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

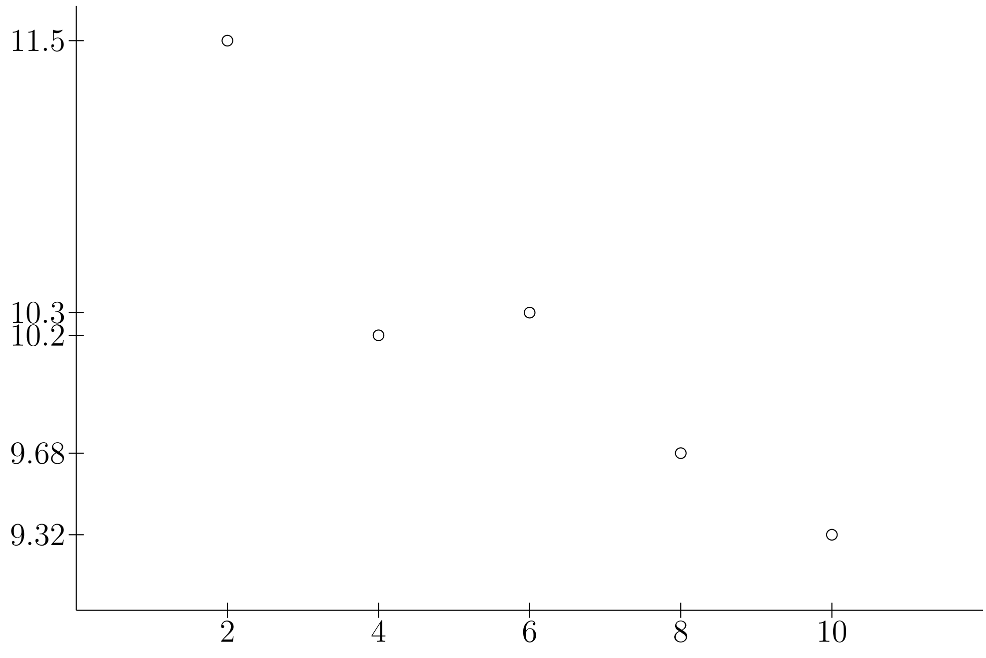
A.A.: 2020/2021
Docente: Gianluca Orlando
Appello: gennaio
Data: 25/01/2022

Esercizio 1. Per lo studio dell'inquinamento di un corso d'acqua, si misura la concentrazione della sostanza inquinante in 5 siti con diverse distanze dalla sorgente inquinante. Nella tabella seguente si presentano le misure effettuate:

km dalla sorgente inquinante	2	4	6	8	10
concentrazione	11.5	10.2	10.3	9.68	9.32

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. Denotiamo con x_i i dati $2, 4, 6, 8, 10$ e con y_i i dati $11.5, 10.2, 10.3, 9.68, 9.32$. 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:

$$\sum_{i=1}^{5} (y_i - \alpha x_i - \beta)^2.$$

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

$$0 = \sum_{i=1}^{5} -2x_i(y_i - \alpha x_i - \beta)$$
$$0 = \sum_{i=1}^{5} 2(y_i - \alpha x_i - \beta)$$

da cui segue

$$\alpha = \frac{\sum_{i=1}^{5} x_i y_i - 5\overline{x}\,\overline{y}}{\sum_{i=1}^{5} x_i^2 - 5\overline{x}^2}$$
$$\beta = \overline{y} - \alpha\overline{x}\,.$$

Riportiamo i dati utili per il calcolo dei coefficienti:

x_i	2	4	6	8	10
y_i	11.5	10.2	10.3	9.68	9.32
$x_i y_i$	23	40.8	61.8	77.44	93.2
x_i^2	4	16	36	64	100
y_i^2	132.25	104.04	106.09	93.7024	86.8624

