Compito scritto dell'esame di Statistica e analisi dei dati 16 giugno 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 e appunti.
- 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. Di due eventi A e B si sa che $P(A \mid B) = \alpha$, $P(B \mid A) = \beta$ e $P(A \cup B) = \gamma$, dove α, β, γ sono valori noti.

(a) Calcolare $P(A)$ e $P(B)$.

Soluzione: Dalle condizioni specificate sappiamo che $A \cap B \neq \emptyset$. Per stabilire una relazione fra le due incognite, usiamo il principio di inclusione/esclusione e scriviamo la somma di $P(A)$ e $P(B)$ come

$$P(A) + P(B) = P(A \cup B) + P(A \cap B) = \gamma + P(A \cap B) \tag{1}$$

Ma per la simmetria della regola del prodotto valgono le seguenti

$$P(A \cap B) = P(A \mid B)P(B) = \alpha P(B) \tag{2}$$

e

$$P(A \cap B) = P(B \mid A)P(A) = \beta P(A). \tag{3}$$

Sostituendo in (1), si ottiene il sistema lineare nelle incognite $P(A)$ e $P(B)$:

$$\begin{cases} P(A) + P(B) & = \gamma + \alpha P(B) \\ \\ P(A) + P(B) & = \gamma + \beta P(A), \end{cases} \tag{4}$$

facilmente risolvibile come

$$\begin{cases} P(A) & = \frac{\alpha\gamma}{\alpha+\beta-\alpha\beta} \\ P(B) & = \frac{\beta\gamma}{\alpha+\beta-\alpha\beta}. \end{cases} \tag{5}$$

Esercizio 2. La durata di funzionamento di una batteria é normalmente distribuita con media $\mu = 500$ ore e deviazione standard $\sigma = 100$ ore. Per incrementare la durata media l'azienda sviluppa un nuovo processo di produzione e decide di adottarlo qualora la durata media osservata $\bar{X}$ in un campione di batterie del nuovo tipo risulti maggiore di 550 ore

(a) Se le batterie del nuovo tipo sottoposte a test sono 9, qual é la probabilitá che l'azienda adotti il nuovo processo anche se esso non é migliore di quello corrente?

Soluzione: La durata X é distribuita con legge normale $X \sim \mathcal{N}(500, 100^2)$.
Siccome, fino a prova contraria, il nuovo processo non puó essere ritenuto migliore di quello corrente, si puó procedere con un test di ipotesi unilaterale

$$\begin{cases} H_0 & : \mu = 500 = \mu_0 \\ H_1 & : \mu > 500. \end{cases} \tag{6}$$

La probabilitá che l'azienda adotti il nuovo processo anche se esso non é migliore di quello attuale corrisponde formalmente alla probabilitá

$$P(\texttt{rifiuto } H_0 \mid H_0 \texttt{ vera})$$

ma dalla teoria dei test di ipotesi sappiamo che tale probabilitá definisce la probabilitá di errore di tipo I (detta anche “rischio”),

$$P(\texttt{rifiuto } H_0 \mid H_0 \texttt{ vera}) = \alpha,$$

ovvero il livello di significativitá del test.
Quindi il problema posto si riduce banalmente a determinare il livello α nelle condizioni date.
La regione critica é

$$Z = \frac{\bar{X} - 500}{100/\sqrt{n}} > z_{1-\alpha} \quad \Longrightarrow \quad \bar{X} > 500 + z_{1-\alpha}\frac{100}{\sqrt{n}} \tag{7}$$

L'azienda é tuttavia disposta a rifiutare H_0 (cioé ad assumersi un rischio) a patto che $\bar{X} > 550$, quindi

$$500 + z_{1-\alpha}\frac{100}{\sqrt{n}} = 550 \tag{8}$$

Nel test effettuato $n = 9$. Utilizzando (8) si puó dunque ricavare l'α cercato.

$$1 - \alpha = \Phi(z_{1-\alpha}) = \Phi(\frac{550 - 500}{100/3}) = \Phi(1.5) \approx 0.9332, \tag{9}$$

da cui

$$\alpha \approx 6.68\%$$

(b) Quante batterie dovrebbero essere sottoposte a test affinché la probabilitá di cui al punto precedente sia al piú 5%?
Soluzione: In questo caso si impone che

$$\alpha \leq 0.05,$$

e la condizione (8) va riscritta come

$$500 + z_{1-0.05}\frac{100}{\sqrt{n}} \leq 550 \tag{10}$$

ovvero

$$n \geq 4z_{0.95}^2 \approx 10.5 \tag{11}$$

Occorre pertanto misurare la durata di almeno 11 batterie per avere una probabilitá d'errore di tipo I massima del 5%.

