I modelli lineari generali

Capitolo 18

Analisi statistica dei dati biologici

Modelli lineari generali (GLM)

I modelli lineari generali prevedono che una variabile risposta Y può essere descritta da un **modello lineare** più **l'errore casuale**.

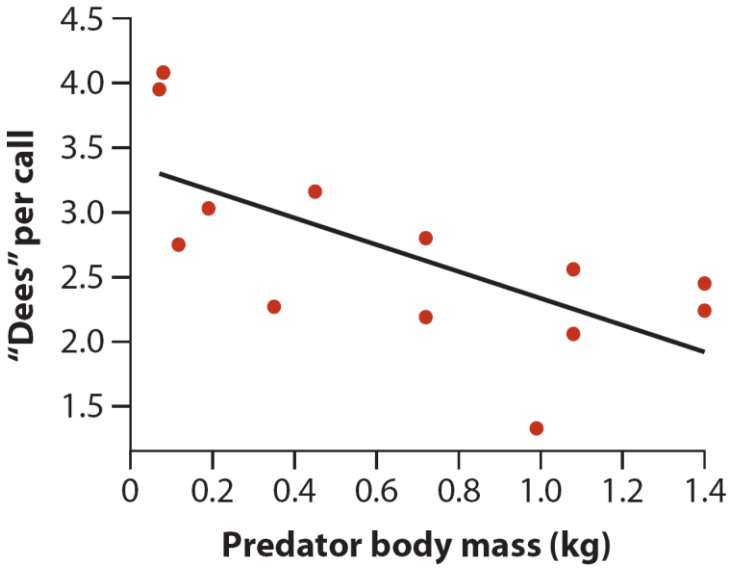
Il **modello** è la rappresentazione matematica della relazione tra una variabile risposta e una, o più, variabili esplicative (ad es $Y=\alpha+\beta X$).

L'**errore casuale** rappresenta la dispersione delle misure di Y rispetto ai valori previsti dal modello .

Il caso della regressione lineare

- Il caso della regressione lineare:
  - **Modello**: Y=α+βX
  - **Errore casuale**: residui

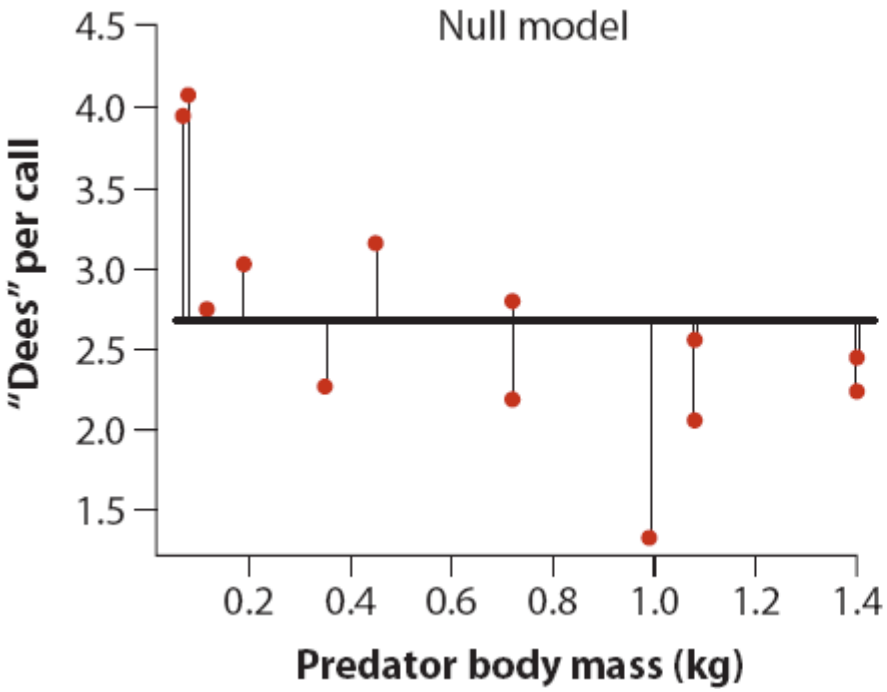
Nell'esempio delle cince la relazione era:
DEE=α+β(PREDMASS)



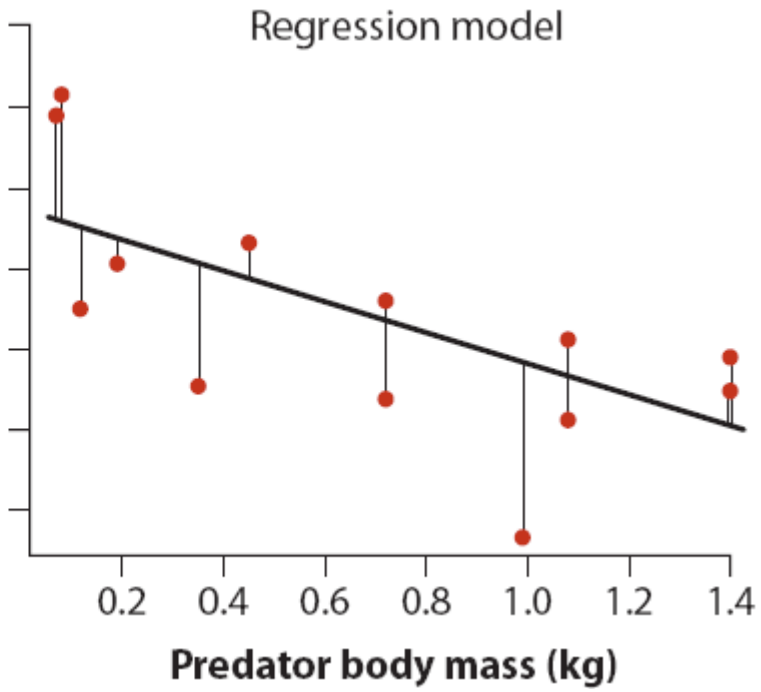
Il caso della regressione lineare

Usando i modelli lineari generali, possiamo confrontare il modello di regressione ($\beta \neq 0$) con il modello nullo ($\beta = 0$).

