STATISTICA

Intervalli di confidenza

Inferenza sulla **media di una popolazione**

$X_1, X_2, \dots, X_n$ i.i.d, $E(X_i) = \mu$, $Var(X_i) = \sigma^2$

μ, σ^2 in generale non noti

media campionaria

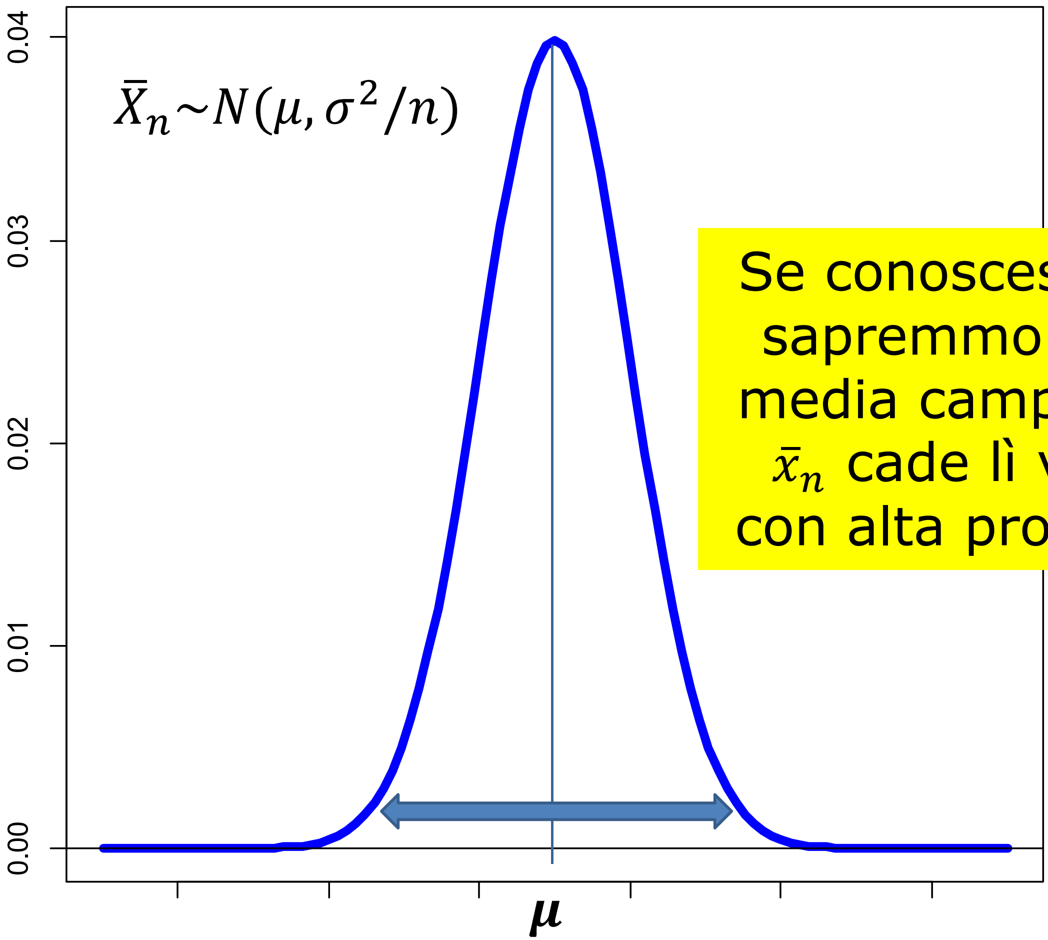
$$\boxed{\bar{X}_n = \frac{X_1 + X_2 + \cdots + X_n}{n}}$$

modello di tutte le possibili stime di μ, ***prima*** di estrarre il campione.

$$(x_1, x_2, \dots, x_n) \Longrightarrow \boxed{\bar{x}_n = \frac{x_1 + x_2 + \cdots + x_n}{n}}$$

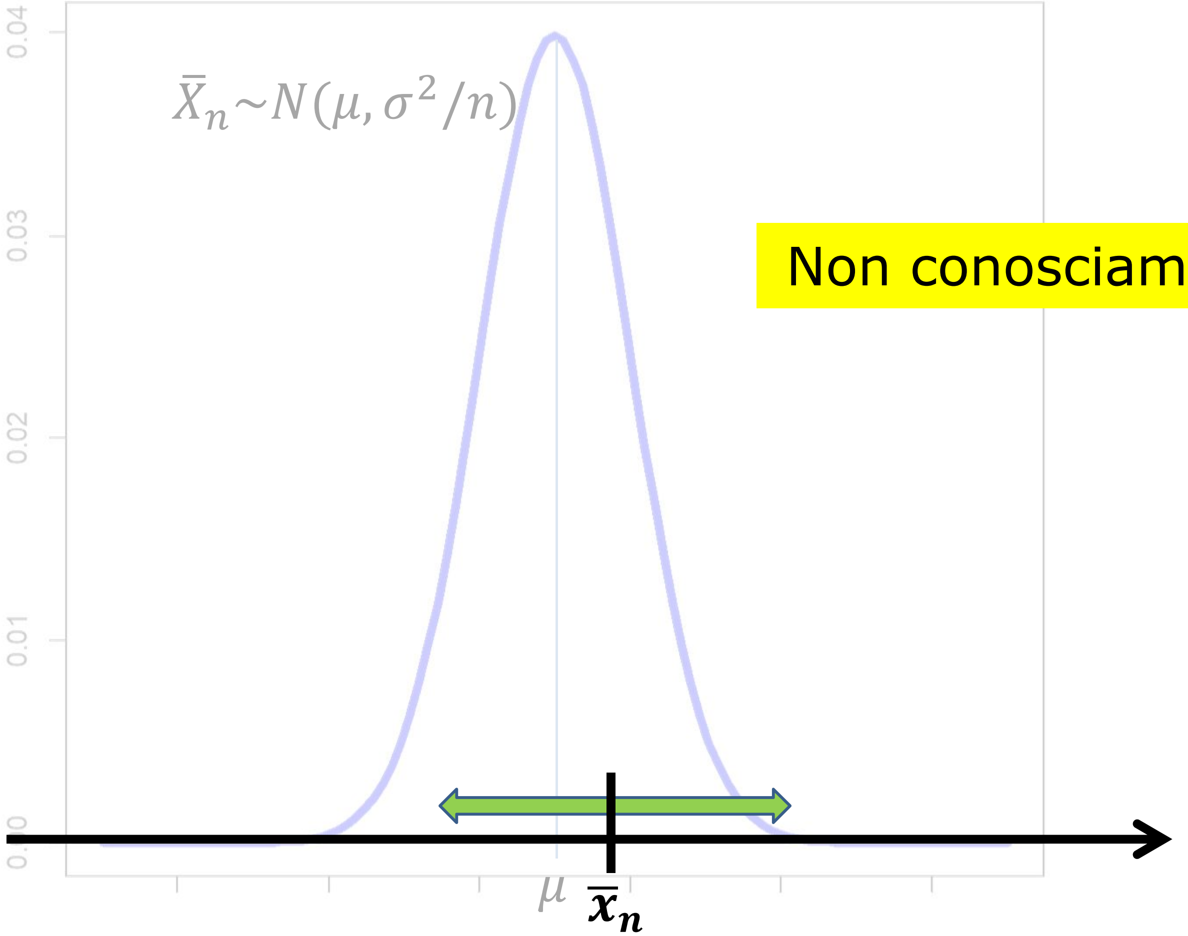
stima, sul campione

LA STIMA FORNITA DALLA MEDIA QUANTO È PRECISA?



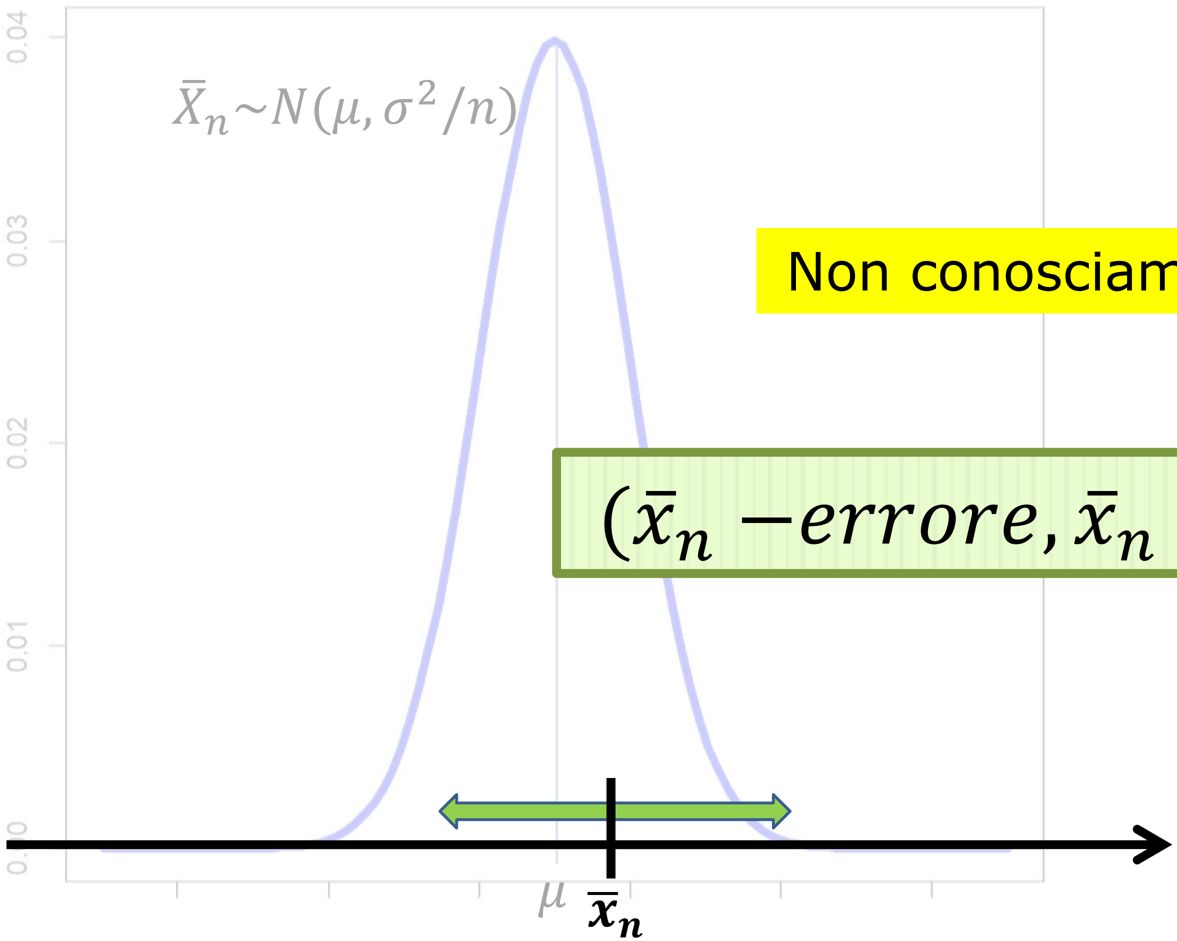
Se conoscessimo $\boldsymbol{\mu}$ sapremmo che la media campionaria $\bar{x}_n$ cade lì vicino, con alta probabilità

$$\color{red}{0.95} = P(\mu - a \leq \bar{X}_n \leq \mu + b) = \ldots$$



Non conosciamo $\boldsymbol{\mu}$!

