Esercizi di Statistica della 3^a settimana (Corso di Laurea in Biotecnologie, Università degli Studi di Padova).

Esercizio 1. Supponiamo che il numero giornaliero di persone ricoverate per attacchi violenti di asma in un dato ospedale sia una variabile aleatoria di Poisson di parametro $\lambda = 1.5$.

1. Come sarà distribuito il numero di persone ricoverate per attacchi violenti di asma in due giorni?
2. Qual è la probabilità di osservare 5 o più casi in un periodo di 2 giorni?

In un particolare periodo di 2 giorni, i livelli di inquinamento dell'aria aumentano e la distribuzione di attacchi su 1 giorno diventa una legge di Poisson con parametro $\lambda_1 = 3$.

3. Rispondere al punto 2 in questa nuova situazione.
4. Se in ogni anno 10 giorni sono di alto inquinamento, qual è il numero atteso di ricoveri per asma in un dato anno? Qual è la varianza?

Esercizio 2. Nei ragazzi (maschi) di 17 anni, la media $\pm$ deviazione standard della pressione diastolica, misurata su un grande campione, risulta essere 63.7 ± 11.4 mm Hg. Supponendo che la pressione X in un dato ragazzo sia una variabile aleatoria gaussiana, ci sono diversi approcci per diagnosticare una "pressione elevata".

1. Un approccio diagnostica "pressione elevata" se la pressione diastolica supera il 90esimo percentile della distribuzione citata sopra. Dire a quanto corrisponde questa soglia x, che deve quindi essere tale che $\mathbb{P}\{X \leq x\} = 0.9$.
2. Un altro approccio diagnostica "pressione elevata" se la pressione diastolica supera i 90 mm Hg. Che percentuale di ragazzi di 17 anni avrà pressione elevata con questo approccio?

Esercizio 3. La distribuzione dei livelli di alfa-tocoferolo (vitamina E) nel sangue è approssimativamente normale con media 860 μg/dl e deviazione standard 340 μg/dl.

1. Che percentuale di persone ha livelli di alfa-tocoferolo compresi tra 400 e 1000 μg/dl?
2. Supponiamo che una persona abbia livelli tossici di alfa-tocoferolo se il suo livello nel sangue è > 1800 μg/dl. Qual è la probabilità che questo accada?

Esercizio 4. Un certo test nazionale di matematica viene proposto in tutte le ultime classi delle scuole secondarie. Esso produce punteggi che hanno distribuzione normale con media 500 e deviazione standard 100. Si scelgano a caso 5 studenti che hanno affrontato il test. Calcola le probabilità che:

1. i loro punteggi siano tutti inferiori a 600;
2. esattamente tre punteggi siano superiori a 640.

Soluzioni su `http://www.math.unipd.it/~vargiolu/Statistica/`

Soluzioni

Esercizio 1.

1. Definiamo X_i, $i = 1, 2$, il numero di persone ricoverate per asma nei giorni 1 e 2. Allora X_1 e X_2 sono indipendenti e con legge di Poisson di parametro 1.5. Allora $X_1 + X_2 \sim Po(3)$.

2. Abbiamo

$$
\begin{aligned}
\mathbb{P}\{X = 0\} &= e^{-3} = 0.04978 \\
\mathbb{P}\{X = 1\} &= \mathbb{P}\{X = 0\}\frac{3}{1} = 0.04978 \cdot \frac{3}{1} = 0.14936 \\
\mathbb{P}\{X = 2\} &= \mathbb{P}\{X = 1\}\frac{3}{2} = 0.14936 \cdot \frac{3}{2} = 0.22404 \\
\mathbb{P}\{X = 3\} &= \mathbb{P}\{X = 2\}\frac{3}{3} = 0.22404 \cdot \frac{3}{3} = 0.22404 \\
\mathbb{P}\{X = 4\} &= \mathbb{P}\{X = 3\}\frac{3}{4} = 0.22404 \cdot \frac{3}{4} = 0.16803
\end{aligned}
$$

quindi

$$\mathbb{P}\{X \geq 5\} = 1 - \sum_{k=0}^{4} \mathbb{P}\{X = k\} = 1 - 0.04978 - 0.14936 - 0.22404 - 0.22404 - 0.16803 = 0.18473$$

