Statistica I - A.A. 2016-2017

Prova scritta - 19 luglio 2017

Problema 1. (pt 21) Un laboratorio ha il compito di testare la robustezza dei pezzi prodotti dal reparto produzione. Sottopone i pezzi a sollecitazioni periodiche, che chiameremo oscillazioni. Sia N il numero di oscillazioni a cui resiste il pezzo prima di rompersi; tale numero è aleatorio, generalmente molto elevato; essendo molto grande si usa farne il logaritmo, cioè calcolare $X = \log_{10} N$. Da osservazioni passate emerge che X ha distribuzione simile ad una gaussiana, per cui verrà considerata una $N(\mu_0, \sigma_0^2)$.

1. Si vuole stimare μ_0. Vengono sottoposti a oscillazioni 15 pezzi fino a portarli a rottura, registrando i corrispondenti valori $x_1, ..., x_{15}$, aventi $\overline{x} = 5.3$, $s = 1.7$. Scrivere un intervallo di fiducia per μ_0 a livello di confidenza 0.9.

2. Se volessimo, sempre a livello di confidenza 0.9, un errore assoluto non superiore a 0.6, risparmiando però il più possibile sul numero di pezzi da esaminare, quanti pezzi dovremmo rompere? Si osservi che la risposta a questa domanda può essere piuttosto articolata (c'è il problema dei gradi di libertà...).

3. Se prendiamo come veri i valori del punto 1, di fronte ad un nuovo pezzo quante oscillazioni ci aspettiamo che resista, salvo il 10% dei casi peggiori? Si vuole il valore soglia di N, non di X; lo si indichi con n_0.

4. Un anno più tardi, un certo numero di contestazioni da parte dei clienti inducono l'azienda ad eseguire un controllo, per capire se, rispetto ai valori del punto 1 (presi come veri), i pezzi ora si rompono prima. Nessuno dubita che durino di più, l'unica possibilità ragionevole è che si rompano prima oppure che siano invariati. Vengono allora portati a rottura 10 pezzi trovando i nuovi valori $\overline{x}_1 = 4.7$, $s = 1.6$. Possiamo affermare, al 90%, che l'apparente diminuzione è dovuta ad un peggioramento della produzione e non è una semplice fluttuazione casuale? Dichiarare con chiarezza le ipotesi e la regione di rifiuto.

5. Il test precedente non è risultato significativo. Se la vera nuova media fosse davvero calata al valore 4.5 (accettando deviazione 1.6), con che probabilità un campione cadrebbe nella regione di rifiuto?

Problema 2. (pt 9)

1. Si consideri la situazione descritta nell'esercizio precedente. Coi valori $\mu_0 = 5$, $\sigma_0^2 = 1$, calcolare $P(N > 15849)$.

2. Trovare il valore delle costanti $C, D > 0$ che rendono una densità di probabilità continua la funzione che vale $f(x) = \frac{C}{(x+2)^2}$ per $x > -1$, $f(x) = \frac{D}{x^4}$ per $x \leq -1$. Tracciarne il grafico.

3. Se Z è una v.a. con densità f data dal punto precedente, esiste finito $E[Z]$?

1 Soluzioni

Esercizio 1.

1. $\mu_0 = 5.3 \pm \frac{1.7 \cdot t_{14,0.05}}{\sqrt{15}} = 5.3 \pm \frac{1.7 \cdot 1.761}{\sqrt{15}} = 5.3 \pm 0.773.$

2. Si può agire utilizzando i quantili gaussiani, per evitare la ricerca manuale di n; però il risultato è molto impreciso, perché la numerosità in gioco è molto bassa. Volendo risolvere il problema con le t di Student, ipotizzando $s = 1.7$ come deviazione sperimentale indipendentemente dal numero di prove e dall'esito aleatorio delle prove stesse, cerchiamo il più piccolo n tale che

$$\frac{1.7 \cdot t_{n-1,0.05}}{\sqrt{n}} \leq 0.6$$

ovvero $\frac{t_{n-1,0.05}}{\sqrt{n}} \leq \frac{0.6}{1.7} = 0.35294$. La ricerca va fatta per tentativi. Ora, $t_{n-1,0.05}$ varia tra 1.761 (certamente il numero di pezzi dovrà essere maggiore di 15) e 1.65 circa. Prendiamo il numero di riferimento 1.7 e cerchiamo n: $\frac{1.7}{\sqrt{n}} \leq 0.35294$, $\sqrt{n} \geq \frac{1.7}{0.35294} = 4.8167$, $n \geq 23.201$. Con questa indicazione sommaria, andiamo per tentativi: per $n = 25$: $\frac{t_{24,0.05}}{\sqrt{25}} = \frac{1.711}{\sqrt{25}} = 0.3422$. Questo valore va bene. Con $n = 24$: $\frac{t_{23,0.05}}{\sqrt{24}} = \frac{1.714}{\sqrt{24}} = 0.34987$; va ancora bene. Per $n = 23$: $\frac{t_{22,0.05}}{\sqrt{23}} = \frac{1.717}{\sqrt{23}} = 0.35802$; non va bene. Quindi il numero minimo è 24.

