Compito scritto dell'esame di Statistica e analisi dei dati 18 gennaio 2022	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. In un recente sondaggio sul ministro della salute il 58% dei rispondenti ha dichiarato di aver fiducia nel suo operato. Le donne costituiscono il 58% del campione, e tra queste il 46% ha riposto positivamente (fiducia).

(a) Si seleziona a caso una persona tra quelle intervistate. Qual é la probabilitá che la persona selezionata sia maschio?

Soluzione: Gli unici eventi che ci interessano sono

- $M =$ "rispondente maschio"
- $D =$ "rispondente donna"

Sappiamo che $P(D) = 0.58$. Dunque:

$$P(M) = 1 - P(D) = 1 - 0.58 = 0.42$$

Esercizio 2.

Due ditte forniscono il medesimo prodotto. Se esso proviene dalla ditta A, la probabilitá che si guasti prima dell'istante t vale $1 - e^{-t}$; se invece proviene dalla ditta B questa probabilitá vale $1 - e^{-2t}$. Il prodotto puó essere acquistato con uguale probabilitá da A o da B, e non é nota la ditta fornitrice. Tuttavia, é stato osservato che il prodotto si guasta in un intervallo di tempo $1 \leq t \leq 2$.

(a) Determinare la probabilitá che esso sia stato acquistato dalla ditta A.

Soluzione: (Teorema di Bayes)

I dati del problema ci dicono che, a priori, le probabilitá che il prodotto provenga da A o da B valgono

$$P(A) = P(B) = 0.5.$$

Possiamo ragionare in due modi:

1. Approccio puramente probabilistico

Definiamo l'evento

$G =$ "guasto in $1 \leq t \leq 2$"

Nel caso di A $1 - e^{-t} = P($ "guasto prima di$t \mid A)$. L'evento G é dato da

$$G = \{ \text{"guasto prima di} t = 2\} - \{ \text{"guasto prima di} t = 1\}$$

La probabilitá di guasto del prodotto A nell'intervallo di tempo $1 \leq t \leq 2$ vale allora

$$P(G \mid A) = 1 - e^{-2} - (1 - e^{-1}) = e^{-1} - e^{-2}.$$

2. Approccio basato su modello di probabilitá e introduzione di una variabile aleatoria X

Dalla natura del problema e dalla forma $P(X < t \mid A) = 1 - e^{-t}$, si riconosce che tale probabilitá definisce la classica **funzione di ripartizione** $F_X(t) = 1 - e^{-t}$ di una v.a. $X \sim Exp(x; \lambda)$ con **densità** esponenziale negativa e^{-t} con $\lambda = 1$.

Pertanto,

$$P(1 \leq X \leq 2 \mid A) = F_X(2) - F_X(1) = 1 - e^{-2} - (1 - e^{-1}) = e^{-1} - e^{-2}$$

.

Oppure, in modo equivalente, integrando direttamente la **densità**:

$$P(1 \leq X \leq 2 \mid A) = \int_1^2 e^{-t} dt = e^{-1} - e^{-2}$$

,

Stesso ragionamento vale per determinare $P(1 \leq X \leq 2 \mid B)$ (in tal caso $\lambda = 2$).
Dunque, per il prodotto B nello stesso intervallo é

$$P(G \mid B) = 1 - e^{-2 \cdot 2} - (1 - e^{-2 \cdot 1}) = e^{-2} - e^{-4}.$$

Applicando la regola di Bayes:

$$P(A \mid G) = \frac{P(G \mid A)P(A)}{P(G \mid A)P(A) + P(G \mid B)P(B)} = \frac{e^{-1} - e^{-2}}{e^{-1} - e^{-2} + e^{-2} - e^{-4}} = \frac{e^2(e-1)}{e^3 - 1} \approx 0.6652$$

Esercizio 3. Sia X la variabile che conta il numero di volte in cui si ottiene una doppia testa in 100 lanci di una coppia di monete regolari.

(a) Calcolare la probabilitá di ottenere una doppia testa almeno 20 volte.

Soluzione: Una doppia Testa (= "successo") nel lancio di due monete regolari si ottiene con probabilitá 1/4.
La risposta richiederebbe l'uso della distribuzione binomiale $Bin(n,p)$, con parametri $n = 100$ e $p = 1/4 = 0.25$.
Poiché n è grande posso approssimare con una normale di parametri $\mu = np = 25, \sigma^2 = npq = 18.75$:

$$Bin(x; 0.25, 100 \approx \mathcal{N}(x; 25, 18.75)$$

Pertanto, usando normale standard e tabelle, con correzione di continuità $19.5 < 20$:

$$P(X \geq 20) = 1 - P(X < 20) \approx 1 - \Phi(\frac{19.5 - 25}{\sqrt{18.75}}) \approx 0.90$$

(con la Binomiale si otterrebbe 0.9000469)

Esercizio 4. Sia $X \sim \mathcal{N}(x; 1, 2^2)$ e si definisca la v.a. $Y = X^2 - 3X$.

(a) Calcolare il valore atteso $E[Y]$ e la probabilità $P(Y > 0)$.

Soluzione: Sappiamo che per $\mathcal{N}(x; 1, 2^2)$, $E[X] = 1$, $var(X) = 2^2$
Per il calcolo del valore atteso si ha:

$$E[Y] = E\left[X^2 - 3X\right] = E\left[X^2\right] - 3E[X]$$

Poichè $var(X) = E\left[X^2\right] - E[X]^2$,

$$E\left[X^2\right] - 3E[X] = var(X) + E[X]^2 - 3E[X] = 4 + 1 - 3 = 2$$

.

