Esame di Probabilità e Statistica [3231]

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: ______________________ Docente: Gianluca Orlando
Nome: ______________________ Appello: luglio 2023 - turno 1
Matricola: ______________________ Data: 20/07/2023

Tempo massimo: 2 ore.

Esercizio 1. (6 punti) Si analizzano i ritardi di un treno all'arrivo in una stazione in diverse giornate. I dati misurati in minuti sono i seguenti:

$$5 \quad 1 \quad 7 \quad 0 \quad 15 \quad 2 \quad 31 \quad 5$$

1. Determinare i quartili (esclusivi) dei dati.
2. Determinare eventuali dati anomali e sospetti.
3. Tracciare un box plot.

Soluzione. 1. Per prima cosa ordiniamo i dati:

$$0 \quad 1 \quad 2 \quad 5 \quad 5 \quad 7 \quad 15 \quad 31$$

Abbiamo $n = 8$ dati.
Calcoliamo il primo quartile: $\frac{n+1}{4} = \frac{9}{4} = 2 + 0.25$. Allora

$$Q_1 = (1 - 0.25)x_2 + 0.25x_3 = 0.75 \cdot 1 + 0.25 \cdot 2 = 1.25\,.$$

Calcoliamo il secondo quartile: $(n+1)\frac{2}{4} = 9\frac{2}{4} = 4 + 0.5$. Allora

$$Q_2 = (1 - 0.5)x_4 + 0.5x_5 = 0.5 \cdot 5 + 0.5 \cdot 5 = 5\,.$$

Calcoliamo il terzo quartile: $(n+1)\frac{3}{4} = 9\frac{3}{4} = 6 + 0.75$. Allora

$$Q_3 = (1 - 0.75)x_6 + 0.75x_7 = 0.25 \cdot 7 + 0.75 \cdot 15 = 13\,.$$

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

$$IQR = Q_3 - Q_1 = 13 - 1.25 = 11.75\,.$$

I dati anomali apparterrebbero agli intervalli

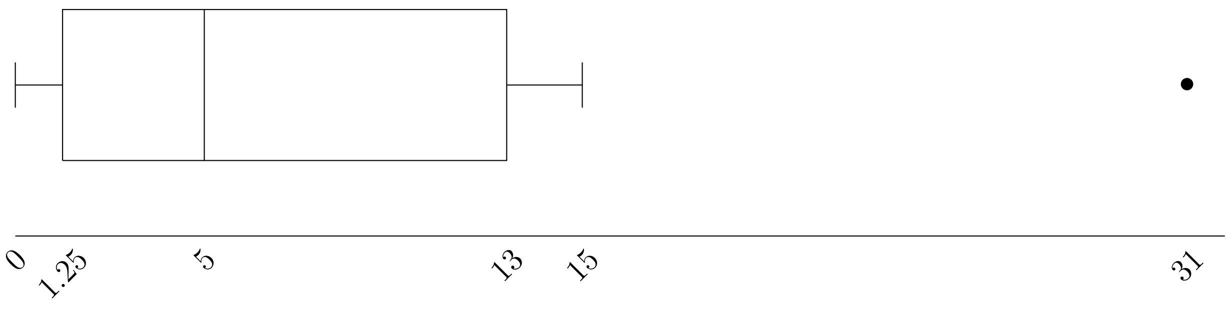
$$(-\infty, Q_1 - 3 \cdot IQR] \cup [Q_3 + 3 \cdot IQR, +\infty) = (-\infty, -34] \cup [48.25, +\infty)\,,$$

quindi non ci sono dati anomali. I dati sospetti appartengono agli intervalli

$$(Q_1 - 3 \cdot IQR, Q_1 - 1.5 \cdot IQR] \cup [Q_3 + 1.5 \cdot IQR, Q_3 + 3 \cdot IQR) = (-34, -16.375] \cup [30.625, 48.25)\,,$$

quindi 31 è un dato sospetto.

3. Segue il box-plot.



Esercizio 2. (8 punti) Durante l'esame di Probabilità e Statistica, alcune persone dimenticano di portare con sè le tavole delle distribuzioni. Si assuma che, durante un appello, il numero di persone che dimenticano di portare le tavole sia distribuito con una legge di Poisson e che, in media, 6 persone dimentichino le tavole.

1. Viene svolto un appello. Qual è la probabilità che almeno 4 persone dimentichino le tavole?

2. Vengono svolti 3 appelli. Si assumano i tre appelli indipendenti. Qual è la probabilità che nella totalità dei 3 appelli esattamente 20 persone dimentichino le tavole?