Esercizio 3. Abbiamo investito in un fondo di investimento il cui portafoglio é garantito avere dal gestore un rendimento medio dell'8.14% e una deviazione standard del 5.12%. L' andamento medio di rendimento ha una distribuzione non nota e la volatilitá del mercato potrebbe escludere che tale distribuzione segua una legge normale. A fine anno, dai quattro rapporti trimestrali, stimiamo una media di rendimento del 6,2%

(a) Determinare se il rendimento teorico dichiarato dal gestore é plausibile considerando un intervallo di confidenza al 95%. Discutere inoltre la rilevanza concreta dell'intervallo determinato.
Soluzione:
I dati a disposizione ci dicono che $\mu = 0.0814, \sigma = 0.0512$. La media da noi stimata é la variabile aleatoria $\bar{X} = 0.062$.
Per un IC al 95%, $\alpha = 0.05$.
La distribuzione di X é ignota, non possiamo assumerla Gaussiana, e su 4 misure non possiamo applicare il Teorema Limite Centrale.
Abbiamo dunque solo a disposizione la disuguaglianza di Chebychev.
Un IC di confidenza $I = [I_{inf}, I_{sup}]$ di livello $1 - \alpha$ per la media μ é definito teoricamente come

$$P(\mu \in I) = P(I_{inf} < \mu < I_{sup}) \geqslant 1 - \alpha. \tag{12}$$

La disuguaglianza di Chebychev ci assicura che nel nostro caso

$$P(|\bar{X} - \mu| < k) \geqslant 1 - \frac{\sigma^2}{k^2}. \tag{13}$$

Dunque (14) e (13) insieme ci dicono che la disuguaglianza di Chebychev definisce un IC di livello

$$1 - \alpha = 1 - \frac{\sigma^2}{k^2},$$

dunque

$$k^2 = \frac{\sigma^2}{\alpha}$$

da cui

$$k = \frac{\sigma}{\sqrt{\alpha}} \simeq 0.23$$

Sostituendo nella (13) ed esplicitando la disuguaglianza, l'IC cercato é definito da

$$P(\bar{X} - \frac{\sigma}{\sqrt{\alpha}} < \mu < \bar{X} + \frac{\sigma}{\sqrt{\alpha}}) \geqslant 1 - \alpha \tag{14}$$

La stima di μ in tali condizioni risulta essere

$$\mu \in \left[\bar{X} - k, \bar{X} + k\right].$$

ovvero, sostituendo i valori noti

$$\mu \in [-0.17, 0.29]$$

Statisticamente non possiamo affermare che il gestore abbia detto il falso, poiché il rendimento medio teorico dichiarato ricade nell'intervallo. Tuttavia é importante notare che la disuguaglianza di Chebychev nelle condizioni date determina un IC molto ampio: infatti considerando il rapporto $\frac{k}{\sigma} \simeq 4.5$, l'intervallo puó riscriversi come

$$\mu \in \left[\bar{X} - 4.5\sigma, \bar{X} + 4.5\sigma\right]$$

ben oltre il classico 2σ solitamente preso in considerazione. Con una variabilitá (σ) cosí elevata, poche misurazioni e un intervallo di tale ampiezza, non é plausibile trarre conclusioni affidabili sull'onestá del gestore.

Esercizio 4.

(a) Si enunci la Legge dei grandi numeri (in forma debole)

Soluzione: Vedere appunti di teoria

(b) Si dimostri l'enunciato

Soluzione: Vedere appunti di teoria

Esercizio 5. 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

$$\begin{cases} H_0 & : \sigma \geq 8 = \sigma_0 \\ H_1 & : \sigma < 8. \end{cases} \tag{15}$$

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.

Esercizio 6. Un apparecchio elettronico é composto da due elementi in parallelo, l'uno indipendente dall'altro e ciascuno con un tempo di vita di media 8 giorni.

(a) Con quale probabilitá l'apparecchio durerá un tempo non superiore a 12 giorni, supposto che esso funzioni se una almeno delle due componenti funziona e sotto l'ipotesi che i guasti seguano legge poissoniana?

Soluzione: Il problema concerne un calcolo di tempo d'attesa ($T \leq 12$) La teoria dei processi poissoniani ci dice che se i guasti seguono la legge di Poisson, la durata T (in giorni), cioé il tempo di attesa al primo guasto, segue una legge esponenziale negativa $Exp(T;\lambda)$. Poiché una variabile aleatoria esponenziale ha media uguale all'inverso del parametro λ, nel nostro caso si ha

$$\lambda = \frac{1}{8};$$

Di conseguenza ciascuna componente ha un tempo di vita $T_i, i = 1{,}2$ avente densitá.

$$f_{T_i} = \begin{cases} 0, & t < 0 \\ \frac{1}{8}e^{-\frac{1}{8}t} & t \geq 0. \end{cases} \tag{16}$$

Indicato quindi con T il tempo di vita dell'apparecchio, si ha $T = \max\{T_1, T_2\}$. Sapendo poi che T_1 e T_2 sono indipendenti, si ha semplicemente che

$$P(T \leq t) = P(T_1 \leq t, T_2 \leq t) = P(T_1 \leq t)P(T_2 \leq t) = (1 - e^{-\frac{1}{8}t})^2$$

dunque

$$P(T \leq 12) = (1 - e^{-\frac{1}{8}\cdot 12})^2 \simeq (1 - 0.223)^2 \simeq 0.6035$$