Statistica I - A.A. 2017-2018

Prova scritta - 30 aprile 2018

Exercise 1 *(pt 20, $\sim$ 4+4+4+4+4) Un ufficio del comune è incaricato di fare statistiche sull'intensità della pioggia relativamente al mese di novembre, allo scopo di programmare gli interventi di pulitura della rete. Come quantità di studio Q prende i millimetri (mm) d'acqua caduti in una giornata di pioggia. Ci sono poi le giornate senza pioggia, in cui ovviamente $Q = 0$, ma tali giorni vengono considerati a parte.*

Ci sono quindi due aspetti casuali in gioco, quando si ragiona su un giorno generico futuro: da un lato, la probabilità $p \in [0, 1]$ che esso sia di pioggia; dall'altro, nel caso sia giorno di pioggia, la quantità Q di acqua che cadrà.

1. Le statistiche degli scorsi 5 anni, relative ai trenta giorni del mese di novembre, forniscono un dataset di 150 valori. Di questi, 80 sono la sigla NON, che indica "non pioggia", i rimanenti 70 sono numeri positivi, i valori di Q in quei giorni. Di tali numeri, la media aritmetica è pari a 30 mm con una deviazione standard di 15 mm. Supponendo Q gaussiana, dare un intervallo in cui cade la quantità media vera giornaliera (per i giorni di pioggia), al 90%. Raffigurare la risoluzione con un disegno.

2. Che probabilità c'è che, in un certo giorno futuro di novembre, cadano più di 50 mm d'acqua? Raffigurare la risoluzione con un disegno. Dare sia una risposta "media" sia quella più pessimistica, sulla base dei dati noti e delle analisi svolte precedentemente.

3. L'accumulo su più giorni di acqua caduta è uno dei problemi principali. Supponiamo che piova per 10 giorni di seguito e che i valori si possano ritenere indipendenti. Che probabilità c'è che, nel complesso, siano caduti (almeno) 40 centimetri d'acqua? Giustificare con precisione la risposta dal punto di vista teorico.

4. Sulla base dei dati illustrati sopra relativi ai cinque anni passati, si pensi al numero di giorni di pioggia del prossimo mese di novembre. Che probabilità c'è che sia superiore a 20? Anche qui, giustificare con precisione la risposta dal punto di vista teorico.

5. In vista del mese di novembre di quest'anno, progettare un test atto a capire se esso sia stato nella media dal punto di vista della quantità d'acqua caduta: scrivere ipotesi e regione di rifiuto, al 90%. Supponiamo poi che il mese di novembre di quest'anno sia trascorso e ci siano stati 12 giorni di pioggia con una media di 25 mm e deviazione standard di 10 mm. Cosa concludiamo?

Exercise 2 *(pt 10, $\sim$ 3+4+3) Ricordiamo che per v.a. U uniforme su $[-1, 1]$ si intende una v.a. la cui densità sia pari a zero fuori da $[-1, 1]$ e sia costante in $[-1, 1]$.*

1. Dopo aver calcolato tale constante, si trovi il numero q tale che $P(U \leq q) = 0.9$. Illustrare la risoluzione con un disegno.

2. Posto $X = U^2$, trovare la densità di probabilità di X.

3. Calcolare media e varianza di U e di X.

1 SOLUZIONI

Esercizio 1.

1)

$$30 \pm \frac{15 \cdot q_{0.95}}{\sqrt{70}} = 30 \pm \frac{15 \cdot 1.64}{\sqrt{70}} = 30 \pm 2.9403$$
$$= [27.06, 32.94] \,.$$

2)

$$P(Q > 50) = P\left(\frac{Q-m}{15} > \frac{50-m}{15}\right) = 1 - \Phi\left(\frac{50-m}{15}\right).$$

Usando $m = 30$, troviamo

$$P(Q > 50) = 1 - \Phi\left(\frac{20}{15}\right) = 1 - \Phi(1.33) = 1 - 0.908 = 0.092.$$

Usando invece $m = 32.94$,

$$P(Q > 50) = 1 - \Phi\left(\frac{50-32.94}{15}\right) = 1 - \Phi(1.14) = 1 - 0.872 = 0.128.$$

3) Indichiamo con Q_1,...,Q_{10} le quantità cadute nei 10 giorni, che sono gaussiane indipendenti. La loro somma $Q = Q_1 + ... + Q_{10}$ è una gaussiana (perché somma di gaussiane indipendenti) di media $10m$ e varianza $10 \cdot 15^2 = 2250$ ovvero deviazione standard 47.434. Pertanto,

$$P(Q > 400) = P\left(\frac{Q-10m}{47.434} > \frac{400-10m}{47.434}\right) = 1 - \Phi\left(\frac{400-10m}{47.434}\right).$$