$Y = \alpha$



$Y = \alpha + \beta X$



Il caso della regressione lineare

Nel caso del test t sulle pendenze verificavamo se β era significativamente diverso da 0.

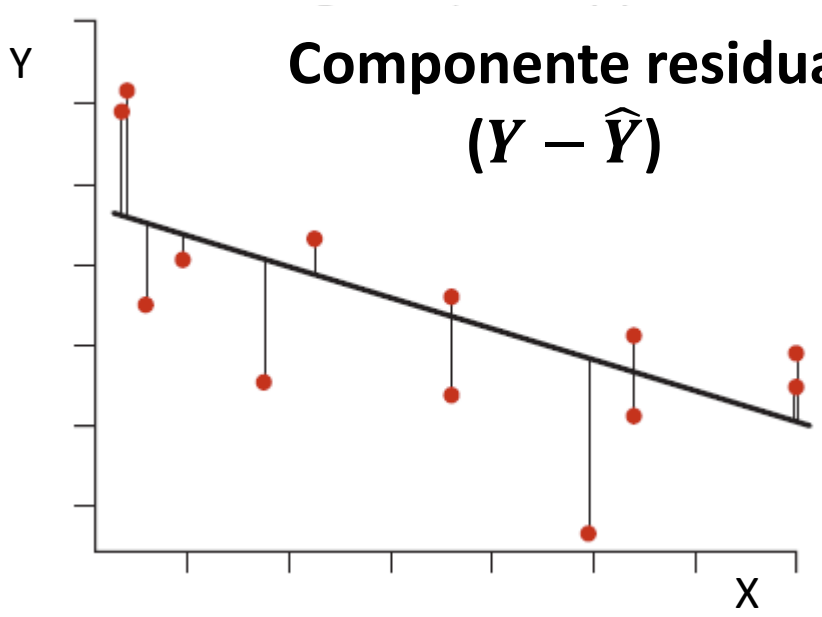
Nei modelli lineari generali bisogna verificare se il modello di regressione si adatti ai dati ***significativamente*** meglio rispetto al modello nullo.

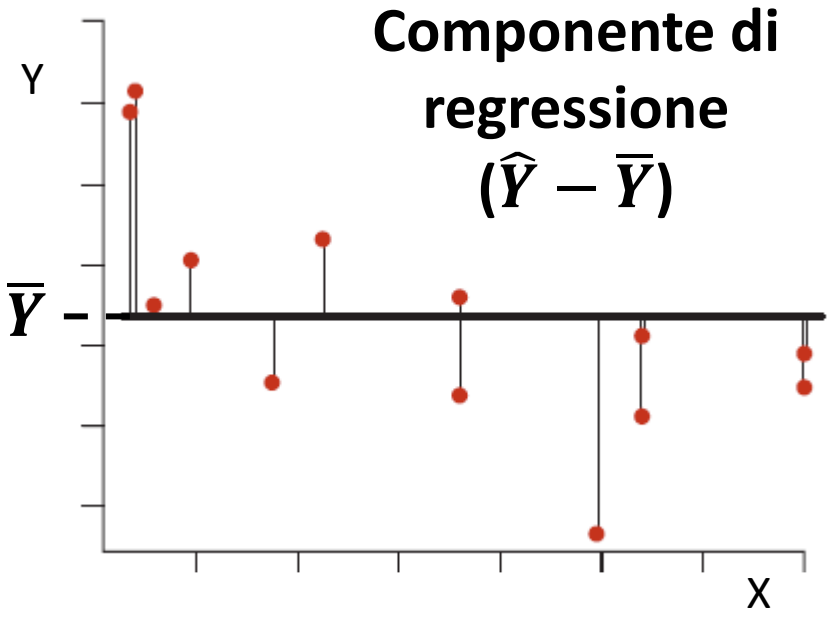
Per verificare questa ipotesi si usa l'approccio dell'ANOVA.

Il caso della regressione lineare

- L'approccio dell'ANOVA

Questo approccio scompone la distanza tra Y e $\overline{Y}$ in due fonti di variazione: la **componente residua** e la **componente di regressione.**





Il caso della regressione lineare

- L'approccio dell'ANOVA

H0: $\beta = 0$, ovvero il dati sono descritti dal modello nullo (Y=a)

HA: $\beta \neq 0$, ovvero i dati sono descritti dal modello di regressione (Y=a+bX)

Se H0 è vera, allora le medie dei quadrati corrispondenti alle due componenti (varianza residua e varianza di regressione) **dovrebbero essere uguali,** tranne per effetto del caso.

Se HA è vera, allora **la varianza di regressione è significativamente maggiore della varianza residua**.

La statistica test è il rapporto ***F*** tra le due varianze (test a due code)

Il caso della regressione lineare

- L'approccio dell'ANOVA

La tabella dell'ANOVA:

Fonte di variabilità	**Somma dei quadrati**	**df**	**Media dei quadrati**	**F**
Regressione	$SS_{regressione} = \sum (\hat{Y} - \bar{Y})^2$	1	$\frac{SS_{regressione}}{df_{regressione}}$	$\frac{MS_{regressione}}{MS_{residua}}$
Residua	$SS_{residua} = \sum (Y - \hat{Y})^2$	n-2	$\frac{SS_{residua}}{df_{residua}}$	
Totale	$SS_{totale} = \sum (Y - \bar{Y})^2$	**n-1**		

Il caso della regressione lineare

- L'approccio dell'ANOVA

Risultato esempio delle cince:

Fonte di variabilità	**Somma dei quadrati**	**df**	**Media dei quadrati**	**F**	***P***
Regressione	3,1268	1	3,1268	9,3106	0,011
Residua	3,6942	11	0,3358		
Totale	**6,8210**	**12**			

Esempio delle cince e GLM con R

```
#H0: B=0, modello Y=a, dove a= Ȳ
#HA: B≠0, modello Y=a+BX
#caricare il file «Regressione_esercizio_cince.txt»
dati<-read.table(choose.files(),header=T)
#visualizzo i dati in un diagramma a dispersione
matplot(dati$mass,dati$dees,pch=16,col="red")
#aggiungo la retta del modello di regressione e del modello nullo
fit<-lm(dati$dees~dati$mass,data=dati)
abline(fit,lwd=2)#se è vera HA
abline(h=mean(dati$dees),lwd=2,col="green")#se è vera H0
#eseguo l'analisi ANOVA
ris<-anova(fit)
```

Il caso dell'ANOVA

I modelli lineari generali permettono di includere **variabili esplicative multiple**, anche di tipo **categorico**.

