Soluzioni Esame di Probabilità e Statistica [3231]

Soluzioni Esame di Calcolo delle Probabilità e Statistica [2959]

Corso di Studi di Ingegneria Gestionale (D.M.270/04) (L)

Dipartimento di Meccanica, Matematica e Management
Politecnico di Bari

Cognome: ______________	A.A.: 2021/2022
Nome: ______________	Docente: Gianluca Orlando
Matricola: ______________	Appello: luglio 2022
Corso di studi: ______________	Data: 18/07/2022

Tempo massimo: 2 ore.

Esercizio 1. I risultati dei test di adesione a trazione su 22 provini di lega U-700 mostrano i seguenti carichi di rottura (in megapascal):

23.1	10.1	15.4	18.5	11.4	14.1	19.5	8.8	14.9	7.5	7.9
12.7	15.4	15.4	11.9	11.4	17.6	16.7	15.8	13.6	11.9	11.4

1. Determinare i quartili dei dati.
2. Determinare eventuali dati anomali o sospetti.
3. Tracciare un box-plot.

Soluzione. 1. Ordiniamo i dati:

7.5	7.9	8.8	10.1	11.4	11.4	11.4	11.9	11.9	12.7	13.6
14.1	14.9	15.4	15.4	15.4	15.8	16.7	17.6	18.5	19.5	23.1

Denotiamo con $x_1, \dots, x_{22}$ i dati ordinati. L'ampiezza del campione è $n = 22$.
Calcolo di Q_1. Per trovare il primo quartile calcoliamo $\frac{n+1}{4} = \frac{23}{4} = 5 + \frac{3}{4} = 5 + 0.75$. Quindi

$$Q_1 = (1 - 0.75)x_5 + 0.75x_6 = 11.4\,.$$

Calcolo di Q_2. Per trovare il secondo quartile calcoliamo $(n+1)\frac{2}{4} = \frac{23}{2} = 11 + \frac{1}{2} = 11 + 0.5$. Quindi

$$Q_2 = 0.5x_{11} + 0.5x_{12} = 13.85\,.$$

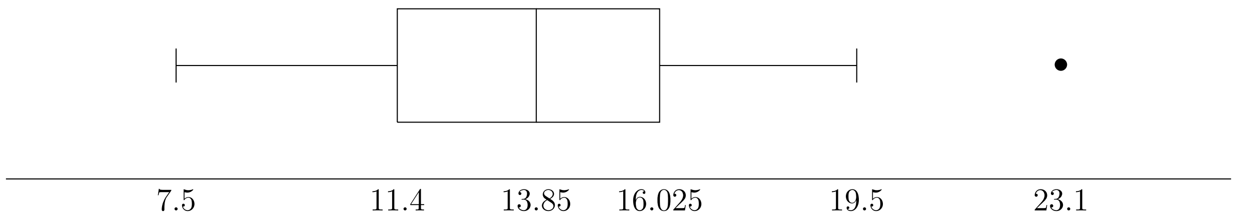
Calcolo di Q_3. Per trovare il terzo quartile calcoliamo $(n+1)\frac{3}{4} = 23\frac{3}{4} = 17 + \frac{1}{4} = 17 + 0.25$. Quindi

$$Q_3 = (1 - 0.25)x_{17} + 0.25x_{18} = 16.025\,.$$

2. Per determinare i dati anomali e sospetti calcoliamo il range interquartile:

$$IQR = Q_3 - Q_1 = 4.625\,.$$

I dati anomali sono più grandi di $Q_3 + 3IQR = 29.9$ o più piccoli di $Q_1 - 3IQR = -2.475$. Quindi non ci sono dati anomali.
I dati sospetti cadono tra $Q_3 + 1.5IQR = 22.9625$ e $Q_3 + 3IQR = 29.9$ oppure tra $Q_1 - 3IQR = -2.475$ e $Q_1 - 1.5IQR = 4.4625$. Quindi 23.1 è un dato anomalo.
3. Segue il box-plot:



Esercizio 2. Sei un ingegnere gestionale e fai parte di un gruppo di esperti selezionati per formare una commissione giudicatrice. Oltre a te ci sono: 2 ingegneri gestionali, 6 ingegneri elettrici, 4 ingegneri civili. Tra gli ingegneri elettrici ci sono tua sorella e tuo fratello.