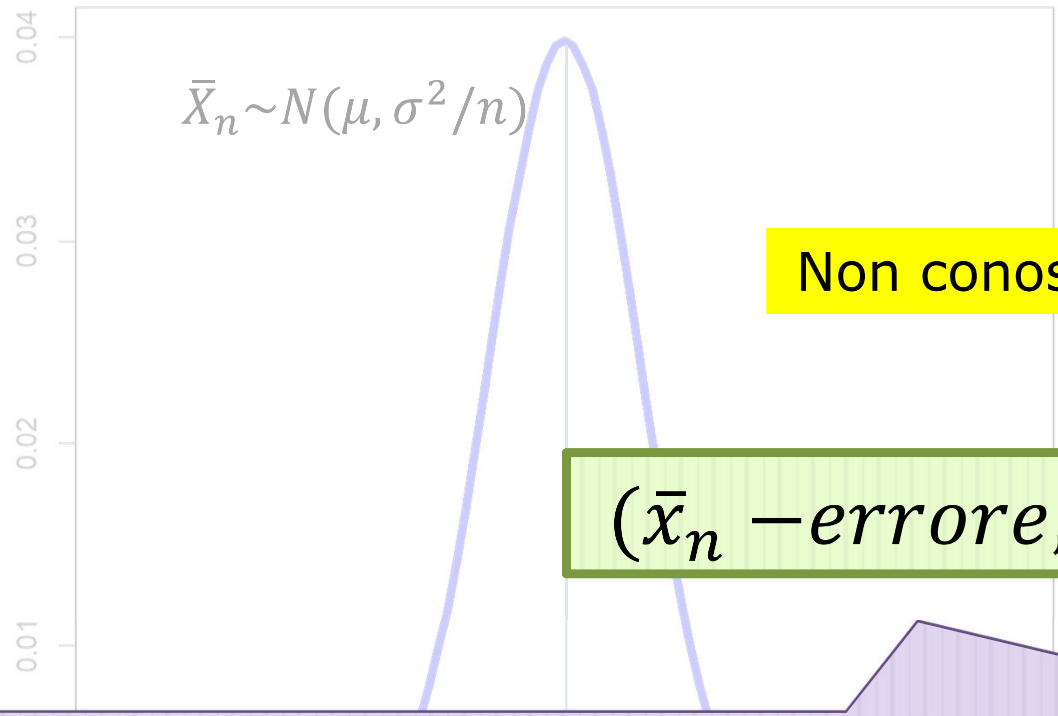
Se la teoria su $\bar{X}_n$ (lo stimatore) ci garantisce che $\bar{x}_n$ è una *buona stima*, possiamo essere fiduciosi che spostandoci un po' attorno a $\bar{x}_n$, anche se non ce ne accorgiamo, troviamo μ.



Non conosciamo $\boldsymbol{\mu}$!

$$(\bar{x}_n - errore, \bar{x}_n + errore)$$

Se la teoria su $\bar{X}_n$ (lo stimatore) ci garantisce che $\bar{x}_n$ è una *buona stima*, possiamo essere fiduciosi che spostandoci un po' attorno a $\bar{x}_n$, anche se non ce ne accorgiamo, troviamo μ.



Non conosciamo $\boldsymbol{\mu}$!

$$(\bar{x}_n - errore, \bar{x}_n + errore)$$

NON E’ POSSIBILE STABILIRE UN INTERVALLO “UTILE” IN CUI ESSERE SICURI AL 100% CHE POSSIAMO TROVARE IL VERO VALORE DI $\boldsymbol{\mu}$

POSSIAMO TROVARE UN INTERVALLO CHE CONTENGA IL VERO VALORE DI $\boldsymbol{\mu}$ CON UNA CONFIDENZA DEL 95%, per esempio.

PER FARE UNA VALUTAZIONE DI PROBABILITA’ SULL’INTERVALLO, CI SERVE IL “MODELLO PROBABILISTICO” PER L’INTERVALLO

Inferenza sulla **media di una popolazione**

$X_1, X_2, \ldots, X_n$ i.i.d, $\sim N(\mu, \sigma^2)$ **con σ^2 nota**

$\Rightarrow \; (\bar{X}_n - ERRORE \, , \, \bar{X}_n + ERRORE)$

E' il modello che prevede tutti i possibili **intervalli di confidenza** di livello $1-\alpha$ per μ

ERRORE : $z_{1-\frac{\alpha}{2}}\sqrt{\frac{\sigma^2}{n}}$

α valore tra 0 e 1, di solito **piccolo**

$z_{1-\frac{\alpha}{2}}$ quantile di ordine 1-$\alpha/2$ di una **gaussiana standard**

perchè $\frac{\bar{X}_n-\mu}{\sqrt{\frac{\sigma^2}{n}}} \sim N(0,1)$

Inferenza sulla **media di una popolazione**

$X_1, X_2, \dots, X_n$ i.i.d, $\sim N(\mu, \sigma^2)$ **con σ^2 nota**

$\Rightarrow \left(\bar{X}_n - ERRORE\,,\, \bar{X}_n + ERRORE\right)$

E' il modello che prevede tutti i possibili **intervalli di confidenza** di livello $1-\alpha$ per μ

ERRORE : $z_{1-\frac{\alpha}{2}}\sqrt{\frac{\sigma^2}{n}}$

DAL NOSTRO CAMPIONE:

$$\left(\bar{x}_n - z_{1-\frac{\alpha}{2}}\sqrt{\frac{\sigma^2}{n}}\,,\, \bar{x}_n + z_{1-\frac{\alpha}{2}}\sqrt{\frac{\sigma^2}{n}}\right)$$

α valore tra 0 e 1, di solito **piccolo**

$z_{1-\frac{\alpha}{2}}$ quantile di ordine 1-$\alpha/2$ di una **gaussiana standard**

perchè $\dfrac{\bar{X}_n - \mu}{\sqrt{\frac{\sigma^2}{n}}} \sim N(0,1)$

Inferenza sulla **media di una popolazione**

$X_1, X_2, \dots, X_n$ i.i.d, $\sim N(\mu, \sigma^2)$ **con σ^2 nota**

$$\Rightarrow \left(\bar{X}_n - z_{1-\frac{\alpha}{2}} \sqrt{\frac{\sigma^2}{n}}, \bar{X}_n + z_{1-\frac{\alpha}{2}} \sqrt{\frac{\sigma^2}{n}} \right)$$

E' il modello che prevede tutti i possibili **intervalli di confidenza** di livello $1-\alpha$ per $\boldsymbol{\mu}$

100 campioni
100 intervalli di conf. di liv. $1-\alpha = 0.95$

95 contengono il vero valore di μ

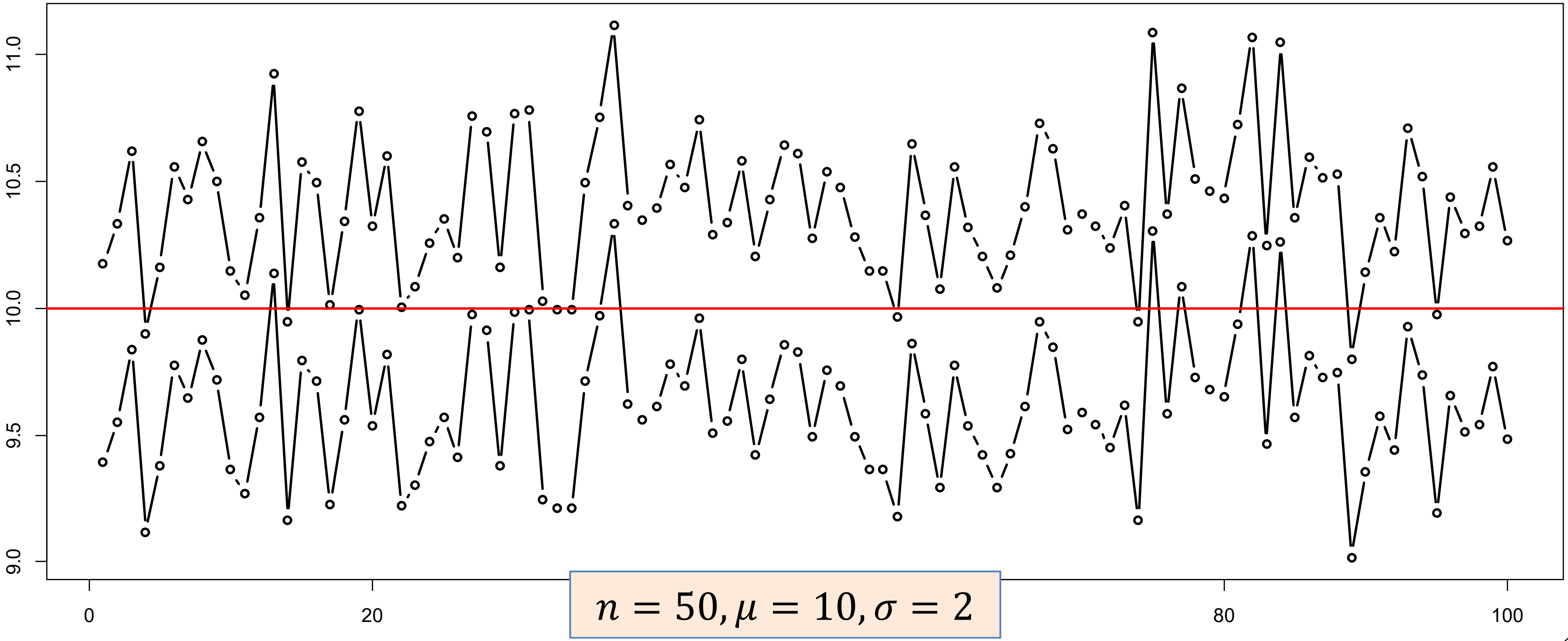
possiamo avere il 95% di fiducia che **il "nostro"** intervallo sia uno di questi.

