

Intervalli di confidenza

	Task Description	Due Date	Status
☐	Stima puntuale e stima intervallare.		
☐	Stima intervallare del parametro μ di una popolazione normale (caso di varianza σ^2 della popolazione nota): intervallo casuale, intervallo di confidenza e sua interpretazione alla luce del principio del campionamento ripetuto.		
☐	Stima intervallare del parametro μ di una popolazione normale (caso di varianza σ^2 della popolazione incognita).		
☐	Intervalli asintotici per la media di una generica popolazione.		
☐	Riepilogo stima intervallare per la media: caso di varianza nota e varianza incognita.		
☐	Stima intervallare del parametro σ^2 di una popolazione normale (caso di media incognita).		
☐	Stima intervallare (per grandi campioni) del parametro π di una popolazione bernoulliana		
☐	Determinazione della numerosità necessaria per ottenere una precisione data degli intervalli di confidenza per la media e la proporzione		
☐	Riepilogo stima intervallari (1): casi ad un campione		
☐	Riepilogo stima intervallare (2): casi a due campioni		

STIMA PARAMETRO ~ stima puntuale

- selezionare il campione (osservato)
- selezionare la statistica opportuna
- calcoliamo la statistica sul campione osservato

↓
stima

STIMA puntuale

Obiettivo: a partire da un campione di n osservazioni (oppure da due campioni nel caso di problemi su due popolazioni) voglio ottenere un valore "plausibile" per θ (oppure per $\theta_x - \theta_y$ o θ_x/θ_y)



Considero uno stimatore T_n che soddisfa le proprietà introdotte, lo calcolo sul campione osservato e ottengo la stima puntuale t_n



sarebbe utile avere un'informazione circa l'accuratezza di t_n

$$P(T_n = \theta)$$

Poiché le distribuzioni campionarie degli stimatori sono continue
(normale, t di Student, chi-quadrato, F di Fisher)

$$P(T_m = \theta) = 0$$

↙
l'unica possibilità è di fare riferimento ad un intervallo
di valori

↓
 $P(T_m \text{ cade in un intervallo che contiene } \theta)$ → STIMA
INTERVALLARE

Stima intervallare $\rightarrow$ per ottenere una misura del grado di precisione della stima puntuale si affianca una stima su intervallo

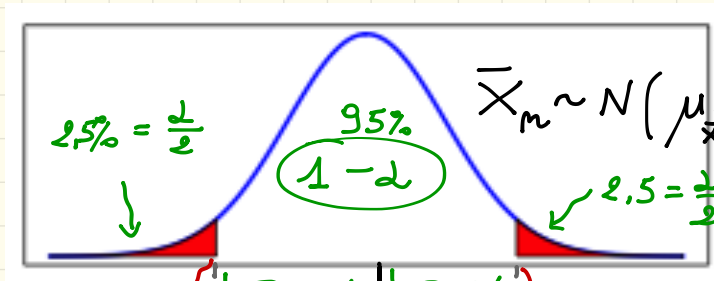
$P(\text{stimatore} \in \text{intervallo dato})$ è calcolabile

$\downarrow$
in base alla distrib.
campionaria di T_n

ESEMPIO $\rightarrow$ problema "deduttivo": ragioniamo su quello che ci possiamo aspettare per T_n (caso di $\bar{X}_n$)

Una macchina produce bulbini il cui peso ha una distribuzione normale con media $\mu = 63$ gr e varianza $\sigma^2 = 0,8$.
Scegliendo a caso $n = 8$ bulbini, qual è l'intervallo che contiene il 95% delle possibili medie campionarie?

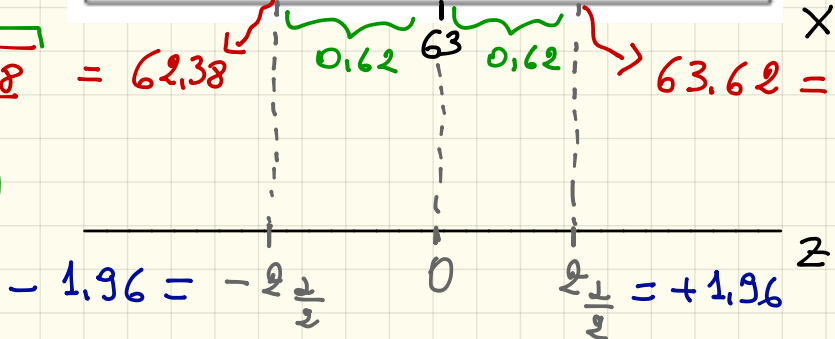
$$\alpha = 5\% = 0,05$$



$$\bar{X}_n \sim N\left(\mu_{\bar{X}} = 63, \sigma_{\bar{X}}^2 = \frac{0,8}{8}\right)$$

