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.: 2020/2021
Docente: Gianluca Orlando
Appello: febbraio 2022
Data: 21/02/2022

Esercizio 1. Si sa che lo 0,1% degli individui di una popolazione ha una malattia. Un test permette di identificare la malattia con le seguenti sensibilità e specificità:

- se l'individuo ha la malattia, il test risulta positivo il 99% delle volte;
- se l'individuo è sano, il test risulta negativo il 98% delle volte.

Quesiti:

1. Un individuo con la malattia effettua k test indipendenti. Qual è la probabilità che i k test siano tutti positivi?
2. Un individuo sano effettua k test indipendenti. Qual è la probabilità che i k test siano tutti positivi?
3. Un individuo non sa se è sano o malato. Effettua k test, che risultano tutti positivi. Quanto deve essere grande k affinché la probabilità che abbia effettivamente la malattia sia almeno del 90%?

Soluzione. Descriviamo lo spazio degli eventi elementari:

$$\Omega = \{(\sigma, \omega_1, \dots, \omega_k) \ : \ \sigma \in \{M, S\}, \ \omega_i \in \{1, 0\}\} ,$$

dove $M =$ "malato", $S =$ "sano", $1 =$ "positivo", $0 =$ "negativo".
Introduciamo gli eventi

$$\begin{aligned}
B &= \text{"l'individuo ha la malattia"} = \{(M, , \omega_1, \dots, \omega_k) \ : \ \omega_i \in \{1, 0\}\} , \\
B^c &= \Omega \setminus B = \text{"l'individuo è sano"} , \\
A_j &= \text{"il } j\text{-esimo test è positivo"} \\
&= \{(\sigma, \omega_1, \dots, \omega_j, \dots, \omega_k) \ : \ \sigma \in \{M, S\}, \ \omega_i \in \{1, 0\} \text{ per } i \neq j \, , \ \omega_j = 1\} , \\
A &= \text{"tutti i } k \text{ test sono positivi"} = A_1 \cap \dots \cap A_k \, .
\end{aligned}$$

L'ipotesi sul numero di individui che ha la malattia è

$$\mathbf{P}(B) = \frac{1}{1000} \, .$$

L'ipotesi sulla sensibilità è

$$\mathbf{P}(A_j | B) = \frac{99}{100} \, .$$

L'ipotesi sulla specificità è

$$\mathbf{P}(A_j|B^c) = 1 - \frac{98}{100} = \frac{2}{100}\,.$$

1. La funzione $\mathbf{P}(\cdot|B)$ è una probabilità. Poiché i k test sono indipendenti, abbiamo che

$$\mathbf{P}(A|B) = \mathbf{P}(A_1 \cap \cdots \cap A_k|B) = \mathbf{P}(A_1|B)\cdots\mathbf{P}(A_k|B) = \Big(\frac{99}{100}\Big)^k.$$

2. Anche la funzione $\mathbf{P}(\cdot|B^c)$ è una probabilità. Poiché i k test sono indipendenti, abbiamo che

$$\mathbf{P}(A|B^c) = \mathbf{P}(A_1 \cap \cdots \cap A_k|B^c) = \mathbf{P}(A_1|B^c)\cdots\mathbf{P}(A_k|B^c) = \Big(\frac{2}{100}\Big)^k.$$

3. Utilizziamo il Teorema di Bayes:

$$\mathbf{P}(B|A) = \frac{\mathbf{P}(A|B)\mathbf{P}(B)}{\mathbf{P}(A|B)\mathbf{P}(B) + \mathbf{P}(A|B^c)\mathbf{P}(B^c)} = \frac{\big(\frac{99}{100}\big)^k \frac{1}{1000}}{\big(\frac{99}{100}\big)^k \frac{1}{1000} + \big(\frac{2}{100}\big)^k \frac{999}{1000}} = \frac{1}{1 + \big(\frac{2}{99}\big)^k 999}\,.$$

Calcoliamo

$$\begin{aligned}
k = 1: &\quad \frac{1}{1 + \frac{2}{99}999} = \frac{11}{233} \sim 4,72\%\,,\\
k = 2: &\quad \frac{1}{1 + \big(\frac{2}{99}\big)^2 999} = \frac{363}{511} \sim 71,04\%\,,\\
k = 3: &\quad \frac{1}{1 + \big(\frac{2}{99}\big)^2 999} = \frac{35937}{36233} \sim 99,18\%\,.
\end{aligned}$$

Quindi $k = 3$.

