Pavia, 9 aprile 2003

Brevissime note di Statistica inferenziale

A cura di Luca La Rocca

Campione casuale estratto da una popolazione normale

Si assume che le osservazioni $x_1, \ldots, x_n$ siano realizzazioni delle variabili aleatorie $X_1, \ldots, X_n$ indipendenti e identicamente distribuite secondo la curva normale:

$$\mathbf{P}\{a \leq X_1 \leq b\} = \int_a^b \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx.$$

Parametri di popolazione: $\mu = \mathbf{E}[X_1]$ e $\sigma^2 = \mathbf{V}[X_1] > 0$.

Taglia del campione: $n \to \infty$.

Stima puntuale dei parametri di una popolazione normale

$$\bar{X} = \frac{X_1 + \cdots + X_n}{n}$$

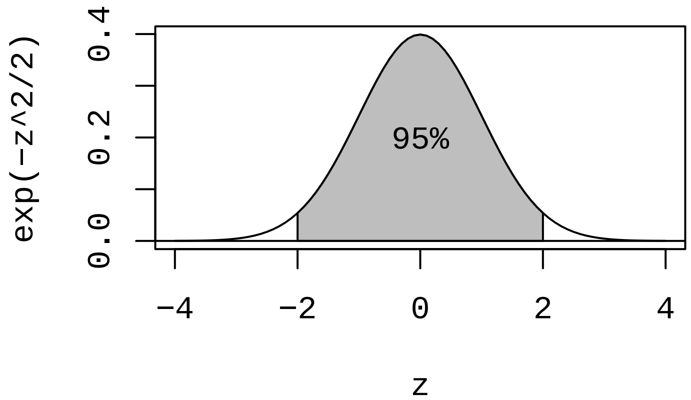
$$\mathbf{E}[\bar{X}] = \mu \quad \& \quad \mathbf{V}[\bar{X}] = \frac{\sigma^2}{n} \to 0$$

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

$$\mathbf{E}[s_X^2] = \sigma^2 \quad \& \quad \mathbf{V}[s_X^2] = \frac{2\sigma^4}{n-1} \to 0$$

Distribuzione della media campionaria

$$Z = \frac{\sqrt{n}(\bar{X} - \mu)}{\sigma} \sim \mathcal{N}(0,1)$$



$$\mathbf{P}\left\{-z_\alpha \leq Z \leq +z_\alpha\right\} = 1 - \alpha$$

Intervalli di confidenza per la media di una normale

Se σ^2 è nota, allora

$$\mathbf{P}\left\{\bar{X} - z_\alpha \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X} + z_\alpha \frac{\sigma}{\sqrt{n}}\right\} = 1 - \alpha$$

altrimenti si considera

$$T = \frac{\sqrt{n}(\bar{X} - \mu)}{s_X} \sim \mathcal{T}(n-1)$$

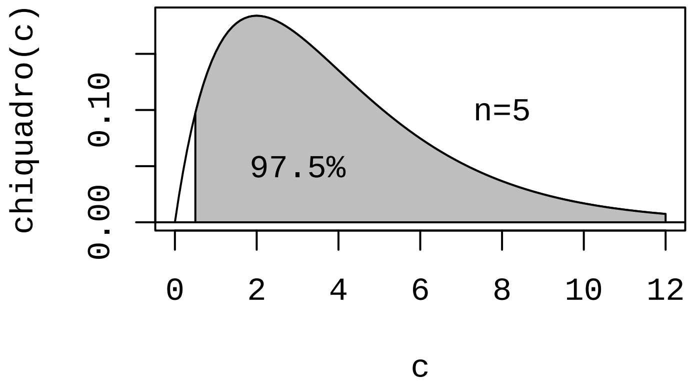
e si trova

$$\mathbf{P}\left\{\bar{X} - t_{\alpha,n-1} \frac{s_X}{\sqrt{n}} \leq \mu \leq \bar{X} + t_{\alpha,n-1} \frac{s_X}{\sqrt{n}}\right\} = 1 - \alpha$$

che converge al precedente per $n \to \infty$.

