Compito scritto dell'esame di Statistica e analisi dei dati 15 febbraio 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. Il giocatore A lancia un dado non truccato per 4 volte, e vince se esce almeno una volta il 6. Il giocatore B lo lancia 8 volte, e vince se il 6 esce almeno due volte.

(a) Chi ha maggiore probabilitá di vincere e perché?

Soluzione: (Distribuzione binomiale) In ogni lancio la probabilitá che esca il 6 vale $p = 1/6$ (equiprobabilitá di 6 eventi). La probabilitá di avere $X = 0$ successi in $n = 4$ prove indipendenti vale, si determina in base alla distribuzione binomiale:

$$Bin\left(X = 0 \mid n = 4, p = \frac{1}{6}\right) = \binom{4}{0}\left(\frac{1}{6}\right)^0\left(\frac{5}{6}\right)^4 = \left(\frac{5}{6}\right)^4 \approx 0.48226$$

Dunque, la probabilitá di vittoria per A é

$$P(A) = 1 - 0.48226 \approx 0.51774$$

Per il giocatore B, la probabilitá di avere non piú di $X = 1$ successi in $n = 8$ prove (perdendo cosí la scommessa) é

$$P_X(0 \le X \le 1) = Bin\left(X = 0 \mid n = 8, p = \frac{1}{6}\right) + Bin\left(X = 1 \mid n = 8, p = \frac{1}{6}\right)$$

ovvero

$$P_X(0 \le X \le 1) = \binom{8}{0}\left(\frac{1}{6}\right)^0\left(\frac{5}{6}\right)^8 + \binom{8}{1}\left(\frac{1}{6}\right)^1\left(\frac{5}{6}\right)^7 \approx 0.6046$$

Pertanto, la probabilitá di vittoria per B é

$$P(B) = 1 - P_X(0 \le X \le 1) \approx 0.3936$$

Esercizio 2. Una variabile aleatoria X ha densitá

$$f_X(x) = \begin{cases} x - \frac{1}{4}x^3, & per \quad 0 \le x \le 2 \\ 0 & altrove \end{cases}$$

(a) Determinare valore atteso, varianza e mediana di X.

Soluzione:

$$E[X] = \int_0^2 x\left(x - \frac{1}{4}x^3\right)dx = \left[\frac{x^3}{3} - \frac{x^5}{20}\right]_0^2 = \frac{16}{15}$$

La varianza é $\sigma_X^2 = E[X^2] - (E[X])^2 = E[X^2] - \left(\frac{16}{15}\right)^2$, dove

$$E[X^2] = \int_0^2 x^2\left(x - \frac{1}{4}x^3\right)dx = \left[\frac{x^4}{4} - \frac{x^6}{24}\right]_0^2 = \frac{4}{3}$$

Pertanto,

$$\sigma_X^2 = \frac{4}{3} - \left(\frac{16}{15}\right)^2 = 16\left(\frac{1}{12} - \frac{16}{225}\right) \approx 0.195$$

Per calcolare la mediana x_{med}, si impone $F_X(x_{med}) = \frac{1}{2}$, dunque:

$$F_X(x_{med}) = \int_0^{x_{med}} \left(x - \frac{1}{4}x^3\right) dx = \left[\frac{x^2}{2} - \frac{x^4}{16}\right]_0^{x_{med}} = \frac{1}{2}\left(x_{med}^2 - \frac{x_{med}^4}{16}\right) = \frac{1}{2}$$

Si deve pertanto risolvere l'equazione

$$x_{med}^4 - 8x_{med}^2 + 8 = 0$$

cercando la radice che appartiene all'intervallo $0 \leq x \leq 2$.
Ponendo $t = x_{med}^2$

$$t = \begin{cases} 4 + 2\sqrt{2} \\ 4 - 2\sqrt{2} \end{cases}$$

da cui

$$\begin{cases} x_{med}^{(1,2)} = \pm\sqrt{4 + 2\sqrt{2}} \approx \pm 2.613 \\ x_{med}^{(3,4)} = \pm\sqrt{4 - 2\sqrt{2}} \approx \pm 1.0924 \end{cases}$$

Scartando le soluzioni $x_{med}^{(1,2)}$ giacché fuori dell'intervallo $0 \leq x \leq 2$, l'unica soluzione ammissibile é

$$x_{med}^{(3)} = \sqrt{4 - 2\sqrt{2}} \approx 1.0924$$

Esercizio 3. Una linea di produzione fabbrica resistori da 1000 ohm (Ω). La fabbrica adotta una tolleranza del 10% (ovvero sono scartati i resistori con resistenza $R < 900\ \Omega$ e $R > 1100\ \Omega$).