1. Si deve formare una commissione composta da 2 ingegneri gestionali, 4 ingegneri elettrici, 2 ingegneri civili scegliendo in modo casuale (e uniformemente rispetto alle possibili commissioni realizzabili) tra i possibili esperti. Qual è la probabilità che tu non venga selezionato?

2. È stata selezionata la commissione come nel punto 1. Ti hanno detto che il tuo cognome (che è anche quello di tua sorella e tuo fratello) compare esattamente due volte, ma non sai di preciso chi di voi tre è stato selezionato. Qual è la probabilità che tu sia stato selezionato?

Soluzione. 1. Lo spazio degli eventi elementari è dato da

$$\begin{aligned}\Omega = \{(\{\omega_1,\omega_2\},\{\omega_3,\omega_4,\omega_5,\omega_6\},\{\omega_7,\omega_8\})\ :\ &\\ &\omega_1,\omega_2 \in \{1,2,3\},\ \omega_3,\omega_4,\omega_5,\omega_6 \in \{4,5,6,7,8,9\},\ \omega_7,\omega_8 \in \{10,11,12,13\},\\ &\omega_i \neq \omega_j \text{ per } i \neq j\}\,,\end{aligned}$$

dove abbiamo indicato con $\{1,2,3\}$ gestionali, $\{4,5,6,7,8,9\}$ elettrici, $\{10,11,12,13\}$ civili. Nella formazione di una commissione si scelgono 2 da 3 elementi (gestionali), 4 da 6 elementi (elettrici), 2 da 4 elementi (civili), quindi le possibilità sono

$$\#\Omega = \binom{3}{2}\binom{6}{4}\binom{4}{2}\,.$$

L'evento "vengo selezionato/a nella commissione" è, indicando con 1 l'elemento corrispondente a me,

$$\begin{aligned}A = \{(\{1,\omega_2\},\{\omega_3,\omega_4,\omega_5,\omega_6\},\{\omega_7,\omega_8\})\ :\ &\\ &\omega_2 \in \{2,3\},\ \omega_3,\omega_4,\omega_5,\omega_6 \in \{4,5,6,7,8,9\},\ \omega_7,\omega_8 \in \{10,11,12,13\},\ \omega_i \neq \omega_j \text{ per } i \neq j\}\end{aligned}$$

che è composto da

$$\#A = \binom{2}{1}\binom{6}{4}\binom{4}{2}$$

elementi, poiché per i gestionali si sceglie solamente tra 2.
La probabilità di non essere selezionato è

$$\mathbb{P}(\Omega \setminus A) = 1 - \mathbb{P}(A) = 1 - \frac{\#A}{\#\Omega} = 1 - \frac{\binom{2}{1}}{\binom{3}{2}} = 1 - \frac{2}{3} = \frac{1}{3}\,.$$

2. Chiamiamo 4 mia sorella e 5 mio fratello. Consideriamo l'evento $B =$ "esattamente due dei tre della famiglia sono selezionati". Questo evento si spezza nell'unione dei tre eventi $B_1 =$ "io e mia sorella veniamo selezionati (mio fratello no)", $B_2 =$ "io e mio fratello veniamo selezionati (mia sorella no)", $B_3 =$ "mia sorella e mio fratello vengono selezionati (io no)". Utilizzando la notazione degli eventi elementari otteniamo

$$\begin{aligned} B &= B_1 \cup B_2 \cup B_3 \\ &= \{(\{1,\omega_2\},\{4,\omega_4,\omega_5,\omega_6\},\{\omega_7,\omega_8\}) \ : \\ &\qquad \omega_2 \in \{2,3\},\ \omega_4,\omega_5,\omega_6 \in \{6,7,8,9\},\ \omega_7,\omega_8 \in \{10,11,12,13\},\ \omega_i \neq \omega_j \text{ per } i \neq j\} \cup \\ &\quad \cup \{(\{1,\omega_2\},\{5,\omega_4,\omega_5,\omega_6\},\{\omega_7,\omega_8\}) \ : \\ &\qquad \omega_2 \in \{2,3\},\ \omega_4,\omega_5,\omega_6 \in \{6,7,8,9\},\ \omega_7,\omega_8 \in \{10,11,12,13\},\ \omega_i \neq \omega_j \text{ per } i \neq j\} \cup \\ &\quad \cup \{(\{\omega_1,\omega_2\},\{4,5,\omega_5,\omega_6\},\{\omega_7,\omega_8\}) \ : \\ &\qquad \omega_1,\omega_2 \in \{2,3\},\ \omega_5,\omega_6 \in \{6,7,8,9\},\ \omega_7,\omega_8 \in \{10,11,12,13\},\ \omega_i \neq \omega_j \text{ per } i \neq j\}. \end{aligned}$$

