

Analisi Statistica dei Dati: foglio di lavoro sull'inferenza asintotica tramite punteggio

Luca La Rocca

13 novembre 2018

Example 4.8

1. Leggete i dati presentati nella sezione 1.1.6 del testo di riferimento dal file `dataScotland.csv` in un data frame denominato `scotland` e controllate che i conteggi attesi e quelli osservati abbiano circa la stessa somma.
2. Calcolate, per ciascuna delle 56 regioni cui i dati si riferiscono, il valore osservato delle statistiche test

$$T_1 = \frac{S(\lambda_0)}{\sqrt{J(\lambda_0)}} = \frac{Z - e\lambda_0}{\sqrt{e\lambda_0}} \quad \text{e} \quad T_2 = \frac{S(\lambda_0)}{\sqrt{I(\lambda_0)}} = \frac{Z - e\lambda_0}{\sqrt{Z}},$$

dove Z è il conteggio della regione e $\lambda_0 = 1$ il valore nullo del rischio relativo $\lambda = \mathbb{E}_\lambda(Z)/e$, avendo supposto $Z \sim \text{Pois}(e\lambda)$, mentre $e = e\lambda_0$ è il conteggio atteso nella regione (sotto l'ipotesi nulla $\lambda = \lambda_0$).

3. Rappresentate graficamente la relazione tra T_1 e T_2 nei dati analizzati per mezzo di un diagramma di dispersione.
4. Calcolate i p-value bilaterali corrispondenti ai valori osservati di T_1 (o T_2) e rappresentateli graficamente mediante un istogramma; in quale proporzione di regioni T_1 (o T_2) rifiuta l'ipotesi nulla al livello di significatività del 10%?

Example 4.9

5. Calcolate, per ciascuna delle 56 regioni, il rapporto tra conteggio osservato e conteggio atteso, salvando i risultati in un vettore denominato `SIR`.
6. Calcolate, per ciascuna delle 56 regioni, l'intervallo di confidenza dato da

$$\frac{Z + q^2/2}{e} - \frac{q}{2e}\sqrt{4Z + q^2} \leq \lambda \leq \frac{Z + q^2/2}{e} + \frac{q}{2e}\sqrt{4Z + q^2},$$

dove q è il quantile di livello 95% della distribuzione normale standard.

7. Rappresentate graficamente gli intervalli di confidenza calcolati al punto 6, mettendo in ascissa le regioni ordinate per Standardized Incidence Ratio (trovando prima con `order` la permutazione `SIRord` che ordina `SIR`).

Testo di riferimento

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