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

A.A.: 2021/2022
Docente: Gianluca Orlando
Appello: aprile 2022
Data: 22/04/2022

Esercizio 1. Nella progettazione del propellente composito per un razzo si analizza un campione di ampiezza $n = 9$ di leganti polimerici. Il punto di fusione medio calcolato sul campione di leganti è 67.5 $°C$ (gradi Celsius). Si sa che il punto di fusione ha distribuzione normale con deviazione standard 1.8 $°F$ (Attenzione: espressa in gradi Fahrenheit!)

1. Ricordando che x $°F$ corrispondono a $(x - 32)\frac{5}{9}$ $°C$, qual è la deviazione standard della distribuzione normale espressa in gradi Celsius $°C$?

2. Testare l'ipotesi $H_0 : \mu = 68$ $°C$ contro $H_1 : \mu \neq 68$ $°C$ con un livello di significatività dell'1%.

3. Quanto dovrebbe essere grande l'ampiezza del campione n per rifiutare l'ipotesi H_0 con significatività del 5%? (Si assuma che la media calcolata sul campione sia sempre 67.5)

Soluzione. 1. Chiamiamo Y la variabile aleatoria con distribuzione normale con media μ_F e deviazione standard $\sigma_F = 1.8$ $°F$ che rappresenta il punto di fusione in gradi Fahrenheit e X la variabile aleatoria con distribuzione normale con media μ_C e deviazione standard σ_C che rappresenta il punto di fusione in gradi Celsius. Dalla formula di conversione abbiamo che

$$X = (Y - 32)\frac{5}{9} = \frac{5}{9}Y - 32\frac{5}{9} = aY + b\,.$$

Poiché $Y \sim \mathcal{N}(\mu_F, \sigma_F^2)$ abbiamo che $X \sim \mathcal{N}(a\mu_F + b, a^2\sigma_F^2)$, cioè

$$\mu_C = a\mu_F + b = (\mu_F - 32)\frac{5}{9}\,, \quad \sigma_C = a\sigma_F = \frac{5}{9}\sigma_F = \frac{5}{9}1.8 = 1\ °C\,.$$

2. La variabile aleatoria X da cui è estratto il campione ha distribuzione normale $\mathcal{N}(\mu, \sigma^2)$ con μ incognita e $\sigma = 1$. Per effettuare il test di ipotesi su μ possiamo sfruttare la statistica

$$Z = \frac{\overline{X} - \mu}{\sigma/\sqrt{n}} \sim \mathcal{N}(0, 1)\,,$$

dove $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ è la media campionaria.
Assumiamo vera l'ipotesi H_0, ovvero $\mu = 68$. Il test ha significatività $\alpha = 1\%$ se viene commesso un errore del I tipo con probabilità α, ovvero

$$\alpha = \mathbb{P}(|Z| > z_{\alpha/2})$$

dove $z_{\alpha/2}$ è il quantile gaussiano tale che $\mathbb{P}(Z > z_{\alpha/2}) = \frac{\alpha}{2}$. Calcoliamolo con le tavole:

$$\mathbb{P}(Z > z_{\alpha/2}) = \frac{\alpha}{2} \implies \mathbb{P}(Z \leq z_{\alpha/2}) = 1 - \frac{\alpha}{2} = 0.995 \implies z_{0.005} \simeq 2.575\,.$$

Studiamo ora la media campionaria calcolata sul campione di dati (realizzazione della variabile aleatoria):

$$\left|\frac{\overline{x} - \mu}{\sigma/\sqrt{n}}\right| = \left|\frac{67.5 - 68}{1/3}\right| = 1.5 < 2.575\,.$$

Pertanto non si rifiuta l'ipotesi nulla con un livello di significatività dell'1%.

3. Calcoliamo con le tavole

$$1 - 2.5\% = 1 - 0.025 = 0.975 = \mathbb{P}(Z > z_{0.025}) \implies z_{0.025} \simeq 1.96\,.$$

Cerchiamo n tale che

$$\left|\frac{\overline{x} - \mu}{\sigma/\sqrt{n}}\right| > 1.96 \iff n > 1.96^2\left|\frac{\sigma}{\overline{x} - \mu}\right|^2 = 1.96^2\left|\frac{1}{67.5 - 68}\right|^2 = 15.3664 \implies n \geq 16\,.$$

Esercizio 2. Si consideri il seguente campione di dati:

$$8, 18, 28, 1, 9, 21, 36, 15, 22, 30, 54, 1, 61, 58.$$

Calcolare i quartili, determinare eventuali dati anomali e sospetti e tracciare un box-plot. Se non ci sono dati anomali, aggiungere al campione un nuovo dato con il più piccolo valore che risulti anomalo più grande del terzo quartile (specificato in aula: più grande di 61).