Esercizio 2. Sia $X = (X_1, X_2)$ una variabile aleatoria bidimensionale di tipo discreto con range

$$\{(-1,-1), (0,-1), (1,-1), (-1,0), (1,0), (0,1), (1,1)\}\,.$$

Si sa che $\mathbf{P}(X_1 = -1, X_2 = -1) = \frac{1}{6}$, $\mathbf{P}(X_1 = 1, X_2 = 0) = \frac{1}{3}$, $\mathbf{P}(X_1 = 0, X_2 = 1) = \frac{1}{4}$ e che gli altri valori che possono essere assunti sono equiprobabili.

1. Calcolare la probabilità che $X_1 > 0$ sapendo che $X_2 < 0$.
2. Stabilire se X_1 e X_2 sono indipendenti.
3. Calcolare il valore atteso della variabile aleatoria $Y = \max\{X_1, X_2\}$.

Soluzione. Sistemiamo le probabilità dei possibili valori in una tabella:

X_2 \ X_1	-1	0	1
-1	1/6	p	p
0	p	0	1/3
1	0	1/4	p

Imponiamo che la probabilità dell'evento certo sia 1:

$$4p + \frac{1}{6} + \frac{1}{3} + \frac{1}{4} = 1 \implies p = \frac{1}{16}.$$

Quindi possiamo completare la tabella

X_2 \ X_1	-1	0	1
-1	1/6	1/16	1/16
0	1/16	0	1/3
1	0	1/4	1/16

1. Usiamo la definizione di probabilità condizionata per calcolare

$$\mathbf{P}(X_1 > 0|X_2 < 0) = \frac{\mathbf{P}(X_1 > 0, X_2 < 0)}{\mathbf{P}(X_2 < 0)} = \frac{\mathbf{P}(X_1 = 1, X_2 = -1)}{\mathbf{P}(X_2 = -1)} = \frac{1/16}{1/6 + 1/16 + 1/16} = \frac{3}{14}.$$

2. Poiché

$$\mathbf{P}(X_1 > 0) = \mathbf{P}(X_1 = 1) = \frac{1}{16} + \frac{1}{3} + \frac{1}{16} = \frac{11}{24}$$

si ha che

$$\mathbf{P}(X_1 = 1|X_2 = -1) \neq \mathbf{P}(X_1 = 1),$$

quindi gli eventi non sono indipendenti e le variabili aleatorie non sono indipendenti.
3. Osserviamo che il range di Y è $\{-1, 0, 1\}$. Calcoliamo

$$\begin{aligned}
\mathbf{P}(Y = -1) &= \mathbf{P}(X_1 = -1, X_2 = -1) = \frac{1}{6}, \\
\mathbf{P}(Y = 0) &= \mathbf{P}(X_1 = 0, X_2 = -1) + \mathbf{P}(X_1 = -1, X_2 = 0) = \frac{1}{16} + \frac{1}{16} = \frac{1}{8}, \\
\mathbf{P}(Y = 1) &= \mathbf{P}(X_1 = 1, X_2 = -1) + \mathbf{P}(X_1 = 1, X_2 = 0) + \mathbf{P}(X_1 = 1, X_2 = 1) \\
&\quad + \mathbf{P}(X_1 = 0, X_2 = 1) = \frac{1}{16} + \frac{1}{3} + \frac{1}{16} + \frac{1}{4} = \frac{17}{24}.
\end{aligned}$$

Segue che

$$\mathbf{E}(Y) = -\mathbf{P}(Y = -1) + \mathbf{P}(Y = 1) = -\frac{1}{6} + \frac{17}{24} = \frac{13}{24} \sim 0.54.$$

Esercizio 3. Sia X un punto scelto sull'intervallo $[-2, 2]$ secondo una legge uniforme.