N.B.: Nella prossima richiesta, gli appelli non sono indipendenti.

Vengono svolti 2 appelli. Al primo appello il numero di persone che dimenticano di portare le tavole è distribuito con una legge di Poisson con media 6. Per il secondo appello:

- Se al primo appello il numero di persone che dimenticano di portare le tavole è maggiore o uguale a 4, allora il docente comunica che chi non porterà le tavole all'appello successivo avrà un voto più basso. L'effetto è che, nel secondo appello, il numero di persone che dimenticano le tavole è distribuito con una legge di Poisson con media 3.

- Se, invece al primo appello il numero di persone che dimenticano di portare le tavole è strettamente minore di 4, il docente non dice nulla e al secondo appello il numero di persone che dimenticano di portare le tavole è ancora distribuito con legge di Poisson con media 6.

Nelle ipotesi precedenti, rispondere alla seguente richiesta.

3. Sappiamo che al secondo appello esattamente 2 persone hanno dimenticato le tavole. Con che probabilità almeno 4 persone hanno dimenticato le tavole?

Soluzione. 1. Consideriamo la variabile aleatoria

$$X_1 = \text{“numero di persone che non portano le tavole in un appello”} \sim \mathrm{P}(\lambda)\,.$$

Ricordiamo che $\mathbb{P}(\{X = k\}) = e^{-\lambda}\frac{\lambda^k}{k!}$. Per la legge di Poisson si ha che $\mathbb{E}(X_1) = \lambda$, si ha che $\lambda = 6$.

Viene chiesto di calcolare

$$\begin{aligned}
\mathbb{P}(\{X_1 \geq 4\}) &= 1 - \mathbb{P}(\{X_1 < 4\}) = 1 - \mathbb{P}(\{X_1 \leq 3\}) \\
&= 1 - \mathbb{P}(\{X_1 = 0\}) - \mathbb{P}(\{X_1 = 1\}) - \mathbb{P}(\{X_1 = 2\}) - \mathbb{P}(\{X_1 = 3\}) \\
&= 1 - e^{-\lambda}\frac{\lambda^0}{0!} - e^{-\lambda}\frac{\lambda^1}{1!} - e^{-\lambda}\frac{\lambda^2}{2!} - e^{-\lambda}\frac{\lambda^3}{3!} \\
&= 1 - e^{-\lambda}\Big(\frac{\lambda^0}{0!} + \frac{\lambda^1}{1!} + \frac{\lambda^2}{2!} + \frac{\lambda^3}{3!}\Big) \\
&= 1 - e^{-6}\Big(1 + 6 + \frac{6^2}{2} + \frac{6^3}{6}\Big) \simeq 84.88\% \,.
\end{aligned}$$

2. Consideriamo tre variabili aleatorie indipendenti

$$X_1 = \text{“numero di persone che non portano le tavole al primo appello”} \sim \mathrm{P}(6)\,.$$

$$X_2 = \text{“numero di persone che non portano le tavole al secondo appello”} \sim \mathrm{P}(6)\,.$$

$$X_3 = \text{“numero di persone che non portano le tavole al terzo appello”} \sim \mathrm{P}(6)\,.$$

Il numero totale di persone che non portano le tavole nei tre appelli è

$$X = X_1 + X_2 + X_3\,.$$

Poiché la somma di variabili aleatorie di Poisson indipendenti è distribuita con legge di Poisson con parametro dato dalla somma dei parametri, otteniamo che

$$X \sim \mathrm{P}(3 \cdot 6) = \mathrm{P}(18)\,.$$

Calcoliamo

$$\mathbb{P}(\{X = 20\}) = e^{-18}\frac{18^{20}}{20!} \simeq 7.98\%\,.$$

3. Consideriamo due variabili aleatorie

$$Y_1 = \text{“numero di persone che non portano le tavole al primo appello”} \sim \mathrm{P}(6)\,.$$

$$Y_2 = \text{“numero di persone che non portano le tavole al secondo appello”}\,.$$

La distribuzione di Y_2 dipende da Y_1, quindi le due variabili non sono indipendenti. Consideriamo due variabili aleatorie ausiliari:

$$Z_2 = \text{“numero di persone che non portano le tavole al secondo appello se } \{Y_1 \geq 4\}\text{”} \sim \mathrm{P}(3)\,.$$

$$W_2 = \text{“numero di persone che non portano le tavole al secondo appello se } \{Y_1 < 4\}\text{”} \sim \mathrm{P}(6)\,.$$