3. Definiamo Y_i, $i = 1, 2$, il numero di persone ricoverate per asma nei due giorni con alto inquinamento. Allora Y_1 e Y_2 sono indipendenti e con legge di Poisson di parametro 3, e $Y_1 + Y_2 \sim Po(6)$. Abbiamo poi

$$
\begin{aligned}
\mathbb{P}\{X = 0\} &= e^{-6} = 0.00247 \\
\mathbb{P}\{X = 1\} &= \mathbb{P}\{X = 0\}\frac{6}{1} = 0.00247 \cdot \frac{6}{1} = 0.01487 \\
\mathbb{P}\{X = 2\} &= \mathbb{P}\{X = 1\}\frac{6}{2} = 0.01487 \cdot \frac{6}{2} = 0.04461 \\
\mathbb{P}\{X = 3\} &= \mathbb{P}\{X = 2\}\frac{6}{3} = 0.04461 \cdot \frac{6}{3} = 0.08923 \\
\mathbb{P}\{X = 4\} &= \mathbb{P}\{X = 3\}\frac{6}{4} = 0.08923 \cdot \frac{6}{4} = 0.13385
\end{aligned}
$$

quindi

$$\mathbb{P}\{X \geq 5\} = 1 - \sum_{k=0}^{4} \mathbb{P}\{X = k\} = 1 - 0.00247 - 0.01487 - 0.04461 - 0.08923 - 0.13385 = 0.71494$$

4. In un dato anno abbiamo 355 giorni ad inquinamento "normale", a cui corrispondono le variabili aleatorie $X_1, \ldots, X_{355}$ i.i.d. di legge $Po(1.5)$, e 10 giorni ad alto inquinamento, a cui corrispondono le $Y_1, \ldots, Y_10$ i.i.d. di legge $Po(3)$. Abbiamo quindi che il numero di ricoveri in un dato anno è $Z := \sum_{i=1}^{355} X_i + \sum_{i=1}^{10} Y_i$, e quindi

$$E[Z] = \sum_{i=1}^{355} E[X_i] + \sum_{i=1}^{10} E[Y_i] = 355 \cdot 1.5 + 10 \cdot 3 = 562.5$$

Esercizio 2.

1. Dobbiamo calcolare il quantile di livello $\alpha = 0.9$ per una variabile aleatoria $X \sim N(63.7, (11.4)^2)$, cioè trovare x tale che $\mathbb{P}\{X \leq x\} = \alpha$. Abbiamo

$$0.9 = \mathbb{P}\{X \leq x\} = \mathbb{P}\left\{\frac{X - 63.7}{11.4} \leq \frac{x - 63.7}{11.4}\right\}$$

   Poichè $\frac{X-63.7}{11.4} \sim N(0,1)$, e per una variabile aleatoria $N(0,1)$ si ha $q_{0.9} = 1.28$, abbiamo che $\frac{x-63.7}{11.4} = 1.28$, e quindi $x = 63.7 + 11.4 \cdot 1.28 = 78.3$.

2. Dobbiamo calcolare

$$\begin{aligned}\mathbb{P}\{X > 90\} &= \mathbb{P}\left\{\frac{X - 63.7}{11.4} > \frac{90 - 63.7}{11.4}\right\} = \mathbb{P}\{Z > 2.30\} = 1 - \mathbb{P}\{Z \leq 2.30\} = \\ &= 1 - F_Z(2.30) = 1 - 0.98928 = 0.01072\end{aligned}$$

   dove $Z := \frac{X-63.7}{11.4} \sim N(0,1)$.

3. Il numero di ragazzi con "pressione alta" è una variabile aleatoria binomiale con parametri $n = 200$ e $p = 0.1$ (se si segue il criterio del punto 1.). Siccome $np(1-p) = 200 \cdot 0.1 \cdot 0.9 = 18 > 5$, possiamo applicare l'approssimazione normale, e abbiamo che

$$\begin{aligned}\mathbb{P}\{S_{200} \geq 25\} &= \mathbb{P}\left\{\frac{S_{200} - np}{\sqrt{np(1-p)}} \geq \frac{24.5 - np}{\sqrt{np(1-p)}}\right\} \simeq \mathbb{P}\left\{Z \geq \frac{24.5 - 20}{\sqrt{18}}\right\} = \\ &= \mathbb{P}\{Z \geq 1.06\} = 1 - \mathbb{P}\{Z < 1.06\} = 1 - F_Z(1.06) = 1 - 0.85543 = 0.14457\end{aligned}$$

   applicando anche la correzione di continuità.

Esercizio 3. Rappresentiamo il livello di alfa-tocoferolo in un generico individuo con una variabile aleatoria $X \sim N(860, 340^2)$.