1. Calcolare la probabilità che il triangolo equilatero di lato $|X|$ abbia area maggiore di 1.
2. Calcolare il valore atteso dell'area del quadrato di lato $|X|$.

Sia Y un punto scelto sulla retta reale secondo una legge normale standard.

3. Calcolare il valore atteso dell'area del cerchio di raggio $|Y|$.

Soluzione.
1. L'area di un triangolo equilatero di base x è $\frac{\sqrt{3}}{4}x^2$. Dobbiamo quindi calcolare

$$\begin{aligned}
\mathbf{P}\Big(\frac{\sqrt{3}}{4}X^2 \geq 1\Big) &= \mathbf{P}\Big(X^2 \geq \frac{4}{\sqrt{3}}\Big) = \mathbf{P}\Big(X \leq -\frac{2}{3^{1/4}}\Big) + \mathbf{P}\Big(X \geq \frac{2}{3^{1/4}}\Big) \\
&= \mathbf{P}\Big(-2 \leq X \leq -\frac{2}{3^{1/4}}\Big) + \mathbf{P}\Big(\frac{2}{3^{1/4}} \leq X \leq 2\Big) \\
&= \int_{-2}^{-\frac{2}{3^{1/4}}} \frac{1}{4}\,\mathrm{d}x + \int_{\frac{2}{3^{1/4}}}^{2} \frac{1}{4}\,\mathrm{d}x \\
&= 1 - \frac{1}{3^{1/4}}\,.
\end{aligned}$$

2. Per calcolare il valore atteso dell'area del quadrato

$$\mathbf{E}(X^2) = \int_{-2}^{2} x^2 \frac{1}{4}\,\mathrm{d}x = \frac{1}{4}\Big[\frac{1}{3}x^3\Big]_{-2}^{2} = \frac{4}{3}\,.$$

3. La variabile aleatoria Y è distribuita secondo una legge normale standard, quindi

$$\mathbf{E}(Y) = 0\,, \quad \mathrm{Var}(Y) = 1\,,$$

Segue che

$$1 = \mathrm{Var}(Y) = \mathbf{E}(Y^2) - \mathbf{E}(Y)^2 = \mathbf{E}(Y^2)\,.$$

L'area del cerchio di raggio $|Y|$ è πY^2, quindi

$$\mathbf{E}(\pi Y^2) = \pi \mathbf{E}(Y^2) = \pi\,.$$

Esercizio 4. Un campione di ampiezza 21 viene estratto da una popolazione avente densità normale con media μ e varianza σ^2. La realizzazione della varianza campionaria risulta uguale a 2.

1. Determinare un intervallo di confidenza al 90% per la varianza σ^2.

Supponiamo di sapere in aggiunta che la media campionaria del campione sia uguale a 0.

2. Determinare un intervallo di confidenza al 95% per la media μ.

Soluzione. 1. Ricordiamo che S^2 è distribuita secondo una legge chi-quadro con $21 - 1 = 20$ gradi di libertà. Più precisamente:

$$X = 20\frac{S^2}{\sigma^2} \sim \chi^2_{20}\,.$$

Per determinare un intervallo di confidenza $[\Theta_1, \Theta_2]$ per σ^2 al $100(1-\alpha)\%$ (con Θ_1, Θ_2 v.a.), osserviamo che

$$\begin{aligned}
1 - \alpha = \mathbf{P}(\Theta_1 \leq \sigma^2 \leq \Theta_2) &= \mathbf{P}\Big(20\frac{S^2}{\Theta_2} \leq X \leq 20\frac{S^2}{\Theta_1}\Big) \\
&= 1 - \mathbf{P}\Big(X < 20\frac{S^2}{\Theta_2}\Big) - \mathbf{P}\Big(X > 20\frac{S^2}{\Theta_1}\Big)
\end{aligned}$$

da cui segue

$$\mathbf{P}\Big(X < 20\frac{S^2}{\Theta_2}\Big) + \mathbf{P}\Big(X > 20\frac{S^2}{\Theta_1}\Big) = \alpha\,.$$