Distribuzione della varianza campionaria

$$C = ns_X^2/\sigma^2 \sim \chi^2_{n-1}$$



$$\mathbf{P}\{C > c_\alpha\} = 1 - \alpha$$

Intervallo di confidenza per la varianza di una normale

Supponendo senz'altro che μ non sia nota, allora

$$\mathbf{P}\left\{\frac{(n-1)s_X^2}{c_{1-\alpha/2}} \leq \sigma^2 \leq \frac{(n-1)s_X^2}{c_{\alpha/2}}\right\} = 1-\alpha$$

ovvero

$$\mathbf{P}\left\{\sqrt{\frac{(n-1)s_X^2}{c_{1-\alpha/2}}} \leq \sigma \leq \sqrt{\frac{(n-1)s_X^2}{c_{\alpha/2}}}\right\} = 1-\alpha$$

per la deviazione standard.

Verifica dell'ipotesi di nullità della media

Il campione fornisce sufficiente evidenza statistica contro l'ipotesi che la media della popolazione sia nulla?

$$H_0 : \mu = 0$$

Non si può rispondere affermativamente se lo zero appartiene all'intervallo di confidenza per la media!

$$\bar{X} - t_{\alpha,n-1}\frac{s_X}{\sqrt{n}} \leq 0 \leq \bar{X} + t_{\alpha,n-1}\frac{s_X}{\sqrt{n}}$$

Significatività e potenza della verifica

La significatività α è la probabilità di rifiutare H_0 se essa è vera:

$$\alpha = \mathbf{P}_0 \left\{ \frac{\sqrt{n}|\bar{X}|}{s_X} > t_{\alpha,n-1} \right\}$$

La potenza β è la probabilità di rifiutare H_0 se essa è falsa:

$$\beta(\mu) = \mathbf{P}_\mu \left\{ \frac{\sqrt{n}|\bar{X}|}{s_X} > t_{\alpha,n-1} \right\}$$

Si noti come la potenza dipenda dall'effettivo valore di $\mu \neq 0$.

Il cosiddetto "p-value"

Se si è osservato il valore

$$t = \frac{\sqrt{n}|\bar{x}|}{s_x}$$

per la statistica test

$$T = \frac{\sqrt{n}|\bar{X}|}{s_X}$$

il "p-value" del campione

$$p = \mathbf{P}_0 \left\{ \frac{\sqrt{n}|\bar{X}|}{s_X} > \frac{\sqrt{n}|\bar{x}|}{s_x} \right\}$$

è più piccolo valore di α per cui H_0 viene rifiutata.

È un dado onesto?

Si considerano i conteggi relativi a n lanci di un dado: N_k, $k = 1 \ldots 6$.

Il dado è descritto dalle probabilità di faccia: π_k, $k = 1 \ldots 6$.

L'ipotesi nulla è che il dado sia onesto, cioè $H_0 : \pi_1 = \cdots = \pi_6$.

Si rifiuta H_0 quando la statistica test

$$C = \sum_{k=1}^{6} \frac{(N_k - n/6)^2}{n/6} \sim> \chi^2_5, \quad n \to \infty$$

è "grande" rispetto al livello di significatività prescelto.

Sono indipendenti?

Si considerano i conteggi congiunti N_{ij}, $i = 1 \dots I$,$j = 1 \dots J$, relativi a due variabili nominali aventi rispettivamente I e J livelli.

Il modello statistico è descritto dalle probabilità congiunte π_{ij}.

L'ipotesi nulla è che le variabili siano indipendenti, cioè $H_0 : \pi_{ij} = \pi_i \pi_j$.

Si rifiuta H_0 quando la statistica test

$$C = \sum_{i=1}^{I} \sum_{j=1}^{J} \frac{(N_{ij} - N_i N_j / N)^2}{N_i N_j / N} \sim> \chi^2_{(I-1)(J-1)}, \quad N \to \infty$$

è "grande" rispetto al livello di significatività prescelto.

Queste note sono state composte in LaTeX, consultando il libro

Keeping (1962), "Introduction to Statistical Inference", Dover

e impiegando il linguaggio R per la realizzazione dei grafici.