Le ipotesi del problema ci permettono di stabilire che

$$\mathbb{P}(\{Y_2 = k\}|\{Y_1 \geq 4\}) = \mathbb{P}(\{Z_2 = k\})\,,$$

$$\mathbb{P}(\{Y_2 = k\}|\{Y_1 < 4\}) = \mathbb{P}(\{W_2 = k\})\,.$$

Ci viene chiesto di calcolare

$$\begin{aligned}
\mathbb{P}(\{Y_1 \geq 4\}|\{Y_2 = 2\}) &= \frac{\mathbb{P}(\{Y_1 \geq 4\} \cap \{Y_2 = 2\})}{\mathbb{P}(\{Y_2 = 2\})} \\
&= \frac{\mathbb{P}(\{Y_2 = 2\}|\{Y_1 \geq 4\})\mathbb{P}(\{Y_1 \geq 4\})}{\mathbb{P}(\{Y_2 = 2\}|\{Y_1 \geq 4\})\mathbb{P}(\{Y_1 \geq 4\}) + \mathbb{P}(\{Y_2 = 2\}|\{Y_1 < 4\})\mathbb{P}(\{Y_1 < 4\})} \\
&= \frac{\mathbb{P}(\{Z_2 = 2\})\mathbb{P}(\{Y_1 \geq 4\})}{\mathbb{P}(\{Z_2 = 2\})\mathbb{P}(\{Y_1 \geq 4\}) + \mathbb{P}(\{W_2 = 2\})\mathbb{P}(\{Y_1 < 4\})} \\
&= \frac{e^{-3}\frac{3^2}{2!}84.88\%}{e^{-3}\frac{3^2}{2!}84.88\% + e^{-6}\frac{6^2}{2!}(1 - 84.88\%)} \simeq 96.57\% \,.
\end{aligned}$$

Esercizio 3. (7 punti) Devi prendere un mezzo pubblico per arrivare ad un appuntamento. Puoi prendere il bus 13 oppure il bus 17.

- Il bus 13 arriva alla fermata tra le 19:20 e le 19:40 e l'orario di arrivo ha distribuzione uniforme.
- Il bus 17 arriva alla fermata in media alle 19:40, con una varianza di 12 min^2, e l'orario di arrivo ha distribuzione uniforme.

Rispondere ai seguenti quesiti.

1. Calcolare la probabilità che il bus 13 arrivi prima delle 19:25 oppure dopo le 19:35.
2. Calcolare la probabilità che il bus 17 arrivi prima delle 19:38.
3. Calcolare la probabilità che uno dei due bus arrivi prima delle 19:35 (è indifferente se il 13 o il 17). Si assumano i due orari di arrivo indipendenti.

Soluzione. Consideriamo le due variabili aleatorie

$$X_{13} = \text{“minuto di arrivo del bus 13”} \sim \mathrm{U}(20, 40)\,,$$

$$X_{17} = \text{“minuto di arrivo del bus 17”} \sim \mathrm{U}(a, b)\,.$$

Ricordiamo che

$$\mathbb{E}(X_{17}) = \frac{a+b}{2}\,, \quad \mathrm{Var}(X_{17}) = \frac{(b-a)^2}{12}\,.$$

Imponiamo che

$$\begin{cases} \dfrac{a+b}{2} = 40 \\ \dfrac{(b-a)^2}{12} = 12 \end{cases} \implies \begin{cases} a+b = 80 \\ b-a = 12 \end{cases} \implies \begin{cases} 2a = 68 \\ 2b = 92 \end{cases} \implies \begin{cases} a = 34 \\ b = 46\,. \end{cases}$$

Quindi $X_{17} \sim \mathrm{U}(34, 46)$.

1. Poiché gli eventi $\{X_{13} \leq 25\}$ e $\{X_{13} \geq 35\}$ sono disgiunti, otteniamo che

$$\begin{aligned}
&\mathbb{P}(\{X_{13} \leq 25\} \cup \{X_{13} \geq 35\}) = \mathbb{P}(\{X_{13} \leq 25\}) + \mathbb{P}(\{X_{13} \geq 35\}) \\
&= \int_{20}^{25} \frac{1}{40-20}\, \mathrm{d}x + \int_{35}^{40} \frac{1}{40-20}\, \mathrm{d}x \\
&= \frac{5}{20} + \frac{5}{20} = \frac{1}{2} = 50\%\,.
\end{aligned}$$