$$63 - \sqrt{0,62} \times 1,96 \times \sqrt{\frac{0,8}{8}} = 62,38$$

$$\mu_{\bar{X}} - z_{\frac{\alpha}{2}} \sigma_{\bar{X}}$$



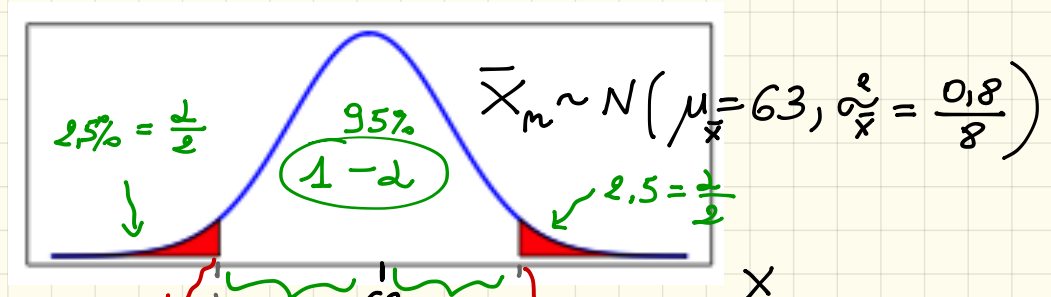
$$63 + \sqrt{0,62} \times 1,96 \times \sqrt{\frac{0,8}{8}} = 63,62$$

$$\mu_{\bar{X}} + z_{\frac{\alpha}{2}} \sigma_{\bar{X}}$$

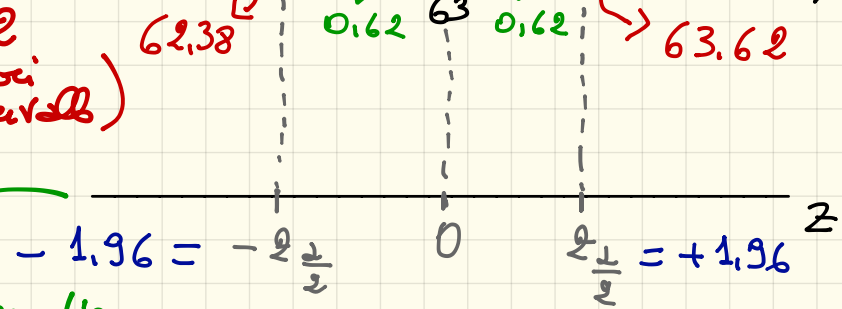
ESEMPLO $\rightarrow$ problema "deduttivo": ragioniamo su quello che ci possiamo aspettare per T_n (caso di $\bar{X}_n$)

Cosa possiamo concludere?

$\alpha = 5\% = 0,05$



il 95% delle medie campionarie si trova tra 62.38 e 63.62 (e solo il 5% fuori da questo intervallo)



il 95% delle medie campionarie sulla scala z (opportunamente standardizzate) si trovano tra -1.96 e $+1.96$

PROBLEMA INVERSO $\rightarrow$ tipico dell'inferenza

Una macchina produce bulloni il cui peso ha una distribuzione normale con media μ incognita e varianza $\sigma^2 = 0.8$.

Scegli a caso 8 bulloni, il loro peso medio $\bar{x}$ risultato pari a 62.6 gr. Quel $\bar{x}$ è l'intervallo che, con probabilità 0.95, contiene il parametro incognito μ ?

(con un grado di fiducia pari a 0.95)

POPOLAZIONE

$$X \sim N(\mu, \sigma^2 = 0.8)$$

STATISTICA

$$\bar{X}_n \sim N\left(\mu, \frac{\sigma^2}{n} = \frac{0.8}{8}\right)$$

proprietà
ripresolutive

$\bar{X}_n$ è uno
stimatore non
distorto

da quanto noto
sullo stimatore $\bar{X}_n$

$$Z \sim N(0, 1)$$

$$P(-1.96 \leq Z \leq +1.96) = 0.95$$

scegliendo $\alpha = 0.05$

$$-\frac{z_{\frac{\alpha}{2}}}{1}$$

$$+\frac{z_{\frac{\alpha}{2}}}{1}$$

$$1 - \alpha$$

livello di confidenza o
di fiducia

questo vale per qualunque
 $N(0, 1)$

↓ e quindi anche per $\bar{X}_n$ opportunamente standardizzate

$$P\left(-1.96 \leq \frac{\bar{X}_n - E(\bar{X}_n)}{\sqrt{\text{Var}(\bar{X}_n)}} \leq +1.96\right) = 0.95$$

$$P\left(-1.96 \leq \frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \leq +1.96\right) = 0.95$$

