Metodologia sperimentale per le scienze agrarie

Andrea Onofri e Dario Sacco

Update: v. 1.0 (15/03/2021), compil. 2021-04-28

Indice

Premessa

Placeholder

Obiettivi

Organizzazione

Software statistico

The authors

Ringraziamenti

Capitolo 1

Scienza e pseudo-scienza

Placeholder

1.1 Scienza = dati

1.2 Dati 'buoni' e 'cattivi'

1.3 Dati 'buoni' e metodi 'buoni'

1.4 Il principio di falsificazione

1.5 Falsificare un risultato

1.6 Elementi fondamentali del disegno sperimentale

1.6.1 Controllo degli errori

1.6.2 Replicazione

1.6.3 Randomizzazione

1.6.4 Esperimenti invalidi

Cattivo controllo degli errori

'Confounding' e correlazione spuria

Pseudo-repliche e randomizzazione poco attenta

1.7 Chi valuta se un esperimento è attendibile?

1.8 Conclusioni

1.9 Altre letture

Capitolo 2

Progettare un esperimento

Placeholder

2.1 Gli elementi della ricerca

2.2 Ipotesi scientifica → obiettivo dell'esperimento

2.3 Identificazione dei fattori sperimentali

2.3.1 Esperimenti (multi-)fattoriali

2.3.2 Controllo o testimone

2.4 Le unità sperimentali

2.5 Allocazione dei trattamenti

2.6 Le variabili sperimentali

2.6.1 Variabili nominali (categoriche)

2.6.2 Variabili ordinali

2.6.3 Variabili quantitative discrete

2.6.4 Variabili quantitative continue

2.6.5 Rilievi visivi e sensoriali

2.6.6 Variabili di confondimento

2.7 Esperimenti di campo

2.7.1 Scegliere il campo

2.7.2 Le unità sperimentali in campo

2.7.3 Numero di repliche

2.7.4 La mappa di campo

2.7.5 Lay-out sperimentale

Capitolo 3

Richiami di statistica descrittiva

Placeholder

3.1 Dati quantitativi

3.1.1 Indicatori di tendenza centrale

3.1.2 Indicatori di dispersione

3.1.3 Incertezza delle misure derivate

3.1.4 Relazioni tra variabili quantitative: correlazione

3.2 Dati qualitativi

3.2.1 Distribuzioni di frequenze e classamento

3.2.2 Statistiche descrittive per le distribuzioni di frequenze

3.2.3 Distribuzioni di frequenza bivariate: le tabelle di contingenze

3.2.4 Connessione

3.3 Statistiche descrittive con R

3.3.1 Descrizione dei sottogruppi

3.3.2 Distribuzioni di frequenze e classamento

3.3.3 Connessione

3.4 Altre letture

Capitolo 4

Modelli statistici ed analisi dei dati

Placeholder

4.1 Verità 'vera' e modelli deterministici

4.2 Genesi deterministica delle osservazioni sperimentali

4.3 Errore sperimentale e modelli stocastici

4.3.1 Funzioni di probabilità

4.3.2 Funzioni di densità

4.3.3 La distribuzione normale (curva di Gauss)

4.4 Modelli 'a due facce'

4.5 E allora?

4.6 Le simulazioni Monte Carlo

4.7 Analisi dei dati e 'model fitting'

4.8 Modelli stocastici non-normali

4.9 Altre letture

Capitolo 5

Stime ed incertezza

Placeholder

Capitolo 6

Decisioni ed incertezza

Placeholder

6.1 Confronto tra due medie: il test t di Student

6.1.1 L'ipotesi nulla e alternativa

6.1.2 La statistica T

6.1.3 Simulazione Monte Carlo

6.1.4 Soluzione formale

6.1.5 Interpretazione del P-level

6.1.6 Tipologie alternative di test t di Student

6.2 Confronto tra due proporzioni: il test χ^2

6.2.1 Simulazione Monte Carlo

6.2.2 Soluzione formale

6.3 Conclusioni e riepilogo

6.4 Altre letture

Capitolo 7

Modelli ANOVA ad una via

Placeholder

7.1 Caso-studio: confronto tra erbicidi in vaso

7.2 Descrizione del dataset

7.3 Definizione di un modello lineare

7.4 Parametrizzazione del modello

7.5 Assunzioni di base

