Statistica - Introduzione a R

a cura di Antonio Iovanella

iovanella@disp.uniroma2.it
http://www.disp.uniroma2.it/Users/iovanella

Intervalli di confidenza

Introduzione

Note generali

La stima puntuale permette di ottenere valori per i parametri di una funzione ma in alcuni casi può risultare insoddisfacente.

Per esempio, a volte è auspicabile avere la stima puntuale associata ad un intervallo centrato sulla stima di θ che misuri il possibile errore di stima con in più una misura della fiducia che il valore stimato cada entro tale intervallo.

Gli ***intervalli di confidenza per la media*** forniscono un campo di variazione (centrato sulla media campionaria) all'interno del quale ci si aspetta di trovare il parametro incognito μ.

Ad ogni intervallo di confidenza viene associato un ***livello di confidenza*** ($1 - \alpha$) che rappresenta il grado di attendibilità del nostro intervallo.

Introduzione

Obbiettivo

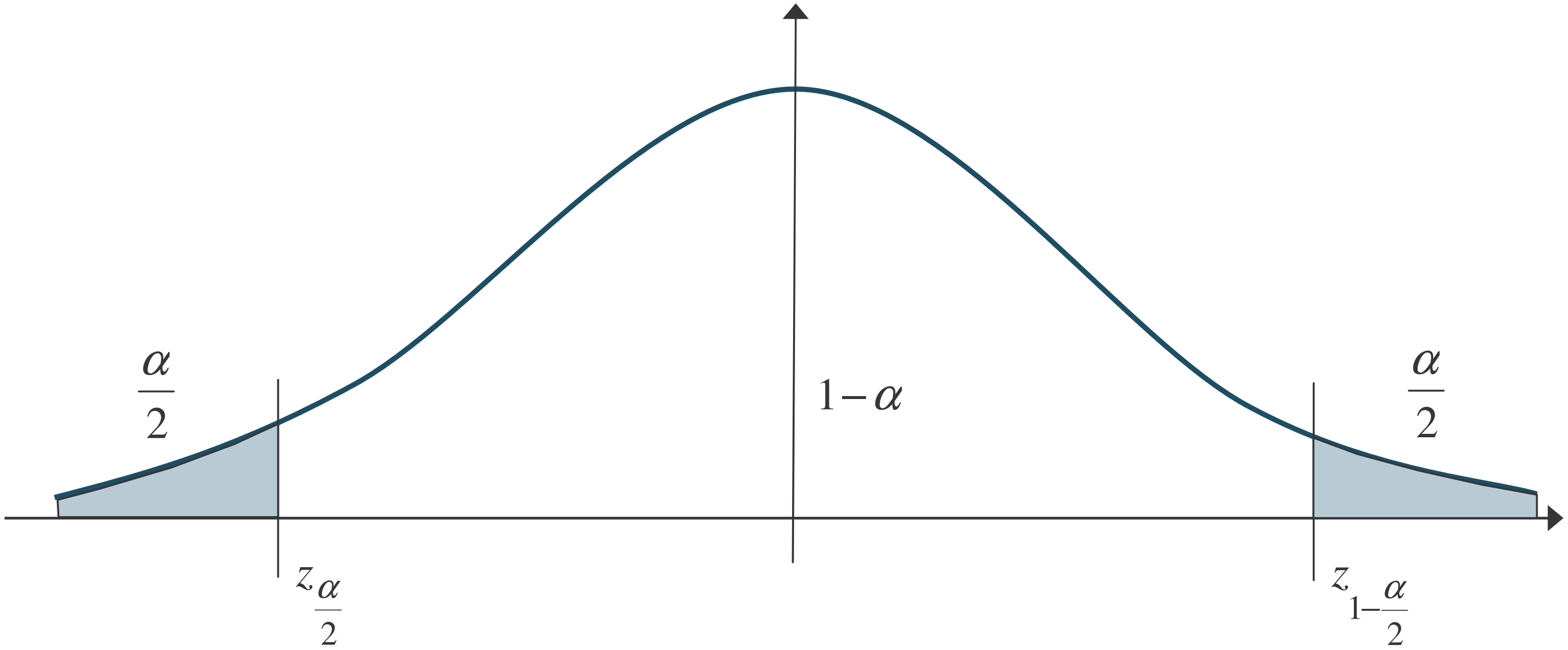
Sia X_1, X_2, …, X_n un campione i.i.d. di variabili casuali gaussiane di media e varianza (incognite) μ, σ^2.

Se consideriamo la media campionaria $\overline{X}_n$, questa ha media μ e varianza σ^2/n.

Il nostro obbiettivo è quello di determinare un intervallo di valori (a, b) che contenga il valore incognito μ, ovvero:

$$P\left(a < \frac{\overline{X}_n - \mu}{\frac{\sigma}{\sqrt{n}}} < b \right) = 1 - \alpha$$

Introduzione



$$\mu \in \left(\overline{X}_n \pm z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \right)$$

Introduzione

Nota 1

Il livello di confidenza è la frequenza degli intervalli aleatori definiti dalla formula precedente che contengono il valore incognito μ. Quindi è scorretto confondere il livello di confidenza con la probabilità che μ sia contenuto nell'intervallo.

