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: aprile 2023
Matricola: ______________________ Data: 25/05/2023

Tempo massimo: 2 ore.

Esercizio 1. (6 punti) Si vuole studiare dopo quanti anni di utilizzo i consumatori sostituiscono i propri smartphone con nuovi dispositivi. Viene effettuato un sondaggio su un campione di persone a cui viene chiesto dopo quanti anni di utilizzo hanno sostituito il loro smartphone precedente. I dati sono raccolti in classi: per ogni intervallo viene indicato il numero di persone che ha sostituito uno smartphone dopo un numero di anni contenuto nell'intervallo:

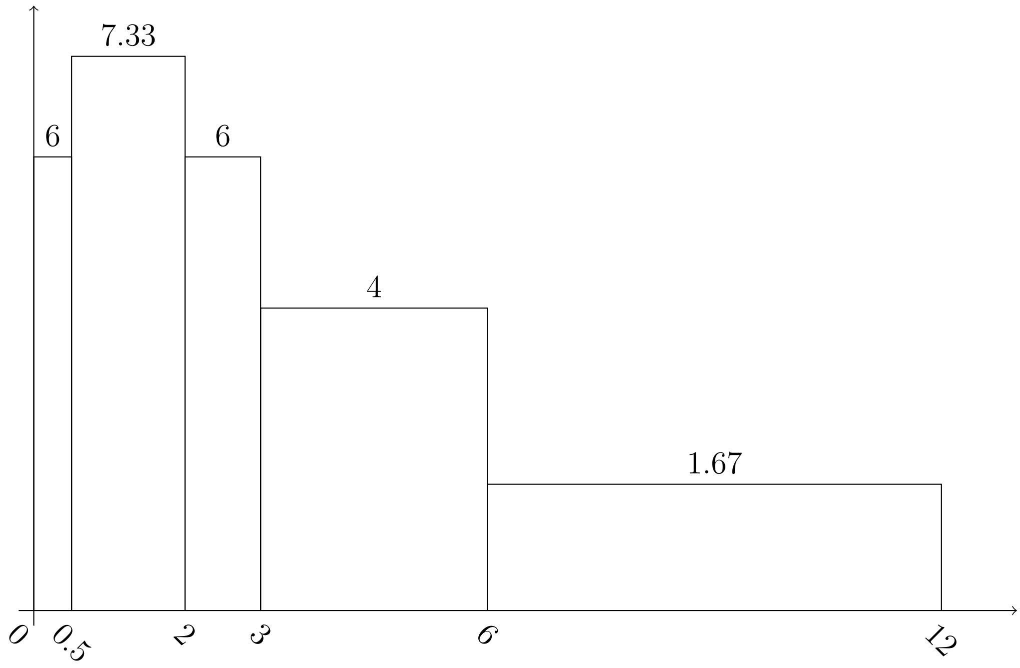
intervallo di anni	frequenze assolute
$[0, 0.5)$	3
$[0.5, 2)$	11
$[2, 3)$	6
$[3, 6)$	12
$[6, 12)$	10

1. Rappresentare un istogramma delle densità di frequenze assolute.
2. Determinare la classe modale.
3. Calcolare un'approssimazione della media e della deviazione standard dei dati.
4. Calcolare un'approssimazione della mediana dei dati.

Soluzione. 1. Calcoliamo le densità di frequenze assolute dividendo le frequenze assolute per l'ampiezza degli intervalli.

intervallo di anni	freq. assolute	densità di freq. ass.	freq. relative	freq. cumulate
$[0, 0.5)$	3	6	7.14%	3
$[0.5, 2)$	11	7.33	26.19%	14
$[2, 3)$	6	6	14.29%	20
$[3, 6)$	12	4	28.57%	32
$[6, 12)$	10	1.67	23.81%	42

Rappresentiamo le densità di frequenze assolute in un istogramma.