7.6 Fitting del modello: metodo manuale

7.6.1 Stima dei parametri

7.6.2 Calcolo dei residui

7.6.3 Stima di σ

7.7 Scomposizione della varianza

7.8 Test d'ipotesi

7.9 Inferenza statistica

7.10 Fitting del modello con R

7.11 Medie marginali attese

7.12 Per concludere ...

7.13 Altre letture

Capitolo 8

La verifica delle assunzioni di base

Placeholder

8.1 Violazioni delle assunzioni di base

8.2 Procedure diagnostiche

8.3 Analisi grafica dei residui

8.3.1 Grafico dei residui contro i valori attesi

8.3.2 QQ-plot

8.4 Test d'ipotesi

8.5 Risultati contraddittori

8.6 'Terapia'

8.6.1 Correzione/Rimozione degli outliers

8.6.2 Correzione del modello

8.6.3 Trasformazione della variabile indipendente

8.6.4 Impiego di metodiche statistiche avanzate

8.6.5 Trasformazioni stabilizzanti

8.7 Esempio 1

8.8 Esempio 2

8.9 Altre letture

Capitolo 9

Contrasti e confronti multipli

Placeholder

Capitolo 10

Modelli ANOVA con fattori di blocco

Placeholder

10.1 Caso-studio: confronto tra erbicidi in campo

10.2 Definizione di un modello lineare

10.3 Stima dei parametri

10.3.1 Coefficienti del modello

10.3.2 Stima di σ

10.4 Scomposizione della varianza

10.5 Adattamento del modello con R

10.6 Disegni a quadrato latino

10.7 Caso studio: confronto tra metodi costruttivi

10.8 Definizione di un modello lineare

Capitolo 11

La regressione lineare semplice

Nei capitoli precedenti abbiamo parlato di modelli basati su una fattori sperimentali in forma di categorie, ad esempio diversi erbicidi o diverse varietà. Abbiamo visto che, con questi fattori sperimentali, si utilizzano i cosiddetti modelli ANOVA.

Nella sperimentazione agronomica e, in genere, biologica, la variabile indipendente (o le variabili indipendenti) può (possono) rappresentare una quantità, come, ad esempio, la dose di un farmaco, il tempo trascorso da un certo evento, la fittezza di semina e così via. In questa condizione, l'analisi dei dati richiede modelli diversi da qualli visti finora, di solito identificati con il nome di modelli di regressione. Questa classe di modelli è estremamente interessante e si presta a sviluppi potentissimi. In questo libro ci accontenteremo di trattare la regressione lineare semplice, cioè un modello lineare (retta) con una variabile dipendente ed un regressore. In un capitolo successivo estenderemo le considerazioni fatte ai modelli non-lineari.

11.1 Caso studio: effetto della concimazione azotata al frumento

Per individuare la relazione tra la concimazione azotata e la produzione del frumento, è stato organizzato un esperimento a randomizzazione completa, con quattro dosi di azoto e quattro repliche. I risultati ottenuti sono riportati nella Tabella 11.1 e possono essere caricati direttamente da gitHub, con il

Tabella 11.1: Dataset relativo ad una prova di concimazione azotata su frumento

Dose	1	2	3	4
0	21.98	25.69	27.71	19.14
60	35.07	35.27	32.56	32.63
120	41.59	40.77	41.81	40.50
180	50.06	52.16	54.40	51.72

codice sottostante. A differenza dei capitoli precedenti, in questo caso il dataset non è ottenuto da una prova vera, ma è stato generato, con il codice riportato nel capitolo 4. Pertanto, l'esempio, pur essendo efficace da un punto di vista didattico, potrebbe non essere assolutamento realistico.

```
fileName <- "https://www.casaonofri.it/_datasets/NWheat.csv"
dataset <- read.csv(fileName, header=T)
```

