Compito scritto dell'esame di Statistica e analisi dei dati 19 luglio 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.
Un robot autonomo controlla il livello di corrosione interna delle tubature dei sistemi di raffreddamento di una centrale nucleare. La corrosione non puó essere osservata direttamente, ma il robot effettua un test che puó fornisce indizi sulla specifica corrosione. Il test non é infallibile: nel 70% dei casi individua una corrosione quando é presente, ma puó anche sbagliare indicando una corrosione inesistente nel 20% dei casi. Risolvere il problema del robot a cui un operatore remoto, sotto l'ipotesi che il 10% delle tubature siano corrose, chiede di:

(a) Comunicargli la probabilitá che una parte delle tubature abbia delle corrosioni interne dopo che il test ha dato esito positivo (presenza di corrosioni). Determinare tale probabilitá.

Soluzione: Definiamo gli eventi $C =$ "la tubatura é corrosa" e $\overline{C} =$ "il test ha identificato la tubatura come corrosa". I dati del problema ci dicono che, a priori,

$$P(C) = 0.1$$

e dunque $P(\sim C) = 1 - 0.1 = 0.9$, dove $\sim C =$ "la tubatura non é corrosa". Inoltre

$$P(\overline{C} \mid C) = 0.7$$
$$P(\overline{C} \mid\sim C) = 0.2$$

Applicando la regola di Bayes:

$$P(C \mid \overline{C}) = \frac{P(\overline{C} \mid C)P(C)}{P(\overline{C})}$$

dove $P(\overline{C}) = P(\overline{C} \mid C)P(C) + P(\overline{C} \mid\sim C)P(\sim C) = 0.25$. Quindi

$$P(C \mid \overline{C}) = \frac{0.7 \times 0.1}{0.25} = 0.28$$

(b) Comunicargli la probabilitá che una parte delle tubature abbia delle corrosioni dopo che il test ha dato esito negativo (nessuna corrosione). Determinare tale probabilitá.

Soluzione: Si ripeta il procedimento precedente per il caso $P(C \mid\sim \overline{C})$

Esercizio 2. La variabile aleatoria $X(w)$ ha densitá di probabilitá

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

(a) Calcolare la probabilitá che $X(w)$ assuma valori in un intorno di raggio $\delta = 0.5$ del suo valor medio.

Soluzione: Il valore atteso di X vale

Soluzione:

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

(infatti nell'intervallo $0 \le x \le 2$, $f_X(x)$ é una parabola con vertice nel punto (1,0)).
Si deve calcolare la probabilitá

$$P(|X - E[X]| < \delta) = P(|X-1| < 0.5)$$

integrando la densitá nell' intervallo $(1-0.5) \le x \le (1+0.5)$:

$$P(|X-1| < 0.5) = \frac{3}{2}\int_{\frac{1}{2}}^{\frac{3}{2}} (x-1)^2 dx = 3\int_1^{\frac{3}{2}} (x-1)^2 dx = \frac{1}{8} = 0.125$$

Esercizio 3. Siano X, Y due variabili aleatorie non necessariamente indipendenti che hanno rispettivamente valore atteso $E[X], E[Y]$ e varianza $var(X), var(Y)$. Si definisca la funzione $g(X,Y) = X + Y$

(a) Si derivi per esteso la varianza della funzione $var(g(X,Y))$.

Soluzione: Vedere appunti di teoria

Esercizio 4. Si estrae un campione $\{X_i\}$ da 50 osservazioni di una caratteristica fisica avente un modello statistico normale $X_i \sim \mathcal{N}(2,1)$, ma é noto soltanto il numero Z di osservazioni dalle quali risulta $X_i \le 0, \quad i = 1, \cdots, 50$.