→ si isola al centro
della disuguaglianza
 μ

↪ statistica pivotale

$$P\left(-1.96 \leq \frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \leq +1.96\right) = 0.95$$

Moltiplichiamo per $\sigma/\sqrt{n}$:

$$P\left(-1.96 \frac{\sigma}{\sqrt{n}} \leq \frac{\bar{X}_n - \mu}{\cancel{\sigma/\sqrt{n}}} \leq +1.96 \frac{\sigma}{\sqrt{n}}\right) = 0.95$$

Sottraiamo $\bar{X}_n$:

$$P\left(-\bar{X}_n - 1.96 \frac{\sigma}{\sqrt{n}} \leq \cancel{\bar{X}_n} - \mu - \cancel{\bar{X}_n} \leq -\bar{X}_n + 1.96 \frac{\sigma}{\sqrt{n}}\right) = 0.95$$

Moltiplichiamo per -1 :

$$P\left(\bar{X}_n + 1.96 \frac{\sigma}{\sqrt{n}} \geq \mu \geq \bar{X}_n - 1.96 \frac{\sigma}{\sqrt{n}}\right) = 0.95$$

Riordiniamo i due membri della disuguaglianza:

$$P\left(\bar{X}_n - 1.96 \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X}_n + 1.96 \frac{\sigma}{\sqrt{n}}\right) = 0.95$$

$\bar{X}_n \pm$ margine
errore

INTERVALLO
CASUALE

← σ dipende da d e da $\sqrt{\text{Var}(\bar{X}_n)}$

$$P\left(\bar{X}_m - 1,96 \frac{\sigma}{\sqrt{m}} \leq \mu \leq \bar{X}_m + 1,96 \frac{\sigma}{\sqrt{m}}\right) = 0,95$$

v.c. limite inferiore

v.c. limite superiore

INTERVALLO CASUALE

↓
coppie di v.c.

A questo punto inseriamo le informazioni campionarie:

~~$$P\left(\bar{X}_m - 1,96 \times \frac{0,8}{\sqrt{8}} \leq \mu \leq \bar{X}_m + 1,96 \times \frac{0,8}{\sqrt{8}}\right) = 0,95$$~~

il valore della v.c. $\bar{X}_m$ sul particolare campione osservato

~~$$P\left(62,6 - 0,62 \leq \mu \leq 62,6 + 0,62\right) = 0,95$$~~

margine errore

margine errore

(realizzazione intervalli casuali)

~~$$P(61,98 \leq \mu \leq 63,22) = 0,95$$~~

INTERVALLO DI CONFIDENZA O DI FIDUCIA

NOTA:

- Il campione osservato $\bar{x}$ è la realizzazione della n -pla di v.c. campione casuale

$$\begin{array}{ccccccc} X_1, & X_2, & \dots, & X_n \\ \downarrow & \downarrow & & \downarrow \\ x_1, & x_2, & \dots, & x_n \end{array}$$

- La stima t_n è la realizzazione della v.c. stimatore

$$\begin{array}{ccccccc} T_n = T(X_1, & X_2, & \dots, & X_n) \\ \downarrow & \downarrow & & \downarrow \\ t_n = T(x_1, & x_2, & \dots, & x_n) \end{array}$$

- L'intervallo di confidenza $\bar{I}$ è la realizzazione della coppia di v.c. intervallo casuale

~~$P(61.98 \leq \mu \leq 63.22) = 0.95$~~ $\rightarrow$ l'interpretaz. probabilistica mi porterebbe a concludere che:

"la media μ si trova tra 61.98 e 63.22 con probab. 0.95"

queste non $\bar{x}$ però
una v.c. ma una costante
(incognita)

costante $\in$ intervallo

la prob. vale 1

costante $\notin$ intervallo

la prob. vale 0

INFORMAZIONE
NON
DISPONIBILE

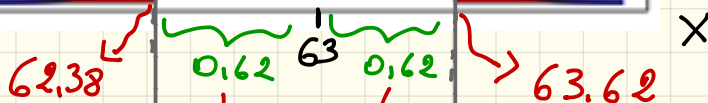
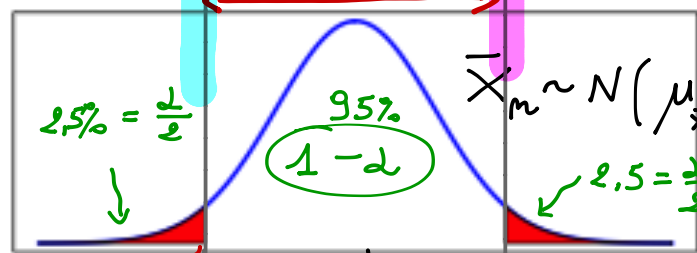
Per evitare confusione nell'interpretazione conviene esprimere l'intervallo di confidenza senza la notazione probabilistica:

$$I_{0.95}(\mu) = [61.98; 63.22]$$

(?) come interpretare allora questo valore (?)
1- α

Nel problema diretto avevamo concluso che il 95% delle possibili medie campionarie cadeva tra 62.38 e 63.62

