Statistica I - A.A. 2017-2018

Prova scritta - 10 gennaio 2018

Exercise 1 *(pt 20, $\sim$ 4+4+4+4+4) Un'azienda farmaceutica produce da tempo un prodotto di successo. Il suo tempo di guarigione, misurato in giorni, può essere ben rappresentato da una v.a. $T \sim N(7,4)$.*

1. Con che probabilità un generico paziente guarisce in un tempo compreso tra 5 e 9 giorni? Raffigurare il calcolo con un disegno.

2. Il farmaco è sempre stato utilizzato per una malattia detta di tipo A. Di recente si è cominciato a somministrarlo anche per la malattia detta di tipo B. Non è chiaro se, per quest'ultima, il tempo medio di guarigione sia lo stesso di quello della malattia di tipo A oppure no. Innanzi tutto, lo si vuole stimare al 95% con precisione pari ad 1 giorno. Quanti pazienti bisogna osservare? Si supponga in questo momento che la varianza sia la stessa, e nota, per le due malattie.

3. L'ipotesi che il tempo medio di guarigione sia lo stesso viene sottoposta a giudizio. Possiamo rifiutarla, al 95%, se un gruppo di 20 pazienti ha mostrato un tempo medio empirico di guarigione pari a 6.5 giorni, con deviazione standard 3? Preliminarmente, scrivere le ipotesi e la regione di rifiuto.

4. Quali cambiamenti della media sarebbero stati rilevati dalla versione gaussiana a varianza nota (pari a 4) del test precedente, con probabilità 0.9?

5. In realtà, osservando un po' ad occhio l'effetto del farmaco in alcuni casi, il dubbio è che sia cambiata molto la varianza, piuttosto che la media; precisamente, il dubbio è che la varianza per i malati di tipo B sia maggiore che per quelli di tipo A. Scrivere le ipotesi e la regione di rifiuto, al 95%, per un campione di 20 pazienti, partendo dal fatto teorico che la v.a. $(n-1)\frac{S^2}{\sigma^2}$ è una chi quadro a $n-1$ gradi di libertà. Si suggerisce il fatto che la regione di rifiuto deve avere la forma $\left\{... : S^2 > ...\right\}$. Nel risultato, lasciare indicato il quantile chi quadro, ricordando che indichiamo con $\chi^2_{n,\alpha}$ l'analogo di $t_{n,\alpha}$, relativo alla distribuzione chi quadro.

Exercise 2 *(pt 10, $\sim$ 3+3.5+3.5) Sia X una v.a. di densità*

$$f(\alpha, x) = \begin{cases} C_\alpha e^{\alpha x} & per \quad x \leq 0 \\ 0 & per \quad x > 0 \end{cases}$$

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

1. Trovare la costante C_α per cui la funzione $f(\alpha, x)$ sia davvero una densità di probabilità.

2. Posto $Y = -X$, trovare la funzione di ripartizione $F_Y(\alpha, y)$ e riconoscere che densità di probabilità $f_Y(\alpha, y)$ ha la v.a. Y.

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

1 SOLUZIONI

Esercizio 1.

1)

$$P(5 \leq T \leq 9) = P\left(-1 \leq \frac{T-7}{2} \leq 1\right) = \Phi(1) - \Phi(-1) = 0.68$$

2) La precisione della stima, al 95%, è data da

$$\delta = \frac{2 \cdot q_{0.975}}{\sqrt{n}} = \frac{2 \cdot 1.96}{\sqrt{n}} = \frac{3.92}{\sqrt{n}}$$

dove n è la numerosità campionaria. Volendo $\delta \leq 1$ abbiamo $\frac{3.92}{\sqrt{n}} \leq 1$, ovvero

$$\begin{aligned} \sqrt{n} &\geq 3.92 \\ n &\geq 15.366 \end{aligned}$$

quindi bisogna osservare 16 pazienti.