Il modello lineare per l'ANOVA diventa:

$$Y = \mu + A$$

dove la **costante $\boldsymbol{\mu}$** rappresenta la media generale tra tutte le osservazioni, e la **variabile A** rappresenta l'effetto del gruppo o del trattamento.

Somiglianze tra il modello di regressione e ANOVA

- Caratteristiche in comune:
  - Includono una variabile risposta e una esplicativa
  - Possiedono un termine costante (Reg: intercetta; ANOVA: media generale)

- Differenza:
  - La variabile esplicativa è numerica nel caso della regressione e categorica nel caso dell'ANOVA

Generalizzazione

I due modelli (regressione e ANOVA) possono essere compresi in un modello unificato:

$$RESPONSE = CONSTANT + VARIABLE$$

dove:

- RESPONSE è la variabile dipendente
- CONSTANT è una costante (intercetta o media generale)
- VARIABLE è una variabile di tipo numerico o categorico

Esempio: analisi di una variabile di tipo categorico

In questo esempio verrà illustrato come si analizzano dati di tipo categorico nel contesto dei modelli lineari generali.

Esempio 18.1

L'empatia nei topi

Langford et al. (2006) hanno scoperto che i topi di laboratorio che provano dolore sembrano condividere anche il dolore dei loro compagni. Le conclusioni dei ricercatori si sono basate in parte sui risultati di un esperimento in cui ad alcuni topi veniva provocato un dolore molto lieve con una iniezione di acido acetico allo 0,9% nell'addome. Questi topi venivano poi sottoposti a uno di tre differenti trattamenti: erano mantenuti in isolamento (1), insieme a un altro topo a cui non era stata fatta l'iniezione (2), oppure insieme a un compagno a cui era stata fatta l'iniezione e che mostrava di provare dolore (3). La variabile risposta era la percentuale di tempo in cui ogni topo sottoposto a iniezione presentava un caratteristico «stiramento» (misurato dalla contrazione addominale), tipico di una situazione di disagio. I topi impiegati nel secondo e nel terzo trattamento provenivano dalla stessa gabbia del topo target. I dati ottenuti per alcuni topi maschi sono riassunti nella Tabella 18.1-1. I risultati indicano che i topi sottoposti a trattamento mostrano il massimo stiramento quando anche il compagno prova il dolore lieve. ■

ossiamo analizzare

Esempio: analisi di una variabile di tipo categorico

Per prima cosa bisogna adattare un modello lineare generale ai dati:

$$RESPONSE = CONSTANT + VARIABLE$$

sostituendo al modello la variabile dipendente «stiramento» e la variabile esplicativa categorica «trattamento compagnia» otteniamo:

$$STRETCHING = CONSTANT + COMPANION$$

Esempio: analisi di una variabile di tipo categorico

Se esiste un effetto del trattamento, allora le medie dei gruppi definiti dalla variabile categorica COMPANION (trattamento) saranno diverse.

FORMULAZIONE TRADIZIONALE

H0: Le medie dei gruppi definiti dal trattamento «compagnia» sono uguali

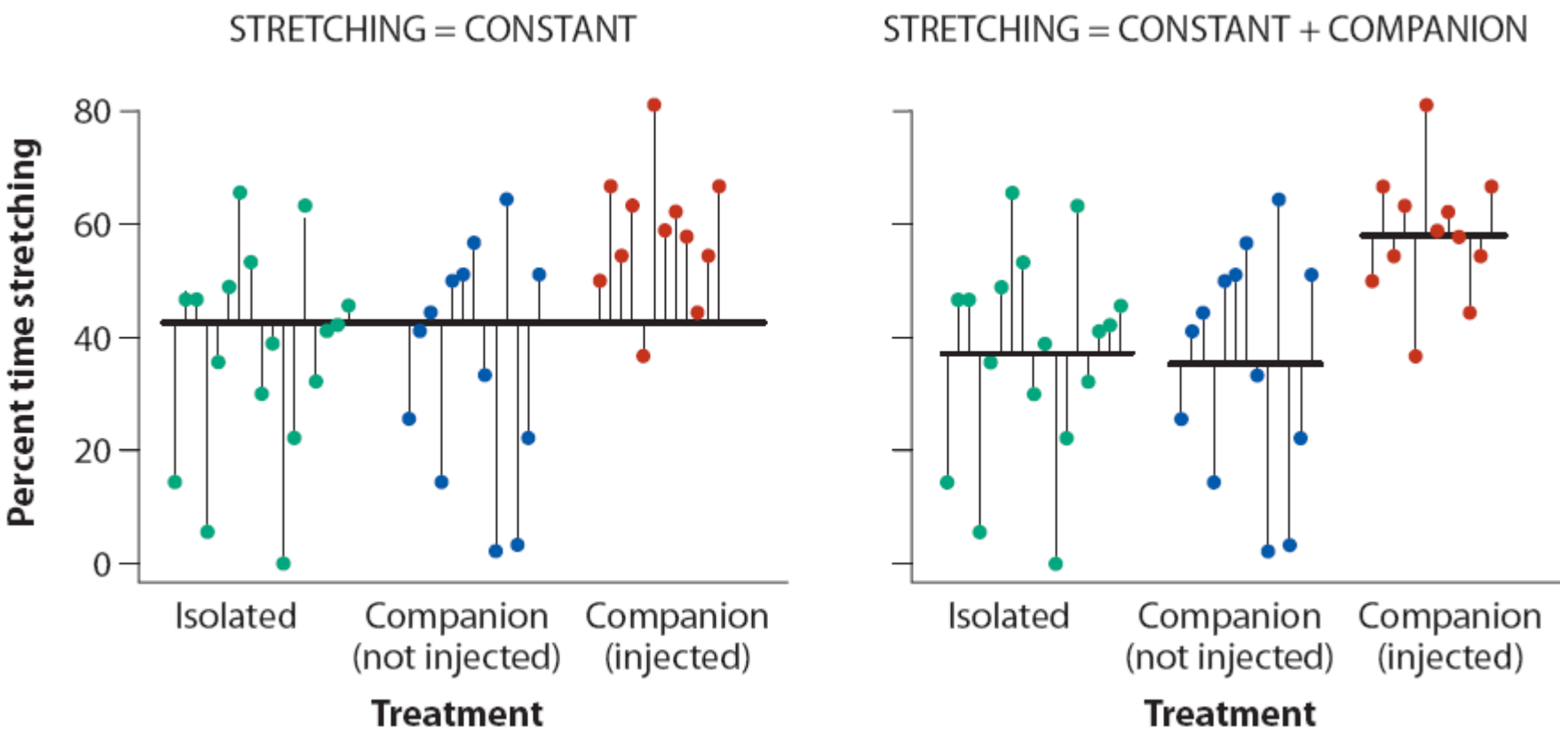
HA: Le medie dei gruppi definiti dal trattamento «compagnia» non sono tutte uguali.

