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: giugno 2023 - turno 1
Matricola: ______________________	Data: 26/06/2023

Tempo massimo: 2 ore.

Esercizio 1. (6 punti) Si misura il carico di rottura a trazione di alcuni provini di acciaio INOX. Le misurazioni in $MPa = N/mm^2$ sono le seguenti:

792 706 737 580 692 720 711 704 749 725 734 702

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:

580 692 702 704 706 711 720 725 734 737 749 792

Abbiamo $n = 12$ dati.

Calcoliamo il primo quartile: $\frac{n+1}{4} = \frac{13}{4} = 3 + 0.25$. Allora

$$Q_1 = (1 - 0.25)x_3 + 0.25x_4 = 0.75 \cdot 702 + 0.25 \cdot 704 = 702.5\,.$$

Calcoliamo il secondo quartile: $(n+1)\frac{2}{4} = 13\frac{2}{4} = 6 + 0.5$. Allora

$$Q_2 = (1 - 0.5)x_6 + 0.5x_7 = 0.5 \cdot 711 + 0.5 \cdot 720 = 715.5\,.$$

Calcoliamo il terzo quartile: $(n+1)\frac{3}{4} = 13\frac{3}{4} = 9 + 0.75$. Allora

$$Q_3 = (1 - 0.75)x_9 + 0.75x_{10} = 0.25 \cdot 734 + 0.75 \cdot 737 = 736.25\,.$$

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

$$IQR = Q_3 - Q_1 = 736.25 - 702.5 = 33.75\,.$$

I dati anomali apparterrebbero agli intervalli

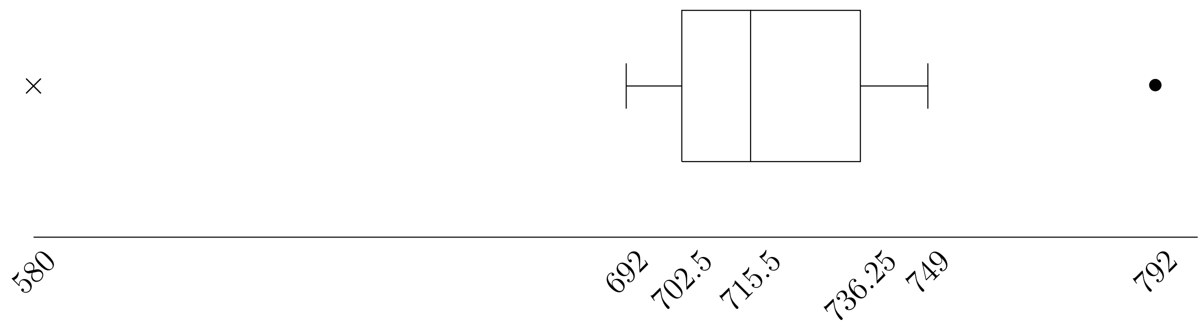
$$(-\infty, Q_1 - 3 \cdot IQR] \cup [Q_3 + 3 \cdot IQR, +\infty) = (-\infty, 601.25] \cup [837.5, +\infty)\,,$$

quindi 580 è un dato anomalo. I dati sospetti appartengono agli intervalli

$$(Q_1 - 3 \cdot IQR, Q_1 - 1.5 \cdot IQR] \cup [Q_3 + 1.5 \cdot IQR, Q_3 + 3 \cdot IQR) = (601.25, 651.875] \cup [786.875, 837.5)\,,$$

quindi 792 è un dato sospetto.

3. Segue il box-plot.



Esercizio 2. (7 punti) Il numero di errori di battitura in una pagina di un libro è distribuito con legge di Poisson. In media, c'è 1 errore di battitura per pagina. Si assumano gli errori in pagine diverse indipendenti.

1. Qual è la probabilità di osservare esattamente 3 errori di battitura in una pagina?
2. Qual è la probabilità di osservare (strettamente) più di 3 errori di battitura nelle prime 4 pagine?
3. Sappiamo che c'è almeno un errore di battitura nelle prime 2 pagine. Tenendo conto di questo, qual è la probabilità che nelle prime 4 pagine ci siano in tutto esattamente 3 errori di battitura?

