Corso di Laurea in Ingegneria Informatica
Anno Accademico 2021/2022
Calcolo delle Probabilità e Statistica Matematica

Nome ....

N. Matricola .... Ancona, 15 novembre 2022

1. (10 punti) Un campione di 100 pazienti viene analizzato per vedere la correlazione tra l'indice di peso corporeo e il contenuto glicemico nel sangue. I dati vengono raccolti nella seguente tabella:

	Indice peso corporeo			
Glicemia	15-20	20-25	25-40	40-50
90-100	4	3	2	1
100-115	3	4	2	1
115-125	0	6	7	3
125-135	0	8	10	9
135-150	1	12	16	8

Prendendo i valori centrali degli intervalli, si chiede di:

- calcolare le densità marginali delle variabili che rappresentano l'indice di peso corporeo e il contenuto glicemico;
- dire se l'indice di peso corporeo e il contenuto glicemico sono indipendenti;
- calcolare il contenuto glicemico medio e l'indice di peso corporeo medio;
- calcolare il coefficiente di correlazione tra l'indice di peso corporeo e il contenuto glicemico.

Soluzione. Sia X la variabile che descrive la glicemia e Y la variabile che descrive il peso. I valori nella tabellina, divisi per 100, non sono altro che i valori della densità congiunta $p_{XY}(i, j)$. Prendendo il valore centrale degli intervalli sia della glicemia che dell'indice, abbiamo che $X \in \{95, 107.5, 120, 130, 142.5\}$ mentre $Y \in \{17.5, 22.5, 32.5, 45\}$. Sommando righe e colonne otteniamo la nuova tabellina:

	Indice peso corporeo				
Glicemia	15-20	20-25	25-40	40-50	
90-100	4	3	2	1	10
100-115	3	4	2	1	10
115-125	0	6	7	3	16
125-135	0	8	10	9	27
135-150	1	12	16	8	37
	8	33	37	22	100

Le marginali si ottengono dividendo per 100 i numeri nell'ultima colonna e nell'ultima riga:

$$p_X(95) = 0.1 \qquad p_X(107.5) = 0.1 \qquad p_X(120) = 0.16 \qquad p_X(130) = 0.27 \qquad p_X(142.5) = 0.37$$
$$p_Y(17.5) = 0.08 \qquad p_Y(22.5) = 0.33 \qquad p_Y(32.5) = 0.37 \qquad p_Y(45) = 0.22$$

Ovviamente le due variabili non sono indipendenti. Le medie sono date da:

$$E[X] = 0.1 \times 95 + 0.1 \times 107.5 + 0.16 \times 120 + 0.27 \times 130 + 0.37 \times 142.5 = 127.275$$
$$E[Y] = 0.08 \times 17.5 + 0.33 \times 22.5 + 0.37 \times 32.5 + 0.22 \times 45 = 30.75$$

Per il calcolo delle varianze, calcoliamo dapprima $E[X^2]$ e $E[Y^2]$:

$$E[X^2] = 0.1 \times 95^2 + 0.1 \times 107.5^2 + 0.16 \times 120^2 + 0.27 \times 130^2 + 0.37 \times 142.5^2 = 16438.4$$
$$E[Y^2] = 0.08 \times 17.5^2 + 0.33 \times 22.5^2 + 0.37 \times 32.5^2 + 0.22 \times 45^2 = 1027.88$$

da cui

$$Var(X) = E[X^2] - E[X]^2 = 16438.4 - 127.275 = 239.512$$
$$Var(Y) = E[Y^2] - E[Y]^2 = 1027.88 - 30.75^2 = 82.3125$$

Per il calcolo del coefficiente di correlazione ci serve ancora $E[XY]$. Abbiamo

$$E[XY] = \sum_{i=1,5} \sum_{j=1,4} x_i \, y_j \, p_{XY}(i,j) = ... = 3950.25$$

La covarianza è quindi

$$Cov(X,Y) = E[XY] - E[X]\,E[Y] = 3950.25 - 127.275 \times 30.75 = 36.5437$$

e finalmente

$$\rho = \frac{Cov(X,Y)}{Var(X)\,Var(Y)} \approx 0.26$$

2. (10 punti) Si deve illuminare un campo sportivo per un evento in notturna della durata di 2 ore. L'impianto di illuminazione possiede 4 fari le cui lampade hanno un tempo di vita medio di 3 ore. Considerando sufficiente l'illuminazione fornita da almeno 2 fari, qual è la probabilità che l'evento non subisca un'interruzione per mancanza di visibilità ?

Soluzione. Sia X la variabile che descrive il tempo di vita di una lampada, cioè $X \sim \lambda\, e^{-\lambda x}$ con $\lambda = 1/3$. Allora la probabilità p che un faro duri almeno 2 ore è

$$p = \int_2^\infty f(x)dx = 1 - \int_0^2 f(x)dx = e^{-2\,\lambda} = e^{-2/3} \sim 0.513$$

Sia ora Y il numero di fari che rimangono accesi per almeno 2 ore. Per Y abbiamo la legge binomiale, quindi $Y \sim B(4,p)$. Quindi la risposta al quesito posto dal problema, indicando con $\mathcal{P}$ la probabilità richiesta, è:

$$\mathcal{P} = \sum_{k=2}^{4} B(4,p)(k) = \sum_{k=2}^{4} \binom{4}{k} p^k\,(1-p)^{n-k} \sim 0.707$$

3. (10 punti) Il consumo giornaliero di gas per riscaldamento (per famiglia) in una città di latitudine media è una variabile casuale X distribuita secondo una legge normale di media $\mu = 2$ metri cubi standard (Smc) e deviazione standard $\sigma = 0.5$ Smc (numeri consistenti con i dati ISTAT). Qual è la probabilità che

   (i) in esattamente 4 giorni nel mese (30 gg) il consumo superi i 2.5 Smc giornalieri?

   (ii) che in tutto il mese di gennaio si consumino almeno 65 Smc di gas?

Soluzione. Innanzitutto standardizziamo la variabile X. Abbiamo che

$$Z = \frac{X - \mu}{\sigma} \sim N(0, 1).$$

(i) la probabilità che in una data giornata i consumi superino i 2.5 Smc è data da

$$p = P(X > 2.5) = 1 - P(X < 2.5) = 1 - P(Z < (2.5 - 2)/0.5) = 1 - P(Z < 1) = 1 - 0.8413 = 0.1587$$

Siccome i giorni in cui questo succede sono distribuiti secondo una legge $B(30, p)$, la risposta è

$$\binom{30}{4} p^4 (1 - p)^{26} = 0.1945$$

(ii) Indichiamo con X_1 il consumo del primo gennaio, con X_2 quello del 2 gennaio, e via dicendo. Dobbiamo calcolare $P(X_1 + X_2 + ... + X_n \geq 650)$, con $n = 31$. Siano

$$\overline{X}_n = \frac{X_1 + X_2 + ... + X_n}{n} \qquad Z = \frac{\overline{X}_n - \mu}{\sigma} \sqrt{n}$$

con μ e σ come dai dati. Usando l'approssimazione normale, abbiamo

$$\begin{aligned} P(X_1 + X_2 + ... + X_n \geq 65) &= P\left(Z \geq \frac{65/n - 2}{0.5} \sqrt{n}\right) \\ &= P(Z \geq 1.08) = 1 - P(Z < 1.08) = 1 - 0.85993 = 0.14 \end{aligned}$$