Statistica I - A.A. 2016-2017

Prova scritta - 28 giugno 2017

Problema 1. (pt 21) Un grande emporio di prodotti di elettronica vende numerosi esemplari di un certo prodotto ogni giorno; sia N tale numero di richieste giornaliere, aleatorio naturalmente.

1. Se descriviamo N con una gaussiana e, da registrazioni recenti, sappiamo che essa ha media 21 con deviazione standard 5, è ragionevole dimensionare il magazzino pensando che 27 giorni ogni 30 le richieste non superano 25 esemplari?

2. Un addetto agli studi teorici delle vendite obietta che una gaussiana assume valori arbitrariamente alti ed arbitrarimente negativi e pertanto non può essere usata per descrivere il numero di pezzi richiesti perché assume anche valori negativi. Gli viene fatto osservare che la probabilità che ciò accada è talmente piccola da essere irrilevante. Siete d'accordo?

3. Viene fatta pubblicità di quel prodotto sul depliant distribuito in giro e questo potrebbe aver aumentato le vendite, ma non ne siamo assolutamente sicuri, guardando i dati così ad occhio. Vengono registrate le richieste per 12 giorni (due settimane lavorative), trovando una media pari a 24, con deviazione standard 6. Possiamo affermare, al 95% che l'apparente aumento da 21 a 24 è dovuto ad una causa reale e non è una semplice fluttuazione casuale? Dichiarare con chiarezza le ipotesi e la regione di rifiuto.

4. Utilizzando per semplicità le statistiche gaussiane con la vecchia deviazione 5, calcolare la potenza del test appena svolto, definendo il concetto di potenza e partendo dalla definizione nello svolgere il calcolo.

5. Torniamo alla situazione del punto 1, senza pubblicità. In 26 giorni (circa un mese lavorativo), in media quanti esemplari vengono richiesti? Il numero N_{tot} dei pezzi richiesti in 26 giorni, che deviazione standard ha? Che probabilità c'è che sia inferiore a $26 \cdot 20$?

Problema 2. (pt 9) Al variare del parametro $\alpha > 0$, si consideri la funzione

$$f_\alpha(x) = \begin{cases} \alpha|x| & \text{per } -C_\alpha \leq x \leq C_\alpha \\ 0 & \text{altrove} \end{cases}$$

dove C_α è una costante positiva, dipendente da α.

1. Tracciare il grafico di $f_\alpha(x)$ e calcolare C_α in modo che $f_\alpha(x)$ sia una densità di probabilità.

2. Se X è una v.a. con densità $f_1(x)$, trovare λ tale che $P(X < \lambda) = 0.4$.
3. Cercare, col metodo che si ritiene più opportuno, uno stimatore di α (illustrare anche solo tentativi parziali).

1 Soluzioni

Esercizio 1.

1. $N \sim N(21, 5^2)$, quindi
$$P(N \leq 25) = P\left(\frac{N-21}{5} \leq \frac{25-21}{5}\right) = \Phi(0.8) = 0.7881.$$
Mentre $\frac{27}{30} = 0.9$. Il magazzino non è ben dimensionato, perché arrivano più di 25 richieste più spesso di quanto previsto.

2. $$P(N < 0) = P\left(\frac{N-21}{5} \leq \frac{0-21}{5}\right) = \Phi(-4.2) = 1 - \Phi(4.2).$$
Il valore 4.2 è fuori dalle tavole, il corrispondente valore $\Phi(4.2)$ è ancor più vicino ad 1 di quanto non sia quello più grande delle tavole e quindi $P(N < 0)$ è davvero piccolissimo. Siamo d'accordo con l'uso della gaussiana.

3. Vista la bassissima numerosità campionaria, eseguiamo sicuramente un test t di Student, per la media. Eseguiamo un test unilaterale, perché non c'è ragione di credere che la pubblicità abbassi le richieste. L'ipotesi nulla è $\mathcal{H}_0$) media ≤ 21; quella alternativa è $\mathcal{H}_1$) media > 21; la regione di rifiuto è $\left\{(x_1, ..., x_{12}) : \frac{\overline{x}_{12}-21}{5}\sqrt{12} > t_{11,0.05}\right\}$. Nel nostro caso vale
$$\frac{\overline{x}_{12} - 21}{5}\sqrt{12} = \frac{24-21}{5}\sqrt{12} = 2.0785$$
che è maggiore di $t_{11,0.05} = 1.796$. Quindi il test è significativo, c'è un aumento delle richieste.