Soluzione. Consideriamo la variabile aleatoria $X \sim \mathrm{P}(\lambda)$. Ci viene detto che $\mathbb{E}(X) = 1$. Ricordando che $\mathbb{E}(X) = \lambda$, concludiamo che $\lambda = 1$.

1. Calcoliamo

$$\mathbb{P}(\{X = 3\}) = e^{-\lambda}\frac{\lambda^3}{3!} = e^{-1}\frac{1}{6} = 6.13\%\,.$$

2. Consideriamo 4 variabili aleatorie indipendenti $X_1, X_2, X_3, X_4 \sim \mathrm{P}(\lambda)$. Ricordando che la somma di variabili aleatorie di Poisson indipendenti è distribuita con legge di Poisson con parametro dato dalla somma dei parametri, otteniamo che

$$Y = X_1 + X_2 + X_3 + X_4 \sim \mathrm{P}(4\lambda) = \mathrm{P}(4)\,.$$

Possiamo calcolare

$$\begin{aligned}\mathbb{P}(\{Y > 3\}) &= 1 - \mathbb{P}(\{Y \leq 3\}) = 1 - \mathbb{P}(\{Y = 0\}) - \mathbb{P}(\{Y = 1\}) - \mathbb{P}(\{Y = 2\}) - \mathbb{P}(\{Y = 3\}) \\ &= 1 - e^{-4}\Big(1 + 4 + \frac{4^2}{2} + \frac{4^3}{3!}\Big) = 56.65\%\,.\end{aligned}$$

3. Consideriamo

$$Z_1 = X_1 + X_2 \sim \mathrm{P}(2\lambda) = \mathrm{P}(2)$$

$$Z_2 = X_3 + X_4 \sim \mathrm{P}(2\lambda) = \mathrm{P}(2)$$

e osserviamo che $Y = Z_1 + Z_2$ e che Z_1 e Z_2 sono indipendenti. Dobbiamo calcolare

$$\begin{aligned}
\mathbb{P}(\{Y=3\}|\{Z_1 \geq 1\}) &= \frac{\mathbb{P}(\{Z_1+Z_2=3\}\cap\{Z_1\geq 1\})}{\mathbb{P}(\{Z_1\geq 1\})} \\
&= \frac{\mathbb{P}(\{Z_2=2\}\cap\{Z_1=1\})+\mathbb{P}(\{Z_2=1\}\cap\{Z_1=2\})+\mathbb{P}(\{Z_2=0\}\cap\{Z_1=3\})}{\mathbb{P}(\{Z_1\geq 1\})} \\
&= \frac{\mathbb{P}(\{Z_2=2\})\mathbb{P}(\{Z_1=1\})+\mathbb{P}(\{Z_2=1\})\mathbb{P}(\{Z_1=2\})+\mathbb{P}(\{Z_2=0\})\mathbb{P}(\{Z_1=3\})}{1-\mathbb{P}(\{Z_1=0\})} \\
&= \frac{e^{-2}\cdot\frac{2^2}{2}\cdot e^{-2}\cdot 2+e^{-2}\cdot 2\cdot e^{-1}\cdot\frac{2^2}{2}+e^{-2}\cdot 1\cdot e^{-2}\cdot\frac{2^3}{3!}}{1-e^{-2}} = 19.77\%\,.
\end{aligned}$$

Esercizio 3. (8 punti) Alice e Bob giocano al seguente gioco. Generano un numero casuale distribuito con legge uniforme nell'intervallo $[0, a]$, dove l'estremo $a > 1$ è da determinare.

1. Calcolare, in funzione del parametro a, la probabilità che il numero generato sia più grande di 1.

Alice e Bob generano tante volte (in modo indipendente) un numero con il procedimento descritto sopra finché non viene generato un numero più grande di 1. Se entro la quarta volta (compresa) viene generato un numero più grande di 1, vince Alice. Altrimenti vince Bob.

