Compito scritto dell'esame di Statistica e analisi dei dati 18 settembre 2023	Prof. Giuseppe Boccignone	Corso di Laurea
Cognome:	Nome:	Matricola:

Istruzioni

- Il tempo riservato alla prova scritta e' di 2 ore. Durante la prova e' possibile consultare il formulario ed utilizzare la calcolatrice. Non e' possibile consultare libri, appunti, cellulari.
- Ogni foglio deve riportare il numero di matricola
- In ogni esercizio occorre indicare chiaramente, per ogni risposta, il numero della domanda corrispondente
- Riportare lo svolgimento degli esercizi per esteso (quando l'esercizio richiede piu' passaggi di calcolo, non sara' preso in considerazione se riporta solo le soluzioni). Se una serie di calcoli coinvolge una o piu' frazioni semplici (numeratore e denominatore interi), per chiarezza, si svolgano i calcoli mantenendo tali numeri in forma frazionaria fin dove possibile (non li si converta nelle loro approssimazioni con virgola e decimali: solo il risultato finale sara' eventualmente rappresentato in quest'ultima forma).

Problemi

Esercizio 1.
Le probabilitá che tre uomini colpiscano un bersaglio sono, rispettivamente $\frac{1}{6}$ $\frac{1}{4}$ e $\frac{1}{3}$.

(a) Determinare la probabilitá che uno solo colpisca il bersaglio

Soluzione: Definiamo i seguenti eventi indipendenti:

- $A =$ `"il primo uomo colpisce il bersaglio"`
- $B =$ `"il secondo uomo colpisce il bersaglio"`
- $C =$ `"il terzo uomo colpisce il bersaglio"`

Il testo del problema ci dice che:

$$P(A) = \frac{1}{6}, P(B) = \frac{1}{4}, P(C) = \frac{1}{3}$$

Denotiamo con $\sim A, \sim B, \sim C$ gli eventi complementari. Ovviamente, $P(\sim A) = 1 - P(A)$, ecc.
L'evento di interesse é

$T =$ `"uno solo colpisce il bersaglio"`,

ed é definito mediante l'unione di tre eventi mutuamente esclusivi

$$T = \{A \cap \sim B \cap \sim C\} \cup \{\sim A \cap B \cap \sim C\} \cup \{\sim A \cap \sim B \cap C\}.$$

La sua probabilitá é data dunque da:

$$P(T) = P(\{A \cap \sim B \cap \sim C\}) + P(\{\sim A \cap B \cap \sim C\}) + P(\{\sim A \cap \sim B \cap C\})$$

Infine, usando l'indipendenza degli eventi:

$$P(T) = \frac{1}{6} \times (1 - \frac{1}{4}) \times (1 - \frac{1}{3}) + (1 - \frac{1}{6}) \times \frac{1}{4} \times (1 - \frac{1}{3}) + (1 - \frac{1}{6}) \times (1 - \frac{1}{4}) \times \frac{1}{3} = 0.43$$

(b) Determinare la probabilitá che uno solo colpisca il bersaglio e che questi sia il primo uomo.

Soluzione: La probabilitá che l'unico che ha colpito il bersaglio sia il primo uomo é la probabilitá condizionata

$$P(A \mid T) = \frac{P(A \cap T)}{P(T)} = \frac{P(\{A \cap \sim B \cap \sim C\})}{P(T)} = \frac{\frac{1}{6} \times (1 - \frac{1}{4}) \times (1 - \frac{1}{3})}{0.43} = 0.19$$

Esercizio 2. Una variabile aleatoria X ha densitá $f_X(x) = kx$ per $0 \leq x \leq 3$

(a) Trovare x_1 e x_2 tali che $P(X \leq x_1) = 0.1$ e $P(X \leq x_2) = 0.95$.