11.2 Analisi preliminari

Questo esperimento è replicato ed è totalmente analogo a quello presentato nel capitolo 7, con l'unica differenza che, in questo caso, la variabile indipendente è quantitativa. Tuttavia, è del tutto logico considerare la dose di azoto come un predittore qualitativo ('factor') ed utilizzare un modello descrittivo ANOVA ad una via. Eseguiamo il 'fitting' con R, ottenendo i risultati seguenti:

```
dataset$DoseF <- factor(dataset$Dose)
model <- lm(Yield ~ DoseF, data = dataset)
anova(model)
## Analysis of Variance Table
##
## Response: Yield
##           Df  Sum Sq Mean Sq F value    Pr(>F)
## DoseF      3 1725.96  575.32  112.77 4.668e-09 ***
## Residuals 12   61.22    5.10
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

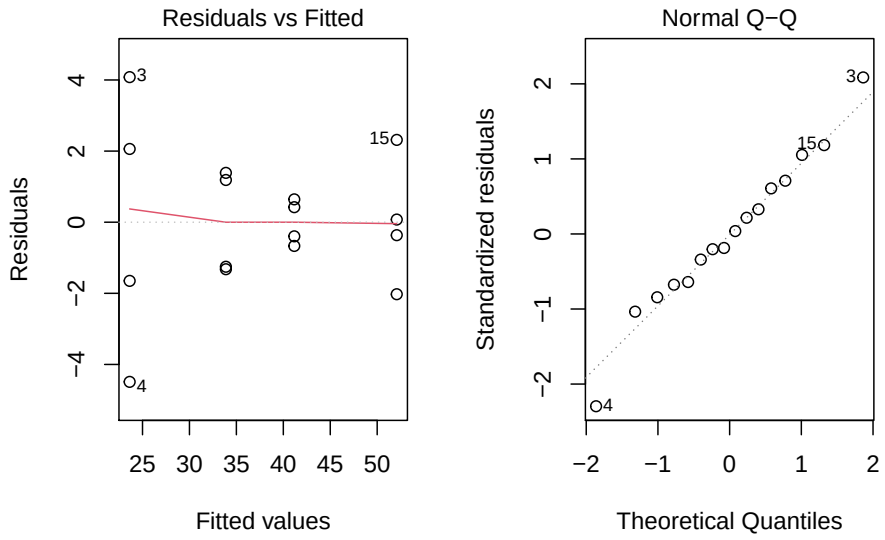

Figura 11.1: Analisi grafica dei residui per una prova di concimazione azotata del frumento

Osserviamo che l'effetto del trattamento è significativo e il SEM è pari a $\sqrt{5.10/4} = 1.129$. Prima di proseguire, verifichiamo che non ci siano problemi relativi alle assunzioni parametriche di base e che, quindi, la trasformazione dei dati non sia necessaria. I grafici dei residui, riportati in Figura 11.1, non mostrano patologie rilevanti.

Da questo momento in avanti, diversamente a quanto abbiamo visto nei capitoli precedenti, l'analisi non prosegue con un test di confronto multiplo, che in questa situazione, se non del tutto errato, sarebbe comunque da considerare 'improprio.' Infatti, quale senso avrebbe confrontare la risposta produttiva a 60 kg N ha^{-1} con quella a 120 kg N ha^{-1}? In realtà noi non siamo specificatamente interessati a queste due dosi, ma a qualunque altra dose nell'intervallo da 0 a 180 kg N ha^{-1}. Abbiamo selezionato quattro dosi per organizzare l'esperimento, ma resta il fatto che siamo interessati a definire una funzione di risposta per tutto l'intervallo delle dosi, non a confrontare le risposte a due dosi in particolare.

Per questo motivo, quando la variabile indipendente è una dose, l'analisi dei dati dovrebbe essere basata sull'impiego di un modello di regressione, in quanto ciò è più coerente con le finalità dell'esperimento, rispetto all'adozione di una procedura di confronto multiplo.