2. La classe modale è quella con maggiore densità di frequenza assoluta, quindi è l'intervallo $[0.5, 2)$.

3. Per calcolare un'approssimazione della media utilizziamo le frequenze relative ottenute da $p_j = f_j/n$ dove $n = 3 + 11 + 6 + 12 + 10 = 42$ e i valori centrali $\tilde{v}_j$ degli intervalli

$$\begin{aligned}\overline{x} &= \frac{1}{n}\sum_{i=1}^{n} x_i \simeq \frac{1}{n}\sum_{j=1}^{k} f_j\tilde{v}_j = \sum_{j=1}^{k} p_j\tilde{v}_j \\ &= 7.14\% \cdot 0.25 + 26.19\% \cdot 1.25 + 14.29\% \cdot 2.5 + 28.57\% \cdot 4.5 + 23.81\% \cdot 9 \simeq 4.13\,.\end{aligned}$$

Calcoliamo un'approssimazione della varianza

$$\begin{aligned}s^2 &= \frac{1}{n-1}\Big(\sum_{i=1}^{n} x_i^2 - n\overline{x}^2\Big) \simeq \frac{1}{n-1}\Big(\sum_{j=1}^{n} f_j\tilde{v}_j^2 - n\overline{x}^2\Big) = \frac{n}{n-1}\Big(\sum_{j=1}^{n} p_j\tilde{v}_j^2 - \overline{x}^2\Big) \\ &= \frac{42}{41}\Big(7.14\% \cdot 0.25^2 + 26.19\% \cdot 1.25^2 + 14.29\% \cdot 2.5^2 + 28.57\% \cdot 4.5^2 + 23.81\% \cdot 9^2 - 4.13^2\Big) \\ &= 9.54\,.\end{aligned}$$

Pertanto la deviazione standard è 3.09.

4. Per calcolare un'approssimazione della mediana dei dati, usiamo le frequenze cumulate. Troviamo l'intervallo I_j tale che $F_j \leq \frac{n}{2} < F_{j+1}$. Si tratta dell'intervallo $[3, 6)$. Approssimiamo la mediana con

$$Q_2 \simeq a_j + \lambda_j(b_j - a_j)$$

dove

$$\lambda_j = \frac{n/2 - F_j}{F_{j+1} - F_j} = \frac{21 - 20}{32 - 20} = \frac{1}{12}\,.$$

Quindi

$$Q_2 \simeq 3 + \frac{1}{12}(6 - 3) = 3 + \frac{1}{4} = 3.25\,.$$

Esercizio 2. (8 punti) Un'azienda produce un componente elettronico. Vengono utilizzate due linee di produzione, una vecchia e una nuova. Per la linea di produzione vecchia, si sa che il numero di componenti difettosi prodotto giornalmente è distribuito con una legge di Poisson con media 5.

1. Qual è la probabilità che in un giorno siano prodotti (strettamente) più di 2 componenti difettosi dalla linea di produzione vecchia?
2. Qual è la probabilità che in 2 giorni indipendenti vengano prodotti (strettamente) più di 4 componenti difettosi dalla linea di produzione vecchia?

Per la linea di produzione nuova, si sa che il numero di componenti difettosi prodotto giornalmente è distribuito con una legge di Poisson, ma non è nota la media. Il 40% dei componenti è prodotto con la linea di produzione nuova. In totale (tenendo conto sia della produzione nuova che di quella vecchia), si è calcolato che la probabilità che in un giorno non venga prodotto alcun componente difettoso è del 5 %.

3. In media, la linea di produzione nuova quanti componenti difettosi produce giornalmente?

Soluzione. Introduciamo la variabile aleatoria

$$X = \text{“numero di componenti difettose in un giorno con linea di produzione vecchia”} \sim P(\lambda)\,.$$

Ricordiamo che $\mathbb{E}(\lambda) = \lambda$, quindi $\lambda = 5$.