Scegliamo di equipartire la probabilità α che l'intervallo $[20\frac{S^2}{\Theta_2}, 20\frac{S^2}{\Theta_1}]$ non contenga σ^2, in modo che

$$\mathbf{P}\Big(X < 20\frac{S^2}{\Theta_2}\Big) = \mathbf{P}\Big(X > 20\frac{S^2}{\Theta_1}\Big) = \frac{\alpha}{2}$$

Definiamo i quantili di una v.a. X distribuita secondo una legge chi-quadro. Per $\alpha \in [0,1]$ definiamo $\chi^2_{\alpha,20}$ il numero tale che

$$\mathbf{P}(X \geq \chi^2_{\alpha,20}) = \alpha\,.$$

Dalla definizione si ha che

$$\mathbf{P}(X < \chi^2_{1-\frac{\alpha}{2},20}) = 1 - \mathbf{P}(X \geq \chi^2_{1-\frac{\alpha}{2},20}) = 1 - \Big(1 - \frac{\alpha}{2}\Big) = \frac{\alpha}{2}\,,$$

$$\mathbf{P}(X > \chi^2_{\frac{\alpha}{2},20}) = \mathbf{P}(X \geq \chi^2_{\frac{\alpha}{2},20}) = \frac{\alpha}{2}\,.$$

Poniamo ora $\alpha = 10\% = 0.1$. Utilizzando il valore della varianza campionaria dato, possiamo determinare una realizzazione numerica degli estremi $20\frac{S^2}{\Theta_2}$ e $20\frac{S^2}{\Theta_1}$. Utilizzando le tabelle per la distribuzione chi-quadro:

$$20\frac{S^2}{\Theta_2} = \chi^2_{0.95,20} \implies \Theta_2 = 20\frac{S^2}{\chi^2_{0.95,20}} = 20\frac{2}{10.851} \sim 3.69\,,$$

$$20\frac{S^2}{\Theta_1} = \chi^2_{0.05,20} \implies \Theta_1 = 20\frac{S^2}{\chi^2_{0.05,20}} = 20\frac{2}{31.410} \sim 1.27\,.$$

Quindi $[1.27, 3.69]$ è un intervallo di confidenza al 90% per σ^2.

2. Ricordiamo che $T = \frac{\overline{X}-\mu}{S/\sqrt{21}}$ è distribuita come una t-Student con $n-1 = 20$ gradi di libertà. Determiniamo $[\Theta_1', \Theta_2']$ tale che

$$\begin{aligned} 1-\alpha = \mathbf{P}(\Theta_1' \leq \mu \leq \Theta_2') &= \mathbf{P}\Big(\frac{\overline{X}-\Theta_2'}{S/\sqrt{21}} \leq T \leq \frac{\overline{X}-\Theta_1'}{S/\sqrt{21}}\Big) \\ &= 1 - \mathbf{P}\Big(T < \frac{\overline{X}-\Theta_2'}{S/\sqrt{21}}\Big) - \mathbf{P}\Big(T > \frac{\overline{X}-\Theta_1'}{S/\sqrt{21}}\Big) \end{aligned}$$

da cui segue, decidendo di equipartire la probabilità,

$$\mathbf{P}\Big(T < \frac{\overline{X}-\Theta_2'}{S/\sqrt{21}}\Big) = \mathbf{P}\Big(T > \frac{\overline{X}-\Theta_1'}{S/\sqrt{21}}\Big) = \frac{\alpha}{2}\,.$$