4. La potenza è la probabilità di accorgersi del cambiamento quando questo c'è, ed è funzione della nuova media $\mu > 21$:
$$\begin{aligned} P^{\mu,5}\left(\frac{\overline{X}_{12}-21}{5}\sqrt{12} > q_{0.95}\right) &= P^{\mu,5}\left(\frac{\overline{X}_{12}-\mu}{5}\sqrt{12} + \frac{\mu-21}{5}\sqrt{12} > q_{0.95}\right) \\ &= 1 - P^{\mu,5}\left(\frac{\overline{X}_{12}-\mu}{5}\sqrt{12} \leq q_{0.95} - \frac{\mu-21}{5}\sqrt{12}\right) \\ &= 1 - \Phi\left(1.64 - \frac{\mu-21}{5}\sqrt{12}\right). \end{aligned}$$

5. Chiamiamo $N_1, ..., N_{26}$ le richieste dei 26 giorni, che supporremo indipendenti; vale $N_{tot} = N_1 + ... + N_{26}$. La media di N_{tot} è $26 \cdot E[N_1] = 26 \cdot 21$. La varianza di N_{tot} è (per l'indipendenza) $26 \cdot Var[N_1] = 26 \cdot 5^2$, quindi la deviazione standard di N_{tot} è $\sqrt{26 \cdot 5^2} = \sqrt{26} \cdot 5 = 25.495$. Infine, osservando

che la somma di v.a. gaussiane è gaussiana, quindi $N_{tot} \sim N\left(26 \cdot 21, 26 \cdot 5^2\right)$, abbiamo

$$P\left(N_{tot} < 26 \cdot 20\right) = P\left(\frac{N_{tot} - 26 \cdot 21}{\sqrt{26} \cdot 5} < \frac{26 \cdot 20 - 26 \cdot 21}{\sqrt{26} \cdot 5}\right)$$
$$= \Phi\left(-1.0198\right) = 1 - \Phi\left(1.0198\right) = 1 - 0.8437 = 0.1563.$$

Esercizio 2.

1.
$$\int f_\alpha\left(x\right) dx = 2 \int_0^{C_\alpha} \alpha x dx = 2\alpha \frac{C_\alpha^2}{2} = \alpha C_\alpha^2$$
e questa vale 1 per $C_\alpha = \frac{1}{\sqrt{\alpha}}$.

2. Essendo l'area divisa a metà dall'origine, il numero λ deve trovarsi nei negativi. Pertanto
$$P\left(X < \lambda\right) = \int_{-C_1}^{\lambda} \alpha \left|x\right| dx = -\int_{-1}^{\lambda} \alpha x dx = -\left[\frac{x^2}{2}\right]_{x=-1}^{x=\lambda} = \frac{1}{2} - \frac{\lambda^2}{2}.$$
Questo deve valere 0.4. Pertanto
$$\frac{1}{2} - \frac{\lambda^2}{2} = 0.4$$
$$\lambda^2 = 0.1 \cdot 2 = 0.2$$
$$\lambda = -\sqrt{0.2} = -0.44721.$$

3.
$$L\left(\alpha, x_1, ..., x_n\right) = \begin{cases} \alpha^n \left|x_1 \cdots x_n\right| & \text{se } -\frac{1}{\sqrt{\alpha}} \leq x_i \leq \frac{1}{\sqrt{\alpha}} \text{ per ogni } i = 1, ..., n \\ 0 & \text{altrimenti} \end{cases}.$$
La regione in cui L vale $\alpha^n \left|x_1 \cdots x_n\right|$ si può riscrivere come $-\frac{1}{\sqrt{\alpha}} \leq \min x_i \leq \max x_i \leq \frac{1}{\sqrt{\alpha}}$. Da qui con un po' di fatica si può arrivare ad uno stimatore, ma c'è il problema di distinguere i casi $\max x_i > 0$ ecc. Invece col metodo dei momenti, stando il fatto che $E\left[X\right] = 0$ per simmetria di $f_\alpha\left(x\right)$, vale però
$$E\left[X^2\right] = \int x^2 f_\alpha\left(x\right) dx = 2 \int_0^{\frac{1}{\sqrt{\alpha}}} \alpha x^3 dx = 2\alpha \frac{\left(\frac{1}{\sqrt{\alpha}}\right)^4}{4} = \frac{1}{2\alpha}$$
per cui
$$\frac{x_1^2 + ... + x_n^2}{n} \sim \frac{1}{2\alpha}$$
ed uno stimatore ragionevole può essere
$$\widehat{\alpha} = \frac{n}{2\left(X_1^2 + ... + X_n^2\right)}.$$