3) $\mathcal{H}_0$) nuova media =7, $\mathcal{H}_1$) nuova media ≠7. Regione di rifiuto:

$$\left\{(t_1, ..., t_{20}) : \left|\frac{\bar{t}-7}{3}\sqrt{20}\right| > t_{19,0.025}\right\}$$

dove $t_{19,0.05} = 2.093$. Il valore empirico è $\bar{t} = 6.5$. Quindi

$$\left|\frac{\bar{t}-7}{3}\sqrt{20}\right| = |-0.74536| = 0.74536$$

che non supera 2.093. Quindi il test non è significativo, non possiamo affermare che la media delle due malattie sia diversa.

4) Nella versione gaussiana, la regione di rifiuto è

$$\left\{(t_1, ..., t_{20}) : \left|\frac{\bar{t}-7}{2}\sqrt{20}\right| > q_{0.975}\right\}.$$

Dobbiamo trovare i valori di μ tali che

$$P^{\mu,2}\left(\left|\frac{\overline{T}-7}{2}\sqrt{20}\right| > q_{0.975}\right) \geq 0.9$$

ovvero

$$P^{\mu,2}\left(-1.96 \leq \frac{\overline{T}-7}{2}\sqrt{20} \leq 1.96\right) \leq 0.1.$$

Siccome $\overline{T} \sim N\left(\mu, \frac{4}{20}\right)$, riscriviamo l'identità precedente come

$$P\left(-1.96 \leq Z + \frac{\mu-7}{2}\sqrt{20} \leq 1.96\right) \leq 0.1$$

$$P\left(-1.96-\frac{\mu-7}{2}\sqrt{20}\leq Z\leq 1.96-\frac{\mu-7}{2}\sqrt{20}\right)\leq 0.1$$

dove $Z\sim N(0,1)$, quindi

$$\Phi\left(1.96-\frac{\mu-7}{2}\sqrt{20}\right)-\Phi\left(-1.96-\frac{\mu-7}{2}\sqrt{20}\right)\leq 0.1.$$

Una delle due probabilità è "estrema" ma questo dipende da μ. Ad esempio, se $\mu>7$, $\Phi\left(-1.96-\frac{\mu-7}{2}\sqrt{20}\right)$ è trascurabile e dobbiamo risolvere la disequazione (approssimata)

$$\Phi\left(1.96-\frac{\mu-7}{2}\sqrt{20}\right)\leq 0.1$$

$$\begin{aligned}1.96-\frac{\mu-7}{2}\sqrt{20} &= q_{0.1}\leq -1.29\\ \frac{\mu-7}{2}\sqrt{20} &= 1.96+1.29\geq 3.25\\ \mu &\geq 7+\frac{2\cdot 3.25}{\sqrt{20}}=8.453\,4.\end{aligned}$$

Questi sono i valori maggiori di 7 rilevabili col test. Simmetricamente si trovano quelli minori di 7.

5) $\mathcal{H}_0$) nuova varianza =2, $\mathcal{H}_1$) nuova varianza >2. Se vale $\mathcal{H}_0$), $(20-1)\,\frac{S^2}{4}$ è una chi quadro a 19 gradi di libertà, quindi supera il quantile $\chi^2_{19,0.05}$ con probabilità 0.05. Pertanto la regione di rifiuto è

$$\left\{(t_1,...,t_{20}):S^2>\frac{\chi^2_{19,0.05}\cdot 4}{19}\right\}.$$

Esercizio 2.

1)

$$1=\int_{-\infty}^{0}C_\alpha e^{\alpha x}dx=C_\alpha\left[\frac{e^{\alpha x}}{\alpha}\right]_{-\infty}^{0}=\frac{C_\alpha}{\alpha}$$

da cui $C_\alpha=\alpha$.

2)

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

Questa è zero per $y<0$, mentre per $y\geq 0$

$$=P(X\geq -y)=\int_{-y}^{0}\alpha e^{\alpha x}dx=\left[e^{\alpha x}\right]_{-y}^{0}=1-e^{-\alpha y}.$$

Quindi la sua derivata è

$$f_Y(\alpha,y)=\alpha e^{-\alpha y}$$

per $y\geq 0$, nulla per $y<0$. La densità $f_Y(\alpha,y)$ è esponenziale di parametro α.

3)

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

$$\log L(\alpha, x_1, ..., x_n) = n \log \alpha + \alpha (x_1 + ... + x_n)$$

$$\frac{n}{\alpha} + (x_1 + ... + x_n) = 0$$

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