Nota 2

Spesso non siamo a conoscenza della varianza σ^2. In questo caso dobbiamo ricorrere ad uno stimatore e l'intervallo di confidenza diventa:

$$\mu \in \left(\overline{X}_n \pm t_{1-\frac{\alpha}{2}}^{(n-1)} \sqrt{\frac{\overline{s}_n^{\,2}}{n}} \right)$$

Esempio

Si voglia determinare il valor medio del peso (in *mg*) di alcuni granelli di polvere asportati da una piastra di silicio, con un livello di confidenza del 95%.

```
x <- c(0.39, 0.68, 0.82, 1.35, 1.38, 1.62, 1.70, 1.71,
1.85, 2.14, 2.89, 3.69)
s2 <- var(x)
mx <- mean(x)
n <- length(x)
a <- qt(0.975, df = n - 1) * sqrt(s2 / n)
l.inf <- mx - a
l.sup <- mx + a
cat("(",l.inf,":",l.sup,")\n")
```

##ESERCIZIO

Esercizio

A partire dall'esempio precedente, calcolare il livello di confidenza del 95% basato sulla distribuzione normale.
I dati sono:

0.39, 0.68, 0.82, 1.35, 1.38, 1.62, 1.70, 1.71, 1.85, 2.14, 2.89, 3.69

##ESEMPIO

Esempio

Ritornando all'esempio precedente, R mette a disposizione una funzione, chiamata **`t.test`** che permette il calcolo diretto degli intervalli di confidenza.

```
t.test(x, con.lev = 0.95)



data:  x
t = 6.3305, df = 11, p-value = 5.595e-05
alternative hypothesis: true mean is not equal to 0
95 percent confidence interval:
 1.099159 2.270841
sample estimates:
mean of x
    1.685
```

Ampiezza del campione

Nota

In generale, all'aumentare di *n* l'intervallo di confidenza si restringe. Nella pratica spesso si vuole restringere l'intervallo ad una larghezza *C*, fermo restando un dato livello di confidenza 1- α. Se indichiamo con L(*n*, α) la lunghezza di un intervallo di confidenza, allora:

$$L(n,\alpha) = 2z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} < C$$

ovvero

$$n > \left(2z_{1-\frac{\alpha}{2}} \frac{\sigma}{C} \right)^2$$

Esempio

In un esame di psicologia vengono misurati i tempi di reazione di 100 individui, riscontrando un tempo medio di 1 secondo. Da studi pregressi, lo scarto quadratico σ è noto essere pari a 0.05 secondi.
Quale deve essere il numero minimo di osservazioni campionarie *n* per avere un'ampiezza dell'intervallo pari al più a 0.02 secondi ed un intervallo di confidenza pari al 99%.

Per la relazione appena vista:

$$2z_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}} < 0.02$$

##ESEMPIO

dato che:

$$z_{1-\frac{\alpha}{2}} = z_{0.995} = 2.63$$

ottengo che:

$$n \geq \left(2 \cdot 2{,}63 \frac{0.05}{0.02} \right)^2 = 172{,}45$$

Con R:

```
n <- 100
a <- qt(0.995, df = n - 1)
n1 <- (2 * a * (0.05/0.02))^2
cat("(",ceiling(n1),")\n")
```

##ESERCIZIO

Esercizio

A partire dall'esempio precedente, Quale deve essere il numero minimo di osservazioni campionarie n per avere un'ampiezza dell'intervallo pari al più a 0.02 milligrammi ed un intervallo di confidenza pari al 99%, basandoci sulla distribuzione normale.

Stima dell'intervallo di confidenza per la varianza

Gli ***intervalli di confidenza per la varianza*** forniscono un campo di variazione all'interno del quale ci si aspetta di trovare il parametro incognito σ^2.

Anche in questo caso ad ogni intervallo di confidenza viene associato un ***livello di confidenza*** (1 - α) che rappresenta il grado di attendibilità del nostro intervallo.

Per la varianza abbiamo che:

$$\frac{(n-1)S_n^2}{\sigma^2} \approx \mathrm{X}_{n-1}^2$$

Nota

Dato che la varianza non può essere negativa, cioè:

$$\sigma^2 \in (0, c)$$

allora:

$$P\left(\frac{(n-1)S_n^2}{\sigma^2} > \frac{(n-1)S_n^2}{c}\right) = 1 - \alpha$$

cioè:

$$\sigma^2 > \frac{(n-1)S_n^2}{\mathrm{X}_{n-1}^2}$$

Esempio

Costruiamo una funzione, chiamata **`ic.var`**, in grado di calcolare, dato un campione in ingresso, l'intervallo di confidenza per la varianza.

```
ic.var <-
function(x, conf.level){
   alfa <- 1 - conf.level
   n <- length(x)
   l.inf <- 0
   l.sup <- (n - 1) * var(x)/qchisq(alfa, df = n - 1)
   c(l.inf, l.sup)
}
```

##ESERCIZIO

Esercizio

Costruire una funzione, in grado di calcolare, dato un campione in ingresso, l’intervallo di confidenza per la media e la varianza.