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 08/05/2009

Esercizio 1(dal Tema d'esame del 25/02/2008)

Una grande catena di *fast-food* analizza mensilmente un campione casuale dei suoi famosi *big* panini, con lo scopo di controllare che mediamente pesino al più 600 grammi (g). In caso contrario, si rendono necessarie delle operazioni di taratura dei macchinari di produzione. Nel mese di gennaio su un campione casuale di 25 *big* panini si è registrato un peso medio campionario pari a 620.0 gr.

Assumendo che il peso dei panini espresso in grammi sia una variabile aleatoria gaussiana, con deviazione standard nota e pari a 80, rispondete alle seguenti domande.

1. Impostate un opportuno test di ipotesi tale che sia al più pari ad $\alpha = 5\%$ la probabilità di commettere l'errore di prima specie di ritenere necessarie operazioni di taratura dei macchinari di produzione che in realtà sono già ben tarati. Alla luce dei dati di gennaio, l'operazione di taratura è necessaria?

2. Sulla base dei dati a disposizione, quanto siete confidenti che la media teorica dei *big* panini di gennaio sia maggiore di 593.68 gr?

3. Fornite analiticamente e rappresentate graficamente la funzione di potenza π del test al punto 1. ($\alpha = 5\%$).

Per il mese di marzo si è deciso di misurare un campione più numeroso di *big* panini al fine di aumentare la potenza del test, ma mantenendo una significatività pari a $\alpha = 5\%$.

4. Se effettivamente il peso medio dei *big* panini è pari a 640 gr, quanti panini bisogna misurare affinché la potenza sia maggiore di 0.94?

Soluzione Se X_i è il peso dell'i-esimo panino, $X_1, \ldots, X_n$ è un campione casuale estratto da una popolazione gaussiana $\mathcal{N}(\mu, 80^2)$ di media μ incognita.

1. Impostiamo il problema di verifica di ipotesi $H_0 : \mu \leq 600$ contro $H_1 : \mu > 600$. Quindi rifiutiamo se $\bar{X} \geq 600 + z_{1-\alpha}80/\sqrt{25} = 600 + 16z_{1-\alpha}$. Il p-value è dato da

$$\text{p-value} = 1 - \Phi\left(\frac{620 - 600}{16}\right) = 1 - \Phi(1.25) \simeq 1 - 0.894 = 0.106 = 10.6\%$$

   Ai consueti livelli di significatività (inferiori al 10%, quindi anche 5%) non possiamo rifiutare l'ipotesi nulla H_0 e concludiamo che non sono necessarie operazioni di taratura.

2. Siamo confidenti al 95% che i *big* panini superino mediamente i 593.68 gr. Infatti, con i dati a disposizione, in virtù della dualità fra la verifica di ipotesi e gli intervalli di confidenza, segue dal punto 1. che $\{\mu : \bar{x} < \mu + 16z_{1-\alpha}\} = \{\mu : \mu > \bar{x} - 16z_{1-\alpha}\} = (620.0 - 16z_{1-\alpha}, \infty)$ è un IC unilatero a una coda superiore per μ di confidenza $1 - \alpha$. Ma, $620.0 - 16z_{1-\alpha} = 593.68$ se, e solo se, $1 - \alpha = 95\%$.

3. La funzione di potenza del test del punto 1. quando si campionano n panini e $\alpha = 5\%$ è data da

$$\pi(\mu) = P_\mu\left(\bar{X} \geq 600 + 1.645 \times \frac{80}{\sqrt{n}}\right) = 1 - \Phi\left(\frac{600 - \mu}{80}\sqrt{n} + 1.645\right) =$$
$$= \Phi\left(\frac{\mu - 600}{80}\sqrt{n} - 1.645\right) = \Phi\left(\frac{\mu}{16} - 39.145\right),\ \mu > 600$$

   Il grafico è nella figura che segue

4. La potenza π calcolata in $\mu = 640$ quando si campionano n panini e $\alpha = 5\%$ diventa $\pi(640) = \Phi\left(\sqrt{n}/2 - 1.645\right)$ che è maggiore di 0.94 se, e solo se, $\sqrt{n}/2 - 1.645 > 1.555$, cioè $n > 40.96$. In definitiva, $\pi(640) > 0.94$ per $n \geq 41$ panini. ∎