FORMULAZIONE MODELLI LINEARI GENERALI

H0: il trattamento non ha effetto; il modello migliore è $STRETCHING = CONSTANT$

HA: il trattamento ha effetto; il modello migliore è $STRETCHING = CONSTANT + COMPANION$

Esempio: analisi di una variabile di tipo categorico



Esempio: analisi di una variabile di tipo categorico

Fonte di variabilità	**Somma dei quadrati**	**df**	**Media dei quadrati**	**F**	***P***
Regressione	4040,92	2	2020,45	6,67	0,0032
Residua	11807,43	39	302,75		
Totale	**15848,35**	**41**			

- Il P-value indica se il miglioramento dell'adattamento sia sufficientemente grande da giustificare il rifiuto di H0 (modello più semplice)
- In questo esempio H0 viene rifiutata; concludiamo quindi che il modello con la variabile COMPANION si adatta ai dati meglio del modello nullo.
- Le medie dei gruppi non sono tutte uguali.

Esempio: analisi di una variabile di tipo categorico con R

```
#carico il file «Modelli_lineari_generali_esercizio_topi.txt»
dati<-read.table(choose.files(),header=T)
#creo un box plot – prima la variabile categorica, poi le osservazioni
plot(dati$trattamento,dati$stretching)
#adatto il modello lineare generale ai dati secondo l'ipotesi alternativa (STRETCHING=CONSTANT+COMPANION)
fit<-lm(stretching~trattamento,data=dati)
#verifico l'ipotesi nulla
ris<-anova(fit)
```

L’analisi di disegni fattoriali

L’approccio dei modelli lineari generali può essere utilizzato per verificare come uno o più **fattori** (e la loro interazione) influenzino una variabile risposta.

Fattori fissi: una variabile esplicativa è un fattore fisso se i gruppi sono predeterminati e di interesse specifico (es: Trattamenti medici alternativi; dosi di una tossina)

Fattori variabili: una variabile esplicativa è un fattore variabile (o casuale) se i gruppi sono campionati casualmente da una popolazione di gruppi possibili (Es: famiglie; aree boschive)

L’analisi di disegni fattoriali: analisi di due fattori fissi

Esempio dell’effetto di due fattori ambientali e della loro interazione sulla crescita delle piante in habitat marino.