Possiamo allora determinare gli estremi dell'intervallo utilizzando i quantili per la distribuzione di t-Student. Da

$$\mathbf{P}(T < -t_{\alpha/2,n-1}) = \frac{\alpha}{2}\,, \quad \mathbf{P}(T > t_{\alpha/2,n-1}) = \frac{\alpha}{2}$$

imponiamo

$$\frac{\overline{X}-\Theta_2'}{S/\sqrt{21}} = -t_{0.025,20} \implies \Theta_2' = \overline{X} + \frac{S}{\sqrt{21}}t_{0.025,20} \sim 0 + \frac{\sqrt{2}}{\sqrt{21}}2.086 \sim 0.64\,,$$

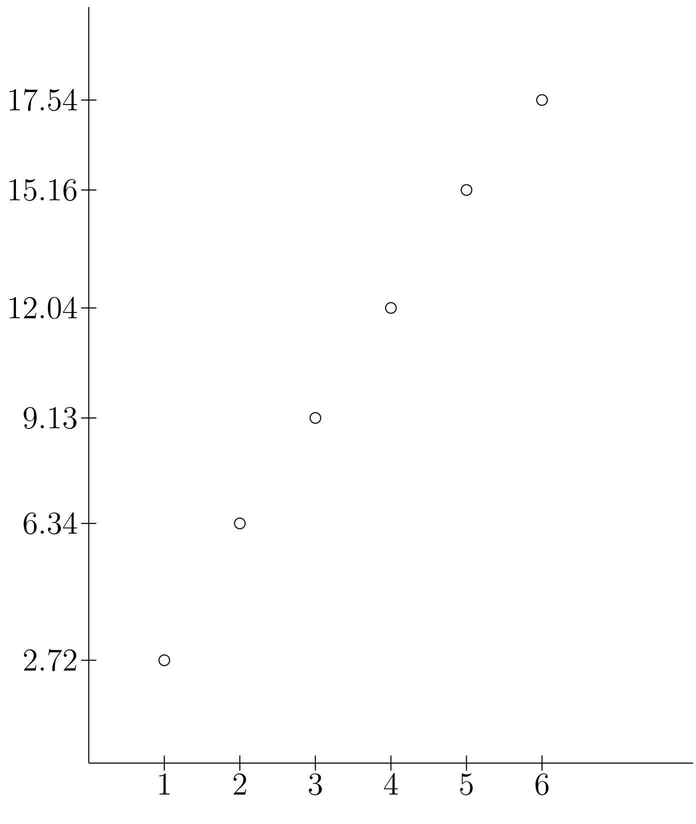
$$\frac{\overline{X} - \Theta_1'}{S/\sqrt{21}} = t_{0.025,20} \implies \Theta_1' = \overline{X} - \frac{S}{\sqrt{21}} t_{0.025,20} \sim 0 - \frac{\sqrt{2}}{\sqrt{21}} 2.086 \sim -0.64\,.$$

Esercizio 5. Si consideri il seguente campione bivariato di dati:

x	1	2	3	4	5	6
y	2.72	6.34	9.13	12.04	15.16	17.54

1. Rappresentare i dati in un diagramma a dispersione (scatterplot).
2. Determinare la retta di regressione lineare.
3. Disegnare la retta di regressione lineare.
4. Calcolare il coefficiente di correlazione lineare.

Soluzione.
1. Segue il diagramma a dispersione.