(a) Sotto l'ipotesi che la resistenza R di un resistore sia una variabile aleatoria distribuita con legge normale di media 1000 Ω e varianza 2500, si calcoli la probabilitá che un resistore preso a caso venga scartato

Soluzione: Sia S l'evento in cui un resistore viene scartato. Allora:

$$S = \{R < 900\} \cup \{R > 1100\}$$

Poiché $\{R < 900\} \cap \{R > 1100\} = \emptyset$, si ha che

$$P(S) = P(\{R < 900\}) + P(\{R > 1100\}) = F_X(900) + (1 - F_X(1100)) = \Phi\left(\frac{900 - \mu}{\sigma}\right) + \left(1 - \Phi\left(\frac{1100 - \mu}{\sigma}\right)\right)$$

Poiché $\mu = 1000$, e $\sigma^2 = 2500 \implies \sigma = 50$, le variabili normalizzate sono

$$\frac{900 - 1000}{50} = -2$$

$$\frac{1100 - 1000}{50} = 2$$

.
Sfruttando la proprietá della cumulativa standardizzata $\Phi(-z) = 1 - \Phi(z)$, la precedente diventa:

$$P(S) = 2\,(1 - \Phi(2))$$

.
Accedendo alla tabella della Normale standard per $z = 2$, $\Phi(2) = 0.9772$ da cui $P(S) = 0.0455$

Esercizio 4. Il numero di automobili che attraversano un particolare incrocio stradale in un'ora é mediamente pari a 30.

(a) Determinare la probabilitá che in un intervallo di tempo di cinque minuti nessuna automobile attraversi l'incrocio in questione.

Soluzione: Sia X la V.A. discreta che indica il numero di veicoli che passa lungo l'incrocio in 5 min. X segue una legge di Poisson, $X \sim Pois(\mu)$. Il numero medio di automobili che passa in 5 min. é $\frac{30}{12} = 2.5$. Dunque $X \sim Pois(2.5)$. Pertanto:

$$P(X = 0) = \frac{\mu^0}{0!}e^{-\mu} = e^{-2.5} = 0.082$$

(b) Qual é la probabilitá che in dieci minuti almeno due automobili passino lungo quel tratto di strada?

Soluzione: Se X denota il numero di veicoli che passa lungo l'incrocio in 5 min, indichiamo con $Y = 2X$ la V.A. discreta che indica il numero di veicoli che passa lungo l'incrocio in 10 min. Allora:

$$E[Y] = E[2X] = 2E[X] = 2\mu = 5$$

Dunque, $Y \sim Pois(5)$. La risposta alla domanda consiste nel calcolare

$$P(Y \geq 2) = 1 - P(Y < 2) = 1 - (P(Y = 0) + P(Y = 1)) = 1 - \left(e^{-5} + \frac{5^1}{1!}e^{-5}\right) = 1 - 6e^{-5} = 0.9595$$

(c) Assumendo che la variabile aleatoria T, rappresentante il tempo (sempre in minuti) trascorso tra il passaggio di un'auto e di quella successiva, si distribuisca con legge esponenziale negativa, qual é la probabilitá che tra il passaggio di un'auto e la successiva trascorra piú di un minuto?

