Statistica I - A.A. 2017-2018

Prova scritta - 6 giugno 2018

Exercise 1 *(pt 20, ∼ 3+4+3+4+3+4) Si osservano i valori dello spread in un certo periodo di tempo ed essi appaiono approssimativamente descrivibili con una v.a.* $S \sim N\left(185, 40^2\right)$.

1. Valori superiori a 250 mettono in allarme. Con che probabilità li osserveremo?

2. In realtà un singolo valore alto non ha effetti così importanti quanto la persistenza di valori alti. Supponendo per semplicità matematica (anche se questo ovviamente non è vero) che i valori giornalieri siano indipendenti, e che siano tutti distribuiti come una $N\left(185, 40^2\right)$, che probabilità c'è che la media aritmetica di 20 giorni consecutivi sia superiore a 220? Descrivere accuratamente la soluzione.

3. Dopo due mesi di altalena dello spread secondo la distribuzione $N(185, 40)$, si ha l'impressione ad occhio che ci sia un miglioramento, anche basandoci su fatti esterni come dichiarazioni positive di politici stranieri. Un peggioramento sembra da escludersi a priori. Si decide di osservare i dati per 14 giorni per capire se i valori sono in accordo coi precedenti oppure mostrano un miglioramento. Scrivere ipotesi e regione di rifiuto del test che si decide di mettere in atto, prendendo la significatività pari al 90%.

4. Nell'eseguire un tale test si potrebbe commettere un errore di seconda specie. Dire innanzi tutto di che errore stiamo parlando (dare la definizione). Calcolarne poi la probabilità, partendo dalla definizione data. Qui per semplicità di usino solo le gaussiane e si supponga nota ed ancora valida la deviazione standard iniziale, pari a 40. Si chiede naturalmente solo di arrivare ad una formula finale teorica (non ad un valore numerico), pur con tutti i valori numerici noti sostituiti.

5. Ed infine, trovata tale probabilità, vedere se 14 giorni di osservazione sono sufficienti a osservare una diminuzione da 185 a 160.

6. Calcolare il valore p del test precedente, dopo i 14 giorni di osservazione, avendo trovato media empirica pari a 165. Partire da una definizione a scelta di valore p, non da una formula prefatta, dichiarando prima a parole quale sia la definizione scelta; di nuovo, si usino solo le gaussiane. Su base intuitiva, senza svolgere ulteriori calcoli, decidere se il test del punto 3 avrebbe dato esito significativo.

Exercise 2 *(pt 10, ∼ 3+3+3) Sia* X *una v.a. di densità*

$$f_\theta(x) = \begin{cases} C_\theta e^{-\theta x} & \textit{per} \quad x \geq 5 \\ 0 & \textit{per} \quad x < 5 \end{cases}$$

dove il parametro θ *varia nell'insieme dei numeri reali positivi.*

1. Trovare la costante C_θ per cui la funzione $f_\theta(x)$ sia davvero una densità di probabilità.

2. Posto $Y = -X$, trovare la densità di probabilità di Y.

3. Trovare lo stimatore di massima verosimiglianza del parametro θ.

1 SOLUZIONI

Esercizio 1.

1)

$$P(S > 250) = P\left(\frac{S-185}{40} > \frac{250-185}{40}\right) = 1 - \Phi(1.625) = 1 - 0.95 = 0.05.$$

2) Indichiamo con $S_1, ..., S_{20}$ i valori dei 20 giorni e con $\overline{S} = \frac{S_1+...+S_{20}}{20}$ la loro media. Sappiamo che, essendo le S_i indipendenti e $N\left(185, 40^2\right)$, la v.a. $\overline{S}$ è una $N\left(185, \frac{40^2}{20}\right)$. Pertanto dobbiamo calcolare

$$\begin{aligned} P\left(\overline{S} > 220\right) &= P\left(\frac{\overline{S}-185}{40}\sqrt{20} > \frac{220-185}{40}\sqrt{20}\right) = 1 - \Phi(3.9131) \\ &= 1 - 0.99995 = 0.00005. \end{aligned}$$

3) $\mathcal{H}_0$) nuova media ≥ 185, $\mathcal{H}_1$) nuova media < 185. Usiamo il test t di Student perché la numerosità è bassa. Regione di rifiuto:

$$\left\{(s_1, ..., s_{14}) : \frac{\overline{s}-185}{\widehat{\sigma}}\sqrt{14} < t_{13,0.1}\right\}$$

dove $t_{13,0.1} = 1.35$ e $\widehat{\sigma}$ è la deviazione standard empirica.