1. Dobbiamo calcolare

$$\begin{aligned}\mathbb{P}\{400 \leq X \leq 1000\} &= \mathbb{P}\left\{\frac{400 - 860}{340} \leq \frac{X - 860}{340} \leq \frac{1000 - 860}{340}\right\} = \\ &= \mathbb{P}\{-1.35 \leq Z \leq 0.41\} = F_Z(0.41) - F_Z(-1.35) = \\ &= F_Z(0.41) - 1 + F_Z(1.35) = 0.65910 - 1 + 0.91149 = 0.57059\end{aligned}$$

   dove $Z := \frac{X-860}{340} \sim N(0,1)$.

2. Dobbiamo calcolare

$$\begin{aligned}\mathbb{P}\{X > 1800\} &= \mathbb{P}\left\{\frac{X - 860}{340} > \frac{1800 - 860}{340}\right\} = \mathbb{P}\{Z > 2.76\} = \\ &= 1 - \mathbb{P}\{Z \leq 2.76\} = 1 - F_Z(2.76) = 1 - 0.99711 = 0.00289\end{aligned}$$

3. Possiamo rappresentare il numero di intossicati con una variabile aleatoria binomiale S_n di parametri $n = 2000$ e $p = 0.00289$ (probabilità calcolata al punto precedente). Dobbiamo quindi calcolare $\mathbb{P}\{S_n > 20\}$. Poichè $np(1-p) = 2000 \cdot 0.00289 \cdot 0.99711 = 5.76 > 5$, possiamo applicare l'approssimazione normale. Calcoliamo

$$\begin{aligned}\mathbb{P}\{S_n > 20\} &= \mathbb{P}\left\{\frac{S_n - np}{\sqrt{np(1-p)}} > \frac{20.5 - np}{\sqrt{np(1-p)}}\right\} \simeq \mathbb{P}\left\{Z > \frac{20.5 - 5.78}{\sqrt{5.76}}\right\} = \\ &= \mathbb{P}\{Z > 6.13\} = 1 - \mathbb{P}\{Z \le 6.13\} = 1 - F_Z(6.13)\end{aligned}$$

dove Z è una (generica) variabile aleatoria di legge $N(0,1)$ e abbiamo usato la correzione di continuità. Nelle tavole troviamo però che l'ultimo valore della funzione di ripartizione è $F_Z(2.99) = 0.99861$, quindi l'unica cosa che possiamo dire è che

$$\mathbb{P}\{S_n > 20\} \simeq 1 - F_Z(6.13) < 1 - F_Z(2.99) = 1 - 0.99861 = 0.00139$$

poichè F_Z è una funzione crescente. Se invece abbiamo a disposizione un calcolatore che implementi la funzione di ripartizione gaussiana, possiamo calcolare $\mathbb{P}\{S_n > 20\} \simeq 1 - F_Z(6.13) = 4.3940 \cdot 10^{-10}$.

Esercizio 4. Possiamo rappresentare il risultato dell'i-esimo ragazzo tramite una variabile aleatoria $X_i \sim N(500; 100^2)$, e supporre che queste variabili aleatorie siano indipendenti.

1. Sfruttando l'indipendenza, si ha

$$\begin{aligned}\mathbb{P}\{X_i < 600 \quad \forall i = 1, \dots, 5\} &= \prod_{i=1}^{5} \mathbb{P}\{X_i < 600\} = \prod_{i=1}^{5} \mathbb{P}\left\{\frac{X_i - 500}{100} \le \frac{600 - 500}{100}\right\} = \\ &= \prod_{i=1}^{5} F_Z(1) = F_Z(1)^5 = 0.84134^5 = 0.42155\end{aligned}$$

dove abbiamo usato il fatto che la legge normale è continua.

2. Per rispondere, conviene definire nuove variabili aleatorie $Y_i := \mathbf{1}_{\{X_i > 640\}}$. Allora le $(Y_i)_i$ sono indipendenti (perchè funzioni di variabili aleatorie indipendenti) e identicamente distribuite con legge di Bernoulli di parametro

$$\begin{aligned}p &:= \mathbb{P}\{X_i > 640\} = 1 - \mathbb{P}\{X_i \le 640\} = 1 - \mathbb{P}\left\{\frac{X_i - 500}{100} \le \frac{640 - 500}{100}\right\} = \\ &= 1 - F_Z(1.4) = 1 - 0.91924 = 0.08076\end{aligned}$$

Allora il numero degli studenti con punteggio maggiore di 640 è $\sum_{i=1}^{5} Y_i \sim B(5, p)$, e dobbiamo calcolare

$$\mathbb{P}\left\{\sum_{i=1}^{5} Y_i = 3\right\} = \binom{5}{3} p^3 (1-p)^2 = 10 \cdot 0.08076^3 \cdot 0.91924^2 = 0.00445$$