Soluzione: Poiché $T \sim Exp(\lambda)$, i dati del problema ci dicono che il rate di passaggio é

$$\lambda = \frac{30}{60}\frac{\text{(auto)}}{\text{(min.)}}.$$

Dunque la V.A. T ha funzione di densitá $f_T(t) = 0.5e^{-0.5t}$, $t \geq 0$. La probabilitá che il tempo di attesa tra il passaggio di un un'auto e la successiva sia > 1 min. é $P(T > 1)$ che si puó ottenere dal calcolo diretto $P(T > 1) = \int_1^{\infty} f_T(t)dt$, oppure usando la funzione di sopravvivenza $S_T(t) = e^{-\lambda t}$:

$$P(T > 1) = 1 - F_T(1) = S_T(1) = e^{-0.5 \cdot 1} = 0.6065$$

Esercizio 5. Una macchina viene utilizzata su una linea di produzione inglese per riempire scatole. La deviazione standard del peso é di 0.3 once. Il problema da risolvere è mantenere bassa la variabilitá del peso di prodotto nella scatola. Dopo aver implementato un miglioramento della linea, si estrae un campione aleatorio di 20 scatole, misurando una varianza campionaria di 0.045 once

(a) Assumendo che i pesi si distribuiscano con legge normale, si tragga una conclusione con un livello di confidenza al 95% sull'efficacia del miglioramento produttivo.

Soluzione: Abbiamo che

$$n = 20, s^2 = 0.045, \sigma = 0.3$$

Utilizziamo l'intervallo di confidenza

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

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 $\nu = n - 1 = 19$. Usando la tabella della distribuzione Chi-quadro, il valore critico per 0.025 con 19 g.d.l. é $\chi^2_{0.025} = 32.825$, e $\chi^2_{1-0.025} = \chi^2_{0.975} = 8.907$.
L'intervallo di interesse é

$$
\begin{gathered}
\frac{(n-1)s^2}{\chi^2_{\alpha/2}} < \sigma^2 < \frac{(n-1)s^2}{\chi^2_{1-\alpha/2}} \\
\frac{(19)\cdot(0.045)}{\chi^2_{0.025}} < \sigma^2 < \frac{(19)\cdot(0.045)}{\chi^2_{0.975}} \\
\frac{(19)\cdot(0.045)}{32.825} < \sigma^2 < \frac{(19)\cdot(0.045)}{8.907} \\
0.012 < \sigma^2 < 0.045
\end{gathered}
$$

Pertanto l'intervallo di confidenza al 95% per σ é approssimabile come

$$0.110 < \sigma < 0.212$$

Poiché 0.3 cade al di fuori di tale intervallo, si puó concludere che da un punto di vista statistico, il nuovo processo abbia significativamente ridotto e dunque migliorata la variabilità

Esercizio 6. Bob ha calcolato un intervallo di confidenza per la media di una popolazione normale, a partire da un campione di 8 osservazioni, dopo averne stimato la varianza con la varianza campionaria $s^2 = 4.2$. Il risultato é $(16.08, 18.82)$.

(a) Con quale confidenza é stato ottenuto l'intervallo?

Soluzione:
La taglia del campione $n = 8$ é piccola e la varianza é ignota essendo stata stimata mediante la varianza campionaria $s^2 = 4.2$. Inoltre, la popolazione da cui proviene il campione é normale. Pertanto l'IC per la media deve essere stato stimato utilizzando la distribuzione t di Student con $\nu = n - 1 = 7$ gradi di libertá:

$$\mu = \bar{x} \pm \frac{s}{\sqrt{n}} t_{\frac{\alpha}{2}}(\nu = 7) = \bar{x} \pm \frac{\sqrt{4.2}}{\sqrt{8}} t_{\frac{\alpha}{2}}(\nu = 7)$$

Gli estremi dell'IC indicato corrispondono a

$$\bar{x} - \frac{\sqrt{4.2}}{\sqrt{8}} t_{\frac{\alpha}{2}}(\nu = 7) = 16.08$$

$$\bar{x} + \frac{\sqrt{4.2}}{\sqrt{8}} t_{\frac{\alpha}{2}}(\nu = 7) = 18.82$$

Risolvendo il sistema a due incognite si ottiene

$$\bar{x} = 17.45$$

$$t_{\frac{\alpha}{2}}(\nu = 7) = 1.895$$

Dalla tabella della t di Student con $\nu = 7$ leggiamo che $t_{\frac{\alpha}{2}} = 1.895$ per $\frac{\alpha}{2} = 0.05$, cioé $\alpha = 0.10$
Dunque l'intervallo é stato ottenuto con una confidenza pari al $(100)(1 - \alpha) = (100)(1 - 0.10) = 90\%$