Statistica inferenziale- II

STATISTICA NUMERICA

A.Y. 2022-2023

Stima di parametri:

STATISTICA NUMERICA, CAP. 6.2.1-6.2.3

Stime di parametri

Generalizziamo ora il concetto di stima di parametri di una distribuzione che abbiamo visto per ora Applicato alla media e alla deviazione standard (o varianza).

La *stima puntuale di un parametro* θ (con le lettere greche indichiamo il parametro di interesse) è un numero che può essere un valore sensibile di θ.

Una stima puntuale è ottenuta scegliendo un'opportuna statistica e calcolando il suo valore a partire da campioni casuali. La statistica scelta è detta *stimatore puntuale di* θ e si indica generalmente con $\hat{\theta}$.

Stime di parametri

La stima di un parametro però non fornisce da sola sufficienti informazioni riguardo alla sua affidabilità. È quindi necssario affiancare alla stima puntuale la stima di un intervallo di valori possibili, detto *intervallo di confidenza*, ottenuto a partire da valori che 'misurano' il *grado di affidabilità* della stima. Più piccolo è l'intervallo stesso e più affidabile è la stima. Un grande intervallo di confidenza è segno di incertezza nella stima calcolata.

Stime di parametri

Supponiamo di voler stimare il parametro θ con lo stimatore θ. Il meglio che si possa ottenere sarebbe che $\hat{\theta} = \theta$ per ogni campione considerato. In realtà, $\hat{\theta}$ è una variabile aleatoria, quindi può capitare che per un certo campione $\hat{\theta} < \theta$ e per un altro $\hat{\theta} > \theta$. In generale, si ha che:

$$\hat{\theta} = \theta + \text{errore di stima.}$$

Stime di parametri

Ovviamente, più è piccolo l'errore, migliore è l'estimatore.

Per valutare l'errore commesso nella stima, si calcolano delle misure di errore come, per esempio, l'errore quadratico medio (Mean Square Error) fra il valore stimato $\hat{\theta}$ e il valore esatto θ:

$$MSE = \frac{\sum_{i=1}^{n} (\hat{\theta} - \theta)^2}{n}.$$

Stime di parametri

Nel confronto fra due stimatori $\hat{\theta}_1$ e $\hat{\theta}_2$, viene considerato migliore quello che ha l'errore quadratico medio minore. È difficile trovare uno stimatore che sia sempre migliore degli altri per ogni θ, visto che lo stimatore stesso dipende da θ.

Uno stimatore puntuale $\hat{\theta}$ è detto *non distorto* se, detta $E(\hat{\theta})$ la media della variabile aleatoria $\hat{\theta}$, $E(\hat{\theta}) = \theta$ per ogni possibile valore di θ.

Se $\hat{\theta}$ è distorto, allora la differenza $E(\hat{\theta}) - E(\theta)$ si dice la *distorsione* di $\hat{\theta}$.

Stime di parametri

Qual è l'importanza della proprietà di distorsione dello stimatore?

Principio della stima non distorta

Quando si deve scegliere fra diversi stimatori, scegliere quello *non distorto*.

Se ci sono più stimatori non distorti di uno stesso parametro allora si sceglie in base alla varianza dello stimatore.

In particolare:

Principio della varianza minima. Fra tutti gli stimatori non distorti di θ, scegliere quello con varianza minima. Il risultante stimatore $\hat{\theta}$ è detto *stimatore non distorto di minima varianza* (MVUE).

SRS(n)

Siano $X_1, X_2, \ldots X_n$ campioni casuali da una distribuzione con media μ. Allora lo stimatore $\bar{X}$ è *uno stimatore non distorto della media* μ.

$$\boxed{\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i}$$

$\bar{X} \sim N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$

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

è *non distorto per stimare la varianza* σ^2. Lo stimatore che ha come denominatore n:

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

è distorto e la sua distorsione è $(n-1)/n\sigma^2 - \sigma^2 = -\sigma^2/n$. Essendo la distorsione di P negativa, lo stimatore P tende a sottostimare la varianza σ^2.

Stimatore non distorto della media e della varianza