1. Viene chiesto di calcolare

$$\begin{aligned}\mathbb{P}(\{X > 2\}) &= 1 - \mathbb{P}(\{X \le 2\}) = 1 - \mathbb{P}(\{X = 0\}) - \mathbb{P}(\{X = 1\}) - \mathbb{P}(\{X = 2\}) \\ &= 1 - e^{-5}\Big(1 + 5 + \frac{5^2}{2}\Big) \simeq 87.53\%\,.\end{aligned}$$

2. Consideriamo due variabili X_1, $X_2 \sim P(5)$ indipendenti. Il numero di componenti difettose in due giorni indipendenti è $X_1 + X_2 \sim P(5+5) = P(10)$. Segue che

$$\begin{aligned}\mathbb{P}(\{X_1 + X_2 > 4\}) &= 1 - \mathbb{P}(\{X_1 + X_2 \le 4\}) \\ &= 1 - \mathbb{P}(\{X_1 + X_2 = 0\}) - \mathbb{P}(\{X_1 + X_2 = 1\}) - \mathbb{P}(\{X_1 + X_2 = 2\}) \\ &\quad - \mathbb{P}(\{X_1 + X_2 = 3\}) - \mathbb{P}(\{X_1 + X_2 = 4\}) \\ &= 1 - e^{-10}\Big(1 + 10 + \frac{10^2}{2} + \frac{10^3}{3!} + \frac{20^4}{4!}\Big) \\ &= 97.07\%\,.\end{aligned}$$

3. Introduciamo le variabili aleatorie

$$Y = \text{“numero di componenti difettose in un giorno con linea di produzione nuova”} \sim P(\mu)\,,$$

$$Z = \text{“numero di componenti difettose in un giorno (entrambe le linee di produzione)”}\,,$$

$$W = \text{“linea di produzione utilizzata (1 è nuova, 0 è vecchia)”} \sim \mathrm{Be}(40\%)\,.$$

Non conosciamo μ. Sappiamo però che

$$\mathbb{P}(\{Z = 0\}) = 5\%\,.$$

Utilizziamo il teorema della probabilità totale e il punto 1. per calcolare

$$\begin{aligned}5\% = \mathbb{P}(\{Z=0\}) &= \mathbb{P}(\{Z=0\}|\{W=1\})\mathbb{P}(\{W=1\}) + \mathbb{P}(\{Z=0\}|\{W=0\})\mathbb{P}(\{W=0\}) \\ &= \mathbb{P}(\{Y=0\})\mathbb{P}(\{W=1\}) + \mathbb{P}(\{X=0\})\mathbb{P}(\{W=0\}) \\ &= e^{-\mu}40\% + e^{-5}\cdot 60\%\,.\end{aligned}$$

Segue che

$$e^{-\mu} = \frac{5\% - e^{-5}\cdot 60\%}{40\%} = \frac{99.72\%}{40\%} \simeq 0.1148$$

da cui

$$\mu = -\log(0.1148) \simeq 2.16\,.$$

Esercizio 3. (7 punti) Un'azienda produce confezioni di erbicida. Il peso netto di una confezione di erbicida è distribuito in modo uniforme tra 20 e 23 kg.

1. Calcolare media e varianza del peso netto di una confezione.

Se una confezione pesa meno di 21 kg, viene chiesto un reclamo. Vengono vendute 15 confezioni.

2. Qual è la probabilità che ci siano (strettamente) più di 3 reclami?

3. Quanti reclami ci saranno in media?

Soluzione. Introduciamo la variabile aleatoria

$$X = \text{“peso netto di una confezione”} \sim \mathrm{U}(20,23)\,.$$

1. La media e la varianza di una variabile aleatoria uniforme $\mathrm{U}(a,b)$ sono date da

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

quindi

$$\mathbb{E}(X) = \frac{20+23}{2} = 21.5\,, \quad \mathrm{Var}(X) = \frac{(23-20)^2}{12} = 0.75\,.$$

2. Introduciamo la variabile aleatoria

$$Y = \text{“numero di confezioni con peso minore di 21 tra le 15 vendute”} \sim \mathrm{B}(15,p)\,,$$

dove p è la probabilità che una singola confezione abbia un peso minore di 21, cioè

$$p = \mathbb{P}(\{X<21\}) = \int_{20}^{21} \frac{1}{23-20}\,\mathrm{d}x = \frac{21-20}{23-20} = \frac{1}{3}\,.$$