Esempio 18.3

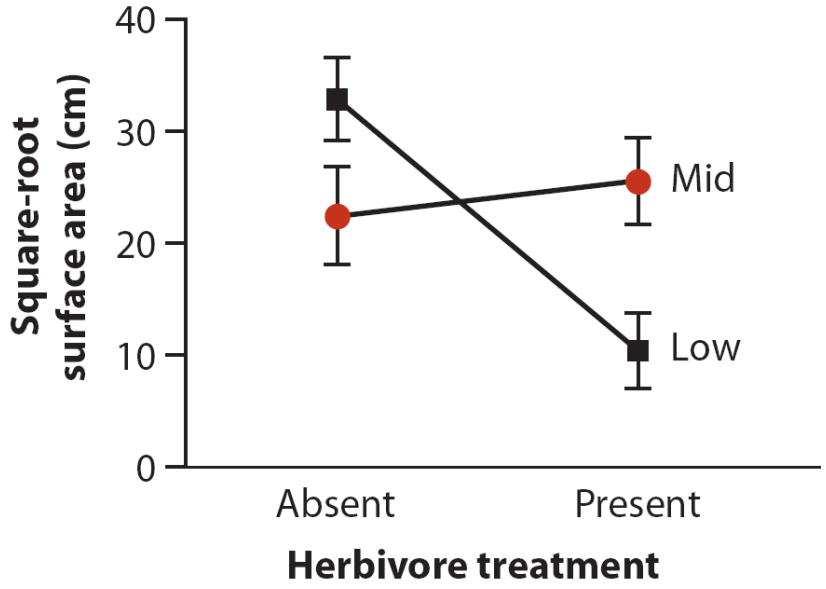
La zona di interazione

Harley (2004) ha indagato sperimentalmente l’impatto degli organismi erbivori sull’abbondanza di piante che vivono nell’habitat intertidale lungo la costa dello Stato di Washington (USA). In particolare, i suoi studi hanno riguardato l’alga rossa *Mazzaella parksii*, e il loro scopo era anche di capire se l’effetto degli erbivori sulle alghe dipendesse dall’area specifica della zona intertidale dove crescono le piante. Sono stati individuati 32 plot immediatamente sopra il livello della bassa marea e altri 32 plot a una quota intermedia tra il livello della bassa marea e quello dell’alta marea. Tutti gli organismi presenti sui plot sono stati rimossi, e in seguito è stata incollata la stessa quantità di alghe sulla superficie rocciosa al centro di ogni area. Usando recinzioni in rame, sono stati esclusi gli organismi erbivori (principalmente patelle e altri gasteropodi) da una metà dei plot scelta a caso, a ogni quota. I restanti plot, invece, sono rimasti accessibili. Nella fotografia qui sopra sono mostrati due gruppi di alghe, di cui uno circondato dall’anello di rame. Il disegno sperimentale era bilanciato e comprendeva ogni combinazione dei trattamenti (quota e presenza di erbivori); i dati sono riassunti nella Tabella 18.3-1. Al termine dell’esperimento, è stata misurata in ciascun plot l’area della superficie coperta da alghe, in centimetri quadrati (cm^2). I dati sono stati trasformati (applicando la radice quadrata) per migliorare l’adattamento alle assunzioni di normalità distributiva con uguale varianza. Le medie e gli errori standard sono riportati in Figura 18.3-1. ■

L'analisi di disegni fattoriali: analisi di due fattori fissi

Considerazioni preliminari:

- L'esperimento comprende due fattori:
  - Presenza/assenza di erbivori
  - Quota rispetto alla marea (bassa, media)
- Potrebbe essere presente interazione tra i fattori



L’analisi di disegni fattoriali: analisi di due fattori fissi

La formulazione del modello

Se rappresentiamo le variabili con la seguente notazione:

- ALGAE: variabile dipendente
- HERBIVORY: variabile esplicativa relativa alla presenza/assenza di erbivori
- HEIGHT: variabile esplicativa relativa alla quota della marea

un modello lineare generale può essere scritto come:

ALGAE=CONSTANT+HERBIVORY+HEIGHT+HERBIVORY*HEIGHT

Variabile dipendente | Effetti Principali | Interazione

L’analisi di disegni fattoriali: analisi di due fattori fissi

L’adattamento del modello ai dati:

Per quantificare il contributo di ciascuno degli effetti principali e della loro interazione, dobbiamo testare (anova) tre coppie di ipotesi:

1) HERBIVORY (effetto principale)

-H0: Non c’è differenza tra i trattamenti presenza/assenza di erbivori sulla copertura algale media

-HA: C’è differenza tra i trattamenti presenza/assenza di erbivori sulla copertura algale media

L’analisi di disegni fattoriali: analisi di due fattori fissi

2) HEIGHT (effetto principale)

-H0: Non c’è differenza fra i trattamenti di quota sulla copertura algale media

-HA: C’è differenza fra i trattamenti di quota sulla copertura algale media

3) HERBIVORY*HEIGTH (effetto di interazione)

-H0: L’effetto della presenza di erbivori sulla copertura algale non dipende dalla quota della zona intertidale

-HA: L’effetto della presenza di erbivori sulla copertura algale dipende dalla quota della zona intertidale

L’analisi di disegni fattoriali: analisi di due fattori fissi

Test delle ipotesi:

Per verificare ciascuna delle coppie di ipotesi **confrontiamo l’adattamento del modello completo ai dati con l’adattamento del modello in cui il termine di interesse viene CANCELLATO**.

Ad esempio per valutare l’interazione (ipotesi 3) viene confrontato il modello completo

