

PRIMO APPELLO

Prova a risolvere i seguenti problemi, giustificando il ragionamento seguito.

La motivazione del processo è molto più importante della risposta numerica.

Puoi usare una calcolatrice o un regolo per i conti, così come R sul calcolatore dell'aula. PUOI CONTROLLARE I TUOI APPUNTI, LE NOTE DEL CORSO O UN LIBRO DI TESTO TRA QUELLI CONSIGLIATI. Lavora per tuo conto, senza aiuto esterno, ma discuti pure i problemi e le tue soluzioni **finita la prova**. I problemi **non** sono in ordine di difficoltà. Indica nome e cognome (e numero di matricola) sui fogli, così come il numero del problema o della domanda. **Lascia un po' di spazio per i commenti**. Le parti in R possono essere copiate sul foglio, oppure salvate come file (.R, .txt, .R-history) indicando con un commento (introdotto dal carattere #) a quale esercizio e domanda il codice si riferisce. Se parte di un problema è svolta in R, indicalo sul foglio in corrispondenza del punto dell'esercizio corrispondente. Non dimenticare di caricare il file nella risorsa esameonline al termine dell'esame.

Buon lavoro!

Problema 1. L'ufficio personale di una grande azienda vuole studiare il rapporto statistico tra le gratifiche e l'anzianità dei e delle dipendenti. Ogni volta che una gratifica viene assegnata a qualche dipendente, sia X la frazione di dipendenti con anzianità maggiore rispetto alla persona premiata. Si pensa di modellizzare X con una funzione di densità lineare, su un opportuno dominio, $f_X(x) = ax + b$.

Siano inoltre μ e σ^2 rispettivamente il valore atteso e la varianza di X .

1. Qual è il supporto di X ?
2. Determinare i coefficienti a e b in funzione di μ .
3. Determinare σ^2 in funzione di μ .
4. Determinare eventuali restrizioni ai possibili valori che possono assumere la speranza e la varianza di questa classe di variabili aleatorie.

Problema 2. Un gruppo di ricercatori ha condotto uno studio per confrontare il tempo che le persone dedicano quotidianamente a interazioni affettive con il proprio animale domestico. I partecipanti sono stati divisi in due gruppi: chi possiede un cane e chi possiede un gatto. Ognuno ha riportato il tempo (in minuti) dedicato in un giorno a interazioni affettive (carezze, giochi, ecc.) e il tempo (in anni) da cui convive con quell'animale.

I dati sono riportati nel file `interazioni_animali.csv`, che include le variabili:

- `specie`: ``cane" o ``gatto"
 - `affetto_minuti`: tempo totale in minuti per le interazioni affettive
 - `anni_insieme`: anni di convivenza con l'animale
1. Esplora, anche con strumenti grafici, la distribuzione delle variabili per i due gruppi. Che differenze noti? Che somiglianze?

2. Come tratteresti i valori mancanti nei dati? Motiva la tua scelta.
3. Formula un test di ipotesi per verificare se, in media, le persone passano più tempo con il proprio cane rispetto al proprio gatto.
4. Applica il test al dataset. Riporta la statistica test, il p-value e un intervallo di confidenza al 98% per la differenza tra le medie.
5. Ripeti il test considerando solo le persone che convivono con l'animale da almeno 5 anni. Cosa cambia?
6. Quali conclusioni puoi trarre dai risultati? Cosa diresti ai ricercatori?

Problema 3. Alex lancia un normale dado a 6 facce, ottenendo un punteggio X . Tira poi X monete bilanciate.

1. Qual è la probabilità, sapendo che non è uscita alcuna testa, che Alex abbia ottenuto 3 sul dado?
2. Se invece X fosse il numero di messaggi di spam che Alex ha ricevuto il 22 giugno (che in media sono α al giorno) e lanciasse una moneta per ciascun messaggio spam ricevuto, quale sarebbe, al variare di k , la probabilità che $X=k$ sapendo che non è uscita alcuna testa?
3. Di che distribuzione si tratta? Qual è il suo valore atteso?