Statistica I - A.A. 2016-2017

Prova scritta - 19 aprile 2017

Problema 1. (pt. 20) Un'azienda che produce ricambi per stampanti esamina la durata di un certo modello di cartuccia d'inchiostro, misurata in numero di copie aventi un livello accettabile di qualità; tale numero è aleatorio per svariate ragioni, che vanno dalla qualità della cartuccia alla densità di parti scure nelle copie, alla qualità della stampante e così via. Supponiamo che la durata, considerata come una variabile aleatoria D, sia gaussiana.

Supponiamo che il modello di cartuccia che si sta esaminando sia in commercio da tempo e si abbia un'ottima stima dei suoi parametri, che possono quindi essere considerati come veri: media pari a 530 copie e deviazione standard pari a 64. Si tratta dei parametri stimati mescolando tutte e possibili situazioni di utilizzo.

1. Un acquirente a cui vengano comunicati tali parametri si chiede: con che probabilità stamperò bene almeno 500 copie? E quante copie sono sicuro di stampare, con un livello di fiducia del 95%?

2. Un aquirente che svolge un ampio lavoro con la stampante compra 30 cartucce e si chiede: con che probabilità stamperò bene almeno 15500 copie?

3. Come dicevamo, l'incertezza nel numero di copie dipende anche dal tipo di stampe: se contengono immagini, che spesso hanno ampie zone scure, il consumo è maggiore. L'azienda svolge uno studio in questa direzione: individua 20 acquirenti che devono continuamente stampare testi contenenti immagini ed esamina se essi abbiano parametri diversi da quelli generali. Attende che ciascuno di essi esaurisca una cartuccia e registra le 20 durate: esse hanno avuto media empirica pari a 470, con deviazione empirica pari a 87. Al 95%, può dire che tali cartucce durano meno di quelle usate in modo del tutto generico?

4. L'azienda prosegue lo studio del punto precedente volendo arrivare ad una dichiarazione circa la durata media delle cartucce soggette a quel tipo d'uso particolare. Sulla base dei dati del punto 3, con quale precisione relativa può dichiarare la media vera, a livello di confidenza 90%?

5. Si scordino i punti 3 e 4 si torni alla situazione generale. Anche la qualità della stampante incide sui parametri di D, in una qualche misura. Un certo modello di stampante è sotto accusa perché sembra consumare troppo in fretta le cartucce. Si vuole svolgere un test per giudicare se la durata media con tale modello sia inferiore, supponendo che la deviazione standard non sia cambiata. Si scriva l'ipotesi nulla, quella alternativa e la regione di rifiuto al 95% nel caso che si possano esaminare 80 stampanti di quel tipo (una

cartuccia per ciascuna stampante). Se la vera durata media delle cartucce per tale modello fosse pari a 490, con che probabilità se ne accorgerebbe questo test (la risposta a questa domanda verrà valutata solo se impostata e sviluppata in modo letterale, non utilizzando formule prefatte)?

Problema 2. (pt. 10) Sia X una v.a. di densità $f_a(x) = C_a \exp(-a|x-1|)$, dove il parametro a varia nell'insieme dei numeri reali positivi.

1. Trovare la costante C_a per cui la funzione $f_a(x)$ 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 a e confrontarlo con quello dei momenti.

1 Soluzioni

Esercizio 1.

1. $$P(D \geq 500) = 1 - P(D < 500) = 1 - P\left(\frac{D-530}{64} < \frac{500-530}{64}\right)$$
$$= 1 - \Phi\left(\frac{500-530}{64}\right) = \Phi\left(\frac{530-500}{64}\right) = \Phi(0.468) = 0.68$$

$$0.95 = P(D \geq n) = \Phi\left(\frac{530-n}{64}\right)$$
$$\frac{530-n}{64} = q_{0.95}$$
$$n = 530 - q_{0.95} \cdot 64 = 530 - 1.64 \cdot 64 = 425.04$$

quindi circa 425 copie.

2. Il numero totale di copie fatte è $D_1 + ... + D_{30}$ che, in ipotesi di indipendenza, è una $N(30 \cdot 530, 30 \cdot 64^2)$. Pertanto

$$P(D_1 + ... + D_{30} \geq 15500) = 1 - P\left(\frac{D_1 + ... + D_{30} - 30 \cdot 530}{64\sqrt{30}} < \frac{15500 - 30 \cdot 530}{64\sqrt{30}}\right)$$
$$= 1 - \Phi\left(\frac{15500 - 30 \cdot 530}{64\sqrt{30}}\right) = \Phi\left(-\frac{15500 - 30 \cdot 530}{64\sqrt{30}}\right)$$
$$= \Phi(1.141) = 0.873.$$

Si noti che, a stretto rigore, non abbiamo applicato il TLC, perché le v.a. in gioco erano già gaussiane, quindi non è stata necessaria alcuna approssimazione.