4) Si tratta della possibilità di non rifiutare l'ipotesi nulla, quando questa è falsa, cioè quando la media vera è inferiore a 185. Non si rifiuta $\mathcal{H}_0$) quando $\frac{\overline{s}-185}{40}\sqrt{14} \geq t_{13,0.1}$, quindi, fissato un valore di riferimento $\mu < 185$, la probabilità di questo errore è

$$\begin{aligned} P^{\mu,40}\left(\frac{\overline{S}-185}{40}\sqrt{14} \geq q_{0.9}\right) &= P^{\mu,40}\left(\frac{\overline{S}-\mu}{40}\sqrt{14} + \frac{\mu-185}{40}\sqrt{14} \geq q_{0.9}\right) \\ &= 1 - \Phi\left(q_{0.9} - \frac{\mu-185}{40}\sqrt{14}\right) \end{aligned}$$

dove $q_{0.9} = 1.29$.

5)

$$\begin{aligned} P^{160,40}\left(\frac{\overline{S}-185}{40}\sqrt{14} \geq q_{0.9}\right) &= 1 - \Phi\left(1.29 - \frac{160-185}{40}\sqrt{14}\right) \\ &= 1 - \Phi(3.6285) = 1 - 0.99986 = 0.00014. \end{aligned}$$

La probabilità di non osservare un tale cambiamento è così bassa che il test deve funzionare sicuramente, se c'è stata una diminuzione di quel tipo.

6) Ad esempio, prendiamo come definizione la probabilità, sotto l'ipotesi nulla, che la grandezza statistica utilizzata per eseguire il test assuma valori più estremi di quello sperimentale; la grandezza statistica è $\overline{S}$, una $N\left(185, \frac{40^2}{14}\right)$, il suo valore sperimentale è 165, per cui la probabilità di valori più estremi (considerando che il test è unilaterale) è (l'esercizio, impostato così, è identico al punto 2)

$$\begin{aligned} P^{185,40}\left(\overline{S} < 165\right) &= P^{185,40}\left(\frac{\overline{S}-185}{40}\sqrt{20} < \frac{165-185}{40}\sqrt{20}\right) = \Phi(-2.2361) \\ &= 1 - \Phi(2.2361) = 1 - 0.98713 = 0.01287. \end{aligned}$$

Anche se abbiamo usato le gaussiane invece che le t di Student, vista l'enorme discrepanza tra 0.01287 e 0.1, riteniamo che sarebbe stato significativo anche il test del punto 3.

Esercizio 2.

1)

$$1 = \int_5^\infty C_\theta e^{-\theta x} dx = C_\theta \left[\frac{e^{-\theta x}}{-\theta} \right]_5^\infty = \frac{C_\theta}{\theta} e^{-5\theta}$$

da cui $C_\theta = \theta e^{5\theta}$.

2)

$$F_Y(\theta, y) = P(Y \leq y) = P(-X \leq y) = P(X \geq -y).$$

Questa è uno per $-y \leq 5$, ovvero $y \geq -5$; mentre per $-y \geq 5$, ovvero $y < -5$ abbiamo

$$= P(X \geq -y) = \int_{-y}^\infty \theta e^{5\theta} e^{-\theta x} dx = e^{5\theta} \left[-e^{-\theta x} \right]_{-y}^\infty = e^{5\theta} e^{\theta y} = e^{\theta(5+y)}.$$

Quindi la sua derivata è

$$f_Y(\theta, y) = \theta e^{\theta(5+y)}$$

per $y < -5$, nulla per $y > -5$.

3)

$$L(\theta, x_1, ..., x_n) = \theta^n e^{5n\theta} e^{-\theta(x_1 + ... + x_n)}$$

$$\log L(\theta, x_1, ..., x_n) = n \log \theta + 5n\theta - \theta(x_1 + ... + x_n)$$

$$\begin{aligned} \frac{n}{\theta} + 5n - (x_1 + ... + x_n) &= 0 \\ \frac{n}{\theta} &= x_1 + ... + x_n - 5n \end{aligned}$$

$$\widehat{\theta} = \frac{n}{x_1 + ... + x_n - 5n}.$$