2. Determinare a in modo che Alice vinca con il 50% di probabilità.

3. In media, qual è la prima volta che viene generato un numero più grande di 1?

Soluzione. 1. Sia $X \sim \mathrm{U}(0, a)$. Allora

$$\mathbb{P}(\{X \geq 1\}) = \int_1^a \frac{1}{a}\,\mathrm{d}x = \frac{a-1}{a}\,.$$

2. Per descrivere il fenomeno, consideriamo una variabile aleatoria con legge geometrica $Y \sim \mathrm{Geo}(p)$ dove p è la probabilità di un successo, ovvero $p = \frac{a-1}{a}$, come calcolato nel punto precedente. Imponiamo che

$$\begin{aligned}
50\% = \mathbb{P}(\{Y \leq 4\}) = 1 - \mathbb{P}(\{Y > 4\}) = 1-(1-p)^4 \implies (1-p)^4 = \frac{1}{2} \\
\implies 1-p = \frac{1}{\sqrt[4]{2}} \implies p = 1 - \frac{1}{\sqrt[4]{2}}\,.
\end{aligned}$$

Segue che

$$\frac{a-1}{a} = p \implies a-1 = pa \implies (1-p)a = 1 \implies a = \frac{1}{1-p} = \sqrt[4]{2} \simeq 1.189\,.$$

3. In media, la prima volta che viene generato un numero più grande di 1 è

$$\mathbb{E}(Y) = \frac{1}{p} = \frac{a}{a-1} \simeq 6.29\,.$$

Esercizio 4. (7 punti) Si vuole stimare la vita media delle batterie per smartphone prodotte da un'azienda. Si misura la vita delle batterie su un campione casuale e si ottengono le seguenti misurazioni:

6.6 6.7 7.9 5.3 3.2 1.8 5.1 5.9

Si supponga che la distribuzione della vita delle batterie sia normale.

1. Calcolare sui dati un intervallo di confidenza bilaterale al 95% per la media.
2. Un intervallo di confidenza calcolato sui dati al 85% sarebbe più grande o più piccolo? (N.B.: Non è richiesto il calcolo esplicito, ma si deve motivare la risposta).

Soluzione. Abbiamo un campione casuale $X_1, \dots, X_n \sim \mathcal{N}(\mu, \sigma^2)$ con $n = 8$. Vogliamo calcolare un intervallo di confidenza di livello $\beta = 95\%$.

1. Dalla definizione di IC si ha che

$$\beta = \mathbb{P}(\{U_n \leq \mu \leq V_n\}) \,.$$

Per stimare μ sfrutteremo lo stimatore media campionaria $\overline{X}_n = \frac{1}{n}\sum_{i=1}^n X_i$. 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 $X_1, \dots, X_n \sim \mathcal{N}(\mu, \sigma^2)$ sono indipendenti, $T_{n-1} = \frac{\overline{X}_n - \mu}{S_n/\sqrt{n}} \sim t(n-1)$. Allora

$$\begin{aligned}
\beta = \mathbb{P}(\{U_n \leq \mu \leq V_n\}) &= \mathbb{P}\Big(\Big\{\frac{\overline{X}_n - V_n}{S_n/\sqrt{n}} \leq \frac{\overline{X}_n - \mu}{S_n/\sqrt{n}} \leq \frac{\overline{X}_n - U_n}{S_n/\sqrt{n}}\Big\}\Big) \\
&= \mathbb{P}\Big(\Big\{\frac{\overline{X}_n - V_n}{S_n/\sqrt{n}} \leq T_{n-1} \leq \frac{\overline{X}_n - U_n}{S_n/\sqrt{n}}\Big\}\Big) \\
&= 1 - \mathbb{P}\Big(\Big\{T_{n-1} < \frac{\overline{X}_n - V_n}{S_n/\sqrt{n}}\Big\}\Big) - \mathbb{P}\Big(\Big\{T_{n-1} > \frac{\overline{X}_n - U_n}{S_n/\sqrt{n}}\Big\}\Big) \,.
\end{aligned}$$