PROBLEMA DIRETTO



corrisponde al margine di errore visto nel problema inverso

$\frac{\alpha}{2}$ delle possibili medie campionarie sono < 62.38

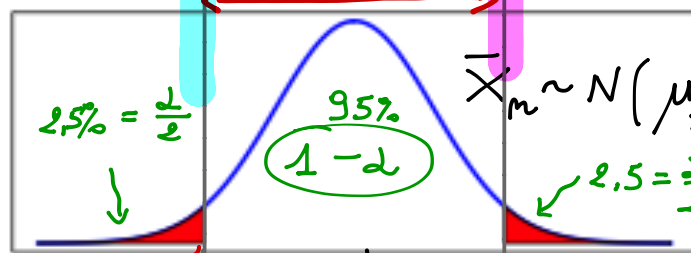
se $\bar{X}_n < 62.38$ l'intervallo non copre μ

$\frac{\alpha}{2}$ delle possibili medie campionarie sono > 63.62

se $\bar{X}_n > 63.62$ l'intervallo non copre μ

Nel problema diretto avevamo concluso che il 95% delle possibili medie campionarie cade tra 62,38 e 63,62

PROBLEMA DIRETTO



$$\bar{X}_n \sim N\left(\mu_{\bar{X}} = 63, \sigma_{\bar{X}}^2 = \frac{0,8}{8}\right)$$

62,38 0,62 63 0,62 63,62

intervalli ottenuti a partire da media campionaria a sinistra di 62,38 non contengono (coprono) μ

il 2,5% delle possibili medie campionarie cade a sinistra di 62,38

intervalli ottenuti a partire da medie campionarie all'interno di questa banda contengono effettivamente μ (coprono μ)

il 95% delle medie campionarie cade in questa banda

intervalli ottenuti a partire da media campionaria a destra di 63,62 non contengono (coprono) μ

il 2,5% delle possibili medie campionarie cade a destra di 63,62

Il metodo di costruzione degli intervalli produce intervalli
"buoni" nell' $1-\alpha\%$ dei casi e intervalli "cattivi" nell' $\alpha\%$
dei casi

↓
che coprono μ

95% nel
caso dell'esempio

↓
che non coprono
 μ

5% nel-
l'esempio

sul mio unico intervallo di confidenza disponibile (unico
perché costruito a partire da un unico campione osservato) non
posso avere nessuna informazione

↓
la fiducia non è sul particolare intervallo ottenuto ma nel
metodo utilizzato per ottenerlo

↳ siccome il metodo produce $(1-\alpha)\%$ intervalli buoni
posso solo sperare che quello effettivamente disponibile
sia tra quelli buoni

$$IC_{(1-\alpha)}[\mu] = \left[\bar{x}_n - z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} ; \bar{x}_n + z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \right]$$

ponendo $\alpha = 0,05$

$$1 - \alpha = 0,95$$

ponendo $\alpha = 0,01$

$$1 - \alpha = 0,99$$

ponendo $\alpha = 0,1$
 $1 - \alpha = 0,90$

1,96

2,58

1,64

LIVELLI DI
CONFIDENZA

TRE LIVELLI COMUNEMENTE
USATI

$\alpha = 0,1$ (10%)

$\alpha = 0,05$ (5%)

$\alpha = 0,01$ (1%)

$1 - \alpha = 0,90$ (90%)

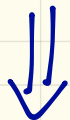
$1 - \alpha = 0,95$ (95%)

$1 - \alpha = 0,99$ (99%)

Se 1-2 cresce (o decresce) aumenta il grado di copertura del mio intervallo, ovvero il livello di fiducia della mia stima (diminuisce l'incertezza legata alla stima intervallo)



aumenta però anche il valore di $z_{\frac{\alpha}{2}}$ (il percentile si sposta verso l'esterno)



però la maggior accuratezza in termini di un margine di errore più alto



in termini pratici la stima ottenuta è meno utile



minore capacità informativa della stima intervallo

È necessario un bilanciamento tra livelli di incertezza e capacità informativa dell'intervallo