lotti (codificati come varietà) annidati in 6 appezzamenti. - Le deviazioni standard nella parte dell'effetto casuale del sommario sono per intercetta e residuo piuttosto che pendenza e residuo, poiché stiamo adattando un modello ad intercetta casuale.

Poiché i termini di interazione non sono significativi, possiamo tralasciarli e ottenere un modello più semplice

```

model.oats.2 <- lme(yield ~ nitro + Variety, data = Oats, random = ~1 | Block/Variety)

summary(model.oats.2)

```

```
## Linear mixed-effects model fit by REML
## Data: Oats
##      AIC      BIC    logLik
## 592.8918 608.4283 -289.4459
##
## Random effects:
## Formula: ~1 | Block
##      (Intercept)
## StdDev:    14.64482
##
## Formula: ~1 | Variety %in% Block
##      (Intercept) Residual
## StdDev:    10.43758 12.86697
##
## Fixed effects: yield ~ nitro + Variety
##              Value Std.Error DF   t-value p-value
## (Intercept)   82.40000  8.058510 53 10.225215  0.0000
## nitro         73.66667  6.781487 53 10.862908  0.0000
## VarietyMarvellous 5.29167  7.078909 10  0.747526  0.4720
## VarietyVictory   -6.87500  7.078909 10 -0.971195  0.3544
## Correlation:
##              (Intr) nitro  VrtyMr
## nitro         -0.252
## VarietyMarvellous -0.439  0.000
## VarietyVictory   -0.439  0.000  0.500
##
## Standardized Within-Group Residuals:
##      Min      Q1      Med      Q3      Max
## -1.62947645 -0.65841389 -0.07207178  0.55784903  1.71462741
##
## Number of Observations: 72
## Number of Groups:
##      Block Variety %in% Block
##      6          18
```

La conclusione del modello più semplice è essenzialmente la stessa: un effetto significativo dell'azoto ma non della varietà. Inoltre, le stime per gli effetti casuali hanno avuto solo modifiche minime (guarda le deviazioni standard per i residui e per l'intercetta).

Possiamo verificare che alcune regressioni abbiano senso tracciando le linee di regressione per diversi lotti.

Abbiamo solo scalfito la superficie di come usare modelli ad effetti misti, e c'è molto di più. Per lo scopo di questo corso ciò che abbiamo fatto è più che sufficiente, ma se sei interessato ci sono molte risorse in giro per diventare un esperto in questo argomento!