2. Calcoliamo
$$\mathbb{P}(\{X_{17} \leq 38\}) = \int_{34}^{38} \frac{1}{46-34}\,\mathrm{d}x = \frac{4}{12} = \frac{1}{3}\,.$$

3. Viene chiesto di calcolare
$$\begin{aligned}
&\mathbb{P}(\{X_{13} \leq 35\} \cup \{X_{17} \leq 35\}) \\
&= \mathbb{P}(\{X_{13} \leq 35\}) + \mathbb{P}(\{X_{17} \leq 35\}) - \mathbb{P}(\{X_{13} \leq 35\} \cap \{X_{17} \leq 35\}) \\
&= \mathbb{P}(\{X_{13} \leq 35\}) + \mathbb{P}(\{X_{17} \leq 35\}) - \mathbb{P}(\{X_{13} \leq 35\})\mathbb{P}(\{X_{17} \leq 35\}) \\
&= \int_{20}^{35} \frac{1}{20}\,\mathrm{d}x + \int_{34}^{35} \frac{1}{12}\,\mathrm{d}x - \Big(\int_{20}^{35} \frac{1}{20}\,\mathrm{d}x\Big)\Big(\int_{34}^{35} \frac{1}{12}\,\mathrm{d}x\Big) \\
&= \frac{15}{20} + \frac{1}{12} - \frac{15}{20} \cdot \frac{1}{12} \simeq 77.08\%\,.
\end{aligned}$$

In alternativa si può anche calcolare il complementare della probabilità che entrambi i bus arrivino dopo le 19:35, ovvero
$$\begin{aligned}
&\mathbb{P}(\{X_{13} \leq 35\} \cup \{X_{17} \leq 35\}) = 1 - \mathbb{P}(\{X_{13} > 35\} \cap \{X_{17} > 35\}) \\
&= 1 - \mathbb{P}(\{X_{13} > 35\})\mathbb{P}(\{X_{17} > 35\}) = 1 - \frac{5}{20} \cdot \frac{11}{12} = 77.08\%\,.
\end{aligned}$$

Esercizio 4. (7 punti) Le capacità (in Ah) di 10 batterie sono state registrate come segue:

140 136 150 144 148 152 138 141 143 151

Si assuma che la capacità di una batteria abbia distribuzione normale.

1. Calcolare sui dati un intervallo di confidenza al 95% bilaterale per la varianza della popolazione.
2. Per 8 volte si ripete la misurazione dei dati su un campione della stessa ampiezza di quello di sopra e su ogni campione si calcola sui dati un intervallo di confidenza al 95% bilaterale per la varianza della popolazione con il procedimento del punto 1. Si assumano le misurazioni sui campioni indipendenti. Qual è la probabilità che almeno 6 volte la varianza appartenga all'intervallo calcolato?

Soluzione. 1. Abbiamo una popolazione $X \sim \mathcal{N}(\mu, \sigma^2)$ e un campione casuale $X_1, \ldots, X_n$ estratto dalla popolazione con $n = 10$. Dalla definizione di IC si ha che (con $\beta = 95\%$)
$$\beta = \mathbb{P}(\{U_n \leq \sigma^2 \leq V_n\})\,.$$

Per stimare σ^2 sfrutteremo lo stimatore varianza campionaria $S_n^2 = \frac{1}{n-1}\sum_{i=1}^n (X_i - \overline{X}_n)^2$. Poiché $X_1, \ldots, X_n \sim \mathcal{N}(\mu, \sigma^2)$ sono indipendenti, $Q_{n-1} = \frac{(n-1)S_n^2}{\sigma^2} \sim \chi^2(n-1)$. Allora
$$\begin{aligned}
\beta &= \mathbb{P}(\{U_n \leq \sigma^2 \leq V_n\}) = \mathbb{P}\Big(\Big\{\frac{(n-1)S_n^2}{V_n} \leq \frac{(n-1)S_n^2}{\sigma^2} \leq \frac{(n-1)S_n^2}{U_n}\Big\}\Big) \\
&= \mathbb{P}\Big(\Big\{\frac{(n-1)S_n^2}{V_n} \leq Q_{n-1} \leq \frac{(n-1)S_n^2}{U_n}\Big\}\Big) \\
&= 1 - \mathbb{P}\Big(\Big\{Q_{n-1} < \frac{(n-1)S_n^2}{V_n}\Big\}\Big) - \mathbb{P}\Big(\Big\{Q_{n-1} > \frac{(n-1)S_n^2}{U_n}\Big\}\Big)\,.
\end{aligned}$$