Inferenza sulla **media di una popolazione**

$X_1, X_2, \dots, X_n$ i.i.d, $\sim N(\mu, \sigma^2)$ **con σ^2 nota**

$$\Rightarrow \left(\bar{X}_n - z_{1-\frac{\alpha}{2}}\sqrt{\frac{\sigma^2}{n}}\ ,\ \bar{X}_n + z_{1-\frac{\alpha}{2}}\sqrt{\frac{\sigma^2}{n}}\right)$$

E' il modello che prevede tutti i possibili **intervalli di confidenza** di livello $1-\alpha$ per μ



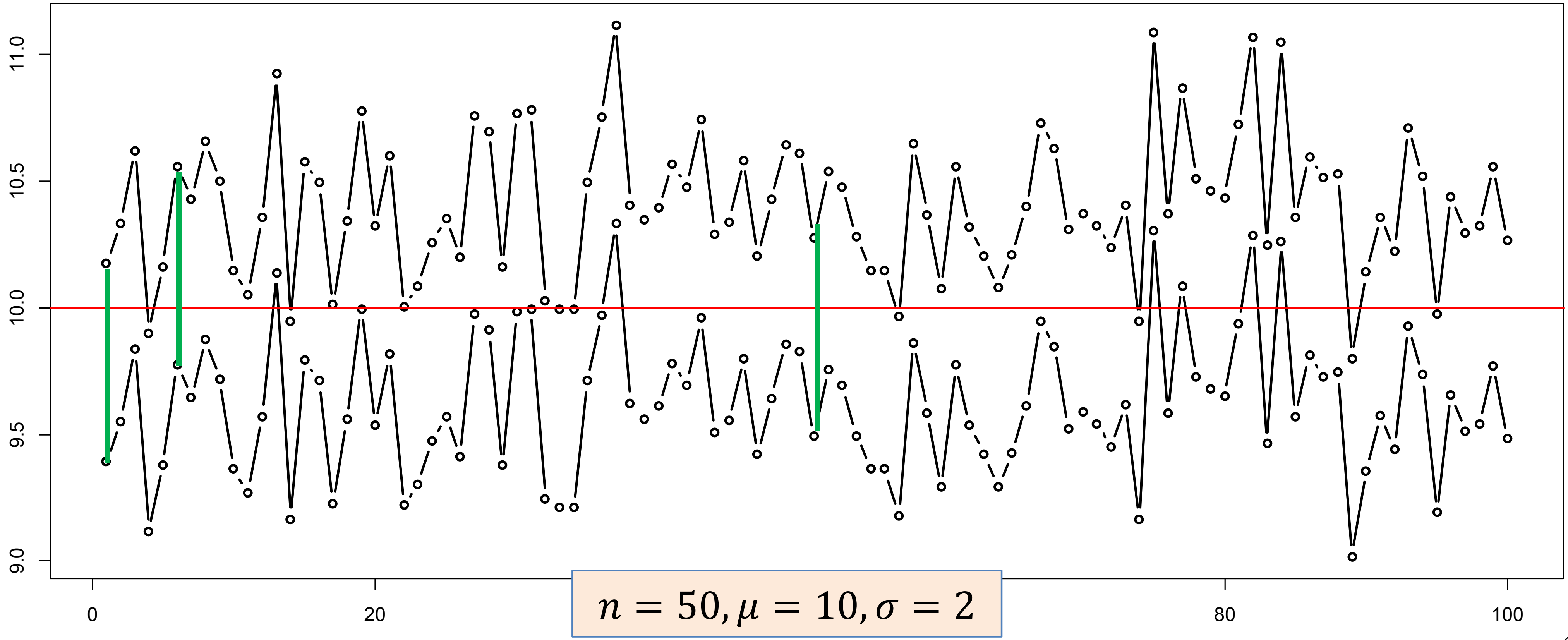
$n = 50, \mu = 10, \sigma = 2$

Inferenza sulla **media di una popolazione**

$X_1, X_2, \dots, X_n$ i.i.d, $\sim N(\mu, \sigma^2)$ **con σ^2 nota**

$$\Rightarrow \left(\bar{X}_n - z_{1-\frac{\alpha}{2}} \sqrt{\frac{\sigma^2}{n}} \, , \bar{X}_n + z_{1-\frac{\alpha}{2}} \sqrt{\frac{\sigma^2}{n}} \right)$$

E' il modello che prevede tutti i possibili **intervalli di confidenza** di livello $1-\alpha$ per $\boldsymbol{\mu}$



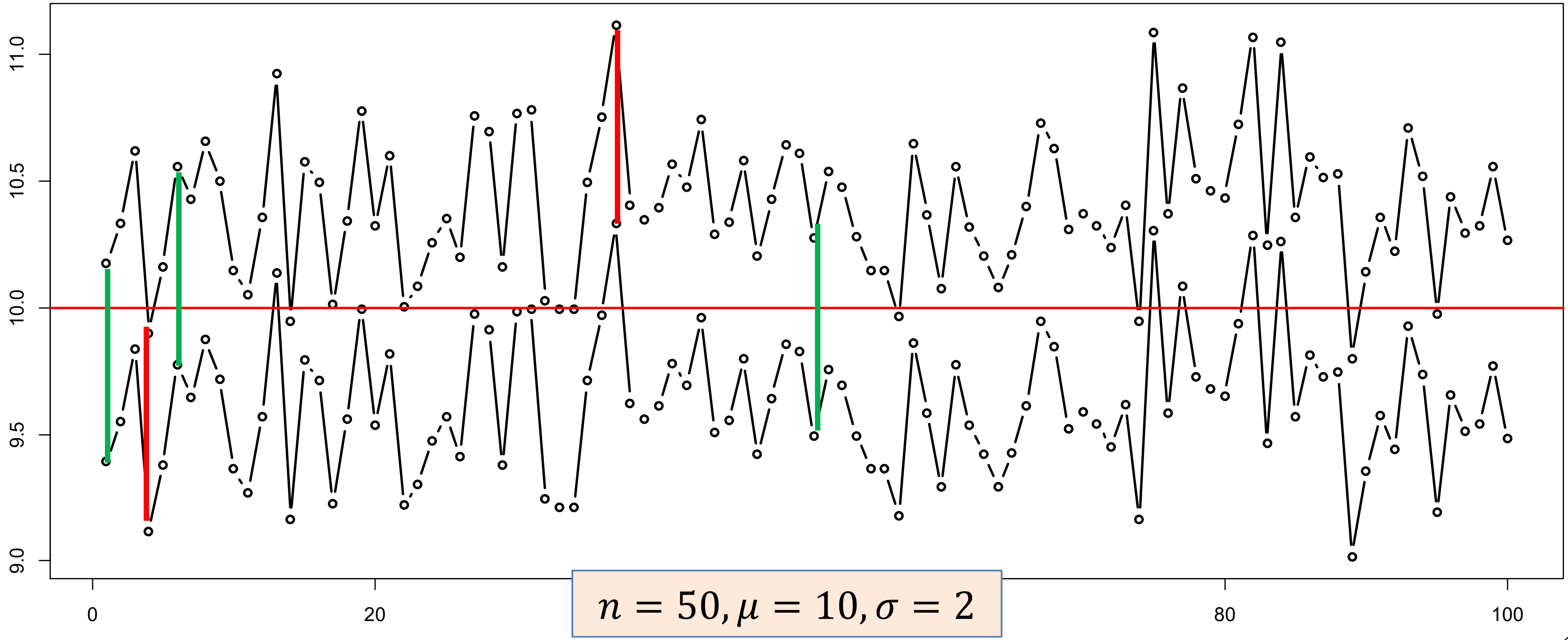
$n = 50, \mu = 10, \sigma = 2$

Inferenza sulla **media di una popolazione**

$X_1, X_2, \dots, X_n$ i.i.d, $\sim N(\mu, \sigma^2)$ **con σ^2 nota**

$$\Rightarrow \left(\bar{X}_n - z_{1-\frac{\alpha}{2}} \sqrt{\frac{\sigma^2}{n}} \, , \bar{X}_n + z_{1-\frac{\alpha}{2}} \sqrt{\frac{\sigma^2}{n}} \right)$$

E' il modello che prevede tutti i possibili **intervalli di confidenza** di livello $1-\alpha$ per μ



$n = 50, \mu = 10, \sigma = 2$

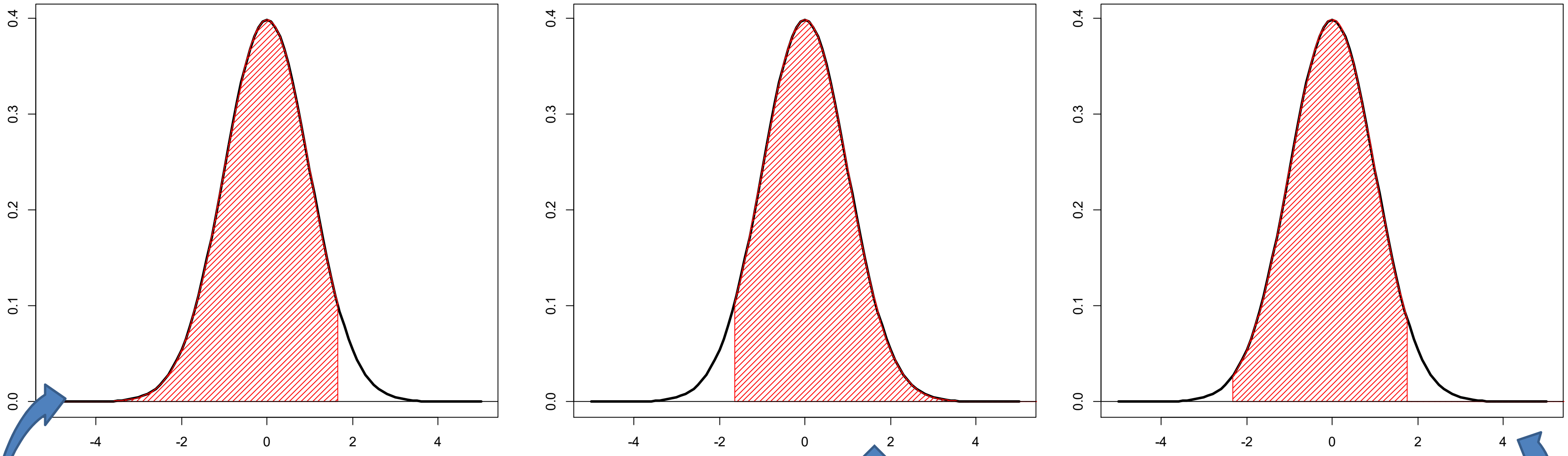
tanto per saperlo...

$$0.95 = P(\mu - a \leq \bar{X}_n \leq \mu + b) = P(\bar{X}_n - b \leq \mu \leq \bar{X}_n + a) =$$

$$= P\left(\frac{\mu - a - \mu}{\sqrt{\frac{\sigma^2}{n}}} \leq \frac{\bar{X}_n - \mu}{\sqrt{\frac{\sigma^2}{n}}} \leq \frac{\mu + b - \mu}{\sqrt{\frac{\sigma^2}{n}}}\right) = P\left(\frac{-a}{\sqrt{\frac{\sigma^2}{n}}} \leq Z \leq \frac{b}{\sqrt{\frac{\sigma^2}{n}}}\right)$$

tanto per saperlo...