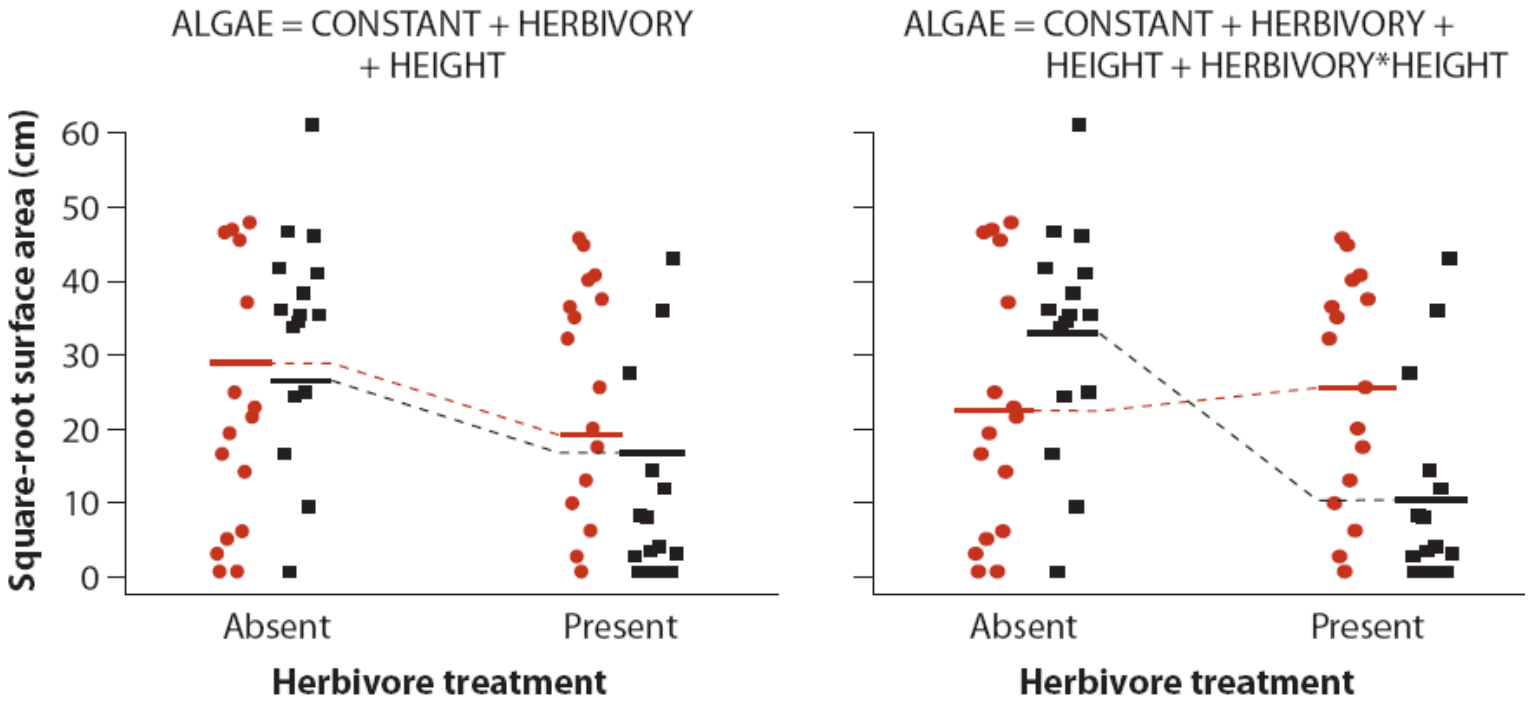
ALGAE=CONSTANT+HERBIVORY+HEIGHT+HERBIVORY*HEIGHT

con il modello nullo

ALGAE=CONSTANT+HERBIVORY+HEIGHT

L’analisi di disegni fattoriali: analisi di due fattori fissi

Confronto grafico dell’adattamento



Il modello nullo forza la differenza tra i valori previsti a differenti quote a essere costante nei casi di presenza/assenza di erbivori

L’analisi di disegni fattoriali: analisi di due fattori fissi

Il modello completo si adatta significativamente meglio ai dati osservati rispetto al modello nullo di assenza di interazione?

Fonte di variabilità	**Somma dei quadrati**	**df**	**Media dei quadrati**	**F**	***P***
HERBIVORY	1512,18	1	1512,18	6,36	0,014
HEIGHT	88,97	1	88,97	0,37	0,543
HERBIVORY* HEIGHT	2616,96	1	2616,96	11,00	0,002
Residuo	14270,52	60	237,84		
Totale	**18488,63**	**63**			

L’analisi di disegni fattoriali: analisi di due fattori fissi

Interpretazione

- L’ipotesi di assenza di interazione può essere esclusa (P=0,002)
- Viene rifiutata l’ipotesi nulla di nessun effetto principale della presenza di erbivori (P=0,014)
- Nessun effetto significativo della quota come effetto principale (P=0,543)
- La variabile quota però sembra esercitare la sua influenza sulla variabile dipendente in maniera indiretta tramite l’interazione con la variabile presenza/assenza di erbivori.

L’analisi di disegni fattoriali: analisi di due fattori fissi

```
#carico il dataset dal file «Modelli_linerai_generali_Alghe.txt»
dati<-read.table(choose.files(),header=T)
#verifico se sembra esserci interazione tra i fattori
interaction.plot(dati$herbivores,dati$height,dati$sqrtarea)
#adatto il modello lineare generale completo
fit<-lm(sqrtarea~herbivores+height+herbivores*height, data=dati)
#eseguo il test
ris<-anova(fit)
```

Analisi di disegni fattoriali: considerare gli effetti di una covariata

In molti casi, potremmo essere interessati a verificare **l'effetto di un fattore** sulla variabile dipendente **tenendo però in considerazione una (o più) variabile di «confondimento».**

Questo tipo di variabile viene chiamato **«covariata»**

Importante negli studi osservazionali perché l'inclusione di queste variabili nelle analisi permette di correggere la loro influenza sulla stima degli effetti.

In molti casi infatti non è possibile effettuare uno studio sperimentale dove è possibile minimizzare l'effetto delle covariate.

Analisi di disegni fattoriali: considerare gli effetti di una covariata

Esempio 18.3

I ratti talpa fannulloni

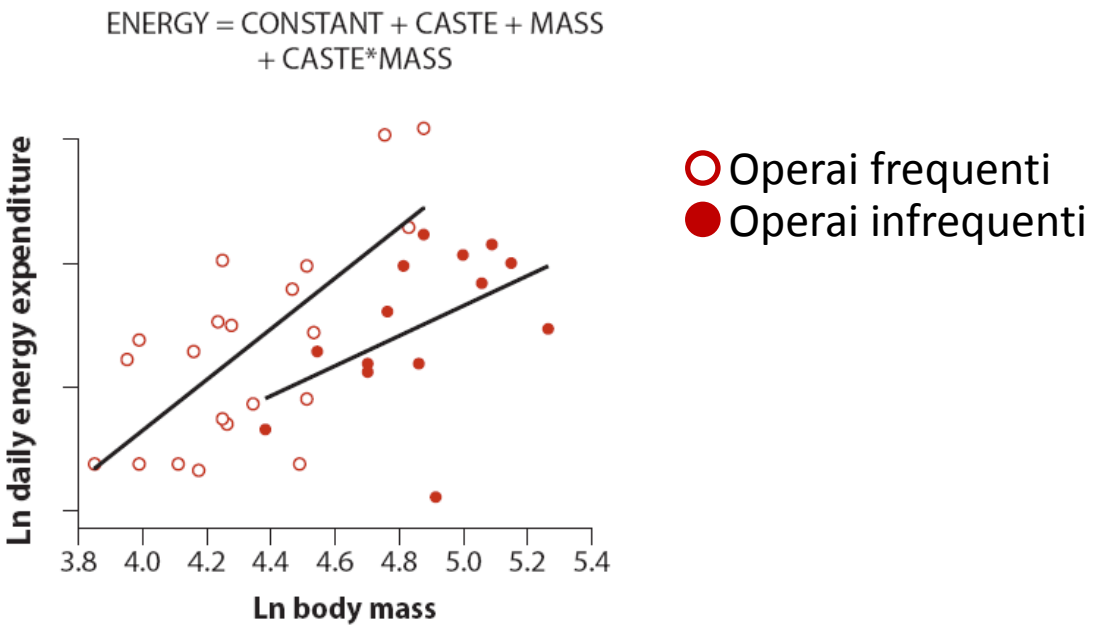
I ratti talpa sono gli unici mammiferi conosciuti che hanno un'organizzazione sociale suddivisa in caste. Una singola regina e un piccolo numero di maschi sono i soli individui riproduttivi all'interno di una colonia. Tutti gli altri, detti operai, raccolgono il cibo, difendono la colonia, si prendono cura dei piccoli e provvedono alla manutenzione delle gallerie. È stato scoperto recentemente che nel ratto talpa di Damara (*Cryptomys damarensis*), che vive nell'Africa subsahariana, potrebbero esservi due caste di operai: la casta degli «operai frequenti», che compiono quasi tutto il lavoro nella colonia, e quella degli «operai infrequenti», che lavorano poco, tranne che in rare occasioni: dopo le piogge, quando c'è bisogno di ampliare il sistema delle gallerie della colonia. Per valutare le differenze fisiologiche tra i due gruppi, Scantlebury et al. (2006) hanno confrontato il consumo energetico giornaliero dei ratti talpa selvatici durante la stagione secca. Il consumo energetico varia con la massa corporea in entrambi i gruppi di operai (Figura 18.4-1), ma gli infrequenti hanno una massa corporea maggiore di quella degli operai frequenti. Quanto è diverso il consumo energetico giornaliero medio tra i due gruppi, quando questo viene corretto per le differenze nella massa corporea? ■