Segue che

$$\mathbb{P}\Big(\Big\{T_{n-1} < \frac{\overline{X}_n - V_n}{S_n/\sqrt{n}}\Big\}\Big) + \mathbb{P}\Big(\Big\{T_{n-1} > \frac{\overline{X}_n - U_n}{S_n/\sqrt{n}}\Big\}\Big) = 1 - \beta = \alpha \,.$$

Decidiamo di equipartire α:

$$\mathbb{P}\Big(\Big\{T_{n-1} < \frac{\overline{X}_n - V_n}{S_n/\sqrt{n}}\Big\}\Big) = \mathbb{P}\Big(\Big\{T_{n-1} < \frac{\overline{X}_n - V_n}{S_n/\sqrt{n}}\Big\}\Big) = \frac{\alpha}{2} \,.$$

Definiamo $t_{n-1,\alpha/2}$ come il punto tale che

$$\mathbb{P}(\{T_{n-1} \geq t_{n-1,\alpha/2}) = \frac{\alpha}{2} \,.$$

Scegliendo

$$\begin{aligned}
\frac{\overline{X}_n - U_n}{S_n/\sqrt{n}} = t_{n-1,\alpha/2} &\implies U_n = \overline{X}_n - \frac{S_n}{\sqrt{n}} t_{n-1,\alpha/2} \,, \\
\frac{\overline{X}_n - V_n}{S_n/\sqrt{n}} = -t_{n-1,\alpha/2} &\implies V_n = \overline{X}_n + \frac{S_n}{\sqrt{n}} t_{n-1,\alpha/2} \,,
\end{aligned}$$

si ottiene la condizione che definisce il livello di confidenza. In conclusione

$$\Big[\overline{X}_n - \frac{S_n}{\sqrt{n}} t_{n-1,\alpha/2}, \overline{X}_n + \frac{S_n}{\sqrt{n}} t_{n-1,\alpha/2}\Big]$$

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

Calcoliamo l'intervallo di confidenza sui dati

$$\Big[\overline{x}_n - \frac{s_n}{\sqrt{n}} t_{n-1,\alpha/2}, \overline{x}_n + \frac{s_n}{\sqrt{n}} t_{n-1,\alpha/2}\Big]$$

dove

$$\overline{x}_n = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{1}{8}(6.6 + 6.7 + 7.9 + 5.3 + 3.2 + 1.8 + 5.1 + 5.9) = 5.3125$$

è la media calcolata sui dati e

$$\begin{aligned} s_n^2 &= \frac{1}{n-1}\Big(\sum_{i=1}^{n} x_i^2 - n\overline{x}_n^2\Big) \\ &= \frac{1}{7}\Big(6.6^2 + 6.7^2 + 7.9^2 + 5.3^2 + 3.2^2 + 1.8^2 + 5.1^2 + 5.9^2 - 8 \cdot 5.3125^2) \\ &\simeq 3.9241 \end{aligned}$$

è la varianza calcolata sui dati, da cui $s_n \simeq 1.98$. Dalla tavola della t-Student calcoliamo

$$t_{n-1,\alpha/2} = t_{7,0.05} = 2.365\,.$$

Quindi l'intervallo di confidenza calcolato sui dati è

$$\Big[\overline{x}_n - \frac{s_n}{\sqrt{n}} t_{n-1,\alpha/2}, \overline{x}_n + \frac{s_n}{\sqrt{n}} t_{n-1,\alpha/2}\Big] \simeq [3.66, 6.97]\,.$$

2. L'intervallo di confidenza sarebbe più piccolo in quanto la probabilità che μ sia nell'intervallo $[U_n, V_n]$ sarebbe più piccola. Si può anche osservare dalla tavola della t-Student che per $\alpha = 7.5\%$ si otterrebbe un valore più piccolo di $t_{n-1,\alpha/2}$.