

Stima dei parametri di modelli lineari

Indice

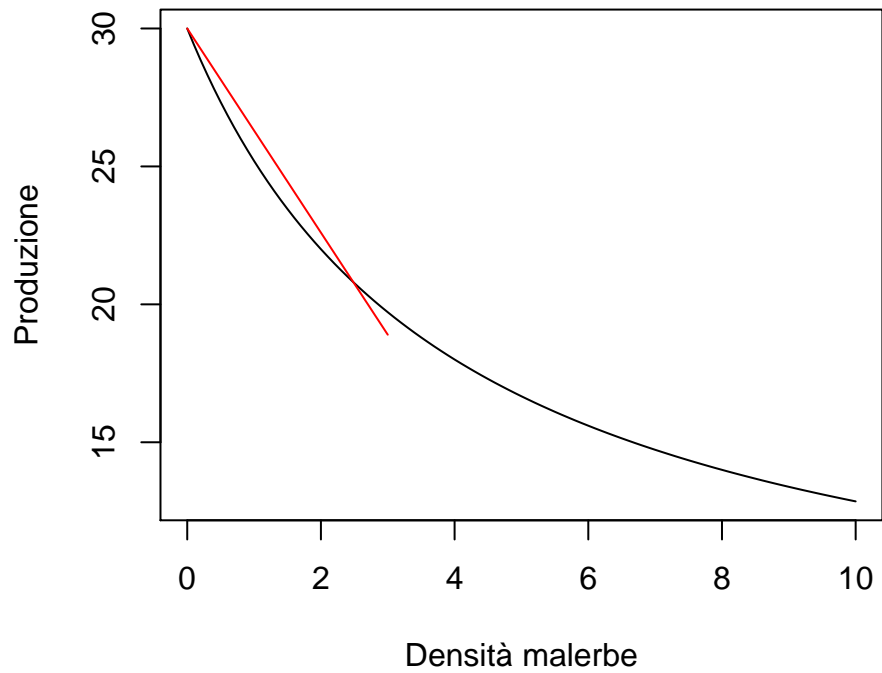
Introduzione	1
Il caso studio	2
Stima dei parametri	3
Bontà delle stime	4
Per approfondimenti	5

Introduzione

In alcune situazioni un modello lineare è più che sufficiente per descrivere un fenomeno biologico, soprattutto quando il nostro interesse risiede in un intervallo abbastanza ristretto della variabile indipendente (approssimazione locale). Ad esempio, se stiamo studiando la relazione tra la densità delle piante infestanti e la produttività di una coltura, è ragionevole attendersi che essa segua un andamento iperbolico decrescente, con un certo asintoto inferiore (figura sottostante). Tuttavia, se vogliamo studiare la soglia economica d'intervento, che in genere si trova a valori piuttosto bassi di densità, il nostro interesse risiede solo nella prima parte della curva, che potremmo quindi approssimare con una retta:

$$P = aD + b + \epsilon$$

dove la produzione P è funzione lineare della densità D , con due parametri a (pendenza) e b (intercetta). Le produzioni effettivamente osservate saranno tuttavia influenzate anche dalla variabilità casuale ϵ (errore sperimentale), normalmente distribuita con media 0 e varianza pari a σ^2 .



Il caso studio

Per parametrizzare la relazione lineare densità-produzione esposta in precedenza, è stato organizzato un esperimento a blocchi randomizzati, dove sono stati inclusi sette diversi livelli di infestazione di *Sinapis arvensis* ed è stata rilevata la produzione di acheni del girasole. I risultati sono:

##	density	block	yield
## 1	0	1	36.630
## 2	14	1	29.730
## 3	19	1	32.120
## 4	28	1	30.610
## 5	32	1	27.700
## 6	38	1	27.430
## 7	54	1	24.790
## 8	0	2	36.110
## 9	14	2	34.720
## 10	19	2	30.120
## 11	28	2	30.800
## 12	32	2	26.530
## 13	38	2	27.600
## 14	54	2	23.310
## 15	0	3	38.350

```

## 16      14      3 32.160
## 17      19      3 31.720
## 18      28      3 28.690
## 19      32      3 25.880
## 20      38      3 28.430
## 21      54      3 30.260
## 22       0      4 36.740
## 23      14      4 32.566
## 24      19      4 29.570
## 25      28      4 33.663
## 26      32      4 28.751
## 27      38      4 27.114
## 28      54      4 24.664

```

Iniziamo questo studio verificando il rispetto delle assunzioni di base ed eseguendo l'ANOVA, che permette di calcolare il SEM e il coefficiente di variabilità, da utilizzare come 'descrittori' della bontà della prova e dell'incertezza dei risultati. In questo caso il test d F e i confronti multipli sono abbastanza irrilevanti, rispetto agli obiettivi dell'esperimento.

```

## Analysis of Variance Table
##
## Response: yield
##           Df  Sum Sq Mean Sq F value    Pr(>F)
## density    1 310.127 310.127 78.2306 3.593e-09 ***
## block      1   2.438   2.438  0.6149  0.4403
## Residuals 25  99.107   3.964
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

Dando per scontato il rispetto degli assunti di base, possiamo quindi calcolare le medie e procedere alla parametrizzazione del modello (si potrebbe anche operare sui dati tal quali, comunque!).

