Statistica I - A.A. 2017-2018

Prova scritta - 12 settembre 2018

Exercise 1 *(pt 20, $\sim$ 6+3+4+3+4) Un'azienda utilizza un certo materiale grezzo il cui costo per kilogrammo è, in euro, una v.a. $C \sim N\left(1.6, 0.2^2\right)$; tale costo cambia di giorno in giorno, perciò viene considerato aleatorio.*

1. Con che probabilità il costo giornaliero supera 1.9 euro? Ed invece, con che probabilità la media aritmetica dei costi di 30 giorni è superiore a 1.7?
2. Dopo due mesi si ha l'impressione ad occhio che il costo, finalmente, inizi a calare. Si decide di osservare i dati per 10 giorni per capire se i valori sono in accordo coi precedenti oppure mostrano un calo. Scrivere ipotesi e regione di rifiuto del test che si decide di mettere in atto, prendendo la significatività pari al 90%.
3. 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 0.2. Si chiede naturalmente solo di arrivare ad una formula finale teorica (non ad un valore numerico), pur con tutti i valori numerici noti sostituiti.
4. Ed infine, trovata tale probabilità, vedere se 10 giorni di osservazione sono sufficienti a osservare una diminuzione da 1.6 a 1.4.
5. Calcolare il valore p del test precedente, dopo i 10 giorni di osservazione, avendo trovato media empirica pari a 1.5. 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, $\sim$ 4+3+3) Sia X una v.a. di densità*

$$f_\mu(x) = \begin{cases} C_\mu e^{\frac{x}{\mu}} & \textit{per} \quad x \leq 1 \\ 0 & \textit{per} \quad x > 1 \end{cases}$$

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

1. Trovare la costante C_μ per cui la funzione $f_\mu(x)$ sia davvero una densità di probabilità.
2. Trovare per quale valore del parametro μ la probabilità che X superi il valore -1 vale 0.5.
3. Trovare lo stimatore di massima verosimiglianza del parametro μ.

1 SOLUZIONI

Esercizio 1.

1)

$$P(C > 1.9) = P\left(\frac{C-1.6}{0.2} > \frac{1.9-1.6}{0.2}\right) = 1 - \Phi(1.5) = 1 - 0.93319 = 0.06681.$$

Indichiamo poi con $C_1, ..., C_{30}$ i valori dei 30 giorni e con $\overline{C} = \frac{C_1+...+C_{30}}{30}$ la loro media. Decidiamo che le C_i siano indipendenti e $N\left(1.6, 0.2^2\right)$; allora la v.a. $\overline{C}$ è una $N\left(1.6, \frac{0.2^2}{30}\right)$. Pertanto dobbiamo calcolare

$$\begin{aligned} P\left(\overline{C} > 1.7\right) &= P\left(\frac{\overline{C}-1.6}{0.2}\sqrt{30} > \frac{1.7-1.6}{0.2}\sqrt{30}\right) = 1 - \Phi(2.7386) \\ &= 1 - 0.99693 = 0.00307. \end{aligned}$$

2) $\mathcal{H}_0$) nuova media ≥ 1.6, $\mathcal{H}_1$) nuova media < 1.6. Usiamo il test t di Student perché la numerosità è bassa. Regione di rifiuto:

$$\left\{(s_1, ..., s_{10}) : \frac{\overline{s}-1.6}{\widehat{\sigma}}\sqrt{10} < -t_{9,0.1}\right\}$$

dove $t_{9,0.1} = 1.383$ e $\widehat{\sigma}$ è la deviazione standard empirica.