2. Utilizzando il metodo dei minimi quadrati si trovano i coefficienti α e β della retta di regressione lineare $y = \alpha x + \beta$. L'obiettivo è minimizzare:

$$\sum_{i=1}^{6} (y_i - \alpha x_i - \beta)^2.$$

Imponiamo che il gradiente rispetto ad α e β sia zero:

$$0 = \sum_{i=1}^{6} -2x_i(y_i - \alpha x_i - \beta)$$
$$0 = \sum_{i=1}^{6} 2(y_i - \alpha x_i - \beta)$$

da cui segue

$$\alpha = \frac{\sum_{i=1}^{6} x_i y_i - 6\overline{x}\,\overline{y}}{\sum_{i=1}^{6} x_i^2 - 6\overline{x}^2}$$
$$\beta = \overline{y} - \alpha\overline{x}\,.$$

Riportiamo i dati utili per il calcolo dei coefficienti:

x_i	1	2	3	4	5	6
y_i	2.72	6.34	9.13	12.04	15.16	17.54
$x_i y_i$	2.72	12.68	27.39	48.16	75.8	105.24
x_i^2	1	4	9	16	25	36
y_i^2	7.3984	40.1956	83.3569	144.9616	229.8256	307.6516

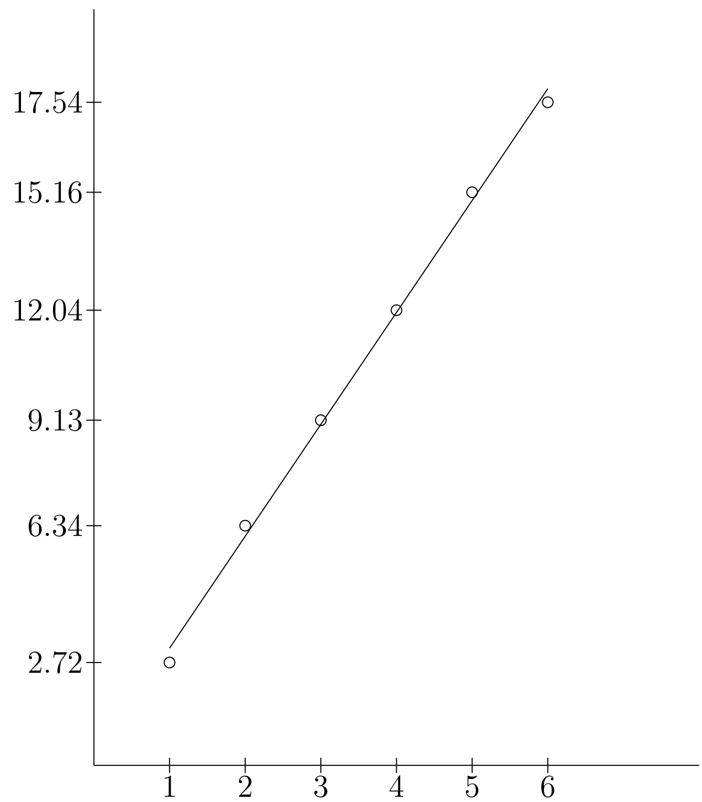
e

$$\overline{x} = \frac{1}{6}\sum_{i=1}^{6} x_i = 3.5 \quad \overline{y} = \frac{1}{6}\sum_{i=1}^{6} y_i \sim 10.49\,.$$

Segue che $\alpha \sim 2.96$ e $\beta \sim 0.14$ (si è arrotondato solo alla fine del conto), quindi la retta di regressione lineare è

$$y = 2.96x + 0.14\,.$$

3. Segue il grafico.



4. Ricordiamo che il coefficiente di correlazione lineare è dato da:

$$\rho = \frac{\sum_{i=1}^{6} x_i y_i - 6\overline{x}\,\overline{y}}{\sqrt{\sum_{i=1}^{6} x_i^2 - 6\overline{x}^2}\sqrt{\sum_{i=1}^{6} y_i^2 - 6\overline{y}^2}} = \alpha \frac{\sqrt{\sum_{i=1}^{6} x_i^2 - 6\overline{x}^2}}{\sqrt{\sum_{i=1}^{6} y_i^2 - 6\overline{y}^2}} = \alpha \frac{S_x}{S_y}.$$

Utilizzando una delle formule precedenti si ha che $\rho \sim -0,99864$.