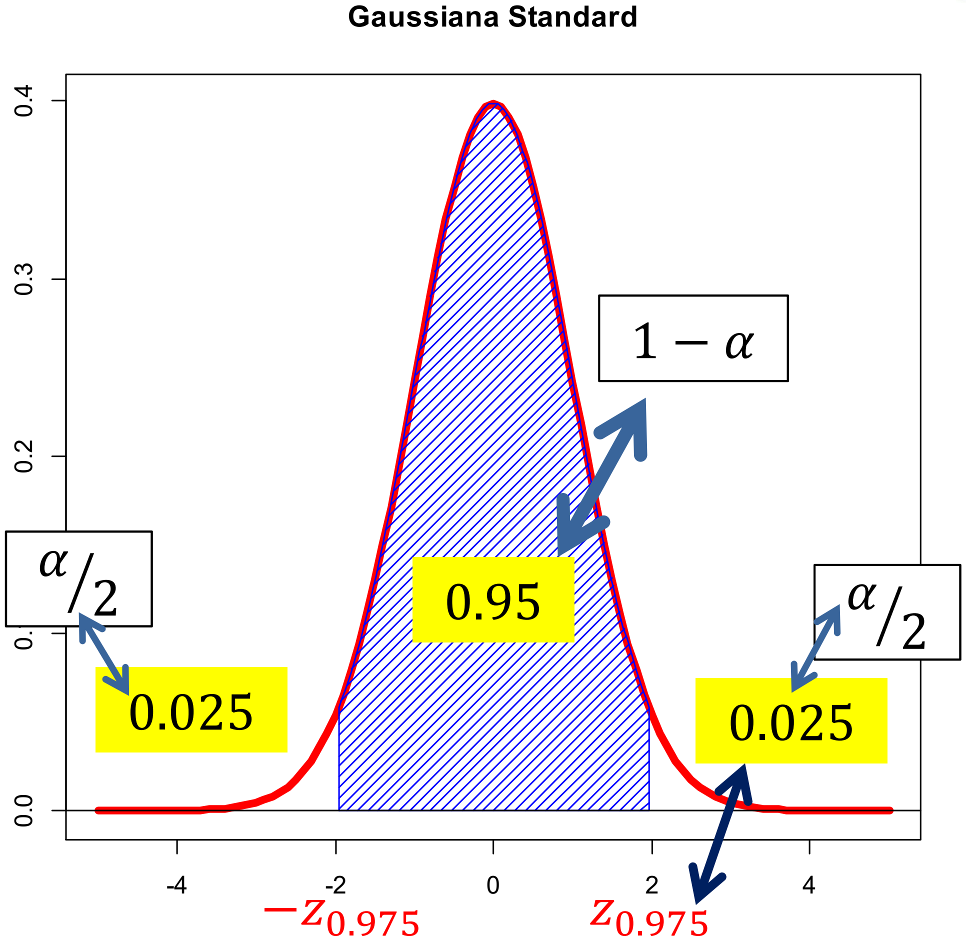
$(-\infty, 1.644854)$ $(-1.644854, +\infty)$ $(-2.326248, 1.750686)$



$$\color{red}{0.95} = P(\mu - a \leq \bar{X}_n \leq \mu + b) = \color{red}{P(\bar{X}_n - b \leq \mu \leq \bar{X}_n + a)} =$$

$$= P\left(\frac{\mu - a - \mu}{\sqrt{\frac{\sigma^2}{n}}} \leq \frac{\bar{X}_n - \mu}{\sqrt{\frac{\sigma^2}{n}}} \leq \frac{\mu + b - \mu}{\sqrt{\frac{\sigma^2}{n}}}\right) = P\left(\frac{-a}{\sqrt{\frac{\sigma^2}{n}}} \leq Z \leq \frac{b}{\sqrt{\frac{\sigma^2}{n}}}\right)$$

tanto per saperlo...



$$a = z_{0.975} \times \sqrt{\frac{\sigma^2}{n}}$$

$$b = z_{0.975} \times \sqrt{\frac{\sigma^2}{n}}$$

$$\frac{-a}{\sqrt{\frac{\sigma^2}{n}}} = -z_{0.975}$$

$$\frac{b}{\sqrt{\frac{\sigma^2}{n}}} = z_{0.975}$$

$$0.95 = P(\mu - a \le \bar{X}_n \le \mu + b) = P\left(\frac{\mu - a - \mu}{\sqrt{\frac{\sigma^2}{n}}} \le \frac{\bar{X}_n - \mu}{\sqrt{\frac{\sigma^2}{n}}} \le \frac{\mu + b - \mu}{\sqrt{\frac{\sigma^2}{n}}}\right) =$$

$$= P\left(\frac{-a}{\sqrt{\frac{\sigma^2}{n}}} \le Z \le \frac{b}{\sqrt{\frac{\sigma^2}{n}}}\right)$$

Esercizio 1 – di compito

Tempi di reazione in un esame di psicologia di un campione di 100 individui: $\bar{x}_{100} = 1$ secondo. Si sa da studi precedenti che lo scarto quadratico medio (*dev. standard*) è $\sigma = 0.07$ secondi.

a) IC(0.95) del tempo medio di reazione

b) IC(0.99) del tempo medio di reazione

c) come in b) con $n = 150$

$$\left(\overline{\boldsymbol{x}}_{\boldsymbol{n}} - \boldsymbol{z}_{\mathbf{1}-\frac{\boldsymbol{\alpha}}{\mathbf{2}}}\sqrt{\frac{\boldsymbol{\sigma}^{\mathbf{2}}}{\boldsymbol{n}}}\ ,\ \overline{\boldsymbol{x}}_{\boldsymbol{n}} + \boldsymbol{z}_{\mathbf{1}-\frac{\boldsymbol{\alpha}}{\mathbf{2}}}\sqrt{\frac{\boldsymbol{\sigma}^{\mathbf{2}}}{\boldsymbol{n}}}\right)$$

$$z_{1-\frac{\alpha}{2}} = \text{qnorm}(1-\tfrac{\alpha}{2}, 0, 1)$$

$$z_{0.975} = \text{qnorm}(0.975,0,1) = 1.959964$$

Inferenza sulla **media di una popolazione**

$X_1, X_2, \dots, X_n$ i.i.d possibili osservazioni

$X_i \sim N(\mu, \sigma^2)$ modello, anche $\boldsymbol{\sigma^2}$ **non nota**

varianza campionaria

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

stimatore di σ^2

$$\frac{\bar{X}_n - \mu}{\sqrt{\frac{{S_n}^2}{n}}} \sim t(n-1)$$

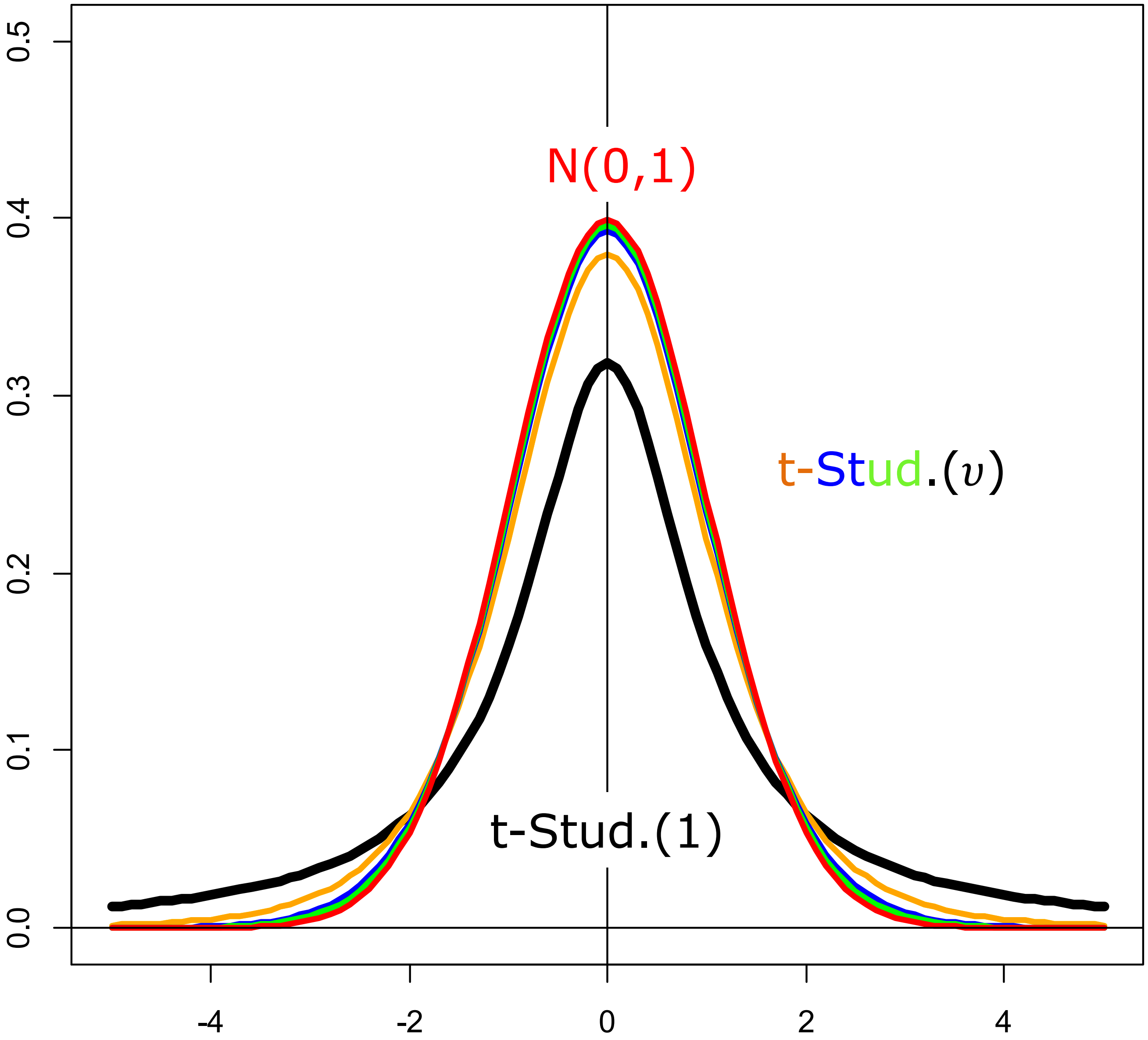
$$\left(\bar{X}_n - t(n-1)_{1-\frac{\alpha}{2}} \times \sqrt{\frac{{S_n}^2}{n}}, \bar{X}_n + t(n-1)_{1-\frac{\alpha}{2}} \times \sqrt{\frac{{S_n}^2}{n}}\right)$$

$$\left\{ \frac{\bar{X}_n - \mu}{\sqrt{\frac{\sigma^2}{n}}} \sim N(0,1) \qquad \left(\bar{X}_n - z_{1-\frac{\alpha}{2}}\sqrt{\frac{\sigma^2}{n}}, \bar{X}_n + z_{1-\frac{\alpha}{2}}\sqrt{\frac{\sigma^2}{n}}\right) \right\}$$

σ^2 nota

Inferenza sulla **media di una popolazione**

Ricordiamoci che (Script2.R):



Esercizio 2

Si vuole sapere se le donne tendono a sposare uomini che sono più alti di loro. Per questo si è scelto un campione di 50 coppie sposate, e si sono confrontate le altezze di moglie e marito calcolando la differenza di altezza: (h del marito – h della moglie). La differenza media è risultata pari a $\bar{x}_n = 11.2$ cm con una deviazione standard $s_n = 10.7$ cm.
Costruire un intervallo di confidenza del 95% per la differenza media dell'altezza ed usare questo risultato per rispondere al quesito.