11.3 Definizione del modello lineare

Immaginiamo che, almeno nell'intervallo di dosi incluso nell'esperimento, l'effetto della concimazione azotata sulla produzione sia lineare. In effetti, l'andamento dei dati conferma questa impressione e, pertanto, poniamo il modello lineare nei termini usuali:

$$Y_i = b_0 + b_1 X_i + \varepsilon_i$$

dove Y è la produzione della parcella i, trattata con la dose X_i, b_0 è l'intercetta (produzione a dose di azoto pari a 0) e b_1 è la pendenza, cioè l'incremento di produzione per ogni incremento unitario della dose. La componente stocastica ε viene assunta omoscedastica e normalmente distribuita, con media 0 e deviazione standard σ.

11.4 Stima dei parametri

Dobbiamo a questo punto individuare i parametri b_0 e b_1 in modo tale che la retta ottenuta sia la più vicina ai dati, cioè in modo da minimizzare gli scostamenti tra i valori di produzione osservati e quelli stimati dal modello (soluzione dei minimi quadrati). La funzione dei minimi quadrati è:

$$\begin{aligned}
Q &= \sum_{i=}^{N} \left(Y_i - \hat{Y}\right)^2 = \sum_{i=}^{N} \left(Y_i - b_0 - b_1 X_i\right)^2 = \\
&= \sum_{i=}^{N} \left(Y_i^2 + b_0^2 + b_1^2 X_i^2 - 2Y_i b_0 - 2Y_i b_1 X_i + 2b_0 b_1 X_i\right) = \\
&= \sum_{i=}^{N} Y_i^2 + N b_0^2 + b_1^2 \sum_{i=}^{N} X_i^2 - 2b_0 \sum_{i=}^{N} Y_i^2 - 2b_1 \sum_{i=}^{N} X_i Y_i + 2b_0 b_1 \sum_{i=}^{N} X_i
\end{aligned}$$

Calcolando le derivate parziali rispetto a b_0 e b_1 che, al momento, sono le nostre incognite, ed eguagliandole a 0 si ottengono le seguenti formule risolutive:

$$b_1 = \frac{\sum_{i=1}^{N} \left[\left(X_i - \mu_X\right)\left(Y_i - \mu_Y\right)\right]}{\sum_{i=1}^{N} \left(X_i - \mu_X\right)^2}$$

e:

$$b_0 = \mu_Y - b_1 \mu_X$$

Invece che svolgere i calcoli a mano, possiamo eseguire il fitting ai minimi quadrati con R. Possiamo notare che l'unica differenza tra questo modello di regressione e il modello ANOVA utilizzato poco sopra è che qui utilizziamo la variabile 'Dose' come tale, senza prima trasformarla in un 'factor.'