Per calcolare $P(Y > 0)$ si ha che

$$Y > 0 \implies X^2 - 3X = X(X-3) > 0,$$

e la disuguaglianza è valida per i casi (disgiunti) $X < 0$ e $X > 3$ Pertanto:

$$P(Y > 0) = P(X < 0) + P(X > 3)$$

La distribuzione di X é normale, dunque standardizzando e usando le tabelle:

$$P(Y > 0) = \Phi(\frac{0-1}{2}) + 1 - \Phi(\frac{3-1}{2}) \approx 0.47$$

Esercizio 5. Una popolazione assume un farmaco per mantenere il valor medio del battito cardiaco (battito/min, bpm) tra 80 e 90 bpm. Il bpm medio rilevato su un campione di 60 soggetti é pari a 91 bpm. Da dati storici, si sa che la variazione del bpm della popolazione segue una legge normale con deviazione standard $\sigma = 5$.

(a) Si verifichi se, relativamente al bpm medio, il farmaco possa ritenersi efficace con livello di confidenza al 90% e 95%

Soluzione:

Sappiamo che $n = 60$ e $\bar{x} = 91$.

L' IC al 90% si ottiene per

$$\begin{aligned}(100)(1-\alpha) &= 90\\ 1-\alpha &= 0.9\\ \alpha &= 0.1\\ \alpha/2 &= 0.05\end{aligned}$$

Il valore critico $z_{0.05}$ é il valore che lascia un'area a destra pari a 0.05 (a sinistra pari a $1 - 0.05 = 0.95$).

Usando la tabella della Normale standard lo z-value che lascia a destra 0.05 é $z_{0.05} = 1.64$.

L'IC cercato é:

$$91 - 1.64\frac{5}{\sqrt{60}} < \mu < 91 + 1.64\frac{5}{\sqrt{60}}$$
$$89.94 < \mu < 92.06$$

Analogamente, per l'IC al 95%, $\alpha = 0.05$ e $z_{\alpha/2} = 1.96$.

Pertanto,

$$91 - 1.96\frac{5}{\sqrt{60}} < \mu < 91 + 1.96\frac{5}{\sqrt{60}}$$
$$89.74 < \mu < 92.26$$

Sulla base di tali risultati i farmaci con IC al 90% e al 95% sono al limite dell'efficacia. Considerando che in entrambi i casi si ha un intervallo di incertezza molto ampio, complessivamente, l'ipotesi di efficacia puó essere rifiutata.

Esercizio 6. Un produttore di robot per le pulizie domestiche garantisce che il suo prodotto dura in media 3 anni con una deviazione standard di 1 anno. Si campionano casualmente 5 robottini e ne risulta che abbiano una durata di $1.9, 2.4, 3.0, 3.5, 4.2$ anni.

(a) Assumendo che la vita della popolazione di robot sia distribuita con legge normale, si tragga una conclusione con un livello di confidenza al 95% sull'asserzione del produttore che $\sigma = 1$.

Soluzione:

Utilizziamo l'intervallo di confidenza

$$\frac{(n-1)s^2}{\chi^2_{\alpha/2}} < \sigma^2 < \frac{(n-1)s^2}{\chi^2_{1-\alpha/2}}$$

con $\sigma^2 = \sigma = 1$

Determiniamo la varianza del campione s^2 con

$$s^2 = \frac{n\sum_{i=1}^n x_i^2 - (\sum_{i=1}^n x_i)^2}{n(n-1)}$$

Per $n = 5$:

$$\sum_{i=1}^5 x_i = 1.9 + 2.4 + 3.0 + 3.5 + 4.2 = 15$$

e

$$\sum_{i=1}^5 x_i^2 = 1.9^2 + 2.4^2 + 3.0^2 + 3.5^2 + 4.2^2 = 48.26$$

Pertanto

$$\begin{aligned}s^2 &= \frac{n\sum_{i=1}^5 x_i^2 - (\sum_{i=1}^5 x_i)^2}{n(n-1)}\\ &= \frac{5\cdot 48.26 - (15)^2}{5(5-1)}\\ &= \frac{16.3}{20}\\ &= 0.815\end{aligned}$$

Determiniamo $\alpha/2$. Da $95\% = 100(1-\alpha)\%$, si ha:

$$\begin{aligned}(100)(1-\alpha) &= 95\\ 1-\alpha &= 0.95\\ \alpha &= 0.05\\ \alpha/2 &= 0.025\end{aligned}$$

I g.d.l. sono $\nu = n - 1 = 4$. Usando la tabella della distribuzione Chi-quadro, il valore critico per 0.025 con 4 g.d.l. é $\chi^2_{0.025} = 11.143$, e $\chi^2_{1-0.025} = \chi^2_{0.975} = 0.484$.
L'intervallo di interesse é

$$\begin{aligned}\frac{(n-1)s^2}{\chi^2_{\alpha/2}} &< \sigma^2 < \frac{(n-1)s^2}{\chi^2_{1-\alpha/2}}\\ \frac{(5-1)\cdot 0.815}{\chi^2_{0.025}} &< \sigma^2 < \frac{(5-1)\cdot 0.815}{\chi^2_{0.975}}\\ \frac{4\cdot 0.815}{11.143} &< \sigma^2 < \frac{4\cdot 0.815}{0.484}\\ 0.29 &< \sigma^2 < 6.74\end{aligned}$$

Poiché 1 cade nell'intervallo di confidenza al 95% si puó concludere che quanto dichiarato dal produttore é plausibile statisticamente.