Esercizio 2

Si vuole sapere se le donne tendono a sposare uomini che sono più alti di loro. Per questo si è scelto un campione di 50 coppie sposate, e si sono confrontate le altezze di moglie e marito calcolando la differenza di altezza: (h del marito – h della moglie). La differenza media è risultata pari a $\bar{x}_n = 11.2$ cm con una deviazione standard $s_n = 10.7$ cm.
Costruire un intervallo di confidenza del 95% per la differenza media dell'altezza ed usare questo risultato per rispondere al quesito.

$X_1, X_2, \dots, X_n$ i.i.d, $X_i \sim N(\mu, \sigma^2), X_i =$ differenza nell'h della coppia *i*

$$\left(\bar{X}_n - \boldsymbol{t(n-1)}_{\mathbf{1}-\frac{\boldsymbol{\alpha}}{\mathbf{2}}} \times \sqrt{\frac{\boldsymbol{S_n}^{\mathbf{2}}}{n}}, \bar{X}_n + \boldsymbol{t(n-1)}_{\mathbf{1}-\frac{\boldsymbol{\alpha}}{\mathbf{2}}} \times \sqrt{\frac{\boldsymbol{S_n}^{\mathbf{2}}}{n}}\right)$$

Esercizio 2

Si vuole sapere se le donne tendono a sposare uomini che sono più alti di loro. Per questo si è scelto un campione di 50 coppie sposate, e si sono confrontate le altezze di moglie e marito calcolando la differenza di altezza: (h del marito – h della moglie). La differenza media è risultata pari a $\bar{x}_n = 11.2$ cm con una deviazione standard $s_n = 10.7$ cm.
Costruire un intervallo di confidenza del 95% per la differenza media dell'altezza ed usare questo risultato per rispondere al quesito.

$X_1, X_2, \dots, X_n$ i.i.d, $X_i \sim N(\mu, \sigma^2), X_i =$ differenza nell'h della coppia *i*

$$\left(\bar{X}_n - \boldsymbol{t(n-1)}_{1-\frac{\alpha}{2}} \times \sqrt{\frac{\boldsymbol{S_n}^2}{n}}, \bar{X}_n + \boldsymbol{t(n-1)}_{1-\frac{\alpha}{2}} \times \sqrt{\frac{\boldsymbol{S_n}^2}{n}}\right)$$

$$\left(11.2 - qt(0.975,49) \times \frac{10.7}{\sqrt{50}}, 11.2 + qt(0.975,49) \times \frac{10.7}{\sqrt{50}}\right) = (8.16,\ 14.24)$$

Esercizio 2

Si vuole sapere se le donne tendono a sposare uomini che sono più alti di loro. Per questo si è scelto un campione di 50 coppie sposate, e si sono confrontate le altezze di moglie e marito calcolando la differenza di altezza: (h del marito – h della moglie). La differenza media è risultata pari a $\bar{x}_n = 11.2$ cm con una deviazione standard $s_n = 10.7$ cm.
Costruire un intervallo di confidenza del 95% per la differenza media dell'altezza ed **usare questo risultato per rispondere al quesito.**

$X_1, X_2, \dots, X_n$ i.i.d, $X_i \sim N(\mu, \sigma^2), X_i =$ differenza nell'h della coppia *i*

$$\left(\bar{X}_n - \boldsymbol{t(n-1)}_{\mathbf{1-\frac{\alpha}{2}}} \times \sqrt{\frac{\boldsymbol{S_n}^{\mathbf{2}}}{n}}, \bar{X}_n + \boldsymbol{t(n-1)}_{\mathbf{1-\frac{\alpha}{2}}} \times \sqrt{\frac{\boldsymbol{S_n}^{\mathbf{2}}}{n}}\right)$$

$$\left(11.2 - qt(0.975{,}49) \times \frac{10.7}{\sqrt{50}}, 11.2 + qt(0.975{,}49) \times \frac{10.7}{\sqrt{50}}\right) = (8.16,\ 14.24)$$

Inferenza sulla **proporzione**

$X_1, X_2, \dots, X_n$ i.i.d possibili osservazioni

$X_i \sim b(p), E(X_i) = p, Var(X_i) = p(1-p)$ modello

$(x_1, x_2, \dots, x_n)$: $\{1,1,0,0,0,1,0,1,0,0\}$ dati (campione casuale)

$(X_1 + X_2 + \cdots + X_n) \sim Binom(n, p)$ num. totale di «1»

$\bar{X}_n \equiv \hat{p}_n$ stimatore di p

per grandi campioni $(np \geq 5, np(1-p) \geq 5)$: $\bar{X}_n \sim N\left(p, \frac{p(1-p)}{n}\right)$

$$\left(\hat{p}_n - z_{1-\frac{\alpha}{2}}\sqrt{\frac{\hat{p}_n(1-\hat{p}_n)}{n}}, \hat{p}_n + z_{1-\frac{\alpha}{2}}\sqrt{\frac{\hat{p}_n(1-\hat{p}_n)}{n}}\right)$$

Inferenza sulla **media di una popolazione**

$X_1, X_2, \ldots, X_n$ i.i.d possibili osservazioni

$\boldsymbol{n\ grande\ (n \geq 30)}$

$X_i \sim ???????$ modello, $\mu = E(X_i)$, anche σ^2 non nota

$(x_1, x_2, \ldots, x_n)$ dati (campione casuale)

$$\left(\bar{X}_n - \boldsymbol{t(n-1)_{1-\alpha/2}} \times \sqrt{\frac{\boldsymbol{S_n}^2}{n}}, \bar{X}_n + \boldsymbol{t(n-1)_{1-\alpha/2}} \times \sqrt{\frac{\boldsymbol{S_n}^2}{n}} \right)$$

Grazie al TLC vale anche per modello *non* gaussiano

Intervalli di confidenza

$X_i \sim N(\mu, \sigma^2)$ **(o $n \geq 30$, TCL)**

Per μ, con σ^2 nota: $\left(\bar{X}_n - z_{1-\alpha/2}\sqrt{\frac{\sigma^2}{n}}, \bar{X}_n + z_{1-\alpha/2}\sqrt{\frac{\sigma^2}{n}}\right)$

Per μ, con σ^2 non nota: $\left(\bar{X}_n - t(n-1)_{1-\alpha/2}\sqrt{\frac{{S_n}^2}{n}}, \bar{X}_n + t(n-1)_{1-\alpha/2}\sqrt{\frac{{S_n}^2}{n}}\right)$

$X_i \sim Bern(p)$

asintotico: $\left(\hat{p}_n - z_{1-\alpha/2}\sqrt{\frac{\hat{p}_n(1-\hat{p}_n)}{n}}, \hat{p}_n + z_{1-\alpha/2}\sqrt{\frac{\hat{p}_n(1-\hat{p}_n)}{n}},\right)$

Intervalli di confidenza

$\boldsymbol{X_i \sim N(\mu, \sigma^2)}$ **(o $\boldsymbol{n \geq 30}$, TCL)**

Per μ, con σ^2 nota: $$\left(\bar{X}_n - z_{1-\alpha/2}\sqrt{\frac{\sigma^2}{n}}, \bar{X}_n + z_{1-\alpha/2}\sqrt{\frac{\sigma^2}{n}}\right)$$

Per μ, con σ^2 non nota: $$\left(\bar{X}_n - t(n-1)_{1-\alpha/2}\sqrt{\frac{{S_n}^2}{n}}, \bar{X}_n + t(n-1)_{1-\alpha/2}\sqrt{\frac{{S_n}^2}{n}}\right)$$

Per $\boldsymbol{\sigma^2}$, con $\boldsymbol{\mu}$ non nota: $$\left(\frac{(\boldsymbol{n-1})\boldsymbol{S^2}}{\boldsymbol{\chi^2(n-1)}_{\boldsymbol{1-\alpha/2}}}, \frac{(\boldsymbol{n-1})\boldsymbol{S^2}}{\boldsymbol{\chi^2(n-1)}_{\boldsymbol{\alpha/2}}}\right)$$

$\boldsymbol{X_i \sim Bern(p)}$
asintotico: $$\left(\hat{p}_n - z_{1-\alpha/2}\sqrt{\frac{\hat{p}_n(1-\hat{p}_n)}{n}}, \hat{p}_n + z_{1-\alpha/2}\sqrt{\frac{\hat{p}_n(1-\hat{p}_n)}{n}},\right)$$

Intervalli di c

INTERVALLI DI CONFIDENZA
BILATERI
O “A DUE CODE”

$X_i \sim N(\mu, \sigma^2)$ (o $n \geq 30$, TCL)

Per μ, con σ^2 nota: $$\left(\bar{X}_n - z_{1-\alpha/2}\sqrt{\frac{\sigma^2}{n}}, \bar{X}_n + z_{1-\alpha/2}\sqrt{\frac{\sigma^2}{n}}\right)$$

Per μ, con σ^2 non nota: $$\left(\bar{X}_n - t(n-1)_{1-\alpha/2}\sqrt{\frac{{S_n}^2}{n}}, \bar{X}_n + t(n-1)_{1-\alpha/2}\sqrt{\frac{{S_n}^2}{n}}\right)$$

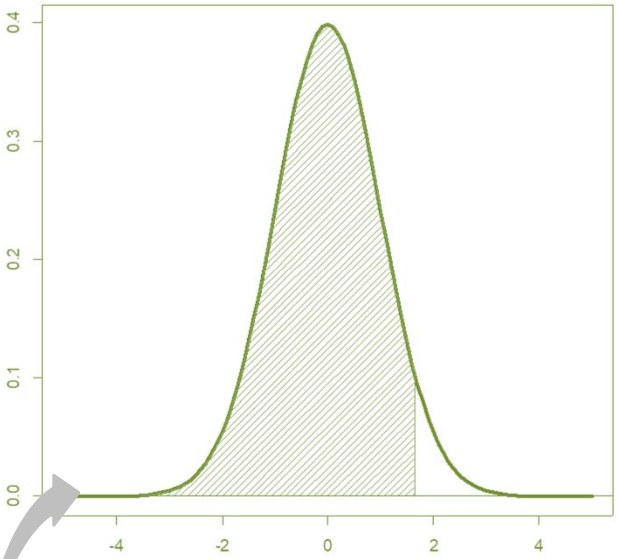
Per σ^2, con μ non nota: $$\left(\frac{(n-1)S^2}{\chi^2(n-1)_{\,1-\alpha/2}}, \frac{(n-1)S^2}{\chi^2(n-1)_{\,\alpha/2}}\right)$$

$X_i \sim Bern(p)$
asintotico: $$\left(\hat{p}_n - z_{1-\alpha/2}\sqrt{\frac{\hat{p}_n(1-\hat{p}_n)}{n}}, \hat{p}_n + z_{1-\alpha/2}\sqrt{\frac{\hat{p}_n(1-\hat{p}_n)}{n}},\right)$$