3) Si tratta della possibilità di non rifiutare l'ipotesi nulla, quando questa è falsa, cioè quando la media vera è inferiore a 1.6. Non si rifiuta $\mathcal{H}_0$) quando $\frac{\overline{s}-1.6}{\widehat{\sigma}}\sqrt{10} \geq q_{0.9}$, quindi, fissato un valore di riferimento $\mu < 1.6$, la probabilità di questo errore è

$$\begin{aligned} P^{\mu,0.2}\left(\frac{\overline{C}-1.6}{0.2}\sqrt{10} \geq -q_{0.9}\right) &= P^{\mu,40}\left(\frac{\overline{C}-\mu}{0.2}\sqrt{10} + \frac{\mu-1.6}{0.2}\sqrt{10} \geq -q_{0.9}\right) \\ &= 1 - \Phi\left(-q_{0.9} - \frac{\mu-1.6}{0.2}\sqrt{10}\right) \end{aligned}$$

dove $q_{0.9} = 1.29$.

4)

$$\begin{aligned} P^{1.4,0.2}\left(\frac{\overline{C}-1.6}{0.2}\sqrt{10} \geq -q_{0.9}\right) &= 1 - \Phi\left(-1.29 - \frac{1.4-1.6}{0.2}\sqrt{10}\right) \\ &= 1 - \Phi(1.8723) \sim 0.031. \end{aligned}$$

5) 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 $\overline{C}$, una $N\left(1.6, \frac{0.2^2}{10}\right)$, il suo valore sperimentale è 1.5, per cui la probabilità di valori più estremi (considerando che il test è unilaterale) è (l'esercizio, impostato così, è identico al punto 2)

$$\begin{aligned} P^{1.6,0.2}\left(\overline{C} < 1.5\right) &= P^{1.6,0.2}\left(\frac{\overline{C}-1.6}{0.2}\sqrt{10} < \frac{1.5-1.6}{0.2}\sqrt{10}\right) = \Phi(-1.5811) \\ &= 1 - \Phi(1.5811) = 1 - 0.94295 = 0.05705. \end{aligned}$$

E' un valore minore di 0.1, quindi il test sarebbe significativo: anche se va tenuto in conto che abbiamo usato le gaussiane invece che le t di Student, con una numerosità davvero bassa.

Esercizio 2.

1)

$$1=\int_{-\infty}^{1} C_\mu e^{\frac{x}{\mu}}\,dx = C_\mu \left[\mu e^{\frac{x}{\mu}}\right]_{-\infty}^{1} = \mu C_\mu e^{\frac{1}{\mu}}$$

da cui $C_\mu = \frac{1}{\mu e^{\frac{1}{\mu}}}$.

2)

$$\begin{aligned} 0.5 &= P\left(X>-1\right) = \int_{-1}^{1} C_\mu e^{\frac{x}{\mu}}\,dx = C_\mu \left[\mu e^{\frac{x}{\mu}}\right]_{-1}^{1} = \mu C_\mu \left(e^{\frac{1}{\mu}} - e^{-\frac{1}{\mu}}\right) \\ &= e^{-\frac{1}{\mu}}\left(e^{\frac{1}{\mu}} - e^{-\frac{1}{\mu}}\right) = 1 - e^{-\frac{2}{\mu}} \end{aligned}$$

da cui

$$\begin{aligned} e^{-\frac{2}{\mu}} &= 0.5 \\ \frac{2}{\mu} &= -\log 0.5 \\ \mu &= \frac{2}{-\log 0.5} = 2.8854. \end{aligned}$$

3)

$$L\left(\mu, x_1, ..., x_n\right) = \mu^{-n} e^{-\frac{n}{\mu}} e^{\frac{x_1+...+x_n}{\mu}}$$

$$\log L\left(\mu, x_1, ..., x_n\right) = -n\log\mu - \frac{n}{\mu} + \frac{x_1+...+x_n}{\mu}$$

$$\begin{aligned} -\frac{n}{\mu} + \frac{n}{\mu^2} - \frac{x_1+...+x_n}{\mu^2} &= 0 \\ -n\mu + n - \left(x_1+...+x_n\right) &= 0 \end{aligned}$$

$$\widehat{\mu} = 1 - \frac{x_1+...+x_n}{n}.$$