Segue che

$$\mathbb{P}\Big(\Big\{Q_{n-1} < \frac{(n-1)S_n^2}{V_n}\Big\}\Big) + \mathbb{P}\Big(\Big\{Q_{n-1} > \frac{(n-1)S_n^2}{U_n}\Big\}\Big) = 1-\beta = \alpha\,.$$

Decidiamo di equipartire α:

$$\mathbb{P}\Big(\Big\{Q_{n-1} < \frac{(n-1)S_n^2}{V_n}\Big\}\Big) = \mathbb{P}\Big(\Big\{Q_{n-1} > \frac{(n-1)S_n^2}{U_n}\Big\}\Big) = \frac{\alpha}{2}\,,$$

da cui

$$\mathbb{P}\Big(\Big\{Q_{n-1} > \frac{(n-1)S_n^2}{U_n}\Big\}\Big) = \frac{\alpha}{2}\,, \quad \mathbb{P}\Big(\Big\{Q_{n-1} \geq \frac{(n-1)S_n^2}{V_n}\Big\}\Big) = 1-\frac{\alpha}{2}\,.$$

Definiamo $\chi^2_{n-1,\alpha/2}$ e $\chi^2_{n-1,1-\alpha/2}$ come i punti tali che

$$\mathbb{P}(\{Q_{n-1} \geq \chi^2_{n-1,\alpha/2}\}) = \frac{\alpha}{2}\,, \quad \mathbb{P}(\{Q_{n-1} \geq \chi^2_{n-1,1-\alpha/2}\}) = 1-\frac{\alpha}{2}\,.$$

Scegliendo

$$\frac{(n-1)S_n^2}{U_n} = \chi^2_{n-1,\alpha/2} \implies U_n = \frac{(n-1)S_n^2}{\chi^2_{n-1,\alpha/2}}\,,$$
$$\frac{(n-1)S_n^2}{V_n} = \chi^2_{n-1,1-\alpha/2} \implies V_n = \frac{(n-1)S_n^2}{\chi^2_{n-1,1-\alpha/2}}\,,$$

si ottiene la condizione desiderata. In conclusione

$$\Big[\frac{(n-1)S_n^2}{\chi^2_{n-1,\alpha/2}}, \frac{(n-1)S_n^2}{\chi^2_{n-1,1-\alpha/2}}\Big]$$

è un intervallo di confidenza bilaterale per σ^2 con livello di confidenza $\beta = 1-\alpha$.

Per calcolarlo sui dati, calcoliamo

$$\overline{x}_n = \frac{1}{n}\sum_{i=1}^n x_i = \frac{1}{10}(140+136+\cdots+143+151) = 144.3\,,$$

$$s_n^2 = \frac{1}{n-1}\Big(\sum_{i=1}^n x_i^2 - n\overline{x}_n^2\Big) = \frac{1}{9}(140^2+\cdots+151^2-10\cdot 144.3^2) \simeq 32.23\,.$$

Calcoliamo l'intervallo di confidenza sui dati

$$\Big[\frac{(n-1)s_n^2}{\chi^2_{n-1,\alpha/2}}, \frac{(n-1)s_n^2}{\chi^2_{n-1,1-\alpha/2}}\Big] \simeq \Big[\frac{9\cdot 32.23}{\chi^2_{9,0.025}}, \frac{9\cdot 32.23}{\chi^2_{9,0.975}}\Big] \simeq \Big[\frac{9\cdot 32.23}{19.023}, \frac{9\cdot 32.23}{2.700}\Big]$$
$$\simeq [15.25, 107.43]\,.$$

2. Consideriamo come successo l'evento "l'intervallo di confidenza contiene la varianza della popolazione". Questo evento ha probabilità 95%. In tutto si fanno 8 tentativi, quindi

$$Y = \text{“numero di successi in 8 tentativi”} \sim \mathrm{B}(8, 95\%)\,.$$

Viene chiesto di calcolare

$$\mathbb{P}(\{Y \geq 6\}) = \mathbb{P}(\{Y=6\}) + \mathbb{P}(\{Y=7\}) + \mathbb{P}(\{Y=8\})$$
$$= \binom{8}{6}(95\%)^6(5\%)^2 + \binom{8}{7}(95\%)^7(5\%)^1 + \binom{8}{8}(95\%)^8(5\%)^0$$
$$\simeq 99.42\%\,.$$