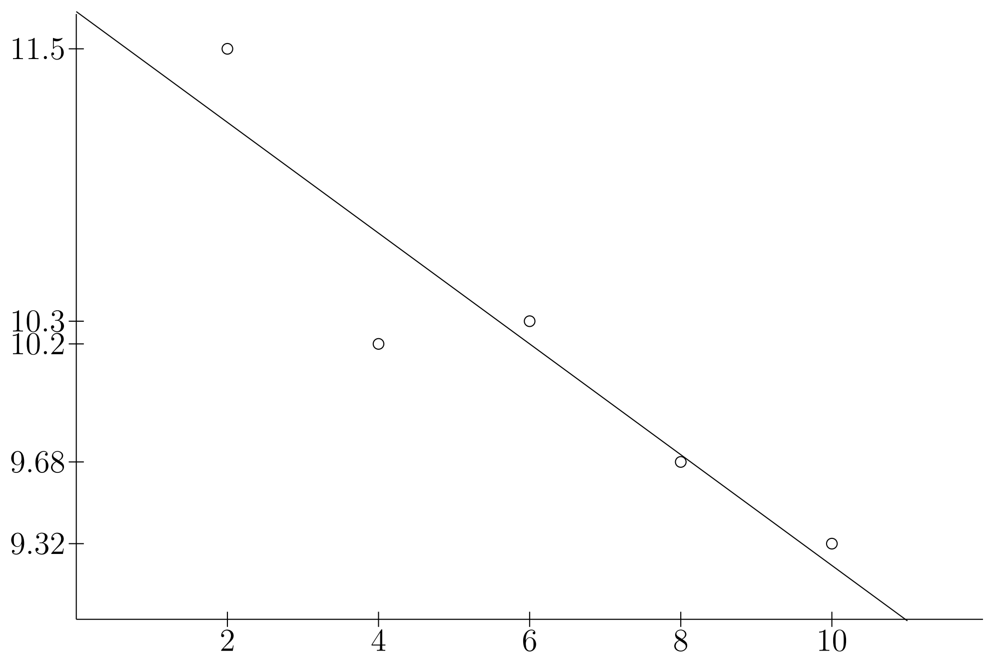
e

$$\overline{x} = \frac{1}{5}\sum_{i=1}^{5} x_i = 6 \quad \overline{y} = \frac{1}{5}\sum_{i=1}^{5} y_i = 10.2\,.$$

Segue che $\alpha = -0.244$ e $\beta = 11.664$, quindi la retta di regressione lineare è

$$y = -0.244x + 11.664\,.$$

3. Segue il grafico.



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

$$\rho = \frac{\sum_{i=1}^{5} x_i y_i - 5\overline{x}\,\overline{y}}{\sqrt{\sum_{i=1}^{5} x_i^2 - 5\overline{x}^2}\sqrt{\sum_{i=1}^{5} y_i^2 - 5\overline{y}^2}} = \alpha \frac{\sqrt{\sum_{i=1}^{5} x_i^2 - 5\overline{x}^2}}{\sqrt{\sum_{i=1}^{5} y_i^2 - 5\overline{y}^2}} = \alpha \frac{S_x}{S_y}.$$

Utilizzando una delle formule precedenti si ha che $\rho \sim -0,93$.

Esercizio 2. Un campione di ampiezza 17 viene estratto da una popolazione avente densità normale con media μ e varianza σ^2. La realizzazione della varianza campionaria risulta uguale a 25.

1. Determinare un intervallo di confidenza al 95% per la varianza σ^2.

Supponiamo di sapere in aggiunta che la media campionaria del campione sia uguale a 9.

2. Determinare un intervallo di confidenza al 99% per la media μ.

Soluzione. 1. Ricordiamo che S^2 è distribuita secondo una legge chi-quadro con $17 - 1 = 16$ gradi di libertà. Più precisamente:

$$X = 16\frac{S^2}{\sigma^2} \sim \chi^2_{16}\,.$$

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

$$\begin{aligned} 1 - \alpha = \mathbf{P}(\Theta_1 \le \sigma^2 \le \Theta_2) &= \mathbf{P}\Big(16\frac{S^2}{\Theta_2} \le X \le 16\frac{S^2}{\Theta_1}\Big) \\ &= 1 - \mathbf{P}\Big(X < 16\frac{S^2}{\Theta_2}\Big) - \mathbf{P}\Big(X > 16\frac{S^2}{\Theta_1}\Big) \end{aligned}$$

da cui segue

$$\mathbf{P}\Big(X < 16\frac{S^2}{\Theta_2}\Big) + \mathbf{P}\Big(X > 16\frac{S^2}{\Theta_1}\Big) = \alpha\,.$$

Scegliamo di equipartire la probabilità α che l'intervallo $[16\frac{S^2}{\Theta_2}, 16\frac{S^2}{\Theta_1}]$ non contenga σ^2, in modo che

$$\mathbf{P}\Big(X < 16\frac{S^2}{\Theta_2}\Big) = \mathbf{P}\Big(X > 16\frac{S^2}{\Theta_1}\Big) = \frac{\alpha}{2}$$