```

## Density YieldM
## 1      0 36.95750
## 2     14 32.29400
## 3     19 30.88250
## 4     28 30.94075
## 5     32 27.21525
## 6     38 27.64350
## 7     54 25.75600

```

Stima dei parametri

Dato che il modello è lineare, la stima dei parametri può essere eseguita con il metodo dei minimi quadrati, che in EXCEL è implementato nella funzione

REGR.LIN(y_nota, x_nota, int, stat). Si tratta di una funzione di matrice, che deve essere immessa in questo modo:

1 - Selezionare un intervallo composto da 5 righe e tante colonne quanti sono i parametri da stimare (in questo caso due) 2 - Immettere la funzione nella cella attiva, mantenendo l'intervallo selezionato 3 - Consolidare la formula premendo contemporaneamente CTRL+SHIFT+INVIO.

Il risultato propone i seguenti valori:

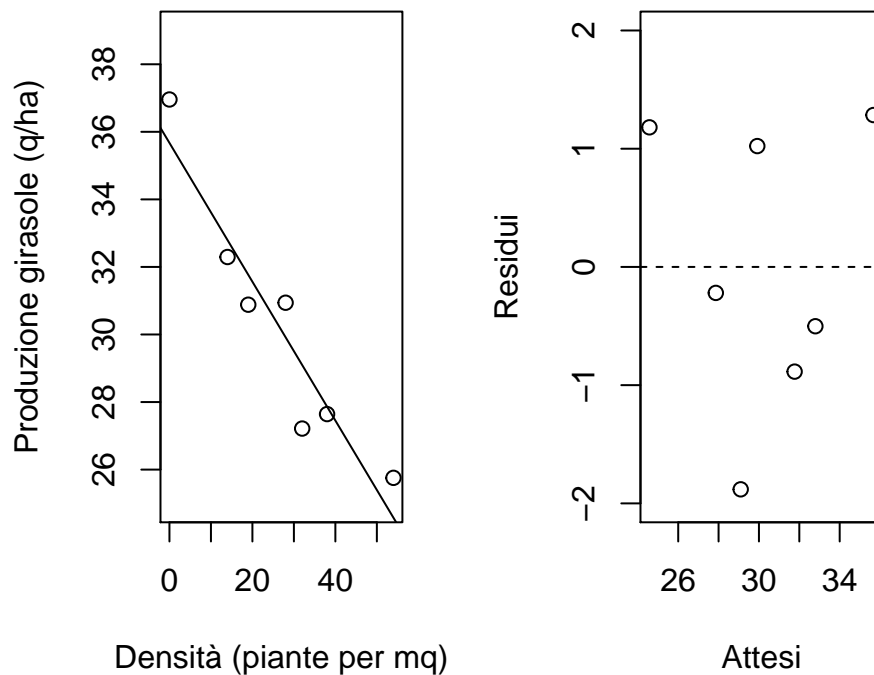
-0.2055121401	35.6727494163
0.0308101753	0.9549734416
0.8989745535	1.3200692458
44.4924811044	5
77.5318329134	8.7129140687

Nella prima riga vi sono le stime dei minimi quadrati (rispettivamente a e b), nella seconda riga gli errori standard, nella terza riga abbiamo il coefficiente di determinazione (R^2) a sinistra e l'errore standard di un valore stimato y (a destra), nella quarta riga abbiamo il valore di F per la regressione, per testare l'ipotesi che la relazione tra x e y è solo casuale (a sinistra) e il numero dei gradi di libertà della regressione (a destra; è pari al numero dei dati meno il numero dei parametri stimati), nell'ultima riga abbiamo la devianza della regressione (a sinistra) e la devianza del residuo (RSS; a destra).

Bontà delle stime

Dopo la stima dei parametri, è necessario valutare la bontà delle stime ottenute. Questo aspetto fa parte della più generale valutazione del modello che sarà trattata in una lezione a parte.

In *primis*, dato che il metodo dei minimi quadrati lavora in base alle stesse assunzioni dell'ANOVA, gli strumenti diagnostici più immediati sono gli stessi già consigliati in precedenza e fondamentalmente basati sull'analisi dei residui. In particolare è necessario valutare il grafico di dispersione dei dati osservati e stimati vs la variabile indipendente e il grafico dei residui vs. gli attesi. Nel caso in studio i grafici sono riportati più sotto.



Possiamo osservare che:

1. il grafico di sinistra mostra che i dati osservati hanno una lieve tendenza curvilinea (coerente con le aspettative sul fenomeno competitivo) e, in particolare, sia il primo che l'ultimo dato sono sopra la retta di regressione. Tuttavia i valori osservati giacciono vicini al modello.
2. Il grafico dei residui rafforza l'impressione precedente: il primo e l'ultimo residuo sono positivi, mentre gli altri, escluso il quarto, sono tutti negativi.

Insomma, l'uso di una regressione lineare per un fenomeno 'curvilineo' non è ottimale in questo caso, ma sufficientemente buono.

Oltre ai metodi grafici di verifica, dobbiamo 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.

Per approfondimenti

CAMUSSI Alessandro , MOELLER Frank , OTTAVIANO Ercole , SARI GORLA Mirella (1995). Metodi Statistici per la sperimentazione biologica. Zanichelli Editore, 496 pp.