(a) Determinare esattamente la legge della variabile aleatoria $Z(\omega)$

Soluzione:
Le 50 osservazioni costituiscono prove ripetute e indipendenti di un esperimento casuale.
Definiamo come "successo" S l'evento

$$S = \{X_i \le 0\}$$

L'evento si verifica con probabilitá

$$p = P(X \le 0) = P(\frac{X-\mu}{\sigma} \le \frac{0-\mu}{\sigma}) = P(Z \le \frac{0-2}{1}) = \Phi(-2) = 1 - \Phi(2) = 1 - 0.9772 = 0.0228$$

dove il valore $\Phi(2)$ é stato ricavato dalla tavola della normale standard.
La variabile aleatoria $Z(\omega)$ che rappresenta il numero di successi S che si verificano in 50 prove ha pertanto distribuzione Binomiale con parametri $n = 50, p = 0.0228$:

$$Z \sim Bin(z; 50, 0.0228)$$

Esercizio 5. Una popolazione assume un farmaco per mantenere il valor medio del battito cardiaco (battito/min, bpm) tra 85 e 95 bpm. Il bpm medio rilevato su un campione di 49 soggetti é pari a 90 bpm. Da dati storici, si sa che la variazione del bpm della popolazione segue una legge normale con deviazione standard $\sigma = 10$.

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

Soluzione:
Sappiamo che $n = 49$ e $\bar{x} = 90$.
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:

$$90 - 1.64\frac{10}{\sqrt{49}} < \mu < 90 + 1.64\frac{10}{\sqrt{49}}$$
$$87.66 < \mu < 92.34$$

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

$$90 - 1.96\frac{10}{\sqrt{49}} < \mu < 90 + 1.96\frac{10}{\sqrt{49}}$$
$$87.2 < \mu < 92.8$$

Sulla base di tali risultati il farmaco si puó ritenere efficace.

Esercizio 6. Le aziende 1 e 2 forniscono un servizio di argentatura. Lo spessore dell'argentatura misurato in mm ha distribuzione normale indipendentemente dal processo di argentatura e i due processi non differiscono significativamente per lo spessore medio di placcatura ottenuto. L'azienda 1 ha un peró un costo di lavorazione lievemente superiore ma sostiene di fornire una placcatura meno variabile, dunque piú affidabile, in termini di spessore finale. Un potenziale compratore seleziona da ciascuna azienda un campione di taglia 12. I due campioni indipendenti forniscono rispettivamente deviazioni standard $s_1 = 0.035$mm e $s_2 = 0.062$mm

(a) Stabilire, mediante un test di ipotesi al livello di significativitá 0.05, se al compratore convenga approvvigionarsi dall'azienda 1 sostenendo una spesa d'acquisto superiore .

Soluzione: Si tratta di stabilire se c'é una minore variabilitá, non dovuta al caso, nell'argentatura eseguita dall' azienda 1 rispetto a quella dell'azienda 2.
La popolazione é normale, utilizziamo un test di ipotesi sulle varianze (test F) che é in questo caso unilaterale

$$\begin{cases} H_0 & : \sigma_1^2 = \sigma_2^2 \\ H_1 & : \sigma_1^2 < \sigma_2^2. \end{cases} \tag{2}$$

Il livello di significativitá é $\alpha = 0.05$.
La statistica di test é

$$F = \frac{S_1^2}{S_2^2} = \frac{0.035^2}{0.062^2} = 0.318$$

Posso rigettare H_0, se

$$F < F_{1-\alpha}(\nu_1{,}\nu_2)$$

.
Dalle tavole della F, con gdl $\nu_1 = \nu_2 = 12 - 1 = 11$:

$$F_{1-\alpha}(\nu_1{,}\nu_2) = F_{0.95}(11{,}11) = \frac{1}{F_{0.05}(11{,}11)} = \frac{1}{2.82} \simeq 0.354$$

Poiché

$$F = 0.318 < 0.354 = F_{0.95}(11{,}11)$$

il compratore puó rigettare l'ipotesi nulla (la placcatura 1 é effettivamente meno variabile) e considerare di acquistare dall'azienda 1.