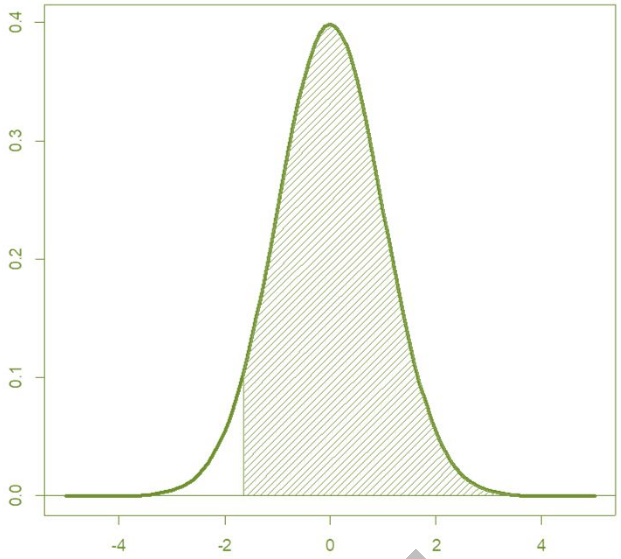
INTERVALLI DI CONFIDENZA **UNI**LATERI O "A UNA CODA"

INTERVALLO DI CONFIDENZA **ASIMMETRICO**

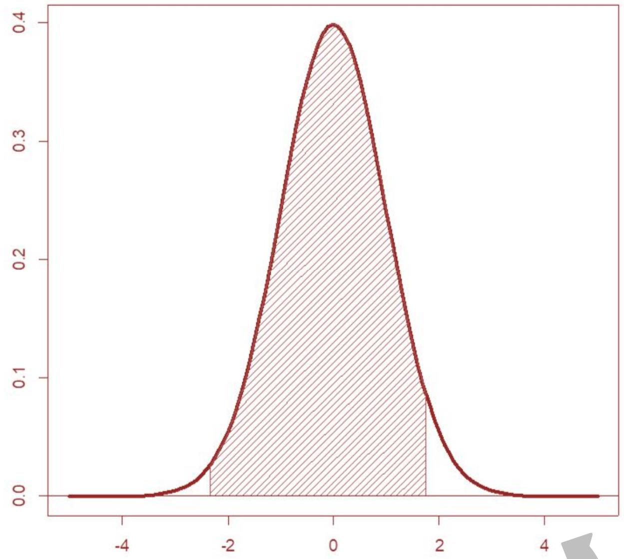
$(-\infty, 1.644854)$



$(-1.644854, +\infty)$



$(-2.326248, 1.750686)$



$$0.95 = P(\mu - a \leq \bar{X}_n \leq \mu + b) = P(\bar{X}_n - b \leq \mu \leq \bar{X}_n + a) =$$

$$= P\left(\frac{\mu - a - \mu}{\sqrt{\frac{\sigma^2}{n}}} \leq \frac{\bar{X}_n - \mu}{\sqrt{\frac{\sigma^2}{n}}} \leq \frac{\mu + b - \mu}{\sqrt{\frac{\sigma^2}{n}}}\right) = P\left(\frac{-a}{\sqrt{\frac{\sigma^2}{n}}} \leq Z \leq \frac{b}{\sqrt{\frac{\sigma^2}{n}}}\right)$$

Esercizio 4

In una indagine della Gallup(*), ai soggetti adulti intervistati fu chiesto se avevano un’arma in casa. Tra coloro che hanno risposto alla domanda 413 hanno detto sì e 646 hanno detto no.

a) quale percentuale di intervistati ha detto sì?
b) IC(95%) per la proporzione negli USA di adulti detentori di un’arma in casa
c) Possiamo concludere tranquillamente che meno del 50% degli adulti intervistati ha risposto sì quando è stato loro chiesto se avevano un’arma in casa?
d) La domanda al punto c) ci interessa? Quale alternativa?!
e) Qual è una risposta ragionevole alla critica che la Gallup non può fornire risultati attendibili perchè il campione è costituito da solo un migliaio di adulti, selezionati da una popolazione enorme di oltre 200 milioni di adulti?

(*) Fondata nel 1935 negli U.S.A, una delle più famose aziende al mondo per le indagini demoscopiche.

Esercizio 4

In una indagine della Gallup, ai soggetti adulti intervistati fu chiesto se avevano un’arma in casa. Tra coloro che hanno risposto alla domanda 413 hanno detto sì e 646 hanno detto.

a) quale percentuale di intervistati ha detto sì?
b) IC(95%) per la proporzione negli USA di adulti detentori di un’arma in casa
c) Possiamo concludere tranquillamente che meno del 50% degli adulti intervistati ha risposto sì quando è stato loro chiesto se avevano un’arma in casa?

Esercizio 4

In una indagine della Gallup, ai soggetti adulti intervistati fu chiesto se avevano un'arma in casa. Tra coloro che hanno risposto alla domanda 413 hanno detto sì e 646 hanno detto.

a) quale percentuale di intervistati ha detto sì?
b) IC(95%) per la proporzione negli USA di adulti detentori di un'arma in casa
c) Possiamo concludere tranquillamente che meno del 50% degli adulti intervistati ha risposto sì quando è stato loro chiesto se avevano un'arma in casa?

$$a)\ \frac{413}{413+646} = 0.39$$

Esercizio 4

In una indagine della Gallup, ai soggetti adulti intervistati fu chiesto se avevano un'arma in casa. Tra coloro che hanno risposto alla domanda 413 hanno detto sì e 646 hanno detto.

a) quale percentuale di intervistati ha detto sì?
b) IC(95%) per la proporzione negli USA di adulti detentori di un'arma in casa
c) Possiamo concludere tranquillamente che meno del 50% degli adulti intervistati ha risposto sì quando è stato loro chiesto se avevano un'arma in casa?

$$a)\ \frac{413}{413+646} = 0.39$$

$$b)\ (0.361, 0.419)$$

Esercizio 4

In una indagine della Gallup, ai soggetti adulti intervistati fu chiesto se avevano un'arma in casa. Tra coloro che hanno risposto alla domanda 413 hanno detto sì e 646 hanno detto.

a) quale percentuale di intervistati ha detto sì?
b) IC(95%) per la proporzione negli USA di adulti detentori di un'arma in casa
c) Possiamo concludere tranquillamente che meno del 50% degli adulti intervistati ha risposto sì quando è stato loro chiesto se avevano un'arma in casa?
d) **La domanda al punto c) ci interessa? Quale alternativa?**
e) Qual è una risposta ragionevole alla critica che la Gallup non può fornire risultati attendibili perchè il campione è costituito da solo un migliaio di adulti, selezionati da una popolazione enorme di oltre 200 milioni di adulti?

(*) Fondata nel 1935 negli U.S.A, una delle più famose aziende al mondo per le indagini demoscopiche.

Esercizio 4

In una indagine della Gallup, ai soggetti adulti intervistati fu chiesto se avevano un’arma in casa. Tra coloro che hanno risposto alla domanda 413 hanno detto sì e 646 hanno detto.

a) quale percentuale di intervistati ha detto sì?
b) IC(95%) per la proporzione negli USA di adulti detentori di un’arma in casa
c) Possiamo concludere tranquillamente che meno del 50% degli adulti intervistati ha risposto sì quando è stato loro chiesto se avevano un’arma in casa?
d) La domanda al punto c) ci interessa?
e) C … non p … tuito c … one e …

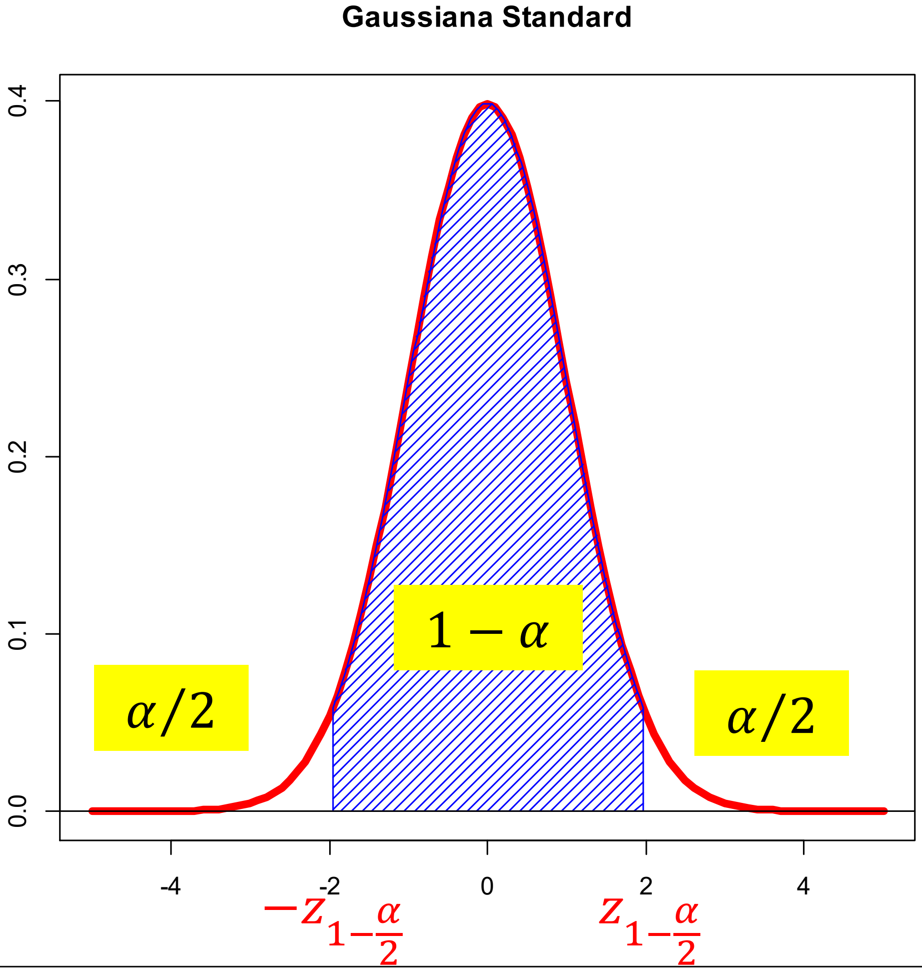
e) La dimensione campionaria **non dipende** dalla dimensione della popolazione di riferimento, ma solo dal livello di confidenza e dalla precisione richiesta.

(*) Fondata nel 1935 negli U.S.A, una delle più famose aziende al mondo per le indagini demoscopiche.