```
modelReg <- lm(Yield ~ Dose, data = dataset)
summary(modelReg)
##
## Call:
## lm(formula = Yield ~ Dose, data = dataset)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -4.6537 -1.5350 -0.4637  1.9250  3.9163
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 23.793750   0.937906   25.37 4.19e-13 ***
## Dose         0.154417   0.008356   18.48 3.13e-11 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.242 on 14 degrees of freedom
## Multiple R-squared:  0.9606, Adjusted R-squared:  0.9578
## F-statistic: 341.5 on 1 and 14 DF,  p-value: 3.129e-11
```

Ora sappiamo che la relazione tra la dose di azoto e la risposta produttiva del frumento è:

$$Y_i = 23.111 + 0.1544 \times X_i$$

L'elemento stocastico ε_i è normalmente distribuito, con media 0 e deviazione standard 2.029 (vedi la voce 'Residual standard error' nell'output sovrstante).

Come al solito, prima di qualunque altra considerazione, dobbiamo verificare la bontà del modello e il rispetto delle assunzioni di base, con una procedura che, per un modello di regressione, deve riguardare un maggior numero di aspetti rispetto ad un modello ANOVA.

11.5 Valutazione della bontà del modello

In primo luogo, è necessario verificare il rispetto delle assunzioni di base di normalità e omoscedasticità dei residui. Per questo, possiamo utilizzare gli stessi metodi impiegati per i modelli ANOVA, vale a dire un grafico dei residui verso gli attesi ed un QQ-plot dei residui standardizzati. In realtà, abbiamo già eseguito questo controllo con il modello ANOVA corrispondente e non vi è la necessità di ripeterlo con questo modello.

Dobbiamo invece assicurarci che i dati osservati siano ben descritti dal modello adottato, senza nessuna componente sistematica di mancanza d'adattamento. In altre parole, le osservazioni non debbono contraddire l'ipotesi che la risposta è lineare, salvo per le eventuali deviazioni casuali insite in qualunque esperimento. Per la verificà della **bontà di adattamento** possiamo utilizzare diverse procedure, che illustreremo di seguito.

11.5.1 Valutazione grafica

Nel modo più semplice, la bontà di adattamento può essere valutata attraverso un grafico dei valori attesi e dei valori osservati, come quello in Figura 11.2. Notiamo che non c'è alcun elemento che faccia pensare ad una sistematica deviazione rispetto alle previsioni fatte dal modello.

11.5.2 Errori standard dei parametri

In secondo luogo, possiamo valutare gli errori standard delle stime dei parametri, che non debbono mai essere superiori alla metà del valore del parametro stimato, cosa che in questo caso è pienamente verificata. Se così non fosse, l'intervallo di confidenza del parametro, usualmente stimato utilizzando il doppio dell'errore standard, conterrebbe lo zero, il che equivarebbe a dire che, ad esempio, la pendenza 'vera' (cioè quella della popolazione da cui il nostro campione è estratto) potrebbe essere nulla. In altre parole, la retta potrebbe essere 'piatta,' dimostrando l'inesistenza di relazione tra la dose di concimazione e la produzione della coltura. Si può notare che, nell'esempio in studio, questo dubbio non sembra sussistere.

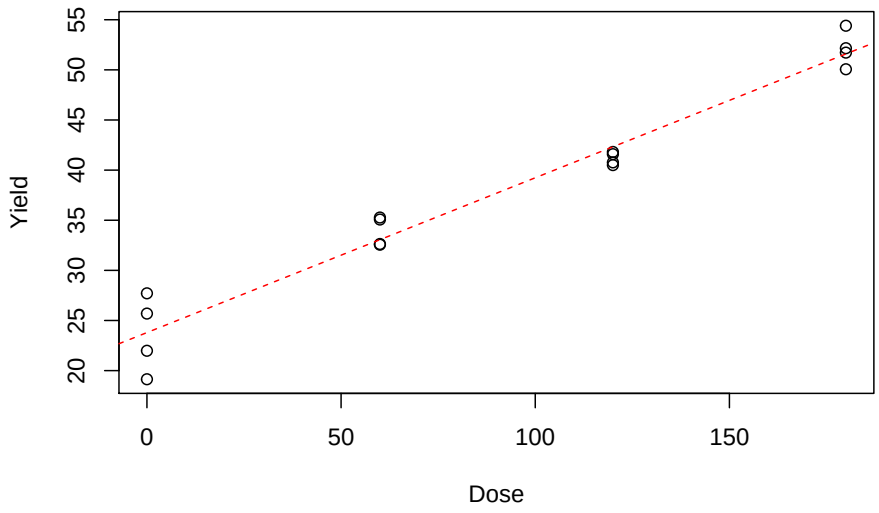

Figura 11.2: Risposta produttiva del frumento alla concimazione azotata: dati osservati (simboli) e valori attesi (linea tratteggiata).

11.5.3 Test F per la mancanza d'adattamento

In terzo luogo, possiamo analizzare i residui della regressione, cioè gli scostamenti dei punti rispetto alla retta e, in particolare, la somma dei loro quadrati. Possiamo vedere che questo valore è pari a 70.37, ed è più alto di quello del corrispondente modello ANOVA, impiegato in precedenza (61.22):

