Analisi Statistica dei Dati: foglio di lavoro sull'inferenza asintotica tramite pivot di Wald e rapporto di verosimiglianza

Luca La Rocca

20 novembre 2018

1. Leggete i dati presentati nella sezione 1.1.6 del testo di riferimento dal file `dataScotland.csv` in un data frame `scotland`; siano $z^\bullet$ il conteggio osservato ed e il conteggio atteso in una regione di interesse.

2. Calcolate, per ciascuna delle 56 regioni, il rapporto tra $z^\bullet$ ed e, salvando i risultati ottenuti in un vettore denominato `SIR`; rappresentate graficamente lo Standardized Incidence Ratio in funzione dei conteggi attesi trasformati su scala logaritmica (in base dieci).

3. Completate il *funnel plot* tracciato al punto 2 con i quantili di livello 0.005 e 0.995 per Z/e, supponendo $Z \sim \mathrm{Pois}(e)$ e facendo variare $\log_{10} e$ su una griglia di valori uniformemente spaziati tra 0 e 2; aggiungete una retta orizzontale in corrispondenza del valor medio $\mathbb{E}(Z/e) = 1$.

   Example 4.14

4. Per ciascuna delle 56 regioni, supponendo $z^\bullet$ osservato da $Z \sim \mathrm{Pois}(e\lambda)$, calcolate l'intervallo di Wald al livello 0.99 per il parametro λ e salvate gli estremi sinistri nel vettore `lambdaLowWald` e quelli destri nel vettore `lambdaUppWald`; tenete presente che $\hat{\Lambda} = Z/e$ mentre $J_\Lambda(\lambda) = e/\lambda$.

5. Nel modello introdotto al punto 4, posto $\varphi = \sqrt{\lambda}$, ricavate lo stimatore di massima verosimiglianza $\hat{\Phi}$ e l'informazione di Fisher attesa $J_\Phi(\varphi)$, quindi calcolate, per ciascuna delle 56 regioni, l'intervallo di Wald al livello 0.99 per il parametro φ e trasformatelo in un intervallo di confidenza allo stesso livello per λ, salvando gli estremi sinistri nel vettore `lambdaLowStab` e quelli destri nel vettore `lambdaUppStab`.

6. Confrontate graficamente gli intervalli ottenuti ai punti 4 e 5, mettendo in ascissa le regioni ordinate per valore di $\hat{\Lambda}$ (Standardized Incidence Ratio).

Example 4.18

7. Per ciascuna delle 56 regioni, calcolate la statistica test di Wald

$$T_3 = \sqrt{J_\Phi(\varphi_0)}(\hat{\Phi} - \phi_0) = 2\left(\sqrt{Z} - \sqrt{e\lambda_0}\right)$$

e la statistica test del rapporto di verosimiglianza

$$T_4 = \text{sign}(\hat{\Lambda} - \lambda_0)\sqrt{2Z\log Z - 2\{\log(e\lambda_0) + 1\}Z + 2e\lambda_0}$$

dove $\varphi_0 = \sqrt{\lambda_0}$ e $\lambda_0 = 1$ è il valore nullo per λ (corrispondente all'ipotesi di regioni omogenee) mentre resta inteso che $0\log 0 = 0$.

8. Calcolate i p-value bilaterali corrispondenti ai valori di T_3 e T_4 osservati nelle 56 regioni e confrontate graficamente le loro radici quadrate mediante un diagramma di dispersione in cui le regioni con $Z < e$ siano rappresentate in rosso e quelle con $Z > e$ siano rappresentate in blu.

   Example 4.19

9. Calcolate, per ciascuna delle 56 regioni, l'intervallo di verosimiglianza al livello 0.99 per il rischio relativo λ e salvate gli estremi sinistri nel vettore `lambdaLowLR` e quelli destri nel vettore `lambdaUppLR`.

10. Verificate che entrambi gli estremi degli intervalli di verosimiglianza ottenuti al punto 9 sono sistematicamente a destra dei corrispondenti estremi degli intervalli di Wald ottenuti al punto 4.

Testo di riferimento

L. Held & D. Sabanés Bové. *Applied Statistical Inference*. Springer, 2014.