Usando $m = 30$,

$$P(Q > 400) = 1 - \Phi\left(\frac{400 - 10 \cdot 30}{47.434}\right) = 1 - \Phi(2.11) = 1 - 0.982 = 0.018.$$

4) Indichiamo con X_1,...,X_{30} delle v.a. che descrivono, per ciascun giorno del prossimo novembre, se piove o meno. Precisamente, X_i vale 1 se piove, 0 se non piove. Abbiamo $P(X_i = 1) = p$. Sulla base dei 5 anni precedenti, usiamo la stima $p = 70/150$. Ora dobbiamo calcolare

$$P(X_1 + ... + X_{30} > 20) = P\left(\frac{X_1 + ... + X_{30} - 30p}{\sqrt{30}\sqrt{p(1-p)}} > \frac{20-30p}{\sqrt{30}\sqrt{p(1-p)}}\right).$$

Per il TLC, questa si approssima a

$$1 - \Phi\left(\frac{20-30p}{\sqrt{30}\sqrt{p(1-p)}}\right) = 1 - \Phi\left(\frac{20 - 30 \cdot 70/150}{\sqrt{30}\sqrt{70/150\,(1-70/150)}}\right)$$
$$= 1 - \Phi(2.1958) = 1 - 0.985 = 0.015.$$

5) $\mathcal{H}_0$) nuova media $= 30$, $\mathcal{H}_1$) nuova media $\neq 30$. Regione di rifiuto (indicando con q_i i valori spermentali, di numerosità n ancora incognita, con $\overline{q}$ la loro media aritmetica e con s la loro deviazione empirica):

$$\left\{ (q_1, ..., q_n) : \left| \frac{\overline{q} - 30}{s} \sqrt{n} \right| > t_{n-1,0.05} \right\}.$$

Trascorso novembre, vale $n = 12$, $\overline{q} = 25$, $s = 10$, $t_{n-1,0.05} = t_{11,0.05} = 1.796$

$$\left| \frac{\overline{q} - 30}{s} \sqrt{n} \right| = \left| \frac{25 - 30}{10} \sqrt{12} \right| = 1.7321$$

quindi il test non è significativo, non abbiamo motivo di ritenere anomalo questo mese di novembre.

Esercizio 2.

1) Base per altezza vale 1, quindi l'altezza, cioè la costante, vale $\frac{1}{2}$. Per la ricerca di q, la base è l'intervallo $[-1, q]$, quindi deve valere $(q+1)\frac{1}{2} = 0.9$, ovvero $q + 1 = 1.8$, $q = 0.8$.

2)

$$F_X(x) = P(X \leq x) = P\left(U^2 \leq x\right) \overset{x \geq 0}{=} P\left(-\sqrt{x} \leq U \leq \sqrt{x}\right).$$

In definitiva

$$F_X(x) = \begin{cases} 0 & \text{se} \quad x < 0 \\ \sqrt{x} & \text{se} \quad x \in [0,1] \\ 1 & \text{se} \quad x > 1 \end{cases}.$$

Pertanto

$$f_X(x) = \begin{cases} 0 & \text{se} \quad x < 0 \\ \frac{1}{2\sqrt{x}} & \text{se} \quad x \in (0,1) \\ 0 & \text{se} \quad x > 1 \end{cases}.$$

3)

$$\begin{aligned} E[U] &= 0 \text{ per simmetria} \\ Var[U] &= E\left[U^2\right] = \frac{1}{2}\int_{-1}^{1} u^2 du = \frac{1}{2}\left[\frac{u^3}{3}\right]_{-1}^{1} = \frac{1}{3}. \end{aligned}$$

$$\begin{aligned} E[X] &= E\left[U^2\right] = \frac{1}{3} \\ Var[X] &= E\left[X^2\right] - \frac{1}{9} = \int_0^1 x^2 \frac{1}{2\sqrt{x}} dx - \frac{1}{9} \\ &= \frac{1}{2}\int_0^1 x^{3/2} dx - \frac{1}{9} = \frac{1}{2}\left[\frac{x^{5/2}}{5/2}\right]_0^1 - \frac{1}{9} \\ &= \frac{1}{2}\frac{2}{5} - \frac{1}{9} = \frac{4}{45} = 0.0888... \end{aligned}$$