Ci viene chiesto di calcolare la probabilità

$$\begin{aligned}&\mathbb{P}(\{Y>3\}) = 1-\mathbb{P}(\{Y\le 3\}) = 1-\mathbb{P}(\{Y=0\})-\mathbb{P}(\{Y=1\})-\mathbb{P}(\{Y=2\})-\mathbb{P}(\{Y=3\}) \\ &= 1-\binom{15}{0}\left(\frac{1}{3}\right)^0\left(\frac{2}{3}\right)^{15} - \binom{15}{1}\left(\frac{1}{3}\right)^1\left(\frac{2}{3}\right)^{14} - \binom{15}{2}\left(\frac{1}{3}\right)^2\left(\frac{2}{3}\right)^{13} - \binom{15}{3}\left(\frac{1}{3}\right)^3\left(\frac{2}{3}\right)^{12} \\ &= 1-\left(\frac{2}{3}\right)^{15} - 15\left(\frac{1}{3}\right)\left(\frac{2}{3}\right)^{14} - \frac{15\cdot 14}{2}\left(\frac{1}{3}\right)^2\left(\frac{2}{3}\right)^{13} - \frac{15\cdot 14\cdot 13}{3\cdot 2}\left(\frac{1}{3}\right)^3\left(\frac{2}{3}\right)^{12} \simeq 79.07\%\,.\end{aligned}$$

3. La media di una variabile aleatoria distribuita con legge binomiale è

$$\mathbb{E}(Y) = np = 15 \cdot \frac{1}{3} = 5\,.$$

Esercizio 4. (7 punti) In un certo procedimento chimico, è di fondamentale importanza che il pH di un reagente sia 8.20. Si sa la distribuzione del pH è normale e ha deviazione standard 0.04. Supponiamo che 10 misurazioni indipendenti abbiano fornito i seguenti valori:

$$8.18 \quad 8.16 \quad 8.17 \quad 8.22 \quad 8.19 \quad 8.17 \quad 8.15 \quad 8.21 \quad 8.16 \quad 8.18\,.$$

1. Qual è il più piccolo livello di significatività per cui i dati permettono di sostenere che il pH è diverso da 8.20?

2. Si può sostenere che il pH è diverso da 8.20 con un livello di significatività del 5%?

Soluzione. Impostiamo un test di ipotesi. Consideriamo un campione casuale di ampiezza $n = 10$ dato da $X_1, \dots, X_n \sim \mathcal{N}(\mu, \sigma^2)$, $\sigma^2 = 0.04$. Il test è

$$H_0 : \mu = 8.20\,, \quad H_1 : \mu \neq 8.20\,.$$

Denotiamo con $\mu_0 = 8.20$.

Definiamo la regione critica per il rifiuto dell'ipotesi nulla:

$$R_C = \{(x_1, \dots, x_n) \in R(X_1, \dots, X_n) \;:\; |\overline{x}_n - \mu_0| > \delta\}\,,$$

dove $\overline{x}_n = \frac{1}{n}\sum_{i=1}^n x_i$ è la media calcolata sui dati (quindi vogliamo rifiutare l'ipotesi nulla quando la media calcolata dai dati è significativamente distante da μ_0.)

La significatività α di un test di ipotesi è la probabilità di commettere un errore del I tipo, ovvero di rifiutare l'ipotesi nulla quando questa è vera. Supponiamo allora che l'ipotesi nulla H_0 sia vera, cioè $\mu = \mu_0$. Allora, introducendo la statistica $\overline{X}_n = \frac{1}{n}\sum_{i=1}^n X_i$,

$$\begin{aligned}\alpha &= \mathbb{P}(\{(X_1, \dots, X_n) \in R_C\}) = \mathbb{P}(\{|\overline{X}_n - \mu_0| > \delta\}) = \mathbb{P}(\{|\overline{X}_n - \mu| > \delta\}) \\ &= \mathbb{P}\Big(\Big\{\frac{|\overline{X}_n - \mu|}{\sigma/\sqrt{n}} > \frac{\delta}{\sigma/\sqrt{n}}\Big\}\Big)\,.\end{aligned}$$

