Prof. Mauro Gasparini - Politecnico di Torino
Modelli statistici - Laurea Magistrale in Ingegneria Matematica
15 Ottobre 2019

Ellissoidi di confidenza e di predizione nel modello lineare

Se C una matrice $q \times p$ con $q \leq p$ di rango di riga pieno, allora

$$\frac{(C\widehat{\boldsymbol{\beta}} - C\boldsymbol{\beta})'(C(X'X)^{-1}C')^{-1}(C\widehat{\boldsymbol{\beta}} - C\boldsymbol{\beta})}{q \text{ MSE}} \sim \mathrm{F}(q, n-p) \tag{1}$$

dove $MSE = (Y - \widehat{Y})'(Y - \widehat{Y})/(n-p)$ è lo stimatore corretto di σ^2.

Da questo risultato si possono ricavare regioni di confidenza congiunte ellissoidali di tipo F per $C\boldsymbol{\beta}$:

$$(C\widehat{\boldsymbol{\beta}} - C\boldsymbol{\beta})'(C(X'X)^{-1}C')^{-1}(C\widehat{\boldsymbol{\beta}} - C\boldsymbol{\beta}) \leq q \text{ MSE } \mathrm{F}_\alpha(q, n-p)$$

Quando $C = I$ è la matrice unitaria $p \times p$:

$$(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})'X'X(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \leq p \text{ MSE } \mathrm{F}_\alpha(p, n-p) \tag{2}$$

cioè un ellissoide di confidenza per $\boldsymbol{\beta}$.
Quando C è un vettore riga di tutti zeri tranne 1 all'$i-$esimo posto, per $i = 1, \ldots, p$, si dimostra agevolmente che la regione di confidenza diventa un intervallo di confidenza di tipo t bilaterale per il singolo coefficiente

$$|\widehat{\beta}_{i-1} - \beta_{i-1}| \leq \ \mathrm{t}_{\alpha/2}(n-p)\sqrt{\text{MSE } (X'X)^{-1}_{i,i}}$$

ovvero

$$\widehat{\beta}_{i-1} \pm \ \mathrm{t}_{\alpha/2}(n-p)\sqrt{\text{MSE } (X'X)^{-1}_{i,i}} \tag{3}$$

grazie alla equivalenza distributiva

$$(\text{Student}(k))^2 \stackrel{d}{=} \chi^2(1, k).$$

Quando C è un vettore riga contenente una nuova configurazione dei predittori x'_f otteniamo il seguente intervallo di confidenza bilaterale per la corrispondente risposta media:

$$x'_f\widehat{\beta} \pm \mathrm{t}_{\alpha/2}(n-p)\sqrt{\text{MSE } x'_f(X'X)^{-1}x_f} \tag{4}$$

Quando C è un vettore riga contenente una nuova configurazione dei predittori x'_f otteniamo il seguente intervallo di predizione bilaterale per la corrispondente risposta:

$$x'_f\widehat{\beta} \pm \mathrm{t}_{\alpha/2}(n-p)\sqrt{\text{MSE } (1 + x'_f(X'X)^{-1}x_f)} \tag{5}$$

Un'applicazione in R

Riconsideriamo il dataset `insulate` e il modello additivo. Vediamo delle istruzioni per trovare la matrice X del piano sperimentale (*design matrix*) e la sua inversa

```
> attach(insulate)
> additivo <- lm(cons ~ quando + temp)
> X <- model.matrix(additivo)
> solve(t(X) %*% X)          ### modo di calcolare l'inversa in R
            (Intercept)  quandoprima         temp
(Intercept)  0.08253950 -0.023558248 -0.011024532
quandoprima -0.02355825  0.073736748 -0.002190086
temp        -0.01102453 -0.002190086  0.002470022
```

1. Calcolare un intervallo di confidenza di livello 95% (il default) per il coefficiente di `quandoprima` usando la formula (3) e verificare che sia uguale al risultato dato da

   ```
   > confint(additivo)
                    2.5 %     97.5 %
   (Intercept)  4.7801676  5.1920806
   quandoprima  1.3705402  1.7598691
   temp        -0.3723252 -0.3010687
   ```

2. calcolare due intervalli di confidenza di livello 99% usando la formula (4) per i valore attesi del consumo in corrispondenza di `quando='prima'` e `temp=3.2` e di `quando='dopo'` e `temp=3.2` e verificare che siano uguali a

   ```
   > predict.lm(additivo,
   +            newdata=data.frame(quando=c("prima","dopo"),temp=rep(3.2,2)),
   +            interval="confidence", level=.99)
          fit      lwr      upr
   1 5.473898 5.260625 5.687172
   2 3.908694 3.724325 4.093063
   ```

3. calcolare un'ellisse di confidenza per la stessa coppia di valori attesi usando la formula (2).

4. calcolare due intervalli di predizione di livello 99% usando la formula (5) per i valore attesi del consumo in corrispondenza di `quando='prima'` e `temp=3.2` e di `quando='dopo'` e `temp=3.2` e verificare che siano uguali a

   ```
   > predict.lm(additivo,
   +            newdata=data.frame(quando=c("prima","dopo"),temp=rep(3.2,2)),
   +            interval="prediction", level=.99)
          fit      lwr      upr
   1 5.473898 4.495432 6.452365
   2 3.908694 2.936118 4.881269
   ```

5. esplorare il pacchetto di R chiamato `ellipse` per fare il grafico di ellissi ed ellissi di confidenza.