L'evento B_1 è composto da un numero di elementi dato dalle scelte di ω_2 per le scelte di $\omega_4, \omega_5, \omega_6$ per le scelte di ω_7, ω_8, quindi è composto da

$$\#B_1 = \binom{2}{1}\binom{4}{3}\binom{4}{2}$$

elementi. Analogamente contiamo il numero di elementi per gli altri eventi:

$$\#B = \#B_1 + \#B_2 + \#B_3 = \binom{2}{1}\binom{4}{3}\binom{4}{2} + \binom{2}{1}\binom{4}{3}\binom{4}{2} + \binom{2}{2}\binom{4}{2}\binom{4}{2}$$

elementi.
Dobbiamo calcolare

$$\begin{aligned} \mathbb{P}(A|B) &= \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} = \frac{\mathbb{P}(B_1) + \mathbb{P}(B_2)}{\mathbb{P}(B)} = \frac{\frac{\#B_1}{\#\Omega} + \frac{\#B_2}{\#\Omega}}{\frac{\#B}{\#\Omega}} = \frac{\#B_1 + \#B_2}{\#B} \\ &= \frac{\binom{2}{1}\binom{4}{3}\binom{4}{2} + \binom{2}{1}\binom{4}{3}\binom{4}{2}}{\binom{2}{1}\binom{4}{3}\binom{4}{2} + \binom{2}{1}\binom{4}{3}\binom{4}{2} + \binom{2}{2}\binom{4}{2}\binom{4}{2}} = \frac{2\binom{2}{1}\binom{4}{3}}{2\binom{2}{1}\binom{4}{3} + \binom{2}{2}\binom{4}{2}} = \frac{2 \cdot 2 \cdot 4}{2 \cdot 2 \cdot 4 + 1 \cdot 6} = \frac{16}{22} = \frac{8}{11} \\ &\simeq 72.72\%. \end{aligned}$$

Osserviamo che è maggiore di $\frac{2}{3}$ (la probabilità di essere scelto senza sapere nulla).

Esercizio 3. Devi aprire un conto alla poste. Prendi il biglietto e vedi che ci sono tre sportelli. A servire lo sportello 1 c'è una persona molto motivata ed efficiente, a servire lo sportello 2 una persona normale, a servire lo sportello 3 una persona evidentemente frustrata e con poca voglia di lavorare. Il tempo che impiegherai per concludere l'apertura del conto è distribuito con una legge esponenziale, ma il tempo medio che impiegherai dipende da quale sportello ti capita: allo sportello 1 la media sarebbe 10 minuti; allo sportello 2, 15 minuti; allo sportello 3, 30 minuti. Nell'attesa ti accorgi che il 65% delle persone viene servito allo sportello 1, il 25% dallo sportello 2, il 10% dallo sportello 3.

1. Qual è la probabilità che impiegherai più di 30 minuti a concludere l'operazione?
2. Finita l'operazione, chiami una tua amica e le dici che hai impiegato più di 30 minuti ad aprire un conto! Lei è stata a quella filiale e conosce le tre persone addette agli sportelli. Con che probabilità è pronta a scommettere che sei stato servito dallo sportello 3?