Esercizio2

Ripetere il punto 1. dell'esercizio 1 senza l'ipotesi che la variaza del peso dei *big* panini sia nota, con l'informazione aggiuntiva che $\sum_{i=1}^{n} x_i = 9764829$

Soluzione La statistica test in questo caso è

$$U := \frac{\bar{X} - \mu_0}{S/\sqrt{n}} \sim t_{n-1}$$

Per calcolare il valore osservato da U bisogna calcolare una stima della varianza tramite la statistica varianza campionaria:

$$S^2 = \frac{1}{n-1}(X_i - \bar{X})^2 = \frac{1}{n-1}(\sum_{i=1}^{n} X_i^2 - n\bar{X}^2)$$

Quindi $s^2 = 6451.2$ e $u = 1.245$ da cui si ricava che p-value$= 1 - \mathbb{P}(U < u) = 1 - 0.0887 = 0.113$ ∎

Esercizio3

Sia X una v.a. con densità

$$f_X(x;\theta) = \begin{cases} \frac{1}{\theta} x^{\frac{1}{\theta}-1} & 0 < x < 1, \\ 0 & \text{altrove.} \quad \theta > 0 \end{cases}$$

Si consideri il problema

$$\text{H}_0\colon\ \theta = 2 \text{ contro } \text{H}_1\colon\ \theta = 0.2 \text{ Sia } X_1, \ldots, X_n \text{ un campione i.i.d. da } f_X \tag{1}$$

1. Si determini la densità di $V = -\sum_{j=1}^{n} \log X_j$ sotto l'ipotesi H_0 e sotto H_1.
2. Si costruisca il test più potente di livello α per il problema (1).
3. Calcolare la potenza del test costruito al punto 2., quando $\alpha = 5\%$ per campione di 3 osservazioni. Come varia la potenza in funzione di α.
4. Siano $n = 3$ e $x_1 \times x_2 \times x_3 = 0.13$. Che decisione prendete per $\alpha = 1\%,\ 5\%,\ 10\%$?

Soluzione

1. Sia $X \in (0,1)$ con densità f_X e poniamo $Y = -\log(X)$; Y a supporto sul semiasse positivo dei reali, inoltre se $y \geq 0$ allora

$$\begin{aligned} F_Y(y) &= \mathbb{P}(Y \leq y) = \mathbb{P}(-\log(X) \leq y) = \mathbb{P}(X \geq \mathrm{e}^{-y}) \\ &= \int_{\mathrm{e}^{-y}}^{1} \frac{1}{\theta} x^{\frac{1}{\theta}-1} dx = \left(x^{\frac{1}{\theta}}\right)_{\mathrm{e}^{-y}}^{1} = 1 - \mathrm{e}^{-\frac{y}{\theta}}. \end{aligned}$$

   Possiamo concludere che Y ha distribuzione *Exp*(θ). Se $Y_1 = -\log(X_1), \ldots, Y_n = -\log(X_n)$, allora $V = \sum_{i=1}^{n} Y_i$ e di conseguenza $V \sim$gamma(n,θ).
   In particolare:

   (a) sotto l'ipotesi H_0: $\theta = 2$, quindi $V_{H_0} \sim \text{gamma}(n,2) \overset{d}{=} \text{gamma}(\frac{2n}{2}, 2) \overset{d}{=} \chi^2_{2n}$;

   (b) sotto l'ipotesi H_1: $\theta = 0.2$, quindi $V_{H_1} \sim \text{gamma}(n, 0.2) \overset{d}{=} \frac{1}{10}\text{gamma}(n,2) \overset{d}{=} \frac{1}{10}\chi^2_{2n}$