Soluzione. Riordiniamo i dati

$$1, 1, 8, 9, 15, 18, 21, 22, 28, 30, 36, 54, 58, 61.$$

L'ampiezza del campione è $n = 14$. Calcoliamo i quartili.

Calcolo di Q_1: $(n+1)\frac{1}{4} = 3.75$. Quindi

$$Q_1 = x_3(1 - 0.75) + x_4(0.75) = 8.75\,.$$

Calcolo di Q_2: $(n+1)\frac{2}{4} = 7.5$. Quindi

$$Q_2 = x_7(1 - 0.5) + x_8(0.5) = 21.5\,.$$

Calcolo di Q_3: $(n+1)\frac{3}{4} = 11.25$. Quindi

$$Q_3 = x_{11}(1 - 0.25) + x_{12}(0.75) = 40.5\,.$$

Il range interquartile è $IQR = Q_3 - Q_1 = 40.5 - 8.75 = 31.75$. Le soglie per determinare i dati anomali sono:

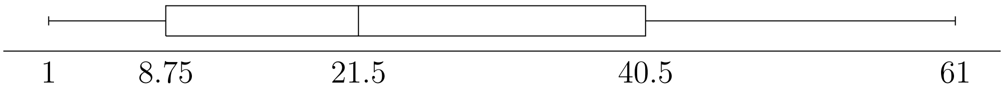
$$Q_1 - 3IQR = -86.5 \quad Q_3 + 3IQR = 135.75\,,$$

quindi non ci sono dati anomali. Le soglie per determinare i dati sospetti sono

$$Q_1 - 1.5IQR = -38.875 \quad Q_3 + 1.5IQR = 88.125\,,$$

quindi non ci sono dati sospetti.

Segue il box-plot:



Vogliamo aggiungere un dato anomalo. Aggiungeremo un dato più grande di 61. Vogliamo quantificare quanto grande. Aggiungendo un dato, l'ampiezza del campione cambie e diventa $n = 15$. I quartili diventano quindi:
Calcolo di Q_1: $(n+1)\frac{1}{4} = 4$. Quindi

$$Q_1 = x_4 = 9\,.$$

Calcolo di Q_3: $(n+1)\frac{3}{4} = 12$. Quindi

$$Q_3 = x_{12} = 54\,.$$

Il range interquartile è $IQR = Q_3 - Q_1 = 45$. La soglia per i dati anomali è $Q_3 + 3IQR = 189$. Quindi il nuovo dato deve essere almeno 189 per essere anomalo.

Esercizio 3. Una variabile aleatoria X rappresenta il peso in grammi di un articolo e ha densità

$$f(x) = \begin{cases} x-8 & \text{se } 8 \le x \le 9\,, \\ 10-x & \text{se } 9 < x \le 10\,, \\ 0 & \text{altrimenti.} \end{cases}$$

1. Calcolare media e varianza di X.
2. Il produttore vende questi articoli a 2 euro l'uno, con la garanzia di restituire i soldi a tutti i clienti che ne trovano uno da meno di 8.25 grammi. Il suo costo di produzione è legato al peso x del pezzo dalla relazione $x/15 + 0.35$. Determina il guadagno medio.

Soluzione. 1. La media è data da

$$\begin{aligned}\mathbb{E}(X) &= \int_{-\infty}^{+\infty} x f(x)\,\mathrm{d}x = \int_8^9 x(x-8)\,\mathrm{d}x + \int_9^{10} x(10-x)\,\mathrm{d}x \\ &= \left[\frac{1}{3}x^3 - 4x^2\right]_8^9 + \left[5x^2 - \frac{1}{3}x^3\right]_9^{10} = 13/3 + 14/3 = 9\,.\end{aligned}$$