Definiamo i quantili di una v.a. X distribuita secondo una legge chi-quadro. Per $\alpha \in [0,1]$ definiamo $\chi^2_{\alpha,16}$ il numero tale che

$$\mathbf{P}(X \ge \chi^2_{\alpha,16}) = \alpha\,.$$

Dalla definizione si ha che

$$\mathbf{P}(X < \chi^2_{1-\frac{\alpha}{2},16}) = 1 - \mathbf{P}(X \ge \chi^2_{1-\frac{\alpha}{2},16}) = 1 - \Big(1 - \frac{\alpha}{2}\Big) = \frac{\alpha}{2}\,,$$

$$\mathbf{P}(X > \chi^2_{\frac{\alpha}{2},16}) = \mathbf{P}(X \ge \chi^2_{\frac{\alpha}{2},16}) = \frac{\alpha}{2}\,.$$

Poniamo ora $\alpha = 5\% = 0.05$. Utilizzando il valore della varianza campionaria dato, possiamo determinare una realizzazione numerica degli estremi $16\frac{S^2}{\Theta_2}$ e $16\frac{S^2}{\Theta_1}$. Utilizzando le tabelle per la distribuzione chi-quadro:

$$16\frac{S^2}{\Theta_2} = \chi^2_{0.975,16} \implies \Theta_2 = 16\frac{S^2}{\chi^2_{0.975,16}} = 16\frac{25}{6.908} \sim 57.90\,,$$

$$16\frac{S^2}{\Theta_1} = \chi^2_{0.025,16} \implies \Theta_1 = 16\frac{S^2}{\chi^2_{0.025,16}} = 16\frac{25}{28.845} \sim 13.87\,.$$

Quindi $[13.87, 57.90]$ è un intervallo di confidenza al 95% per σ^2.

2. Ricordiamo che $T = \frac{\overline{X}-\mu}{S/\sqrt{17}}$ è distribuita come una t-Student con $n-1$ gradi di libertà. Determiniamo $[\Theta_1', \Theta_2']$ tale che

$$\begin{aligned} 1-\alpha = \mathbf{P}(\Theta_1' \le \mu \le \Theta_2') &= \mathbf{P}\Big(\frac{\overline{X}-\Theta_2'}{S/\sqrt{17}} \le T \le \frac{\overline{X}-\Theta_1'}{S/\sqrt{17}}\Big) \\ &= 1 - \mathbf{P}\Big(T < \frac{\overline{X}-\Theta_2'}{S/\sqrt{17}}\Big) - \mathbf{P}\Big(T > \frac{\overline{X}-\Theta_1'}{S/\sqrt{17}}\Big) \end{aligned}$$

da cui segue, decidendo di equipartire la probabilità,

$$\mathbf{P}\Big(T < \frac{\overline{X}-\Theta_2'}{S/\sqrt{17}}\Big) = \mathbf{P}\Big(T > \frac{\overline{X}-\Theta_1'}{S/\sqrt{17}}\Big) = \frac{\alpha}{2}\,.$$

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

$$\frac{\overline{X}-\Theta_2'}{S/\sqrt{17}} = -t_{0.005,16} \implies \Theta_2' = \overline{X} + \frac{S}{\sqrt{17}}t_{0.005,16} \sim 9 + \frac{5}{\sqrt{17}}2.921 \sim 12.54\,,$$

$$\frac{\overline{X}-\Theta_1'}{S/\sqrt{17}} = t_{0.005,16} \implies \Theta_1' = \overline{X} - \frac{S}{\sqrt{17}}t_{0.005,16} \sim 9 - \frac{5}{\sqrt{17}}2.921 \sim 5.46\,.$$

Esercizio 3. Sia X una variabile aleatoria continua avente la seguente funzione di densità di probabilità

$$f(x) = \begin{cases} \dfrac{1}{|x|+k} & \text{se } |x| < a\,, \\ 0 & \text{altrimenti,} \end{cases}$$

dove $k > 0$ e $a > 0$.

1. Determinare il valore dei parametri k e a per cui f sia una densità di probabilità tale che X soddisfi a $\mathbf{P}(X > 1) = 1/3$.

2. Calcolare valore atteso e varianza di X. (Si consiglia di effettuare i conti sostituendo il valore esplicito di k e a solo alla fine.)