Soluzione:
Per procedere, trovo la costante k imponendo la normalizzazione

$$\int_0^3 f_X(x)dx = \int_0^3 kxdx = 1$$

da cui ottengo $k = \frac{2}{9}$
Per qualunque $0 \leq x \leq 3$,

$$P(X \leq x) = \frac{2}{9}\int_0^x tdt = \frac{x^2}{9} = F_X(x)$$

.
Pertanto, $\frac{x_1^2}{9} = 0.1$ e $\frac{x_2^2}{9} = 0.95$. Da cui

$$x_1 = \sqrt{(0.9)} = 0.9487$$

$$x_2 = \sqrt{(8.55)} = 2.9240$$

(b) Determinare $P(|X - 1.8| < 0.6)$.

Soluzione: Usando la cumulativa $F_X(x)$ calcolata precedentemente:

$$P(|X - 1.8| < 0.6) = P(1.2 < X < 2.4) = \frac{1}{9}(2.4^2 - 1.2^2) = 0.48$$

Esercizio 3.
Al servizio di pronto soccorso di un ospedale, aperto 24 ore su 24, arrivano in media 48 chiamate al giorno, due in media all'ora

(a) Calcolare la probabilitá che nella prima ora arrivino almeno due chiamate

Soluzione: (Distribuzione di Poisson) Possiamo assumere che le chiamate si distribuiscano con legge di Poisson. La variabile aleatoria X che descrive il numero di chiamate per ora ha media

$$\mu = 2,$$

pertanto la distribuzione di X si scrive

$$p_X(x) = \frac{2^x e^{-2}}{x!}.$$

Le probabilitá che non arrivino chiamate o che ne arrivi solo una sono date rispettivamente da

$$P(X = 0) = p_X(0) = \frac{2^0 e^{-2}}{0!} = e^{-2},$$

$$P(X = 1) = p_X(1) = \frac{2^1 e^{-2}}{1!} = 2e^{-2}.$$

Usando la probabilitá complementare si ha

$$P(X > 1) = 1 - P(0 \leq X \leq 1) = 1 - 3e^{-2}.$$

(b) Calcolare la probabilitá che il tempo di attesa fino alla prima chiamata di un certo giorno sia di almeno un'ora

Soluzione: Il quesito si risolve immediatamente notando che

$$P(T \geq 1) = P(X = 0) = e^{-2}.$$

Esercizio 4. Si dispone di un campione di 100 misure di una variabile temporale X di una popolazione di cui non conosciamo la distribuzione, ma la cui deviazione standard é nota e vale $\sigma_X = 120$ secondi. .

(a) Qual é la probabilitá che la media campionaria differisca per piú di 3 secondi dal valore atteso teorico (incognito) dei tempi misurati?

Soluzione:
Siano:

$$\frac{S_n}{n} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

la media campionaria;

$$\mu_X = E[X]$$

il valore atteso teorico.
Il problema ci chiede di determinare la probabilitá

$$P_X\left(\left|\frac{S_n}{n} - \mu\right| > 3\right)$$

Non conosciamo la legge di probabilitá secondo cui si distribuisce la popolazione, ma il Teorema Centrale Limite (TCL) ci assicura che per n sufficientemente grande la distribuzione degli scarti fra media campionaria $\frac{S_n}{n}$ e valore atteso teorico μ_X segue una legge Gaussiana:

$$P\left(a \le \frac{\sqrt{n}}{\sigma_X}\cdot\left(\frac{S_n}{n} - \mu\right) \le b\right) \approx \Phi\left(b\right) - \Phi\left(a\right).$$

Nel nostro caso,

$$\frac{\sigma_X}{\sqrt{n}} = \frac{120}{\sqrt{100}} = 12$$

é la deviazione standard del campione.
Usando il TCL con $a = -3, b = 3$ e passando alla variabile standardizzata $Z_n \sim N(0,1)$

$$Z_n = \frac{\frac{S_n}{n} - \mu_X}{\frac{\sigma_X}{\sqrt{n}}},$$

il problema iniziale si riduce a calcolare la seguente:

$$P_X\left(\left|\frac{S_n}{n} - \mu_X\right| > 3\right) = P_Z\left(\frac{\sigma_X}{\sqrt{n}}|Z_n| > 3\right) = P_Z(|Z_n| > 0.25).$$

Nel caso Gaussiano, $P_Z(|Z_n| > z) = 2(1 - \Phi(z))$, dunque :

$$P_Z(|Z_n| > 0.25) = 2(1 - \Phi(0.25)).$$

Usando le tabelle della CDF standard Φ leggiamo che

$$\Phi(0.25) = 0.5987.$$

Pertanto,

$$P_X\left(\left|\frac{S_n}{n} - \mu\right| > 3\right) = 2(1 - 0.5987) = 0.8026.$$

Esercizio 5. Un campione di 100 lampadine della marca A ha mostrato una durata media di 1190 ore ed una deviazione standard di 90 ore; un campione di 75 lampadine della marca B ha mostrato invece una durata media di 1230 ore ed una deviazione standard di 120 ore.

