Politecnico di Milano – Facoltà di Ingegneria Informatica

Anno Accademico 2008/2009

Corso di Statistica (2L) per INF e TEL

Docente: Antonio Pievatolo; esercitazioni: Raffaele Argiento

Esercitazione del 04/06/2009

Esercizio 1

I ricercatori di una società finanziaria vogliono determinare se le scadenze delle obbligazioni di tipo $\mathcal{X}$ e quelle di tipo $\mathcal{Y}$ hanno varianze differenti. Un campione di 16 obbligazioni $\mathcal{X}$ ha fornito una deviazione standard campionaria per le scadenze pari a $s_X = 11.11$, mentre un campione di 11 obbligazioni ha fornito $s_Y = 5.83$. È ragionevole supporre che le scadenze delle obbligazioni di tipo $\mathcal{X}$ e quelle di tipo $\mathcal{Y}$ siano indipendenti e normalmente distribuite.

1. Scrivere la regione di rifiuto di livello α per il test

$$\mathrm{H}_0\text{: } \sigma_X^2 = \sigma_Y^2 \text{ contro } \mathrm{H}_1\text{: } \sigma_X^2 \neq \sigma_Y^2$$

   Rifiutate o meno l'ipotesi nulla al livello $\alpha = 5\%$?

2. Si calcoli il p-value per il test al punto 1.

SOLUZIONE

1. Sotto l'ipotesi nulla

$$U := \frac{S_X^2}{S_Y^2} \sim F_{n_X-1,n_Y-1},$$

   Dove F_{n_X-1,n_Y-1} è la distribuzione di Fisher con $n_X - 1$ gradi di libertà al numeratore e $n_Y - 1$ gradi di libertà al denominatore. La regione critica per il test in esame è quindi

$$C = \left\{ \underline{x}, \underline{y} : \frac{s_X^2}{s_Y^2} \leq F_{n_X-1,n_Y-1}\left(\frac{\alpha}{2}\right); \frac{s_X^2}{s_Y^2} \geq F_{n_X-1,n_Y-1}\left(1 - \frac{\alpha}{2}\right) \right\}$$

   Dove $F_{n_X-1,n_Y-1}(\alpha)$ è il percentile di ordine α della distribuzione di Fisher. Si ricordi che i percentili di tale distribuzione sono tabulati per $\alpha \geq 0.5$, per valori minori di α si usa la formula:

$$F_n, m(\alpha) = \frac{1}{F_{m,n}(1-\alpha)}.$$

   Con i valori forniti nel testo $n_X = 16$, $n_Y = 11$ e $\alpha = 0.05$ otteniamo $F_{15,10}(0.975) = 3.53$ e $F_{15,10}(0.025) = \frac{1}{F_{10.15}(0.975)} = 0.3268$ quindi

$$C = \left\{ \underline{x}, \underline{y} : \frac{s_X^2}{s_Y^2} \leq 0.3267; \frac{s_X^2}{s_Y^2} \geq 3.52 \right\}.$$

   Il valore osservato della statistica test, ricavato dalla realizzazione campionaria, è $u = \frac{11.11^2}{5.83^2} = 4.042$. Dato che $u \in C$ rifiuto H_0 al livello $\alpha = 5\%$.

2. Il p-value per il test in esame è p-value= $\min\{p_1, p_2\}$ dove

$$p_1 = \mathbb{P}_{H_0}(U \leq u); \qquad p_2 = 1 - p_1$$

   dalle tavole si ricava che $0.95 < p_1 < 0.98$ e di conseguenza $0.02 < p_2 < 0.05$.
   In fine $0.02 <$p-value< 0.05.

∎

Esercizio2

I valori che seguono rappresentano i giorni di sopravvivenza di un campione di 6 topi affetti da cancro e curati con una terapia sperimentale:

$$29, \quad 700, \quad 1, \quad 335, \quad 15, \quad 160$$

1. Determinate la funzione di ripartizione empirica $\hat{F}_6$ associata al campione dei 6 topi.
2. Determinate una stima della probabilità che un topo affetto da cancro sottoposto alla terapia viva più di 15 giorni.

Si pensa che la sopravvivenza dei topi malati di cancro e sottoposti alla terapia sperimentale possa essere modellata come una variabile aleatoria X assolutamente continua che ha densità di Weibull:

$$f_0(x) = \frac{1}{20\sqrt{x}} \mathrm{e}^{-\frac{\sqrt{x}}{10}} \mathbf{1}_{(0,\infty)}(x) \tag{1}$$

3. Usate un opportuno test con il 5% di livello di significatività, per stabilire se i dati forniti sui topi possano provenire dalla densità scritta in (1)

La terapia sperimentale viene provata su un nuovo campione di 70 topi.

4. Scrivere la regione di critica dello stesso test studiato al punto 3. per il nuovo campione di 70 topi, sapendo che la realizzazione della statistica di Kolmogorov-Smirnov è $d_{70} = 0.221$. Cosa concludete ad un livello di significatività $\alpha = 5\%$?

