Statistica I - A.A. 2016-2017

Prova scritta - 10 giugno 2017

Problema 1. (pt. 20) Nel reparto produzione, vi viene affidato il compito di controllare saltuariamente la qualità, non attraverso carte di controllo ma tramite singoli controlli su richiesta (ad esempio in seguito a segnalazioni da parte dell'ufficio vendite). I pezzi prodotti, per poter essere assemblati correttamente, devono avere un certo diametro compreso tra 4.8 e 5.2 cm. Supponiamo che il diametro dei pezzi prodotti, sempre un po' diverso da un esemplare ad un altro, si possa considerare una v.a. gaussiana, chiamata D nel seguito. Sottolineiamo che eventuali peggioramenti della produzione possono manifestarsi equivalentemente tramite un aumento o una diminuzione del diametro; non c'è a priori una direzione unilaterale del concetto di peggioramento.

1. Supponiamo che l'attuale media sia pari a 5 cm con deviazione standard pari ad 1 mm. Un generico pezzo, che probabilità ha di rispettare gli standard?

2. Calcolare la probabilità che, in un lotto di 100 pezzi, ce ne siano almeno 10 difettosi.

3. L'ufficio vendite segnala, da parte degli acquirenti, il fatto che troppo frequentemente i pezzi devono essere scartati in fase di assemblaggio. Supponiamo che in prima battuta si ritenga inalterata la deviazione standard e che il problema vada attribuito ad una variazione non voluta del diametro medio. Si prendono 100 pezzi e si calcola il loro diametro medio, trovandolo pari a 4.965. Si può ritenere al 95% che si tratti di una modifica della qualità di produzione e non solo di una fluttuazione casuale di quel campione particolare? Scrivere accuratamente le ipotesi e la regione di rifiuto.

4. Calcolare il valore p del test precedente, partendo da una definizione di valore p e svolgendo i calcoli, e non utilizzando una formuletta mnemonica.

5. Provare anche a calcolare la potenza del test precedente, partendo anche in questo caso dalla definizione. Trovare il valore numerico nel caso in cui la media vera fosse 5.03.

Problema 2. (pt. 10) Sia X una v.a. di densità $f_\theta(x) = C_\theta \exp\left(-\frac{|x+1|}{\theta}\right)$, 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 $Y = |X+1|$, trovare la funzione di ripartizione $F_Y(y)$.

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

1 Soluzioni

Esercizio 1.

1.

$$P(4.8 \le D \le 5.2) = P\left(-\frac{0.2}{0.1} \le \frac{D-5}{0.1} \le \frac{0.2}{0.1}\right) = P(-2 \le Z \le 2)$$
$$= P(Z \le 2) - P(Z \le -2) = \Phi(2) - \Phi(-2)$$
$$= 2\Phi(2) - 1 = 2 \cdot 0.97725 - 1 = 0.9545.$$

2. Introdiciamo le v.a. $X_1, ..., X_{100}$ dove $X_i = 1$ se il pezzo i-esimo è difettoso, $X_i = 0$ altrimenti. Vale $P(X_i = 1) = 1 - 0.9545 = 0.0455$. Il numero di pezzi difettosi è $X_1 + ... + X_{100}$, quindi dobbiamo calcolare

$$P(X_1 + ... + X_{100} \ge 10) = 1 - P\left(\frac{X_1 + ... + X_{100} - 100 \cdot 0.0455}{\sqrt{100 \cdot 0.0455 \cdot 0.9545}} < \frac{10 - 100 \cdot 0.0455}{\sqrt{100 \cdot 0.0455 \cdot 0.9545}}\right)$$

che per il TLC si può approssimare a

$$= 1 - \Phi\left(\frac{10 - 100 \cdot 0.0455}{\sqrt{100 \cdot 0.0455 \cdot 0.9545}}\right) = 1 - \Phi(2.615\,2)$$
$$= 1 - 0.99547 = 0.00453.$$

3. Vista la buona numerosità campionaria e l'ipotesi di invarianza della varianza, eseguiamo un test gaussiano per la media. Eseguiamo un test bilaterale, per le ragioni illustrate nel testo. L'ipotesi nulla è $\mathcal{H}_0$) media $= 5$; quella alternativa è $\mathcal{H}_1$) media $\neq 5$; la regione di rifiuto è $\left\{(x_1, ..., x_{80}) : \left|\frac{\overline{x}_{100}-5}{0.1}\sqrt{100}\right| > q_{0.975}\right\}$. Nel nostro caso vale

$$\frac{\overline{x}_{100} - 5}{0.1}\sqrt{100} = \frac{4.965 - 5}{0.1}\sqrt{100} = -3.5$$

che è, in valore assoluto, di gran lunga maggiore di $q_{0.975} = 1.96$.