(a) Sottoporre a test l'ipotesi che le lampadine della marca B sono superiori a quelle della marca A considerando un livello di significativitá $\alpha = 0.05$

Soluzione:
Si puó procedere con un test di ipotesi unilaterale sulla differenza fra le due medie μ_A,μ_B

$$\begin{cases} H_0 & : \mu_A \ge \mu_B \\ \\ H_1 & : \mu_A < \mu_B. \end{cases} \tag{1}$$

Le varianze teoriche sono ignote ma, considerate la taglia ampia di entrambi i campioni ($n_A, n_B \ge 30$), si utilizza un test per grandi campioni dove σ_A^2 e σ_B^2 possono essere sostituite con le varianze campionarie $s_A^2 \approx \sigma_A^2$ e $s_B^2 \approx \sigma_B^2$, commettendo un errore di approssimazione accettabile.
Si ha che:

$$n_A = 100 \quad \bar{x}_A = 1190 \quad s_A = 90$$

$$n_B = 75 \quad \bar{x}_B = 1230 \quad s_B = 120$$

Per il livello di significatività $\alpha = 0.05$ la regione di rifiuto é data dal valore $Z < -1.645$.
Calcolando la statistica

$$Z = \frac{\bar{x}_A - \bar{x}_B}{\sqrt{\frac{s_A^2}{n_A} + \frac{s_B^2}{n_B}}} = \frac{1190 - 1230}{\sqrt{\frac{90^2}{100} + \frac{120^2}{75}}} = -2.42$$

si apprezza che il valore $Z = -2.424$ appartiene alla regione di rifiuto, perciò si puó rigettare l'ipotesi nulla ($\mu_A \geq \mu_B$), concludendo che la marca B é superiore alla marca A avendo statisticamente durata media maggiore ($\mu_A < \mu_B$).

Esercizio 6. La deviazione standard delle temperature annuali di una cittá in un periodo di 100 anni é stata di 8^0C. Misurando le temperatura medie degli ultimi 10 anni si é rilevato che la deviazione standard delle temperature annuali é di 6^0C.

(a) Sottoporre a test l'ipotesi che la temperatura della cittá sia diventata meno variabile che in passato, usando i livelli di significativitá $\alpha = 0.05$ e $\alpha = 0.01$.

Soluzione: E' ragionevole supporre che la temperatura X sia distribuita con legge normale. I dati sui 100 anni costituiscono i dati della popolazione, le misure rilevate negli ultimi 10 rappresentano il nostro campione.
Usiamo un test di ipotesi unilaterale

$$\begin{cases} H_0 & : \sigma \geq 8 = \sigma_0 \\ H_1 & : \sigma < 8. \end{cases} \tag{2}$$

E' un test effettuato sulla varianza, dunque si deve utilizzare la statistica test χ^2 con $\nu = n - 1 = 9$ gdl che assume il valore

$$\chi^2 = \frac{(n-1)S^2}{\sigma_0^2} = \frac{9 \cdot 6^2}{8^2} \simeq 5.06.$$

Il test é a una coda e la regione di rifiuto é $\chi^2 < \chi^2_{1-\alpha}$.
Dalle tavole statistiche, con $\nu = 9$ per $\alpha = 0.05$, $\chi^2_{0.95} = 3.325$; per $\alpha = 0.01$, $\chi^2_{0.99} = 2.088$
In entrambi i casi la statistica χ^2 non appartiene alla regione di rifiuto, quindi si puó concludere che la variabilitá della temperatura non é cambiata e il risultato é sostanzialmente dovuto al caso.