

Master in Data Science
Statistical Models
Richiami di Inferenza Statistica (prof. Anthony Cossari)
verifica

Per ciascuno degli esercizi proposti, evidenziare la variabile (o le variabili) di popolazione e le ipotesi sul modello distributivo.

Inoltre, discutere i risultati dell'analisi con particolare attenzione all'interpretazione del p-value e/o dell'intervallo di confidenza.

Naturalmente, per gli aspetti computazionali, fare riferimento ai comandi di R e agli script descritti nel corso delle lezioni. Per produrre l'output si può fare uso, in alternativa, di R Commander (impiegato durante le lezioni), di R Studio, o anche della Console originale di R. In ogni modo, salvare l'output come file di testo (formato .txt), inserendo al suo interno anche i commenti e le interpretazioni richieste, ed allegarlo ad un post (conversazione) da inviare nel team di Microsoft Teams.

Esercizio 1

Il valore di ph dell'acqua che fuoriesce da un impianto di depurazione ha un livello nominale, specificato dal produttore, pari a 7. Prelevando dall'impianto, in modo casuale, 15 campioni di acqua ed analizzandone i contenuti, si ottiene per il valore di ph una media campionaria pari a 6.8 e una deviazione standard campionaria pari a 0.9.

Stabilire se la specificazione del livello nominale fornita dal produttore è corretta, attraverso un intervallo di confidenza al 95%.

Esercizio 2

In uno studio per accertare l'efficacia di un programma di esercizi fisici per dimagrire, sono stati ottenuti i valori della riduzione di peso (in kg), relativi ad un campione casuale di 9 persone a cui era stato assegnato il suddetto programma per la durata di un mese. I dati sono riportati nel file Excel allegato (*osservazione*: valori negativi della riduzione di peso indicano aumento dello stesso).

Verificare l'efficacia del programma di dimagrimento, mediante un opportuno test unilaterale (stabilire se a coda destra o a coda sinistra), al livello di significatività del 1%.

(*suggerimento*: fissare l'ipotesi nulla negando l'efficacia del programma).

Esercizio 3

In un campione casuale di 12 pazienti trattati con il farmaco A, media e deviazione standard campionarie del tasso di glicemia sono pari rispettivamente a 105 e 7.5, mentre in un campione casuale, indipendente dal primo, di 10 pazienti trattati con il farmaco B esse sono pari rispettivamente a 112 e 9.2.

Stabilire se il farmaco A è da ritenersi più efficace del farmaco B nel *ridurre* il tasso di glicemia (la media del tasso di glicemia), mediante un opportuno test unilaterale (stabilire se a coda destra o a coda sinistra), al livello di significatività del 5%.

(*suggerimento*: fissare l'ipotesi nulla supponendo che i due farmaci hanno uguale efficacia).

Esercizio 4

Si vuole stabilire se un certo farmaco ha effetto per la cura dell'ipertensione. Si esegue un esperimento su un campione casuale di 11 pazienti, ciascuno dei quali è sottoposto a due trattamenti, il farmaco oggetto di studio ed un placebo, con relativa misurazione della pressione (sistolica). I dati sono riportati nel file Excel allegato.

Per studiare l'effetto del farmaco, eseguire un opportuno test unilaterale (stabilire se a coda destra o a coda sinistra), al livello di significatività del 5%, per verificare l'ipotesi nulla che il farmaco non è efficace.

(*suggerimento*: nel fissare l'ipotesi nulla, e quella alternativa, si noti che la non efficacia del farmaco equivale all'uguaglianza delle medie di pressione relative al farmaco ed al placebo]

Esercizio 5

Da un'indagine di mercato su 600 famiglie è emerso che 390 di esse fanno la spesa alimentare nei grandi supermercati.

Da studi preliminari, si suppone che il 60% delle famiglie, nella popolazione di riferimento, acquista i prodotti alimentari nei grandi supermercati.

Per verificare tale supposizione, ricavare un intervallo di confidenza al 95%.

Ripetere l'analisi mediante un intervallo di confidenza al 99%, e commentare opportunamente.

Esercizio 6

In un campione casuale di 83 studenti universitari è stato osservato un numero di fumatori pari a 23; in un campione casuale, indipendente dal primo, di 108 studentesse, il numero di fumatrici è risultato pari a 26.

Stabilire se, nelle popolazioni di riferimento, le proporzioni dei fumatori e delle fumatrici si equivalgono o meno, attraverso un intervallo di confidenza al 95%.

Esercizio 7

In uno studio sulle famiglie calabresi, è stato registrato il numero di figli per un campione numeroso di famiglie selezionate casualmente. I dati sono riportati nel file Excel allegato.

Per misurare l'ordine di grandezza del numero di figli nelle famiglie calabresi, ricavare un intervallo di confidenza al 95%.

Esercizio 8

Per accertare se c'è differenza di trattamento economico tra gli operai di una fabbrica, in due stabilimenti diversi A e B, è stato esaminato un gruppo numeroso di operai, scelti casualmente tra quelli dei due stabilimenti, registrando per ciascuno di essi lo stabilimento di provenienza e lo stipendio medio mensile. I dati sono riportati nel file Excel allegato.

Verificare se esiste differenza di trattamento economico, eseguendo un test bilaterale al livello di significatività del 1%.