Metodi Matematici e Statistici, Modulo di Statistica
Ing. Gestionale, II compitino, 18/5/04

Esercizio 1 (circa 22 punti). Il benzinaio di un paese con 100 automobili vuole indagare sui rifornimenti mensili di benzina effettuati presso il suo distributore dai 100 proprietari dei veicoli. Fissa allora l'attenzione su un campione di 10 di essi e registra per un mese i loro rifornimenti. Al termine del mese ha ottenuto i valori complessivi $x_1, ..., x_{10}$ di media $\overline{x} = 120$ e deviazione standard empirica $S = 40$. Non ha invece elementi visibili per ipotizzare che il rifornimento mensile del generico proprietario sia una v.a. gaussiana, ma nemmeno il contrario. Sulla base di questi dati sperimentali, il benzinaio tira alcune conclusioni. Descrivere dettagliatamente i ragionamenti che possono giustificare tali conclusioni (specificare in ogni punto che ipotesi si stanno facendo, che teoremi si stanno usando, ecc.). Circa eventuali test, si supponga che il benzinaio non abbia pregiudizi a priori circa scostamenti in positivo o negativo.

1) Consideriamo il rifornimento mensile *totale* del paese (la somma dei rifornimenti mensili delle 100 automobili). Il benzinaio afferma che il rifornimento mensile totale *medio* è pari a 12000 con deviazione standard 400.

2) Con probabilità circa pari a 0.79, il rifornimento mensile totale è compreso tra 11500 e 12500.

3) Per conoscere il rifornimento mensile *medio* del *singolo* veicolo con una precisione pari a 20 e confidenza 99%, non bastano le osservazioni che ha fatto.

4) La multinazionale petrolifera che lo rifornisce, sulla base di dati medi internazionali, gli ha dato l'indicazione che il consumo mensile di una generica automobile ha media 140. Al 95%, il benzinaio non riscontra una differenza significativa tra gli utenti del proprio paese e la media internazionale.

Esercizio 2 (circa 8-10 punti). Relativamente al punto 4 dell'esercizio precedente, aggiungendo per semplicità l'ipotesi che sia $\sigma = 40$, si chiede di determinare la significatività migliore possibile rispetto a cui il benzinaio sarebbe arrivato alla conclusione opposta (chiarendo facoltativamente il legame con uno dei punti già risolti sopra). Si chiede inoltre di calcolare la potenza del punto 4, relativamente ad uno scostamento da 140 a 120. Infine si chiede di determinare quante osservazioni (invece che solamente su 10 automobilisti) sarebbero necessarie per avere una potenza pari ad almeno 0.9.

Soluzioni

Es. 1 Chiamiamo μ la media e σ^2 la varianza della v.a. X che rappresenta il rifornimento mensile di un singolo utente.

Punto 1. Dai valori osservati il benziaio può stimare (stima puntuale) che

$$\mu \simeq \overline{x} = 120, \qquad \sigma^2 \simeq S^2 = 40^2.$$

Da qui può dedurre che il rifornimento totale Y di 100 utenti, che si esprime come somma di 100 v.a. indipendenti e tutte distribuite come X deve avere media e varianza pari a

$$\mathrm{E}[Y] = 100 \cdot E[X] = 12000, \qquad \mathrm{Var}[Y] = 100 \cdot \mathrm{Var}[X] = 400^2.$$

Punto 2. Se continuamo a denotare Y il rifornimento mensile totale e se osserviamo che esso è la somma di $N = 100$ addendi indipendenti e identicamente distribuiti, ciascuno con media μ e varianza σ stimate nel punto precedente, otteniamo che

$$\begin{aligned} P\left(11500 < Y \leq 12500\right) &= P\left(\frac{11500 - N\mu}{\sqrt{N}\sigma} < Z \leq \frac{12500 - N\mu}{\sqrt{N}\sigma}\right) \\ &= P\left(-\frac{5}{4} < Z \leq \frac{5}{4}\right) \end{aligned}$$

dove

$$Z = \frac{Y - N\mu}{\sqrt{N}\sigma}$$

è approssimabile, per il teorema limite centrale, con una gaussiana standard (media nulla e varianza unitaria). Questo senza ulteriori ipotesi circa la distribuzione dei singoli addendi.

Utilizzando la funzione di ripartizione Gaussiana $\Phi(x) = P[Y \leq x]$ si ottiene quindi

$$\begin{aligned} P\left(11500 < Y \leq 12500\right) &\simeq \Phi\left(\frac{5}{4}\right) - \Phi\left(-\frac{5}{4}\right) \\ &= 2\Phi\left(\frac{5}{4}\right) - 1 = 2 \cdot 0.8944 - 1 \simeq 0.79 \end{aligned}$$

Punto 3. Proviamo a stimare con quale confidenza il benzinaio può affermare di conoscere il consumo mensile medio del singolo utente. Lo stimatore

$\overline{X}$ rappresenta la stima del benzinaio. Chiamiamo n le osservazioni effettuate : $n = 10$. Vogliamo quindi calcolare la probabilità che il valore vero μ non disti da $\overline{X}$ più di $\delta = 20$. Osserviamo che

$$P\left(\left|\mu - \overline{X}\right| < \delta\right) = P\left(\left|\sqrt{n}\frac{\mu - \overline{X}}{\sigma}\right| < \sqrt{n}\frac{\delta}{\sigma}\right)$$

e che la v.a. $\sqrt{n}\frac{\mu-\overline{X}}{\sigma}$ è approssimabile con una gaussiana standard Z, quindi la probabilità cercata si ottiene tramite la tabella della funzione di ripartizione gaussiana standard:

$$\begin{aligned} P\left(\left|\mu - \overline{X}\right| < 20\right) &\sim P\left(|Z| < 1.58\right) \\ &= 2\Phi\left(1.58\right) - 1 \sim 2 \cdot 0.9429 - 1 \sim 0.8858. \end{aligned}$$

La confidenza in un errore di 20 è dell' 88% circa. Non è quindi sufficiente per raggiungere il livello desiderato del 99%.