```
anova(modelReg)
## Analysis of Variance Table
##
## Response: Yield
##           Df  Sum Sq Mean Sq F value    Pr(>F)
## Dose       1 1716.80 1716.80  341.54 3.129e-11 ***
## Residuals 14   70.37    5.03
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Il risultato è perfettamente normale, dato che il residuo del modello ANOVA contiene solo la misura dello scostamento di ogni dato rispetto alla media del suo gruppo, che si può considerare 'errore puro,' mentre il residuo del-

la regressione, oltre all'errore puro, contiene anche una componente detta 'mancanza d'adattamento' (*lack of fit*), misurabile con lo scostamento di ogni media dalla linea di regressione. In effetti, la regressione lineare è solo un'approssimazione della reale relazione biologica tra la concimazione e la produzione del frumento.

Insomma, il modello di regressione è un modello che ha sempre minor capacità descrittiva rispetto ad un modello ANOVA. La differenza può essere quantificata utilizzando le devianze dei rispettivi residui:

$$\text{Lack of fit} = 70.37 - 61.22 = 9.15$$

Bisogna però anche dire che il modello di regressione è più parsimonioso, nel senso che ci ha costretto a stimare solo due parametri (b_0 e b_1), mentre il modello ANOVA ce ne ha fatti stimare quattro (μ, α_2, α_3 e α_4, considerando che $\alpha_1 = 0$). Quindi il residuo del modello di regressione ha 14 gradi di libertà (16 dati meno due parametri stimati), mentre il residuo del modello ANOVA ne ha 12 (16 - 4). La componente di lack of fit ha quindi 14 - 12 = 2 gradi di libertà. Ci chiediamo, questa componente di lack of fit è significativamente più grande dell'errore puro?

L'ipotesi nulla di assenza di lack of fit può essere testata con un test di F, per il confronto di due varianze: se questo è significativo allora la componente di mancanza d'adattamento non è trascurabile, ed il modello di regressione dovrebbe essere rifiutato. L'espressione è:

$$F_{lack} = \frac{\frac{RSS_r - RSS_a}{DF_r - DF_a}}{\frac{RSS_a}{DF_a}} = \frac{MS_{lack}}{MSE_a}$$

dove RSS_r è la devianza residua della regressione con i sui gradi di liberta DF_r e RSS_a è la devianza residua del modello ANOVA, con i suoi gradi di libertà DF_a. In R, il test F per la mancanza d'adattamento può essere eseguito con la funzione `anova()`, confrontando i due modelli alternativi:

```
anova(modelReg, model)
## Analysis of Variance Table
##
## Model 1: Yield ~ Dose
## Model 2: Yield ~ DoseF
##   Res.Df    RSS Df Sum of Sq      F Pr(>F)
## 1     14 70.373
## 2     12 61.219  2    9.1542 0.8972 0.4334
```

Vediamo che non otteniamo risultati significativi (P = 0.4334). Ciò supporta l'idea che non vi sia mancanza d'adattamento e quindi la regressione fornisca una descrizione altrettanto adeguata dei dati sperimentali rispetto al più 'complesso' modello ANOVA. Scegliamo quindi il modello di regressione, in quanto più semplice, nel rispetto del principio del rasoio di Occam.

11.5.4 Test F per la bontà di adattamento e coefficiente di determinazione

Abbiamo dimostrato che il modello di regressione non è significativamente peggiore del modello ANOVA corrispondente. Un approccio alternativo per dimostrare la bontà di adattamento è verificare se il modello di regressione è significativamente migliore di un modello 'nullo.' Ricordiamo che con il modello 'nullo' ($Y_i = \mu + \varepsilon_i$) si assume che la risposta sia costante e pari alla media di tutti i dati, escludendo così ogni effetto della dose di concimazione. La devianza del residuo di un modello nullo non è altro che la devianza totale dei dati, che risulta pari a 1787.178:

```
modNull <- lm(Yield ~ 1, data = dataset)
deviance(modNull)
## [1] 1787.178
```

Abbiamo visto che la devianza del modello di regressione è pari a 70.37: la differenza (1716.81) rappresenta la 'bontà di adattamento,' cioè una misura di quanto migliora il potere descrittivo del modello aggiungendo l'effetto 'dose.' Quindi, un test di F per la bontà di adattamento può essere costruito come:

$$F_{good} = \frac{\frac{RSS_t - RSS_r}{DF_t - DF_r}}{\frac{RSS_r}{DF_r}} = \frac{MS_{good}}{MSE_r}$$