Soluzione. 1. Consideriamo due variabili aleatorie: X con range $\{1, 2, 3\}$ che indica lo sportello che ci capita e Y con range $(0, +\infty)$ che indica il tempo in minuti necessario per l'operazione allo sportello. La variabile X è distribuita nel seguente modo:

$$\mathbb{P}(\{X = 1\}) = 65\% \quad \mathbb{P}(\{X = 2\}) = 25\% \quad \mathbb{P}(\{X = 3\}) = 10\%.$$

La distribuzione di Y dipende dalla realizzazione di X. Ricordando che la media di una variabile aleatoria con legge esponenziale è il reciproco del parametro, otteniamo che

$$\mathbb{P}(\{Y \geq t\}|\{X = 1\}) = \int_t^{+\infty} \frac{1}{10} e^{-\frac{1}{10}y}\, dy = \Big[- e^{-\frac{1}{10}x} \Big]_t^{+\infty} = e^{-\frac{1}{10}t},$$

$$\mathbb{P}(\{Y \geq t\}|\{X = 2\}) = \int_t^{+\infty} \frac{1}{15} e^{-\frac{1}{15}y}\, dy = e^{-\frac{1}{15}t},$$

$$\mathbb{P}(\{Y \geq t\}|\{X = 3\}) = \int_t^{+\infty} \frac{1}{30} e^{-\frac{1}{30}y}\, dy = e^{-\frac{1}{30}t}.$$

Possiamo utilizzare il teorema della probabilità totale per calcolare

$$\begin{aligned}\mathbb{P}(\{Y \geq 30\}) &= \mathbb{P}(\{Y \geq 30\}|\{X = 1\})\mathbb{P}(\{X = 1\}) + \mathbb{P}(\{Y \geq 30\}|\{X = 2\})\mathbb{P}(\{X = 2\}) \\ &\quad + \mathbb{P}(\{Y \geq 30\}|\{X = 3\})\mathbb{P}(\{X = 3\}) \\ &= e^{-\frac{30}{10}} 65\% + e^{-\frac{30}{15}} 25\% + e^{-\frac{30}{30}} 10\% = 10.29\%.\end{aligned}$$

2. Calcoliamo con il Teorema di Bayes

$$\mathbb{P}(\{X = 3\}|\{Y \geq 30\}) = \frac{\mathbb{P}(\{Y \geq 30\}|\{X = 3\})\mathbb{P}(\{X = 3\})}{\mathbb{P}(\{Y \geq 30\})} = \frac{e^{-\frac{30}{30}} 10\%}{10.29\%} \sim 35.75\%$$

Esercizio 4. Una riempitrice automatica viene utilizzata per riempire dosatori da 100 ml con gel igienizzante. Viene misurata la differenza tra il volume di riferimento 100 ml e il volume di riempimento effettivo in un campione casuale di dosatori:

$$-0.91 \quad 0.51 \quad 0.85 \quad 0.10 \quad -0.56 \quad 1.10 \quad 0.38 \quad 0.26 \quad 0.30 \quad -0.67 \quad 0.03 \quad -0.31 \quad 0.59$$

(ad esempio, il dato -0.91 indica che il dosatore è stato riempito con 100.91 ml di gel, il dato 0.51 indica che il dosatore è stato riempito con 99.49 ml di gel). Se la deviazione standard del volume di riempimento è superiore a 0.16 ml, la macchina riempitrice deve essere tarata nuovamente. I dati sono significativi all'1% per concludere che si deve tarare la macchina riempitrice? E al 5%? Si assuma che la popolazione sia distribuita con legge normale.
(N.B.: Ricavare le formule!)

Soluzione. (Nota: la deviazione standard 0.16 ml è troppo piccola per un errore di trascrizione. Ad ogni modo i numeri non influiscono sullo svolgimento dell'esercizio.) La differenza tra 100 ml e il volume di riempimento effettivo è una variabile aleatoria con legge normale. Abbiamo un campione casuale $X_1, \ldots, X_n \sim \mathcal{N}(\mu, \sigma)$ dove $n = 13$ e μ e σ non sono note. Il problema può essere posto come un test d'ipotesi (esattamente come fatto per l'Esercizio 10 risolto nella Lezione 23):

$$H_0 : \sigma^2 \leq \sigma_0^2, \quad H_1 : \sigma^2 > \sigma_0^2$$

dove $\sigma_0 = 0.16$. Il test è impostato così perché la domanda è se i dati sono abbastanza significativi da rifiutare l'ipotesi nulla.
Ricordiamo che il livello di significatività α di un test è la probabilità di commettere un errore del I tipo, ovvero di rifiutare l'ipotesi H_0 quando questa è vera. Allora assumiamo che H_0 sia vera: $\sigma^2 \leq \sigma_0^2$. La regione critica per il test è della forma