Un altro modo di procede potrebbe essere quello di calcolare il numero di osservazioni necessarie per conseguire una precisione di 20 con una confidenza del 99%. Oppure infine di calcolare la precisione al 99% con 10 esperimenti.

Punto 4. Secondo i dati della multinazionale la distribuzione del consumo mensile X di un tipico utente internazionale ha media $\mu_0 = 140$. Per confrontare questo dato con il campione considerato dal benzinaio (estratto dalla popolazione locale) possiamo effettuare un test sulla media. Qui supponiamo per semplicità la gaussianità di X, anche se il teorema limite centrale giustificherebbe approssimativamente l'uso del test in generale.

L'ipotesi nulla H_0 è quella secondo cui la media della popolazione locale è μ_0. Scegliamo l'ipotesi alternativa $\mu \neq \mu_0$. La variabile di test T è costruita come

$$T = \sqrt{n}\frac{\overline{X} - \mu_0}{S} = -1.58$$

dove n è ancora il rango del campione ($n = 10$). Fissato il livello di confidenza $1-\alpha = 95\%$ abbiamo che $\alpha = 0.05$. La regione critica è data dalla condizione $|T| > t^{(9)}_{\frac{\alpha}{2}}$, dove $t^{(9)}_{\frac{\alpha}{2}} = t^{(9)}_{0.025} = 2.262$. Siccome per il nostro campione risulta $|T| < t^{(9)}_{\frac{\alpha}{2}}$, il campione è compatibile con l'ipotesi di media 140.

Es. 2 La significatività migliore possibile per la quale il benzinaio avrebbe rigettato l'ipotesi nulla è il valore p dei dati, che nell'ipotesi di varianza nota e test bilaterale è

$$p = 2 - 2\Phi(|z|) = 2 - 2\Phi(1.58) = 2 - 2 \cdot 0.9429 = 0.1142.$$

Si noti che questa domanda è equivalente al punto 3 dell'esercizio precedente. Si noti inoltre che abbiamo usato come z il valore $z = \sqrt{n}\frac{\overline{X}-\mu_0}{\sigma}$ invece che $\sqrt{n}\frac{\overline{X}-\mu_0}{S}$, essendo σ nota (numericamente coincidono).

Calcolare la potenza del test del punto 4 nell'esercizio precedente relativamente a uno scostamento da $\mu_0 = 140$ a $\mu_1 = 120$ significa valutare la probabilità di rifiutare H_0 (ovvero μ_0) quando H_1 (che è l'ipotesi sotto la quale μ_1 è la vera media) è vera. Qui usiamo le gaussiane, quindi il quantile $q_{1-\frac{\alpha}{2}} = 1.96$ relativo alla significatività 95%. Supponiamo quindi che H_1 sia l'ipotesi vera. La probabilità di accettare H_0 è data da

$$\begin{aligned}
\beta &= \Phi\left(\sqrt{n}\frac{\mu_0-\mu_1}{\sigma} + q_{1-\frac{\alpha}{2}}\right) - \Phi\left(\sqrt{n}\frac{\mu_0-\mu_1}{\sigma} - q_{1-\frac{\alpha}{2}}\right) \\
&= \Phi(1.58 + 1.96) - \Phi(1.58 - 1.96) \\
&= \Phi(3.54) - \Phi(-0.38) \sim 1 - (1 - \Phi(0.38)) \\
&= \Phi(0.38) = 0.648.
\end{aligned}$$

La potenza è quindi pari a 0.352, molto bassa.

Per avere una potenza pari ad almeno 0.9, deve essere $\beta \leq 0.1$, ovvero

$$\Phi\left(\sqrt{n}\frac{\mu_0-\mu_1}{\sigma} + q_{1-\frac{\alpha}{2}}\right) - \Phi\left(\sqrt{n}\frac{\mu_0-\mu_1}{\sigma} - q_{1-\frac{\alpha}{2}}\right) \leq 0.1.$$

Qui $\alpha, \sigma, \mu_0, \mu_1$ sono noti, n invece deve essere determinato. Abbiamo quindi il problema

$$\Phi\left(0.5\sqrt{n} + 1.96\right) - \Phi\left(0.5\sqrt{n} - 1.96\right) \leq 0.1.$$

Siccome, anche se non conosciamo ancora n, sarà comunque $n > 10$, sarà anche $0.5\sqrt{n} + 1.96 > 3.54$, quindi $\Phi(0.5\sqrt{n} + 1.96) \sim 1$. Quindi il problema diventa

$$\Phi\left(0.5\sqrt{n} - 1.96\right) \geq 0.9$$

ovvero

$$0.5\sqrt{n} - 1.96 \geq q_{0.9} = 1.28$$

ovvero

$$\sqrt{n} \geq \frac{3.\,24}{0.5} = 6.\,48$$

quindi $n = 42$.