Laboratorio di inferenza statistica (2024-2025)

CAS0 DI STUDIO 1 – 06 aprile 2025

1. (`caso di studio - gasoline`, modello normale). Una società di noleggio di furgoni vuole stimare il consumo medio settimanale per miglio (in galloni) della propria flotta. A tal fine viene rilevato il consumo per miglia settimanale di 24 furgoni. I dati risultanti, che indichiamo con `gasoline`, sono i seguenti:

```
15.5 21.0 18.5 19.3 19.7 16.9 20.2 14.5
16.5 19.2 18.7 18.2 18.0 17.5 18.5 20.5
18.6 19.1 19.8 18.0 19.8 18.2 20.3 21.8
```

I dati vengono considerati realizzazioni di variabili aleatorie iid normali di parametri (θ, σ^2), con θ incognito e $\sigma^2 = 2.8$.

(a) Creare un oggetto `R` contenente il valori dei dati `gasoline`. [Sugg.: si può usare, ad esempio, la funzione `scan()`].

(b) Ottenere e commentare le statistiche descrittive relative ai dai `gasoline`, usando la funzione `summary()`.

(c) Visualizzare l'istogramma delle frequenze relative e discutere se l'ipotesi di normalità sembra appropriata.

(d) Sovrapporre al grafico dell'istogramma quello della stima di densità ottenuto con la funzione `density()`.

(e) Determinare il valore numerico della stima di massima verosimiglianza di θ.

(f) Sovrapporre ai precedenti il grafico della densità normale con valore atteso posto uguale alla stima di massima verosimiglianza di θ basata sui dati e varianza pari a 2.8

(g) Scrivere la funzione `Lik.norm.rel(theta,y,sig2)` per la funzione di verosimiglianza relativa di θ associata a un campione `y` per il modello normale con varianza `sig2`. Disegnare quindi il grafico della funzione nell'intervallo $[17, 20]$ di Θ che si ottiene ponendo `y=gasoline`, inserendo le opportune etichette per gli assi cartesiani.

(h) Scrivere la funzione `L.q.fun(q,y,sig2)` per ottenere l'insieme di verosimiglianza di livello $q \in [0, 1]$ per θ per un generico campione `y`. Determinare l'insieme L_q per θ usando i dati `gasoline` e ponendo $q = 0.75$.

(i) Tracciare nel grafico che rappresenta la fdv relativa le rette verticali in corrispondenza degli estremi dell'intervallo determinato.

(j) Scrivere ora in modo diverso la fdv per θ, indicandola con `Lik.norm.rel.SMV(theta,n,smv,sig2)`. La funzione, questa volta, deve dipendere da `(theta,n,smv,sig2)` [I dati vengono inseriti attraverso la smv e non attraverso il vettore numerico delle osservazioni]. Tracciare, nello stesso sistema di riferimento cartesiano, i grafici della verosimiglianza di θ associata al campione `gasoline` e quello che si ottiene per un nuovo campione, `dati2`, di dimensione $n = 100$, la sui media campionaria coincide con quella dei dati `gasoline`.

(k) Scrivere una nuova funzione `L.q.fun.SMV(q,n,smv,sig2)` per il calcolo degli insiemi di verosimiglianza, questa volta in funzione di `(q,n,smv,sig2)`. Determinare e confrontare gli insiemi di verosimiglianza di livello $q = 0.75$ che si ottengono con `gasoline` e `dati 2`.

(l) Scrivere un breve commento statistico all'analisi effettuata.