$$R_C = \{(x_1, \ldots, x_n) \in R(X_1, \ldots, X_n) : s_n^2 > c\sigma_0^2\}$$

dove $s_n^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \overline{x}_n)^2$ è la varianza campionaria calcolata sui dati e $\overline{x}_n = \frac{1}{n} \sum_{i=1}^n x_i$ è la media campionaria calcolata sui dati. (N.B.: questo è coerente con la domanda del problema: la macchina va tarata se la varianza campionaria calcolata sui dati è esageratamente più grande di quella che vorremmo! Controllare sempre se la regione critica ha senso, altrimenti correggere il test d'ipotesi)
Osserviamo che se $S_n^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \overline{X}_n)^2$ è la varianza campionaria, allora non è detto che $\frac{(n-1)S_n^2}{\sigma_0^2}$ sia distribuita come una chi-quadro, poiché σ_0^2 non è la vera varianza della popolazione. Tuttavia, osserviamo che l'ipotesi nulla $\sigma^2 \leq \sigma_0^2$ implica che

$$S_n^2 > c\sigma_0^2 \implies S_n^2 > c\sigma^2$$

e quindi

$$\mathbb{P}(\{(X_1, \ldots, X_n) \in R_C\}) = \mathbb{P}(\{S_n^2 > c\sigma_0^2\}) \leq \mathbb{P}(\{S_n^2 > c\sigma^2\}) = \mathbb{P}\Big(\Big\{\frac{(n-1)S_n^2}{\sigma^2} > (n-1)c\Big\}\Big).$$

La statistica $\Xi_{n-1} = \frac{(n-1)S_n^2}{\sigma^2}$ è distribuita come una chi-quadro con $n-1$ gradi di libertà, poiché la popolazione è normale e σ^2 è la varianza della popolazione. Pertanto scegliendo $(n-1)c = \chi^2_{n-1,\alpha}$, dove $\chi^2_{n-1,\alpha}$ è il quantile α di una chi-quadro con $n-1$ gradi di libertà, si ha che

$$\mathbb{P}(\{S_n^2 > c\sigma_0^2\}) \leq \mathbb{P}(\{\Xi_{n-1} > \chi^2_{n-1,\alpha}\}) = \alpha\,,$$

cioè la probabilità di commettere un errore del I tipo è più piccola di α.
Utilizziamo le tavole per calcolare il quantile

$$\chi^2_{n-1,\alpha} = \chi^2_{12,0.01} = 26.217\,.$$

L'ipotesi H_0 viene rifiutata a favore di H_1 se i dati ricadono nella regione critica, ovvero

$$s_n^2 > c\sigma_0^2 = \frac{\chi^2_{n-1,\alpha}}{(n-1)}\sigma_0^2 = \frac{26.217}{12} 0.0256 \simeq 0.0559.$$

Calcoliamo la realizzazione della varianza campionaria sui dati:

$$\overline{x}_n = \frac{1}{n}\sum_{i=1}^n x_i = \frac{1}{13}(-0.91 + \cdots + 0.59) \simeq 0.13\,.$$

$$s_n^2 = \frac{1}{n-1}\sum_{i=1}^n (x_i - \overline{x}_n)^2 = \frac{1}{12}(0.91^2 + \cdots + 0.59^2 - 13 \cdot 0.13^2) \simeq 0.36\,.$$

(N.B.: 0.36 è ben più grande di 0.16^2, quindi ha senso un test unilaterale con $H_1 : \sigma^2 > \sigma_0^2$.) Segue che i dati ricadono nella regione di rifiuto e pertanto l'ipotesi nulla è rifiutata con significatività 1%. A maggior ragione viene rifiutata con significatività 5% poiché se aumenta la probabilità di ricadere nella regione critica, la regione critica diventa più grande. In effetti,

$$\frac{\chi^2_{12,0.05}}{12}\sigma_0^2 \simeq \frac{21.026}{12} 0.0256 \simeq 0.044\,.$$