La varianza è

$$\begin{aligned}\mathrm{Var}(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2 &= \int_{-\infty}^{+\infty} x^2 f(x)\,\mathrm{d}x - (27/3)^2 \\ &= \int_8^9 x^2(x-8)\,\mathrm{d}x + \int_9^{10} x^2(10-x)\,\mathrm{d}x - (27/3)^2 \\ &= \left[\frac{1}{4}x^4 - \frac{8}{3}x^3\right]_8^9 - \left[\frac{10}{3}x^3 - \frac{1}{4}x^4\right]_9^{10} - (27/3)^2 = 451/12 + 523/12 - 729/9 = 1/6\,.\end{aligned}$$

2. Se viene prodotto un articolo con peso x, il ricavo è

$$\begin{cases} 2 & \text{se } x \ge 8.25\,, \\ 0 & \text{altrimenti} \end{cases}$$

e, in vista dei costi di produzione, il guadagno è

$$H(x) = \begin{cases} 1.65 - \frac{x}{15} & \text{se } x \geq 8.25\,, \\ -\frac{x}{15} - 0.35 & \text{altrimenti.} \end{cases}$$

Il guadagno medio è quindi

$$\begin{aligned} E(H(X)) &= \int_{-\infty}^{+\infty} H(x) f(x)\,\mathrm{d}x \\ &= \int_{8}^{8.25} \Big(-\frac{x}{15} - 0.35\Big)(x-8)\,\mathrm{d}x + \int_{8.25}^{9} \Big(1.65 - \frac{x}{15}\Big)(x-8)\,\mathrm{d}x \\ &\quad + \int_{9}^{10} \Big(1.65 - \frac{x}{15}\Big)(10-x)\,\mathrm{d}x \\ &\simeq -0.0279514 + 0.501563 + 0.513889 = 0.9875006\,. \end{aligned}$$

Esercizio 4. Un pronto soccorso in un certo ospedale riceve una media di un paziente all'ora. Si assuma che il numero di pazienti ricevuti in un'ora sia distribuito secondo una legge di Poisson e i numeri di pazienti ricevuti in ore distinte siano indipendenti.

1. Un medico vuole conoscere la probabilità che il pronto soccorso riceva almeno 4 pazienti nel suo turno di 8 ore. Determinarla.
2. Sapendo che nella prima ora del turno non ci sono stati pazienti al pronto soccorso, qual è la probabilità di ricevere esattamente 2 pazienti nell'arco delle prime due ore?
3. Calcolare la varianza della variabile aleatoria che descrive il numero di pazienti ricevuti in 8 ore.

Soluzione. 1. Consideriamo la variabile aleatoria

$$X = X_1 + \cdots + X_8$$

dove $X_i \in P(1)$ rappresenta il numero di pazienti ricevuti nella i-esima ora e $X_1, \ldots, X_8$ sono indipendenti. Poiché la somma di variabili aleatorie indipendenti distribuite con legge di Poisson è una variabile aleatoria distribuita con una legge di Poisson, abbiamo che

$$X \sim P(8)\,.$$

Possiamo allora calcolare

$$\begin{aligned} \mathbb{P}(X \geq 4) &= 1 - \mathbb{P}(X \leq 3) = 1 - \mathbb{P}(X=0) - \mathbb{P}(X=1) - \mathbb{P}(X=2) - \mathbb{P}(X=3) \\ &= 1 - e^{-8} - e^{-8}8 - e^{-8}\frac{8^2}{2} - e^{-8}\frac{8^3}{6} \simeq 95.76\%\,. \end{aligned}$$

2. Usiamo il fatto che X_1 e X_2 sono indipendenti per calcolare

$$\begin{aligned} \mathbb{P}(X_1 + X_2 = 2 | X_1 = 0) &= \frac{\mathbb{P}(X_1 + X_2 = 2, X_1 = 0)}{\mathbb{P}(X_1 = 0)} \\ &= \frac{\mathbb{P}(X_2 = 2, X_1 = 0)}{\mathbb{P}(X_1 = 0)} \\ &= \frac{\mathbb{P}(X_2 = 2)\mathbb{P}(X_1 = 0)}{\mathbb{P}(X_1 = 0)} \\ &= \mathbb{P}(X_2 = 2) = e^{-1}\frac{1}{2} \simeq 18.39\%\,. \end{aligned}$$

3. La varianza di una variabile aleatoria di Poisson è uguale alla media, quindi è 8.

Esercizio 5. Ci sono tre forzieri. Ogni forziere contiene delle palline. Il primo forziere ha 13 palline rosse e 12 palline blu. Il secondo ha 20 palline rosse e 3 blu. Il terzo ha 12 palline rosse e 10 palline blu.