3. Calcolare la probabilità che $X < 1$ sapendo che si è verificato l'evento $X > 0$.

Soluzione. 1. Imponiamo che l'integrale di f sia uguale a 1:

$$1 = \int_{-\infty}^{+\infty} f(x)\,\mathrm{d}x = \int_{-a}^{a} \frac{1}{|x|+k}\,\mathrm{d}x = 2\int_0^a \frac{1}{x+k} = 2\log(x+k)\Big|_0^a = 2\log(a+k) - 2\log(k)$$
$$= 2\log\Big(\frac{a+k}{k}\Big).$$

Dalla condizione $\mathbf{P}(X > 1) = 1/3$ risulta evidente che $a > 1$, altrimenti si avrebbe $\mathbf{P}(X > 1) = 0$.

$$\frac{1}{3} = \mathbf{P}(X > 1) = \int_1^a f(x)\,\mathrm{d}x = \log(x+k)\Big|_1^a = \log(a+k) - \log(1+k) = \log\Big(\frac{a+k}{1+k}\Big).$$

Poniamo a sistema le due condizioni.

$$\begin{cases} \log\Big(\dfrac{a+k}{k}\Big) = \dfrac{1}{2} \\ \log\Big(\dfrac{a+k}{1+k}\Big) = \dfrac{1}{3} \end{cases} \implies \begin{cases} a+k = e^{\frac{1}{2}}k \\ a+k = e^{\frac{1}{3}}(1+k) \end{cases} \implies \begin{cases} a = (e^{\frac{1}{2}}-1)k \\ k = \dfrac{e^{\frac{1}{3}}}{e^{\frac{1}{2}} - e^{\frac{1}{3}}} \end{cases} \implies \begin{cases} a = \dfrac{e^{\frac{1}{2}}-1}{e^{\frac{1}{6}}-1} \\ k = \dfrac{1}{e^{\frac{1}{6}}-1}. \end{cases}$$

2. Calcoliamo il valore atteso

$$\mathbf{E}(X) = \int_{-\infty}^{+\infty} x f(x)\,\mathrm{d}x = \int_{-a}^{a} \frac{x}{|x|+k}\,\mathrm{d}x = 0,$$

poiché $\frac{x}{|x|+k}$ è una funzione dispari integrata su un intervallo simmetrico.
Calcoliamo la varianza

$$V(X) = \mathbf{E}(X^2) - \mathbf{E}(X)^2 = \mathbf{E}(X^2) = \int_{-\infty}^{+\infty} x^2 f(x)\,\mathrm{d}x = \int_{-a}^{a} \frac{x^2}{|x|+k}\,\mathrm{d}x = 2\int_0^a \frac{x^2}{x+k}\,\mathrm{d}x$$
$$= 2\int_0^a \frac{x^2+kx}{x+k}\,\mathrm{d}x - 2\int_0^a \frac{kx+k^2}{x+k}\,\mathrm{d}x + 2\int_0^a \frac{k^2}{x+k}\,\mathrm{d}x$$
$$= \Big[x^2 - 2kx + 2k^2\log(x+k)\Big]_0^a = a^2 - 2ka + k^2 = (a-k)^2 = \Big(\frac{e^{\frac{1}{2}}-2}{e^{\frac{1}{6}}-1}\Big)^2.$$

3. Per definizione di probabilità condizionata

$$\mathbf{P}(X < 1 | X > 0) = \frac{\mathbf{P}(0 < X < 1)}{\mathbf{P}(X > 0)} = \frac{\mathbf{P}(X > 0) - \mathbf{P}(X \geq 1)}{\mathbf{P}(X > 0)} = \frac{1/2 - 1/3}{1/2} = 1/3.$$

Esercizio 4. Un'urna contiene 20 palline. Tre delle palline sono bianche, le restanti sono nere.

1. Si supponga di pescare casualmente 1 pallina. Qual è la probabilità di pescare una pallina bianca?

2. Si supponga di pescare casualmente 3 palline (senza reinserimento). Qual è la probabilità che le 3 palline pescate siano bianche?

3. Si supponga di pescare casualmente 4 palline (senza reinserimento). Qual è la probabilità di pescare 3 palline bianche?

Soluzione. Numeriamo le palline da 1 a 20 e supponiamo che le prime 3 siano le palline bianche.
1. Lo spazio campione è dato da
$$\Omega = \{1, \dots, 20\} .$$
Gli eventi elementari sono tutti equiprobabili e pertanto hanno probabilità $1/\#\Omega = 1/20$, dove $\#\Omega$ è il numero di elementi di Ω. Gli eventi favorevoli sono 3 (ovvero $a = 1, 2, 3$), quindi la probabilità è $3/20$.

2. Lo spazio campione è dato da
$$\Omega = \Big\{ \{a, b, c\} \ : \ a, b, c \in \{1, \dots, 20\}, \ a, b, c \text{ distinti} \Big\} .$$
Gli eventi elementari in Ω sono equiprobabili e pertanto la loro probabilità è data da $1/\#\Omega$. Questo numero è dato dalla scelta di 3 elementi tra 20 (l'ordine non è importante) ed è pertanto dato da
$$\#\Omega = \binom{20}{3} = \frac{20!}{3!17!} .$$
L'evento favorevole è uno: $\{1, 2, 3\}$. Quindi la probabilità richiesta è $1/\binom{20}{3} = 3!17!/20!$.