3. Applichiamo un test t unilaterale per la media (scegliamo il test t per via della bassa numerosità campionaria; useremo quindi la deviazione campionaria del piccolo campione). L'ipotesi nulla è $\mathcal{H}_0$) media ≥ 530; $\mathcal{H}_1$) media < 530; la regione di rifiuto è $\left\{(x_1, ..., x_{80}) : \frac{\overline{x}_{80} - 530}{87}\sqrt{20} < -t_{n-1,0.05}\right\}$. Vale

$$\frac{\overline{x} - 530}{s}\sqrt{n} = \frac{470 - 530}{87}\sqrt{20} = -3.0842$$

mentre

$$-t_{n-1,0.05} = -t_{19,0.05} = -1.729$$

quindi il test è significativo, quel tipo di stampe consuma molto più in fretta le cartucce.

4. Dobbiamo dichiarare l'intervallo di confidenza (o meglio la sua semi-ampiezza) trovato usando la distribuzione t di Student. Vale

$$\delta = \frac{s \cdot t_{19,0.05}}{470\sqrt{20}} = \frac{87 \cdot 1.729}{470\sqrt{20}} = 0.0715$$

(si noti che l'intervallo è bilaterale, al 90%, quindi $\alpha = 0.05$) quindi la precisione relativa è circa del 7 per cento.

5. Utilizziamo un test gaussiano unilaterale per la media, con deviazione standard nota: $\mathcal{H}_0$) media ≥ 530; $\mathcal{H}_1$) media < 530; la regione di rifiuto è $\left\{(x_1, ..., x_{80}) : \frac{\overline{x}_{80}-530}{64}\sqrt{80} < q_{0.05}\right\}$. Ci viene chiesto di calcolare

$$P^{(490,64)}\left(\frac{\overline{X}_{80} - 530}{64}\sqrt{80} < q_{0.05}\right).$$

La v.a. $\overline{X}_{80}$ è una $N\left(490, \frac{64^2}{80}\right)$, quindi

$$\begin{aligned} &= P^{(490,64)}\left(\frac{\overline{X}_{80} - 490}{64}\sqrt{80} < -q_{0.95} + \frac{530-490}{64}\sqrt{80}\right) \\ &= \Phi\left(-1.64 + \frac{530-490}{64}\sqrt{80}\right) = \Phi(3.95) = 0.99996. \end{aligned}$$

Quindi la probabilità di accorgersene è molto alta.

Esercizio 2.

1.

$$\begin{aligned} 1 &= \int_{-\infty}^{\infty} C_a \exp(-a|x-1|)\, dx = 2\int_1^{\infty} C_a \exp(-a(x-1))\, dx \\ &= -2C_a\left[\frac{1}{a}\exp(-a(x-1))\right]_1^{\infty} = 2C_a\frac{1}{a} \\ C_a &= \frac{a}{2}. \end{aligned}$$

2.

$$F_Y(y) = P(Y \leq y) = P(|X-1| \leq y)$$

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

$$= P(-y \leq X - 1 \leq y) = P(-y+1 \leq X \leq y+1) = \int_{-y+1}^{y+1} \frac{a}{2}\exp(-a|x-1|)\, dx$$

$$= a\int_1^{y+1} \exp(-a(x-1))\, dx = -\left[\exp(-a(x-1))\right]_1^{y+1} = 1 - \exp(-ay)$$

[La densità è esponenziale di parametro a.]

3.
$$L(a, x_1, ..., x_n) = \frac{a^n}{2^n} \exp\left(-a\left(|x_1 - 1| + ... + |x_n - 1|\right)\right)$$
$$\log L(a, x_1, ..., x_n) = n \log 2 + n \log a - a\left(|x_1 - 1| + ... + |x_n - 1|\right)$$
$$\frac{n}{a} - \left(|x_1 - 1| + ... + |x_n - 1|\right) = 0$$
$$\frac{n}{a} = |x_1 - 1| + ... + |x_n - 1|$$

quindi infine

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

Vediamo infine il metodo dei momenti. Vale

$$E[X] = 1.$$

Il calcolo facile è quello della varianza. Infatti essa è $E\left[|X-1|^2\right]$ ma questa espressione non è altro che il momento secondo di Y, che abbiamo già visto essere esponenziale di parametro a. Pertanto

$$E\left[|X-1|^2\right] = E\left[Y^2\right] = Var[Y] + E[Y]^2 = \frac{1}{a^2} + \frac{1}{a^2} = \frac{2}{a^2}.$$

Troviamo in conclusione

$$a = \sqrt{\frac{2}{E\left[|X-1|^2\right]}}$$

da cui emerge lo stimatore

$$\widehat{a} = \sqrt{\frac{2n}{\sum_{i=1}^{n} |x_i - 1|^2}}.$$