1. Viene scelto uno dei tre forzieri a caso e da questo vengono estratte con reinserimento e indipendentemente 5 palline. Qual è la probabilità che 4 siano rosse e 1 blu?

2. Viene scelto uno dei tre forzieri a caso e da questo vengono estratte con reinserimento e indipendentemente 6 palline, di cui 4 risultano essere rosse. Qual è la probabilità che il forziere scelto fosse il secondo?

Soluzione. 1. Denotiamo con X la variabile aleatoria che fornisce il numero del forziere estratto

$$\mathbb{P}(X = i) = \frac{1}{3} \quad \text{per } i = 1, 2, 3\,.$$

Denotiamo con Y il numero di palline rosse estratte alla fine dell'esperimento. Osserviamo che se è stato pescato il forziere i, la variabili aleatoria Y ha una certa legge binomiale (dipende dal forziere pescato). Ad esempio:

$$\mathbb{P}(Y = k|X = 1) = \binom{5}{k}\left(\frac{13}{25}\right)^k\left(\frac{12}{25}\right)^{5-k}$$

Utilizziamo il Teorema della probabilità totale per calcolare

$$\begin{aligned}
&\mathbb{P}(Y = 4)\\
&\quad= \mathbb{P}(Y = 4|X = 1)\mathbb{P}(X = 1) + \mathbb{P}(Y = 4|X = 2)\mathbb{P}(X = 2) + \mathbb{P}(Y = 4|X = 3)\mathbb{P}(X = 3)\\
&\quad= \binom{5}{4}\left(\frac{13}{25}\right)^4\left(\frac{12}{25}\right)^1\frac{1}{3} + \binom{5}{4}\left(\frac{20}{23}\right)^4\left(\frac{3}{23}\right)^1\frac{1}{3} + \binom{5}{4}\left(\frac{12}{22}\right)^4\left(\frac{10}{22}\right)^1\frac{1}{3}\\
&\quad= \frac{5}{3}\left(\left(\frac{13}{25}\right)^4\left(\frac{12}{25}\right) + \left(\frac{20}{23}\right)^4\left(\frac{3}{23}\right) + \left(\frac{12}{22}\right)^4\left(\frac{10}{22}\right)\right) \simeq 24.98\%\,.
\end{aligned}$$

2. Poiché è cambiato l'esperimento di estrazione delle palline (adesso vengono estratte 6 palline), diamo un nuovo nome alla variabile aleatoria che fornisce il numero di palline rosse estratte alla fine dell'esperimento: Z.
Per calcolare la probabilità richiesta, utilizziamo il Teorema di Bayes:

$$\begin{aligned}
&\mathbb{P}(X = 2|Z = 4)\\
&\quad= \frac{\mathbb{P}(Z = 4|X = 2)\mathbb{P}(X = 2)}{\mathbb{P}(Z = 4|X = 1)\mathbb{P}(X = 1) + \mathbb{P}(Z = 4|X = 2)\mathbb{P}(X = 2) + \mathbb{P}(Z = 4|X = 3)\mathbb{P}(X = 3)}\\
&\quad= \frac{\binom{6}{4}\left(\frac{20}{23}\right)^4\left(\frac{3}{23}\right)^2\frac{1}{3}}{\binom{6}{4}\left(\frac{13}{25}\right)^4\left(\frac{12}{25}\right)^2\frac{1}{3} + \binom{6}{4}\left(\frac{13}{25}\right)^4\left(\frac{12}{25}\right)^2\frac{1}{3} + \binom{6}{4}\left(\frac{20}{23}\right)^4\left(\frac{3}{23}\right)^2\frac{1}{3} + \binom{6}{4}\left(\frac{12}{22}\right)^4\left(\frac{10}{22}\right)^2\frac{1}{3}}\\
&\quad= \frac{\left(\frac{20}{23}\right)^4\left(\frac{3}{23}\right)^2}{\left(\frac{13}{25}\right)^4\left(\frac{12}{25}\right)^2 + \left(\frac{13}{25}\right)^4\left(\frac{12}{25}\right)^2 + \left(\frac{20}{23}\right)^4\left(\frac{3}{23}\right)^2 + \left(\frac{12}{22}\right)^4\left(\frac{10}{22}\right)^2}\\
&\quad= 21.68\%\,.
\end{aligned}$$