Inferenza sulla **media di una popolazione**

Intervallo di confidenza di livello $1-\alpha$

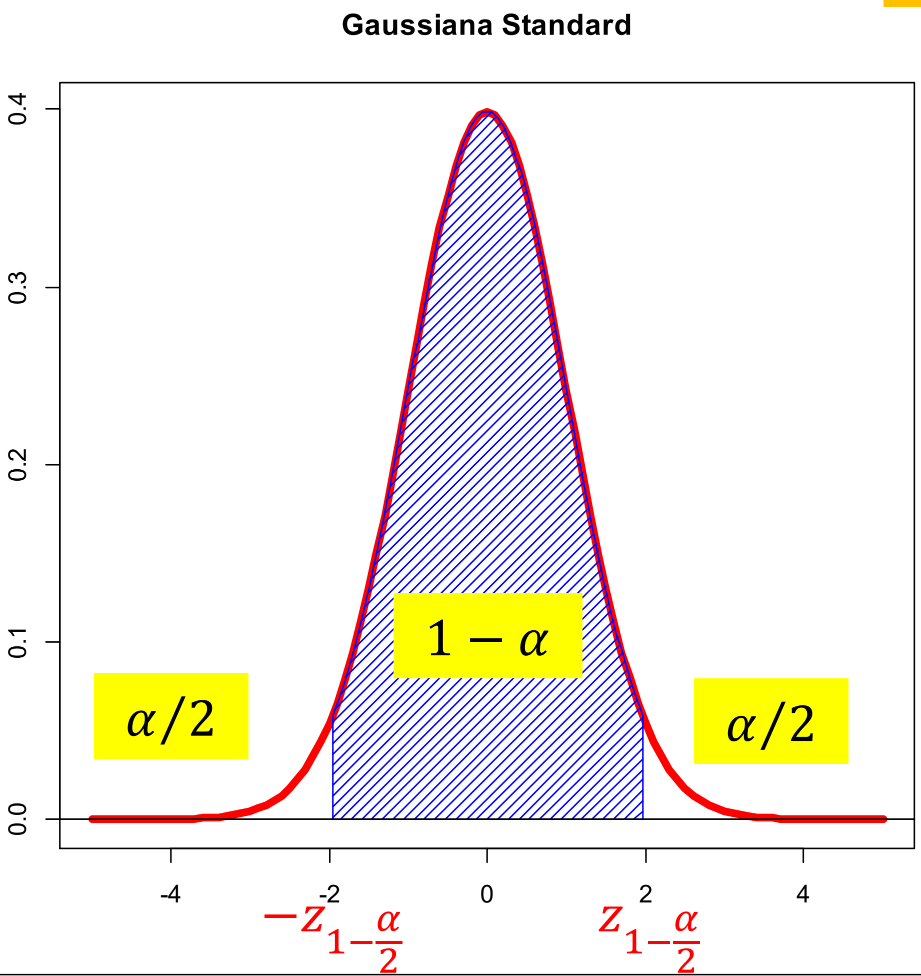
$$\left(\bar{X}_n - z_{1-\alpha/2} \times \sqrt{\frac{\sigma^2}{n}}, \bar{X}_n + z_{1-\alpha/2} \times \sqrt{\frac{\sigma^2}{n}}\right)$$



Inferenza sulla **media di una popolazione**

Intervallo di confidenza di livello $1-\alpha$

$$\left(\bar{X}_n - z_{1-\alpha/2} \times \sqrt{\frac{\sigma^2}{n}}, \bar{X}_n + z_{1-\alpha/2} \times \sqrt{\frac{\sigma^2}{n}}\right)$$



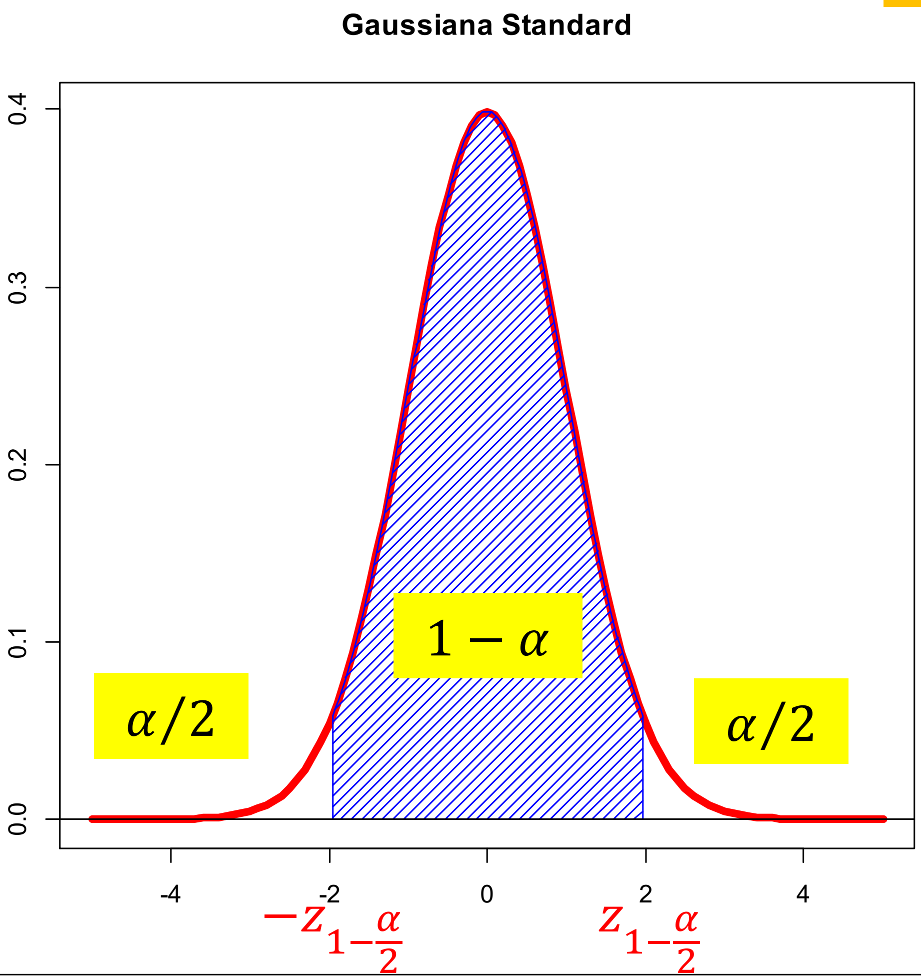
- aumentando n:
  IC si accorcia

- aumentando il livello $1-\alpha$:
  IC si allunga

Inferenza sulla **media di una popolazione**

Intervallo di confidenza di livello $1-\alpha$

$$\left(\bar{X}_n - z_{1-\alpha/2} \times \sqrt{\frac{\sigma^2}{n}}, \bar{X}_n + z_{1-\alpha/2} \times \sqrt{\frac{\sigma^2}{n}}\right)$$



- aumentando n:
  IC si accorcia
- aumentando il livello $1-\alpha$:
  IC si allunga

Si può agire su n per avere il livello di errore desiderato.

Intervalli di confidenza

Margine di errore

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

caso gaussiano (anche per $n \geq 30$)

$$z_{1-\frac{\alpha}{2}}\sqrt{\frac{\hat{p}_n(1-\hat{p}_n)}{n}}$$

caso bernoulliano (n grande)

Fissato il livello $1-\alpha$, al crescere di n il margine di errore diminuisce

Fissata la dimensione campionaria n, al crescere del livello $1-\alpha$ (i.e., al diminuire di α) il margine di errore diminuisce

Dimensione campionaria

Fissato il livello $1-\alpha$, possiamo determinare la dimensione campionaria necessaria ad avere un margine d'errore prefissato

17/10/2016, su un quotidiano on-line:

Sì: 33.8%
No: 37.0%
Indeciso: 29.2%

*«Ed è per questo che, ..., "la sfida" del 4 dicembre(*) "si giocherà **sul filo dei decimali**". Anche perché tutti i sondaggi che vengono pubblicati da giornali e tv in queste settimane hanno come sempre un **margine d'errore** del **3 per cento**.»*

margine di errore :

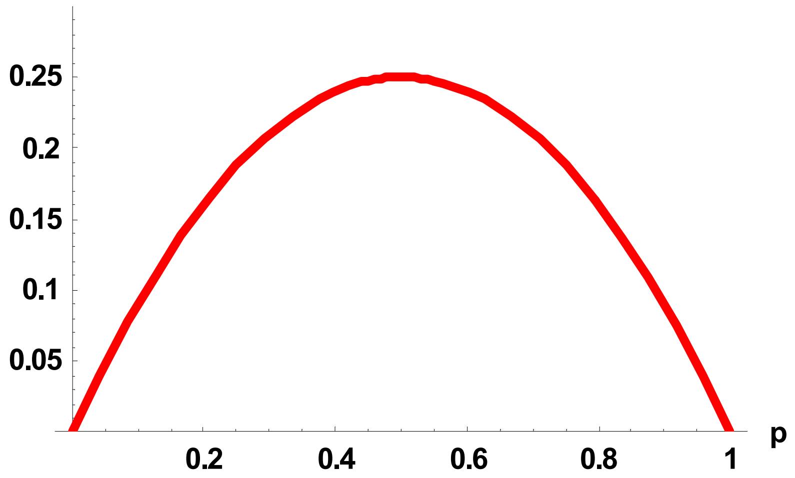
$$z_{1-\frac{\alpha}{2}}\sqrt{\frac{\hat{p}_n(1-\hat{p}_n)}{n}} = 0.03 \quad \Rightarrow \quad n = ??$$

(*) Referendum costituzionale

Dimensione campionaria

Fissato il livello $1-\alpha$, possiamo determinare la dimensione campionaria necessaria ad avere un margine d'errore prefissato

Fissato α



$$p(1-p) \le \frac{1}{4}$$

$$z_{1-\frac{\alpha}{2}}\sqrt{\frac{\hat{p}_n(1-\hat{p}_n)}{n}} \le$$

$$z_{1-\frac{\alpha}{2}}\sqrt{\frac{1}{4n}} \le 0.03$$

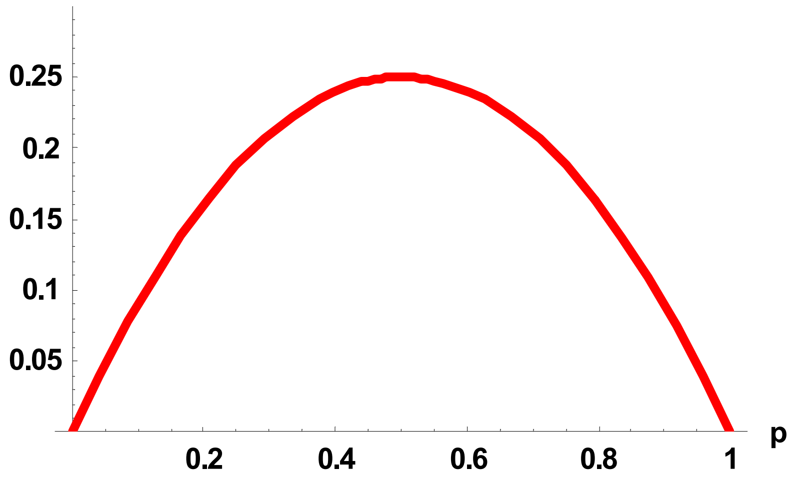
margine di errore :

$$z_{1-\frac{\alpha}{2}}\sqrt{\frac{\hat{p}_n(1-\hat{p}_n)}{n}} = 0.03$$

Dimensione campionaria

Fissato il livello $1-\alpha$, possiamo determinare la dimensione campionaria necessaria ad avere un margine d'errore prefissato

Fissato α



$$p(1-p) \leq \frac{1}{4}$$

$$z_{1-\frac{\alpha}{2}}\sqrt{\frac{\hat{p}_n(1-\hat{p}_n)}{n}} \leq$$

$$z_{1-\frac{\alpha}{2}}\sqrt{\frac{1}{4n}} \leq \varepsilon$$

margine di errore :

$$z_{1-\frac{\alpha}{2}}\sqrt{\frac{\hat{p}_n(1-\hat{p}_n)}{n}} = \varepsilon$$

$$\boxed{n \geq \frac{(z_{1-\alpha/2})^2}{4\varepsilon^2}}$$

Dimensione campionaria

Fissato il livello $1-\alpha$, possiamo determinare la dimensione campionaria necessaria ad avere un margine d'errore prefissato