Analisi di disegni fattoriali: considerare gli effetti di una covariata

- Il modello lineare generale

$$ENERGY = CONSTANT + CASTE + MASS + CASTE*MASS$$

ENERGY	CASTE	MASS	CASTE*MASS
Variabile Dipendente Numerica	Variabile Esplicativa Categorica	Variabile Esplicativa Numerica	Interazione



Analisi di disegni fattoriali: considerare gli effetti di una covariata

- Approccio analitico
  - 1) Verificare se esiste interazione tra le variabili esplicative (cioè se il modello completo è supportato dai dati)
  - 2) Riformulare un nuovo modello completo con o senza iterazione per testare l'effetto del fattore.

Analisi di disegni fattoriali: considerare gli effetti di una covariata

- 1)Verificare se esiste interazione tra le variabili esplicative

Vogliamo capire se il consumo energetico giornaliero sia diverso tra le due caste di operai considerando il fatto che ci sono delle differenze nella massa corporea.

Testiamo se le rette di regressione (consumo energetico in base alla massa corporea) abbiano la stessa pendenza nelle due caste.

Ipotesi nulla relativa al problema biologico:
H0: Non c'è interazione tra casta e massa corporea (stessa pendenza nelle caste)
HA: C'è interazione tra casta e massa corporea (pendenza diversa nelle caste)

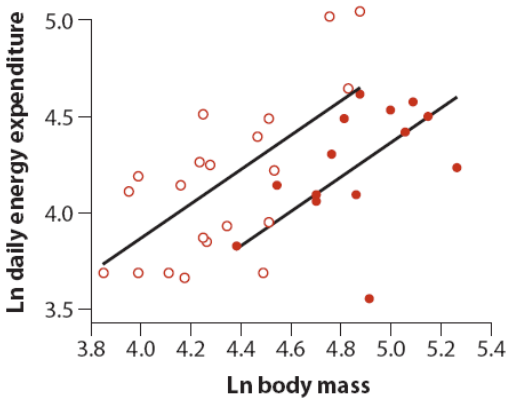
Ipotesi nulla nei modelli lineari generali:
H0: i dati supportano il modello nullo (senza il termine dell'interazione)
HA: i dati supportano il modello completo (con il termine dell'interazione)

Analisi di disegni fattoriali: considerare gli effetti di una covariata

- I risultati del confronto

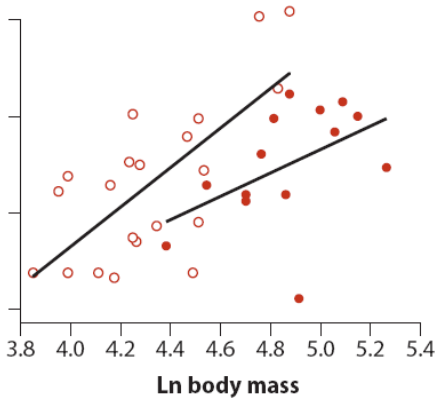
Modello Nullo

ENERGY = CONSTANT + CASTE + MASS



Modello Completo

ENERGY = CONSTANT + CASTE + MASS + CASTE*MASS



NON RIFIUTIAMO L'IPOTESI NULLA: **NESSUNA EVIDENZA DI INTERAZIONE**

Fonte di variabilità	Somma dei quadrati	df	Media dei quadrati	F	*P*
CASTE	0,0570	1	0,0570		
MASS	1,3618	1	1,3618		
CASTE*MASS	0,0896	1	0,0896	1,02	0,32
Residuo	2,7249	31	0,0879		
Totale	**4,2333**	**34**			

Analisi di disegni fattoriali: considerare gli effetti di una covariata

- 2) Riformulare il modello completo

Se l'ipotesi di iterazione tra le variabili può essere esclusa, allora il modello lineare generale completo diventa:
ENERGY=CONSTANT+CASTE+MASS
altrimenti il modello completo conserva il termine di interazione.

A questo punto non resta che testare l'ipotesi nulla:

Ipotesi nulla relativa al problema biologico:
H0: Le caste non differiscono nel consumo energetico
HA: Le caste differiscono nel consumo energetico

Ipotesi nulla nei modelli lineari generali:
H0: i dati supportano il modello nullo (senza il termine CASTE)
HA: i dati supportano il modello completo

Analisi di disegni fattoriali: considerare gli effetti di una covariata

- Risultati del confronto

Fonte di variabilità	Somma dei quadrati	df	Media dei quadrati	F	*P*
CASTE	1,8815	1	1,8815	21,39	<0,001
MASS	0,6375	1	0,6375	7,25	0,011
Residuo	2,8145	32	0,0880		
Totale	**5,3335**	**34**			

Il modello nullo corrisponde a una singola regressione lineare dell'energia rispetto alla massa (infatti il termine CASTE non è incluso nel modello)

Il rapporto F per CASTE è significativo, confermando che le due caste di operai differiscono nel loro consumo energetico giornaliero medio dopo l'aggiustamento per la massa corporea.

I risultati indicano che durante la stagione secca gli operai infrequenti tendono a consumare meno energia rispetto a quelli frequenti.