Un modo alternativo di calcolare la probabilità è il seguente. La probabilità di pescare una pallina bianca alla prima estrazione è $3/20$, di pescarne una alla seconda estrazione $2/19$, di pescarne una alla terza estrazione $1/18$. Facendo il prodotto delle probabilità dei tre eventi si ottiene $3!17!/20!$.

3. Lo spazio campione è dato da
$$\Omega = \Big\{ \{a, b, c, d\} \ : \ a, b, c, d \in \{1, \dots, 20\}, \ a, b, c, d \text{ distinti} \Big\} .$$
Gli eventi elementari in Ω sono equiprobabili e pertanto la loro probabilità è data da $1/\#\Omega$. Questo numero è dato dalla scelta di 4 elementi tra 20 (l'ordine non è importante) ed è pertanto dato da
$$\#\Omega = \binom{20}{4} = \frac{20!}{4!16!} .$$
Restano da contare gli eventi favorevoli. Dobbiamo contare solo il numero di modi in cui si può estrarre una pallina nera tra le $20 - 3 = 17$ disponibili:
$$\binom{17}{1} = 17 .$$
Invece c'è un solo modo di scegliere le tre palline bianche tra le 3 disponibili:
$$\binom{3}{3} = 1 .$$
In conclusione la probabilità cercata è
$$\frac{17}{\binom{20}{4}} = 4 \frac{17}{20} \frac{3}{19} \frac{2}{18} \frac{1}{17} .$$
C'è un modo alternativo per ottenere la stessa formula: date 4 palline estratte, se una delle palline è nera può essere stata pescata per prima, seconda, terza o quarta. In ciascuno dei casi la probabilità di pescare 3 palline bianche e una nera è
$$\frac{17}{20} \frac{3}{19} \frac{2}{18} \frac{1}{17}$$

che quindi deve essere moltiplicato per le 4 possibili posizioni dell'estrazione.
In generale, se le palline totali sono N, le bianche K, ne vengono estratte n, allora la probabilità che k delle estratte siano bianche si calcola in questo modo. Lo spazio campione ha $\binom{N}{n}$ elementi. Gli eventi favorevoli si calcolano contando le possibili estrazioni di k palline bianche tra le K presenti, $\binom{K}{k}$, e le possibili estrazioni di $n-k$ palline nere tra le $N-K$ presenti, $\binom{N-K}{n-k}$. In conclusione la probabilità è

$$\frac{\binom{K}{k}\binom{N-K}{n-k}}{\binom{N}{n}}.$$

Esercizio 5. Bob ha deciso di non studiare il programma di Probabilità ma di provare a passare l'esame comunque. Ad ogni appello ha il 10% di probabilità di passare l'esame (indipendentemente dalle prove svolte precedentemente). Prova e riprova l'esame finché non lo passa.

1. Qual è la probabilità che passi l'esame entro il terzo appello?

2. Entro quale appello può passare l'esame almeno con il 50% di probabilità?

Soluzione. 1. Sia X la v.a. data da

$$X = \text{numero dell'appello in cui Bob passa l'esame.}$$

La probabilità che passi l'esame entro il terzo appello è

$$\begin{aligned}\mathbf{P}(X \le 5) &= \mathbf{P}(X=1) + \mathbf{P}(X=2) + \mathbf{P}(X=3) \\ &= \frac{1}{10} + \frac{9}{10}\frac{1}{10} + \frac{9}{10}\frac{9}{10}\frac{1}{10} = 27.1\%\,.\end{aligned}$$

2. Sia k il numero dell'appello incognito. Cerchiamo k tale che $\mathbf{P}(X \le k) \ge 50\%$.
Come prima calcoliamo

$$\mathbf{P}(X \le k) = \frac{1}{10} + \frac{9}{10}\frac{1}{10} + \frac{9}{10}\frac{9}{10}\frac{1}{10} + \cdots + \frac{1}{10}\left(\frac{9}{10}\right)^k.$$

Con questa formula troviamo che

$$\mathbf{P}(X \le 5) \sim 46.86\% < 50\%\,, \quad \mathbf{P}(X \le 6) \sim 52.17\%\,.$$

Quindi entro 6 appelli passerà l'esame con almeno il 50% di probabilità.