dove $\mathrm{RSS_t}$ è la devianza totale dei dati con i sui gradi di liberta $\mathrm{DF_t}$. In R, il test F per la bontà d'adattamento può essere eseguito con la funzione `anova()`, senza la necessità di includere come argomento il modello 'nullo':

```
anova(modelReg)
## Analysis of Variance Table
##
## Response: Yield
##           Df  Sum Sq Mean Sq F value    Pr(>F)
```

```
## Dose       1 1716.80 1716.80  341.54 3.129e-11 ***
## Residuals 14   70.37    5.03
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Vediamo che in questo caso l'ipotesi nulla deve essere rifiutata: la varianza spiegata dalla regressione è significativamente maggiore di quella del residuo.

Più frequentemente, la devianza spiegata dalla regressione viene rapportata alla devianza totale, per individuare quanta parte della variabilità dei dati è spiegata dal modello prescelto. Questa proporzione definisce il cosidetto **coefficente di determinazione** o R^2:

$$R^2 = \frac{SS_{reg}}{SS_{tot}} = \frac{1716.81}{1787.18} = 0.961$$

Questa statistica varia da 0 ad 1 e la regressione è tanto migliore quanto più essa si avvicina ad 1. In realtà il coefficiente di determinazione è visibile nell'output della funzione `summary()` applicata all'oggetto 'modelReg' (vedi più sopra).

11.6 Previsioni

Dato che il modello ha mostrato di funzionare bene, con prudenza, possiamo utilizzarlo per effettuare due tipi di previsioni: diretta ed inversa. Nel primo caso, possiamo prevedere la risposta per una qualunque dose, nel secondo caso, possiamo prevedere la dose che ha indotto una data risposta (Figure 11.3).

Per entrambi i tipi di previsione ci si dovrebbe mantenere entro i livelli di dosi e risposte massimi e minimi inclusi ed osservati in prova, per evitare pericolose estrapolazioni (la risposta ha mostrato di essere lineare solo nell'intervallo osservato, ma non sappiamo cosa potrebbe capitare fuori da questo).

I parametri stimati possono essere facilmente utilizzati per fare previsioni dirette, come indicato nel box sottostante.

```
23.793750 + 0.154417 * 30
## [1] 28.42626
```

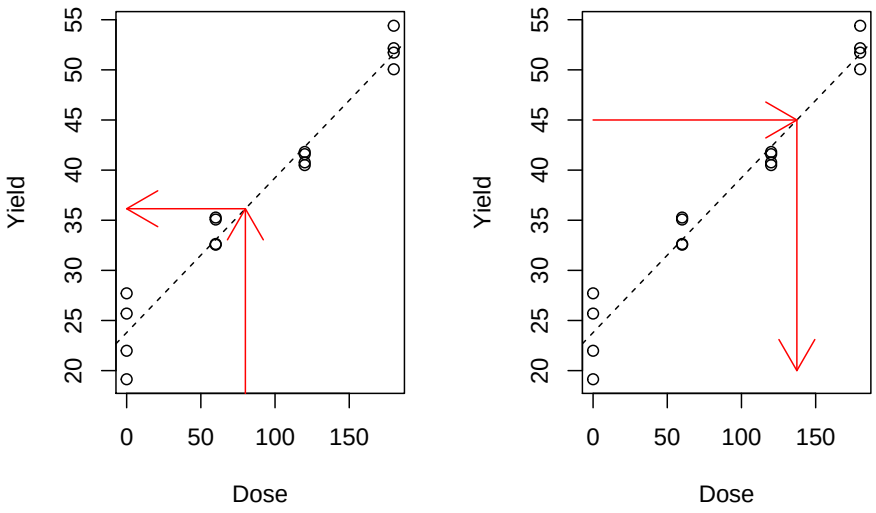

Figura 11.3: Esempio di previsioni: a destra, previsione della risposta per una data dose, a sinistra previsione della dose per ottenere una data risposta

```
23.793750 + 0.154417 * 80
## [1] 36.14711
```

Anche queste stime, come tutte le altre, si riferiscono al campione in prova, mentre noi siamo interessati a fornire una previsione per l'intera popolazione da cui i nostri dati sono capionati. Per questo motivo, abbiamo la necessità di quantificare l'incertezza, attraverso un'opportuna stima dell'errore standard. In R, ciò è possibile utilizzando la funzione `predict()`, passando come argomento le dosi alle quali effettuare la previsione, organizzate in un data frame. Ad esempio, se si vuole prevedere la produzione a 30 e 80 kg N ha^{-1}, il codice è:

```
pred <- predict(modelReg, newdata=data.frame(Dose=c(30, 80)), se=T)
pred
## $fit
##        1        2
## 28.42625 36.14708
##
## $se.fit
##         1         2
## 0.7519981 0.5666999
```

```
##
## $df
## [1] 14
##
## $residual.scale
## [1] 2.242025
```

E'anche possibile effettuare la previsione inversa, cioè chiedere ai dati qual è la dose a cui corrisponde una produzione di 45 q/ha. In questo caso dobbiamo tener presente che l'equazione inversa è:

$$X = \frac{Y - b_0}{b_1}$$

e la previsione corrispondente è:

```
(45 - 23.793750)/0.154417
## [1] 137.3311
```

Per determinare l'errore standard possiamo utilizzare la funzione `deltaMethod()`, nel package 'car,' che ci calcola anche gli errori standard con il metodo della propagazione degli errori:

```
car::deltaMethod(modelReg, "(45 - b0)/b1",
                 parameterNames=c("b0", "b1"))
##               Estimate       SE    2.5 % 97.5 %
## (45 - b0)/b1 137.3314   4.4424 128.6244 146.04
```

Il procedimento sopra descritto è molto comune, per esempio nei laboratori chimici, dove viene utilizzato nella fase di calibrazione di uno strumento. Una volta che la retta di calibrazione è stata individuate, essa viene utilizzata per determinare le concentrazioni incognite di campioni per i quali sia stata misurata la risposta.

11.7 Altre letture

1. Draper, N.R., Smith, H., 1981. Applied Regression Analysis, in: Applied Regression. John Wiley & Sons, Inc., IDA, pp. 224–241.
2. Faraway, J.J., 2002. Practical regression and Anova using R. `http://cran.r-project.org/doc/contrib/Faraway-PRA.pdf`, R.

Capitolo 12

Modelli ANOVA a due vie

Placeholder

Capitolo 13

La regressione non-lineare

Placeholder

13.1 Caso studio: degradazione di un erbicida nel terreno

13.2 Scelta della funzione

13.3 Stima dei parametri

13.3.1 Linearizzazione della funzione

13.3.2 Approssimazione della vera funzione tramite una polinomiale in X

13.3.3 Minimi quadrati non-lineari

13.4 La regressione non-lineare con R

13.5 Verifica della bontà del modello

13.5.1 Analisi grafica dei residui

13.5.2 Test F per la mancanza di adattamento (approssimato)

13.5.3 Errori standard dei parametri

13.5.4 Coefficienti di determinazione

13.6 Funzioni lineari e nonlineari dei parametri

13.7 Previsioni

13.8 Gestione delle situazioni 'patologiche'

13.8.1 Trasformazione del modello

13.8.2 Trasformazione dei dati

13.9 Per approfondire un po'...

Capitolo 14

Esercizi

Placeholder

14.1 Capitoli 1 e 2

14.1.1 Esercizio 1

14.2 Capitolo 3

14.2.1 Esercizio 1

14.2.2 Esercizio 2

14.2.3 Esercizio 3

14.3 Capitolo 4

14.3.1 Esercizio 1

14.3.2 Esercizio 2

14.3.3 Esercizio 3

14.3.4 Esercizio 4

14.3.5 Esercizio 5

14.3.6 Esercizio 6

14.3.7 Esercizio 7

14.3.8 Esercizio 8

14.4 Capitolo 5

14.4.1 Esercizio 1

14.4.2 Esercizio 2

14.4.3 Esercizio 3

14.4.4 Esercizio 4

14.4.5 Esercizio 5

Capitolo 15

Appendice 1: breve introduzione ad R

Placeholder

Cosa è R?

Oggetti e assegnazioni

Costanti e vettori

Matrici

Dataframe

Quale oggetto sto utilizzando?

Operazioni ed operatori

Funzioni ed argomenti

Consigli per l'immissione di dati sperimentali

Immissione di numeri progressivi

Immissione dei codici delle tesi e dei blocchi

Immissione dei valori e creazione del datframe

Leggere e salvare dati esterni

Alcune operazioni comuni sul dataset

Selezionare un subset di dati

Ordinare un vettore o un dataframe

Workspace

Script o programmi

Interrogazione di oggetti