4. Il valore p è dato dalla probabilità che la grandezza statistica superi il suo valore sperimentale, che nel caso bilaterale va interpretato nella forma

$$P^{5,0.1}\left(\left|\frac{\overline{X}_{100} - 5}{0.1}\sqrt{100}\right| > 3.5\right).$$

La grandezza $\frac{\overline{X}_{100}-5}{0.1}\sqrt{100}$ è una $N(0,1)$, quindi dobbiamo calcolare

$$P(|Z| > 3.5) = 2P(Z > 3.5) = 2(1 - \Phi(3.5)) = 2(1 - 0.99977) = 0.00046.$$

5. La potenza è la probabilità di rifiutare quando si deve, ovvero la funzione
$$P^{\mu,0.1}\left(\left|\frac{\overline{X}_{100}-5}{0.1}\sqrt{100}\right|>q_{0.975}\right)$$
al variare di $\mu \neq 5$. Essa vale
$$\begin{aligned}
&= P^{\mu,0.1}\left(\left|\frac{\overline{X}_{100}-\mu}{0.1}\sqrt{100}+\frac{\mu-5}{0.1}\sqrt{100}\right|>q_{0.975}\right)\\
&= P\left(\left|Z+\frac{\mu-5}{0.1}\sqrt{100}\right|>q_{0.975}\right)\\
&= P\left(Z+\frac{\mu-5}{0.1}\sqrt{100}>q_{0.975}\right)+P\left(Z+\frac{\mu-5}{0.1}\sqrt{100}<-q_{0.975}\right)\\
&= P\left(Z>q_{0.975}-\frac{\mu-5}{0.1}\sqrt{100}\right)+P\left(Z<-q_{0.975}-\frac{\mu-5}{0.1}\sqrt{100}\right)\\
&= 1-\Phi\left(1.96-\frac{\mu-5}{0.1}\sqrt{100}\right)+\Phi\left(-1.96-\frac{\mu-5}{0.1}\sqrt{100}\right).
\end{aligned}$$
Nel caso particolare vale
$$\begin{aligned}
&1-\Phi\left(1.96-\frac{5.03-5}{0.1}\sqrt{100}\right)+\Phi\left(-1.96-\frac{5.03-5}{0.1}\sqrt{100}\right)\\
&= 1-\Phi\left(-1.04\right)+\Phi\left(-4.96\right)\\
&= \Phi\left(1.04\right)+\Phi\left(-4.96\right)\sim\Phi\left(1.04\right)=0.85083.
\end{aligned}$$

Esercizio 2.

1.
$$\begin{aligned}
1 &= \int_{-\infty}^{\infty} C_\theta \exp\left(-\frac{|x+1|}{\theta}\right)dx = 2\int_{-1}^{\infty} C_\theta \exp\left(-\frac{x+1}{\theta}\right)dx\\
&= -2C_\theta\left[\theta\exp\left(-\frac{x+1}{\theta}\right)\right]_{-1}^{\infty} = 2C_\theta\theta\\
C_\theta &= \frac{1}{2\theta}.
\end{aligned}$$

2.
$$F_Y(y)=P(Y\leq y)=P(|X+1|\leq y)$$
è zero per $y<0$, mentre per $y\geq 0$
$$= P(-y\leq X+1\leq y)=P(-y-1\leq X\leq y-1)=\int_{-y-1}^{y-1}\frac{1}{2\theta}\exp\left(-\frac{|x+1|}{\theta}\right)dx$$
$$=\frac{1}{\theta}\int_{-1}^{y-1}\exp\left(-\frac{x+1}{\theta}\right)dx=-\left[\exp\left(-\frac{x+1}{\theta}\right)\right]_{-1}^{y-1}=1-\exp\left(-\frac{y}{\theta}\right).$$
[La densità è esponenziale di parametro $\frac{1}{\theta}$.]

3.
$$L\left(\theta, x_1, ..., x_n\right) = \frac{1}{2^n \theta^n} \exp\left(-\frac{|x_1+1| + ... + |x_n+1|}{\theta}\right)$$
$$\log L\left(\theta, x_1, ..., x_n\right) = -n \log 2 - n \log \theta - \frac{|x_1+1| + ... + |x_n+1|}{\theta}$$

da cui derivando rispetto a θ

$$-\frac{n}{\theta} + \frac{|x_1+1| + ... + |x_n+1|}{\theta^2} = 0$$
$$\widehat{\theta} = \frac{|x_1+1| + ... + |x_n+1|}{n}.$$