Analisi di disegni fattoriali: considerare gli effetti di una covariata con R I

```
#caricare il file «Modelli_lineari_generali_RattiTalpa.txt»
dati<-read.table(choose.files(),header=T)
#generiamo un grafico a dispersione con i gruppi worker e lazy distinti
matplot(dati$lnmass,dati$lnenergy,pch=1,col="red",xlab="LnMass",ylab="LnEnergy")
matplot(dati$lnmass[dati$caste=="lazy"],dati$lnenergy[dati$caste=="lazy"],pch=16,col="red",add=T)
#Testo l'ipotesi nulla:
#H0: Non c'è interazione tra massa e casta
#HA: C'è interazione tra massa e casta
#genero il modello lineare completo
#NB: nella formula è meglio inserire prima la variabile numerica, poi il fattore
intfit<-lm(lnenergy~lnmass+caste+lnmass*caste, data=dati)
#verifico se il termine per l'interazione caste*lnmass è significativo
intris<-anova(intfit)
#H0 non viene rifiutata, cioè non ci sono evidenze a favore dell'ipotesi alternativa di presenza di interazione
```

Analisi di disegni fattoriali: considerare gli effetti di una covariata con R II

#testo la seconda ipotesi nulla:

H0: Le caste non differiscono nel consumo energetico (H0: favorito il modello nullo senza il termine caste)

HA: Le caste differiscono nel consumo energetico

(HA: favorito il modello completo)

#genero il modello lineare completo senza il termine dell'interazione (esclusa in precedenza)

```
fit<-lm(lnenergy~lnmass+caste, data=dati)
```

#verifico l'ipotesi nulla

```
ris<-anova(fit)
```

#rifiuto H0 per il termine caste, favorisco il modello completo.

#Le caste differiscono nel consumo energetico

Le assunzioni dei modelli lineari generali

I modelli generali trattano la regressione lineare e l'analisi di fattori in un unico contesto ed effettuano un test d'ipotesi tramite ANOVA.

Per questo motivo, le assunzioni della regressione lineare e dell'ANOVA devono essere rispettate:

- le misure per ogni combinazione di valori delle variabili esplicative sono un campione casuale estratto dalla popolazione di misure possibili
- le misure per ogni combinazione di valori delle variabili esplicative hanno una distribuzione normale nella popolazione corrispondente
- La varianza della variabile risposta è la stessa per tutte le combinazioni delle variabili esplicative.

Esempio di riepilogo

L'uomo di Neanderthal aveva un encefalo più piccolo di quello dell'uomo moderno? Le stime della capacità cranica sulla base dei fossili indicano che l'uomo di Neanderthal aveva una grande massa encefalica, ma aveva anche una grande massa corporea. Il diagramma che accompagna questo problema presenta i dati trattati da Ruff et al. (1977) sulla massa encefalica log-trasformata e sulla massa corporea log-trasformata stimate sulla base di reperti ossei di uomo di Neanderthal (cerchi pieni) e dei primi uomini anatomicamente moderni (cerchi vuoti).

L'analisi si proponeva di determinare se l'uomo anatomicamente moderno e l'uomo di Neanderthal avessero masse encefaliche diverse tenendo in debita considerazione le differenze di massa corporea.

Esempio di riepilogo

- Caricare i dati e creare un diagramma a dispersione che rappresenti le osservazioni per uomo di Neanderthal e uomo anatomicamente moderno con simboli diversi.

```
#carichiamo i dati presenti nel file «Modelli_lineari_generali_Uomo.txt»
dati<-read.table(choose.files(),header=T)
#creiamo il diagramma a dispersione
matplot(dati$lnmass,dati$lnbrain,pch=1,col="red",xlab="LnMass",ylab="Ln Brain")
matplot(dati$lnmass[dati$species=="neanderthal"],dati$lnbrain[dati$species=="neanderthal"],pch=16,col="red",add=T)
```

Esempio di riepilogo

- Per capire se i due uomini avessero masse encefaliche diverse dobbiamo verificare se è presente interazione tra il tipo di uomo (fattore) e la massa corporea, o in altre parole, se le rette di regressione stimate per i due gruppi hanno pendenze significativamente diverse.

```
#formuliamo il modello lineare completo
fit<-lm(lnbrain~lnmass+species+lnmass*species,data=dati)
#definiamo l'ipotesi nulla sull'interazione
#H0: Non esiste interazione tra massa corporea e specie
#HA: Esiste interazione tra massa corporea e specie
#verifico l'ipotesi nulla di assenza di interazione
ris<-anova(fit)
#Response: lnbrain
#                  Df   SumSq     MeanSq   F value  Pr(>F)
#lnmass             1  0.1025    0.1025    23.1465  2.835e-05
#species            1  0.027553  0.027553  6.2203   0.01750
#lnmass:species     1  0.004845  0.004845  1.0938   0.30279
#Residuals          35 0.155033  0.004430
#Non rifiuto H0, l'interazione può essere esclusa (la rette di regressione che mettono in rapporto la massa corporea e la massa celebrale nei due gruppi hanno la stessa pendenza)
```

Esempio di riepilogo

- A questo punto possiamo escludere il termine di interazione, generare il nuovo modello lineare generale e studiare le differenze nella massa celebrale tra le specie controllando per la massa croporea.

```
#formuliamo il nuovo modello lineare completo
fit<-lm(lnbrain~lnmass+species,data=dati)
#definiamo l'ipotesi nulla e alternativa:
#H0: le due specie non hanno masse celebrali diverse (modello senza il termine species)
#HA: le due specie hanno masse celebrali diverse (modello con il termine species)
ris<-anova(fit)
#Response: lnbrain
#            Df     SumSq      MeanSq Fvalue  Pr(>F)
#lnmass       1  0.102528  0.102528   23.08  2.724e-05
#species      1  0.027553  0.027553   6.2041  0.01749
#Residuals   36  0.159878  0.004441

#Pvalue (species)<0.05, Rifiuto H0, le due specie hanno masse celebrali diverse.
```