2. Per costruire il test più potente, per il problema (1), usiamo il lemma di Neyman-Pearson. Siano $\theta_0 = 2$ e $\theta_1 = 0.2$, calcoliamo la statistica

$$\lambda(\underline{x}) = \frac{L(\theta_1, \underline{x})}{L(\theta_2, \underline{x})} = \frac{\prod_{i=1}^n f_X(x_i;\theta_0)}{\prod_{i=1}^n f_X(x_i;\theta_1)} = \left(\frac{\theta_1}{\theta_0}\right)^2 \left(\prod_{i=1}^n x_i\right)^{\frac{1}{\theta_0}-\frac{1}{\theta_1}}.$$

La regine critica del test più potente per il problema (1) è

$$C = \left\{ \underline{x} : \left(\frac{\theta_1}{\theta_0}\right)^2 \left(\prod_{i=1}^n x_i\right)^{\frac{1}{\theta_0}-\frac{1}{\theta_1}} \leq \delta \right\},$$

e bisogna scegliere δ in modo tale che essa sia di livello α. A tal proposito osserviamo che tutte le disuguaglianze che seguono sono equivalenti, dato che le quantità $(\delta, \delta_1, \delta_2, \delta_3)$ alla destra delle disuguaglianze non dipendono dal campione:

$$\begin{aligned}
\left(\frac{\theta_1}{\theta_0}\right)^n \left(\prod_{i=1}^n x_i\right)^{\frac{1}{\theta_0}-\frac{1}{\theta_1}} \leq \delta \quad &\Rightarrow \quad \left(\prod_{i=1}^n x_i\right)^{\frac{1}{\theta_0}-\frac{1}{\theta_1}} \leq \delta_1 \\
&\Rightarrow \quad \prod_{i=1}^n x_i \geq \delta_2 \quad (\text{Nota}\frac{1}{\theta_0} - \frac{1}{\theta_1} < 0) \\
&\Rightarrow \quad -\sum_{i=1}^n \log(x_i) \leq \delta_3
\end{aligned}$$

In conclusione possiamo riscrivere C come:

$$C = \left\{ \underline{x} : -\sum_{i=1}^n \log(x_i) \leq \delta_3 \right\}.$$

Come calcolato al punto 1. dell'esercizio, la distribuzione di $-\sum_{i=1}^n \log(X_i)$, sotto H_0, è quella di una χ^2_{2n} allora C è di livello α se $\delta_3 = \chi^2_{2n}(\alpha)$.

3. La potenza del test è :

$$\pi_\alpha(\theta_1) = \mathbb{P}_{H_1}(\underline{X} \in C) = \mathbb{P}_{H_1}(-\sum_{i=1}^n \log(X_i) \leq \chi^2_{2n}(\alpha)) = \mathbb{P}_{H_1}(-10\log(X_i) \leq 10\chi^2_{2n}(\alpha))$$

Al punto 1. dell'esercizio si è osservato come, sotto H_1 $-10\log(X_i) \sim \chi^2_{2n}$. Sia F_{2n} la sua funzione di ripartizione, allora, $\pi_\alpha(\theta_1) = F_{2n}(\chi^2_{2n}(\alpha))$. Ora si osservi che sia $F_{2n}(\cdot)$, sia $\chi^2_{2n}(\cdot)$, sono funzioni crescenti. Quindi $\pi_\alpha(\theta_1)$ è funzione crescente di α

4. Calcoliamo il p-value del test per $n = 3$ e $x_1 \times x_2 \times x_3 = 0.13$. Il valore osservato della statistica test è $v_{H_0} = -\log(0.13) = 2.04$. Quindi p-value$\mathbb{P}(V_{H_0} \leq v_{H_0}) = \mathbb{P}(V_{H_0} \leq 2.04)$, ricordando che $V_{H_0} \sim \chi^2_{2n}$ si ottiene 0.05 <p-value< 0.1. Quindi:
   (a) Se $\alpha = 0.01$ allora α <p-value $\Rightarrow$ non rifiuto H_0
   (b) Se $\alpha = 0.05$ allora α <p-value $\Rightarrow$ non rifiuto H_0
   (c) Se $\alpha = 0.1$ allora α >p-value $\Rightarrow$ rifiuto H_0

∎