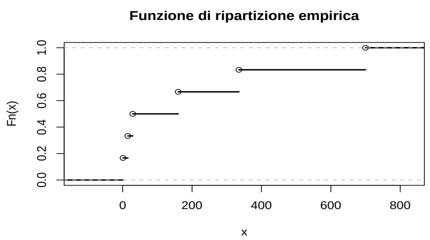
Soluzione

1. Ordinando le osservazioni in modo crescente

$$1, \quad 15, \quad 29, \quad 160, \quad 335, \quad 700$$

otteniamo la seguente realizzazione della funzione di ripartizione empirica

$$\hat{F}_6(x) = \begin{cases} 0 & x < 1 \\ \frac{1}{6} & 1 \leq x < 15 \\ \frac{1}{3} & 15 \leq x < 29 \\ \frac{1}{2} & 29 \leq x < 160 \\ \frac{2}{3} & 160 \leq x < 335 \\ \frac{5}{6} & 335 \leq x < 700 \\ 1 & x \geq 700 \end{cases}$$



2. Si osservi che $\mathbb{P}(X > 15) = 1 - \mathbb{P}(\leq 15) = 1 - F_X(15)$. La funzione di ripartizione empirica, $\hat{F}_6(x)$, è una stima non distorta e consistente in media quadratica di $F_X(x)$ per ogni $x \in \mathbb{R}$ (NB, in realtà $\hat{F}_n(x)$ converge uniformemente a $F_X(x)$ per il teorema di Glivenko-Cantelli). Stimando $F_X(15)$ con $\hat{F}(15) = \frac{2}{3}$ si ottiene $\frac{2}{3}$ come stima di $\mathbb{P}(X > 15)$
3. Si vuole studiare il test:

$$\mathrm{H}_0:\ X \sim F_0 \text{ contro } \mathrm{H}_1:\ X \nsim F_0$$

dove F_0 è la funzione di ripartizione corrispondente alla densità f_0 scritta in (1).

$$\begin{aligned} F_0(x) &= \int_0^x \frac{1}{20\sqrt{u}} \mathrm{e}^{-\frac{\sqrt{u}}{10}} du \quad \text{Sostituendo } \sqrt{u} = y \text{e } \frac{1}{1\sqrt{u}} = dy \\ &= \int_0^{\sqrt{x}} \frac{1}{10} \mathrm{e}^{-\frac{y}{10}} dy = \left(-\mathrm{e}^{-\frac{y}{10}}\right)_0^{\sqrt{x}} = 1 - \mathrm{e}^{\frac{-\sqrt{x}}{10}} \end{aligned}$$

Dato che F_0 è una densità continua ed ho (pochi) dati non raggruppati posso usare il test di Kolmogorov-Smirnov. La statistica test è

$$D_6 = \sup_{x \in \mathbb{R}} |\hat{F}_6(x) - F_0(x)|.$$

La distribuzione di D_n è nota, possiamo scrivere la regione critica di livello α:

$$C = \{x_1, \dots, x_6 : d_6 \geq D_6(1-\alpha)\}$$

Dove $D_6(1-\alpha)$ è il percentile di ordine $1-\alpha$ della distribuzione di Kolmogorov. In particolare se $\alpha = 0.05$ allora $D_6(1-0.05) = 5.193$ e quindi $C = \{x_1, \dots, x_6 : d_6 \geq 6.193\}$. Per calcolare il valore osservato di D_6, poniamo $x_0 = -\infty$ e costruiamo la tabella

	1	15	29	160	335	700
$F_6(x_j)$	1/6	1/3	1/2	2/3	5/6	1
$F_0(x_j)$	0.095	0.321	0.416	0.718	0.840	0.929
$\lvert\hat{F}_6(x_j) - F_0(x_j)\rvert$	0.072	0.013	0.084	0.051	0.006	0.071
$\lvert\hat{F}_6(x_{j-1}) - F_0(x_j)\rvert$	0.095	0.154	0.082	0.218	0.173	0.0956

Si ottiene facilmente che il valore osservato dalla statistica test è $d_6 = 0.218$. Dato che $0.218 \notin C$ non Rifiutiamo l'ipotesi nulla al livello $\alpha = 0.05$.

4. Se D_n è la statistica di Kolmogorov per un campione di n osservazione, allara vale il seguente risultato asintotico:

$$\lim_{n \to \infty} \mathbb{P}_0 \left(\sqrt{n} D_n < z\right) = H(z) \quad \text{per ogni} z \geq 0.$$

dove

$$H(z) = 1 - 2 \sum_{i=0}^{\infty} (-1)^{i-1} \mathrm{e}^{-2(iz)^2}.$$

Sulla base del precedente risultato, per $n > 40$ la tabulazione della statistica D_n é data in funzione di n si ha quindi:

$$D_n(1-\alpha) = \frac{H^{-1}(1-\alpha)}{\sqrt{n}}.$$

per $n = 70$ e $\alpha = 0.05$ otteniamo $\frac{D_{70}(1-0.05)}{\sqrt{70}} = \frac{1.36}{\sqrt{70}} = 0.1625$. Possiamo quindi scrivere la regione critica

$$C = \{x_1, \dots, x_{70} : d_{70} \geq 0.1625\}.$$

In conclusione dato che $d_{70} = 0.221 \in C$ Rifiuto l'ipotesi nulla al livello $\alpha = 0.05$.

∎

Esercizio 3

Esempio 2.13 dalle dispense “Inferenza non parametrica” della Prof.ssa I. Epifani

Esercizio 4

Esempio 2.14 dalle dispense “Inferenza non parametrica” della Prof.ssa I. Epifani