Utilizziamo il fatto che $X_1, \dots, X_n$ sono indipendenti e distribuite con legge normale, quindi $Z = \frac{\overline{X}_n - \mu}{\sigma/\sqrt{n}} \sim \mathcal{N}(0, 1)$. Segue che

$$\alpha = \mathbb{P}\Big(\Big\{|Z| > \frac{\delta}{\sigma/\sqrt{n}}\Big\}\Big) = \mathbb{P}\Big(\Big\{Z < -\frac{\delta}{\sigma/\sqrt{n}}\Big\}\Big) + \mathbb{P}\Big(\Big\{Z > \frac{\delta}{\sigma/\sqrt{n}}\Big\}\Big)\,.$$

Scegliamo di equipartire la probabilità, ovvero

$$\mathbb{P}\Big(\Big\{Z < -\frac{\delta}{\sigma/\sqrt{n}}\Big\}\Big) = \mathbb{P}\Big(\Big\{Z > \frac{\delta}{\sigma/\sqrt{n}}\Big\}\Big) = \frac{\alpha}{2}\,.$$

Introducendo il valore $z_{\alpha/2}$ tale che

$$\frac{\alpha}{2} = \mathbb{P}(\{Z \geq z_{\alpha/2}\})$$

otteniamo che le uguaglianze di sopra sono verificate per

$$\frac{\delta}{\sigma/\sqrt{n}} = z_{\alpha/2} \implies \delta = \frac{\sigma}{\sqrt{n}} z_{\alpha/2}\,.$$

In conclusione, l'ipotesi nulla viene rifiutata quando

$$|\overline{x}_n - \mu_0| > \frac{\sigma}{\sqrt{n}} z_{\alpha/2}\,.$$

Il più piccolo livello di significatività per cui viene rifiutata l'ipotesi nulla è il p-value del test ed è dato da

$$\begin{aligned}
p\text{-value} &= \inf_\alpha \Big\{ |\overline{x}_n - \mu_0| > \frac{\sigma}{\sqrt{n}} z_{\alpha/2} \Big\} = \inf_\alpha \Big\{ \frac{|\overline{x}_n - \mu_0|}{\sigma/\sqrt{n}} > z_{\alpha/2} \Big\} \\
&= \inf_\alpha \Big\{ \Phi\Big(\frac{|\overline{x}_n - \mu_0|}{\sigma/\sqrt{n}}\Big) > \Phi(z_{\alpha/2}) \Big\} = \inf_\alpha \Big\{ \Phi\Big(\frac{|\overline{x}_n - \mu_0|}{\sigma/\sqrt{n}}\Big) > 1 - \frac{\alpha}{2} \Big\} \\
&= \inf_\alpha \Big\{ \alpha > 2\Big(1 - \Phi\Big(\frac{|\overline{x}_n - \mu_0|}{\sigma/\sqrt{n}}\Big)\Big) \Big\} = 2\Big(1 - \Phi\Big(\frac{|\overline{x}_n - \mu_0|}{\sigma/\sqrt{n}}\Big)\Big)\,.
\end{aligned}$$

Calcoliamo

$$\overline{x}_n = \frac{1}{n}\sum_{i=1}^{n} x_i = (8.18+8.16+8.17+8.22+8.19+8.17+8.15+8.21+8.16+8.18)/10 = 8.179\,.$$

$$\frac{|\overline{x}_n - \mu_0|}{\sigma/\sqrt{n}} = \frac{|8.179 - 8.20|}{0.04/\sqrt{10}} = 1.66\,.$$

Quindi, utilizzando la tavola della distribuzione normale,

$$p\text{-value} = 2\Big(1 - \Phi\Big(\frac{|\overline{x}_n - \mu_0|}{\sigma/\sqrt{n}}\Big)\Big) = 2\big(1 - \Phi(1.66)\big) = 2(1 - 0.9515) = 9.7\%\,.$$

2. $5\% < 9.7\% = p$-value, quindi l'ipotesi nulla non viene rifiutata.