17/10/2016, su un quotidiano on-line:

Sì: 33.8%
No: 37.0%
Indeciso: 29.2%

«Ed è per questo che, …, "la sfida" del 4 dicembre "si giocherà ***sul filo dei decimali****". Anche perché tutti i sondaggi che vengono pubblicati da giornali e tv in queste settimane hanno come sempre un* ***margine d'errore*** *del* ***3 per cento****.»*

$\alpha = 0.05$ ⟹ $n \geq \frac{(z_{1-\alpha/2})^2}{4\varepsilon^2} = \frac{1.96^2}{4 \times 0.03^2} = \mathbf{1067.11}$

Dimensione campionaria

Fissato il livello $1-\alpha$, possiamo determinare la dimensione campionaria necessaria ad avere un margine d'errore prefissato

17/10/2016, su un quotidiano on-line:

Sì: 33.8%
No: 37.0%
Indeciso: 29.2%

«Ed è per questo che, …, "la sfida" del 4 dicembre "si giocherà ***sul filo dei decimali****". Anche perché tutti i sondaggi che vengono pubblicati da giornali e tv in queste settimane hanno come sempre un* ***margine d'errore*** *del* ***1 per cento****.»*

$\alpha = 0.05$ ⇒

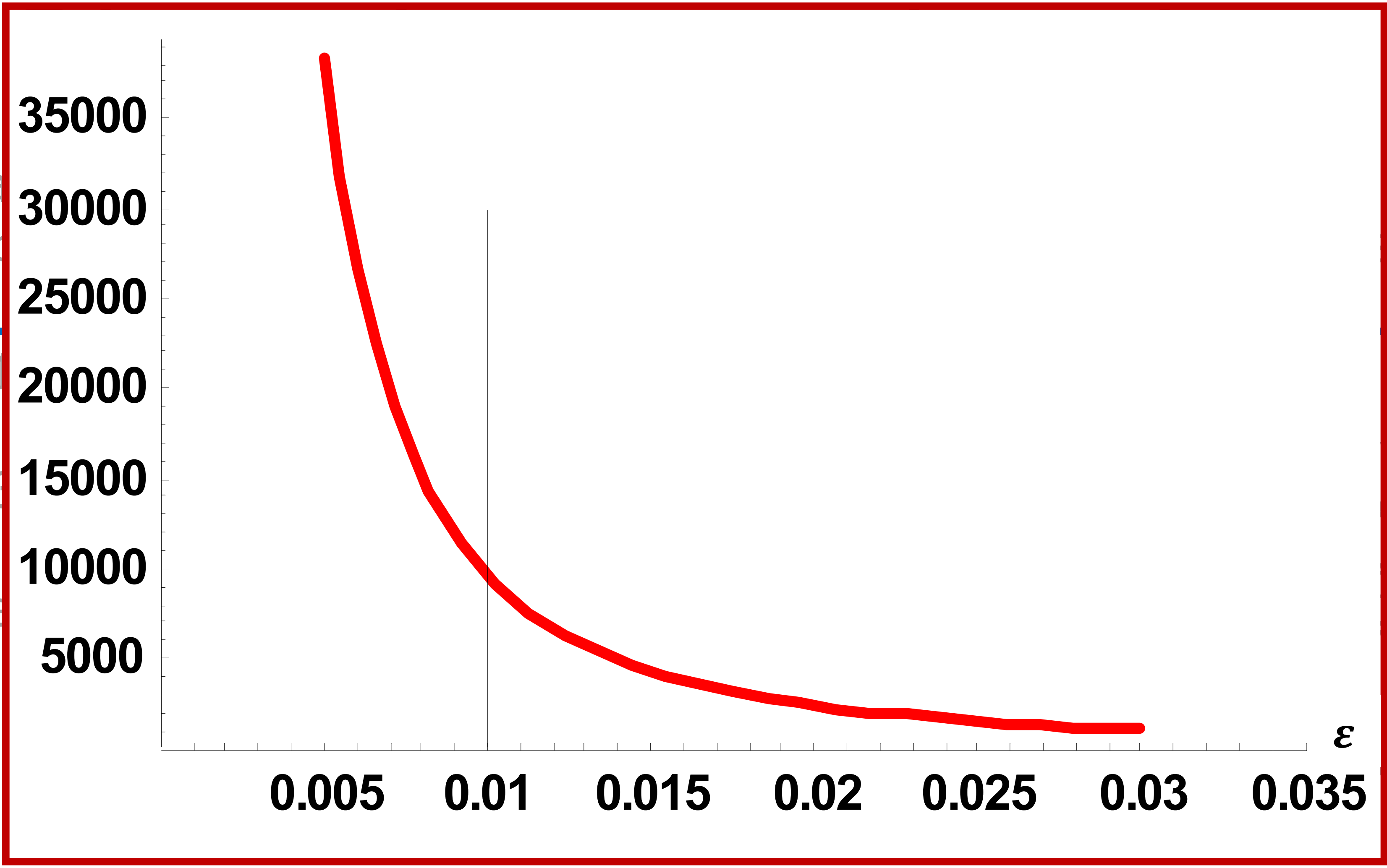
$$n \geq \frac{(z_{1-\alpha/2})^2}{4\varepsilon^2} = \frac{1.96^2}{4 \times \mathbf{0.01^2}} = \mathbf{9604.00}!!$$

Fissa ... sato
camp

17/1

Sì: 3
No:
Inde

a”
ilo
ti i
da
e



$\alpha = 0.05$ ➡ $n \geq \frac{(z_{1-\alpha/2})^2}{4\varepsilon^2} = \frac{1.96^2}{4 \times \mathbf{0.01^2}} = \mathbf{9604.00}!!$

Fissa

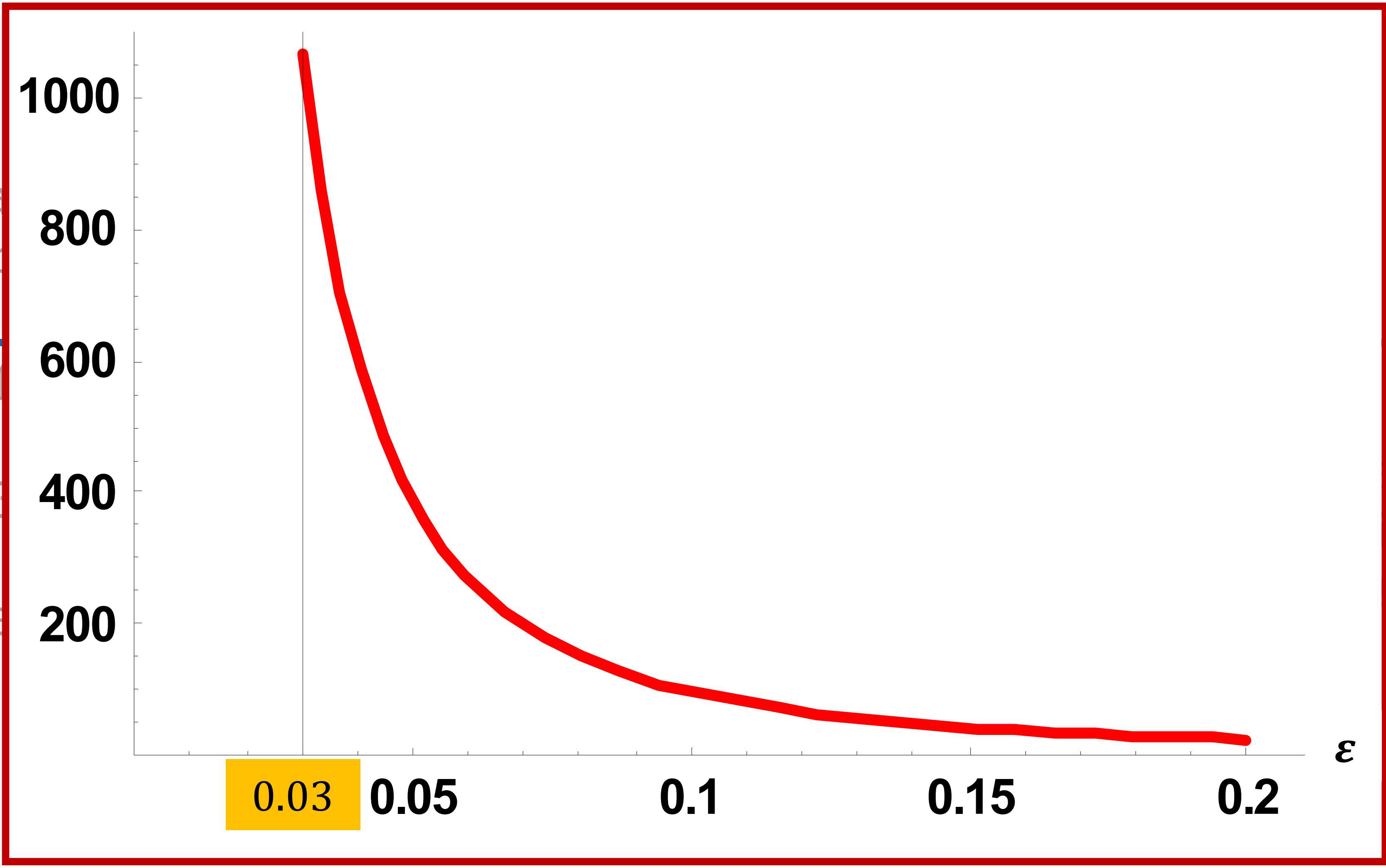
camp ... ato

17/1 ... a"

Sì: 3 ... lo

No: ... ti i

Inde ... da



$\alpha = 0.05$

$$n \geq \frac{(z_{1-\alpha/2})^2}{4\varepsilon^2} = \frac{1.96^2}{4 \times \mathbf{0.01^2}} = \mathbf{9604.00!!}$$

Dimensione campionaria

Un economista vuole fare un'indagine sullo stipendio mensile medio dei laureati in materie legate alle Scienze dell'Ambiente ad un anno dalla laurea.
Quale dovrà essere la dimensione campionaria perchè il margine di errore non superi i 100€ se come riferimento per la deviazione standard σ dello stipendio si usa la stima da un precedente studio, pari a 900€ ?

Dimensione campionaria

Un economista vuole fare un'indagine sullo stipendio mensile medio dei laureati in materie legate alle Scienze dell'Ambiente ad un anno dalla laurea.
Quale dovrà essere la dimensione campionaria perchè il margine di errore non superi i 100€ se come riferimento per la deviazione standard σ dello stipendio si usa la stima da un precedente studio, pari a 900€ ?

$X_1, X_2, \dots, X_n$ i.i.d, $X_i \sim N(\mu, 900^2)$

$\alpha = 0.05$

$$z_{1-\frac{\alpha}{2}}\sqrt{\frac{\sigma^2}{n}} \leq 100 \Leftrightarrow n \geq \frac{\left(z_{1-\alpha/2}\right)^2 \sigma^2}{\varepsilon^2} = \frac{1.96^2 \times 900^2}{100^2} = 311.2 \Rightarrow n \geq 312$$