3. Cerchiamo n_0 tale

$$\begin{aligned}
P(N \geq n_0) &= 0.9 \\
P(\log_{10} N \geq \log_{10} n_0) &= 0.9 \\
P(X \geq \log_{10} n_0) &= 0.9 \\
P\left(\frac{X-5.3}{1.7} \geq \frac{\log_{10} n_0 - 5.3}{1.7}\right) &= 0.9 \\
1 - \Phi\left(\frac{\log_{10} n_0 - 5.3}{1.7}\right) &= 0.9 \\
\Phi\left(\frac{\log_{10} n_0 - 5.3}{1.7}\right) &= 0.1 \\
\frac{\log_{10} n_0 - 5.3}{1.7} &= q_{0.1}
\end{aligned}$$

$$\begin{aligned}
\log_{10} n_0 &= 5.3 + 1.7 \cdot q_{0.1} = 5.3 - 1.7 \cdot q_{0.9} = 5.3 - 1.7 \cdot 1.29 = 3.107 \\
n_0 &= 10^{3.107} = 1279.4.
\end{aligned}$$

4. Vista la bassissima numerosità campionaria, eseguiamo sicuramente un test t di Student, per la media. Eseguiamo un test unilaterale, perché fortemente indicato dal testo. L'ipotesi nulla è $\mathcal{H}_0$) media ≥ 5.3; quella alternativa è $\mathcal{H}_1$) media < 5.3; la regione di rifiuto è $\left\{(x_1, ..., x_{10}) : \frac{\overline{x}_{10}-5.3}{1.6}\sqrt{10} < -t_{9,0.1}\right\}$. Nel nostro caso vale

$$\frac{\overline{x}_{10} - 5.3}{1.6}\sqrt{10} = \frac{4.7 - 5.3}{1.6}\sqrt{10} = -1.1859$$

che è maggiore di $-t_{9,0.1} = -1.383$, cioè vale $\frac{\overline{x}_{10}-5.3}{1.6}\sqrt{10} > -t_{9,0.1}$. Quindi il test non è significativo, non siamo in grado di dire che c'è un peggioramento.

5.

$$\begin{aligned} P^{4.5,1.6}\left(\frac{\overline{X}_{10} - 5.3}{1.6}\sqrt{10} < -t_{9,0.1}\right) &= P^{4.5,1.6}\left(\frac{\overline{X}_{10} - 4.5}{1.6}\sqrt{10} + \frac{4.5 - 5.3}{1.6}\sqrt{10} < -t_{9,0.1}\right) \\ &= P\left(Z < -t_{9,0.1} - \frac{4.5 - 5.3}{1.6}\sqrt{10}\right) \\ &= P\left(Z < -1.383 + 1.5811\right) \\ &= P\left(Z < 0.1981\right) = 0.575. \end{aligned}$$

Esercizio 2.

1.

$$\begin{aligned} P\left(N > 15849\right) &= P\left(\log_{10} N > \log_{10} 15849\right) \\ &= P\left(X > 4.2\right) \\ &= P\left(\frac{X - 5}{1} > \frac{4.2 - 5}{1}\right) \\ &= 1 - \Phi\left(\frac{4.2 - 5}{1}\right) = 1 - \Phi\left(-0.8\right) \\ &= \Phi\left(0.8\right) = 0.788. \end{aligned}$$

2. La condizione di densità impone

$$1 = \int_{-\infty}^{-1} \frac{D}{x^4}dx + \int_{-1}^{\infty} \frac{C}{(x+2)^2}dx = D\left[\frac{x^{-3}}{-3}\right]_{-\infty}^{-1} + C\left[\frac{(x+2)^{-1}}{-1}\right]_{-1}^{\infty} = \frac{D}{3} + C$$

mentre la condizione di continuità impone

$$\begin{aligned} \frac{C}{(-1+2)^2} &= \frac{D}{1^4} \\ C &= D \end{aligned}$$

da cui $\frac{C}{3} + C = 1$, $C = D = \frac{3}{4}$.

3. No: l'integrale

$$\int_{-1}^{\infty} \frac{x}{(x+2)^2} dx = \int_{1}^{\infty} \frac{y-2}{y^2} dx = \int_{1}^{\infty} \frac{1}{y} dx - \int_{1}^{\infty} \frac{2}{y^2} dx$$

diverge, perché $\int_1^\infty \frac{1}{y} dx = +\infty$ (mentre $\int_1^\infty \frac{2}{y^2} dx < \infty$).