Laboratorio di inferenza statistica (2024-2025)

Lab 11 – Intervalli e test (22 maggio 2025)

1. **Test su valore atteso normale**

   I seguenti dati indicano il numero di azioni di diverso tipo scambiate in un'unità di tempo:

   $$(63, 65, 94, 37, 83, 95, 70, 96, 47, 29, 52, 38, 47, 79, 66, 25, 48, 80, 52, 49)$$

   Assumere che i dati siano la realizzazione di un campione casuale da popolazione normale e usare la funzione `t.test()` per rispondere alle seguenti domande.

   (a) Calcolare l'intervallo di confidenza di livello $1 - \alpha = 0.95$ per il valore atteso della popolazione da cui si assume provengano i dati.

   (b) Effettuare un test per stabilire se è possibile accettare l'ipotesi che il numero medio (valore atteso) di azioni scambiato sia pari a 70 contro alternativa che sia **minore di** da 70, assumendo un'ampiezza del test pari a 0.05.

   (c) Con riferimento al test effettuato, usare l'output `R` e indicare il p-value e stabilire se si accetta l'ipotesi nulla.

   (d) Con riferimento al test effettuato, usare l'output `R` e indicare il valore della statistica test, della numerosità campionaria e della media campionaria.

   (e) Con riferimento al test effettuato, usare l'output `R` e calcolare il valore della varianza campionaria corretta.

   (f) Effettuare un test per stabilire se è possibile accettare l'ipotesi che il numero medio (valore atteso) di azioni scambiato sia inferiore a 70 se si sceglie $\alpha = 0.01$.

2. **Inferenza per valori attesi di due popolazioni normali**

   I seguenti sono due campioni relativi all'ammontare in euro delle tasse pagate in un semestre dagli abitanti dei quartieri A e B di una città per lo smaltimento dei rifiuti:

   $\mathbf{x}_{na} = (100, 125, 135, 128, 140, 142, 128, 137, 156, 142, 134, 122, 100, 98, 101)$

   $\mathbf{x}_{nb} = (95, 87, 100, 75, 110, 85, 95, 98, 96, 100, 107)$

   Si ipotizza che i dati provengano da due popolazioni normali indipendenti $\mathrm{N}(\mu_a, \sigma_a^2)$ e $\mathrm{N}(\mu_b, \sigma_b^2)$ (varianze incognite).

   Rispondere ai seguenti quesiti utilizzando le appropriate funzioni `R`, tra le quali `summary` e `t.test`.

   (a) Stimare la spesa media semestrale nei due quartieri. In quale quartiere la spesa risulta più elevata?

   (b) Fornire le stime della variabilità della spesa in A e B (σ_a^2 e σ_b^2). In quale quartiere la spesa è più variabile?

   (c) Fornire gli intervalli di confidenza per μ_a e per μ_b al 95%

   (d) Fornire gli intervalli di confidenza per μ_a e per μ_b all' 80%. Commentare.

   (e) Posso rifiutare l'ipotesi nulla che $\mu_b \geq 100$ (contro alternativa $\mu_b < 100$) in un test di ampiezza 0.05? Motivare la risposta riportando e discutendo il p-value.

   (f) Determinare l'intervallo di confidenza per $\mu_a - \mu_b$ al 95% per cento.

   (g) Sulla base del precedente intervallo posso accettare l'ipotesi che la spesa media nei due quartieri sia uguale (ovvero che la differenza tra le medie sia nulla)? E sulla base del p-value del test bilaterale cosa si conclude?

   (h) Determinare l'intervallo di confidenza per $\mu_a - \mu_b$ al 98%. Cosa si osserva?

   (i) Condurre il test con $\alpha = 0.05$ per stabilire se la differenza tra le spese media possa risultare uguale a 40 euro (contro l'ipotesi bilaterale). Fornire il p-value e stabilire se si può accettare l'ipotesi nulla.

(j) Condurre il test per stabilire se la differenza tra le spese media può risultare maggiore a 45 euro (contro l'ipotesi alternativa che possa essere minore a 40 euro) con $\alpha = 0.05$. Fornire il p-value e stabilire se si può accettare l'ipotesi nulla se $\alpha = 0.05$?

3. **Intervalli e test dati importati da R**

Importare i dati `women` da `R` e considedare le due variabili `height` e `weight`.

(a) Fornire le principali statistiche descrittive e fornire un breve commento.

(b) Stabilire se è possibile accettare l'ipotesi che i due campioni provengano da popolazioni normali.

(c) Trovare gli intervalli di confidenza al 95 % per il valore atteso delle distribuzioni di `height` e `weight`.

(d) Ripetere prendendo un livello di confidenza pari a 0.85.

(e) Effettuare il test $H_0 : \theta \geq 68$ contro l'ipotesi $H_1 < 68$, assumendo $\alpha = 0.1$.

(f) Indicare il p-value e rappresentarlo graficamente, assieme al valore osservato della statistica test e ad α.

(g) Con riferimento al test effettuato, usare l'output `R` e calcolare il valore della varianza campionaria corretta.

4. **Intervallo confidenza Poisson**

In una giornata sono arrivare 27 richieste di risarcimento in 7 agenzie di una compagnia di assicurazioni. Supponendo per i dati una distribuzione di Poisson, calcolare l'intervallo di confidenza al livello 0.95 per il numero medio complessivo di richieste di risarcimenti. Usare `poisson.test` e confrontare con l'intervallo asintotico teorico.

5. **Intervalli di confidenza per la varianza modello normale**

Considerare il modello $\mathrm{N}(\theta_1, \theta_2)$ (entrambi i parametri incogniti). Scrivere la funzione `R` denominata `IC.var(dati,conf.lev)` che restituisca stima puntuale e gli estremi dell'intervallo di confidenza di livello `conf.lev` per θ_2 in corrispondenza del vettore di osservazioni `dati`. Applicare la funzione ai dati dell'Esercizio 2.