STATISTICA

Intervalli di confidenza

Inferenza sulla **media di una popolazione**

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

μ, σ^2 in generale **non noti**

media campionaria

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

stimatore di μ

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

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

stima di μ

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**

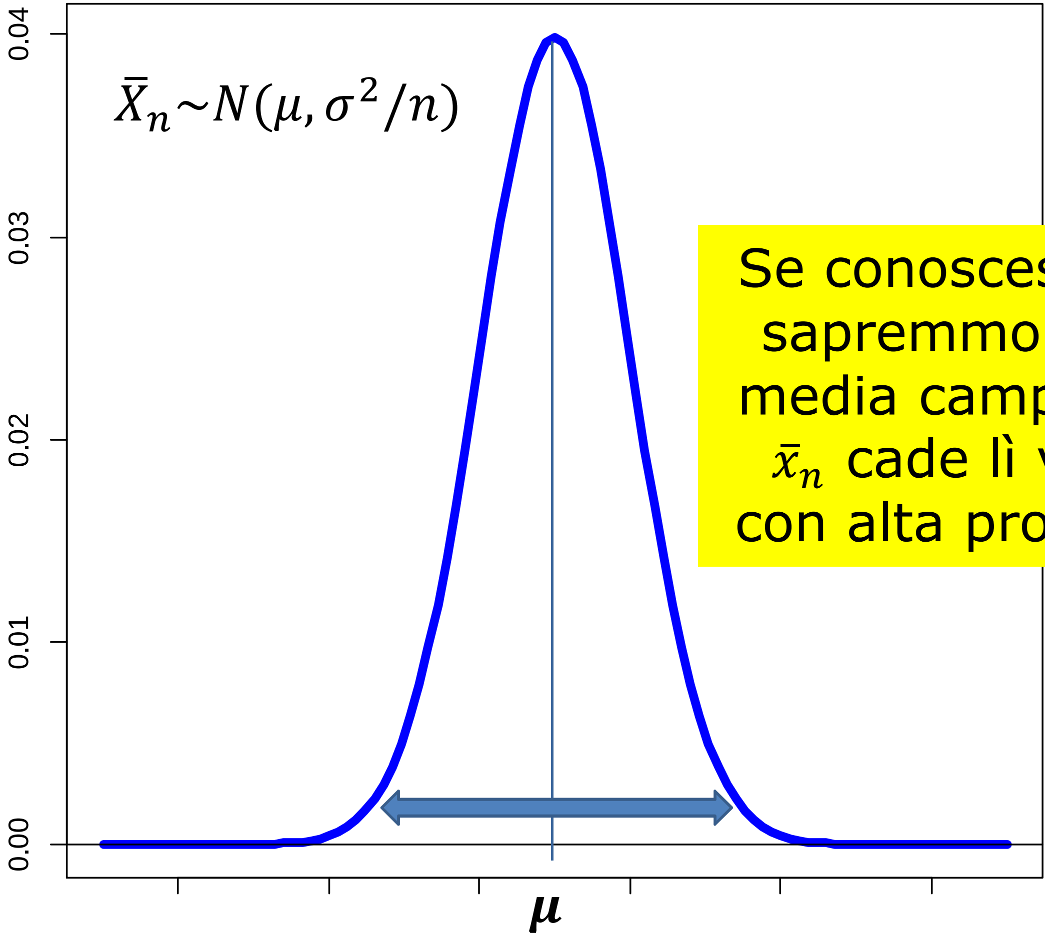
$$\Rightarrow \left(\bar{X}_n - z_{\frac{\alpha}{2}} \sqrt{\frac{\sigma^2}{n}} \, , \, \bar{X}_n + z_{\frac{\alpha}{2}} \sqrt{\frac{\sigma^2}{n}} \right)$$

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

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

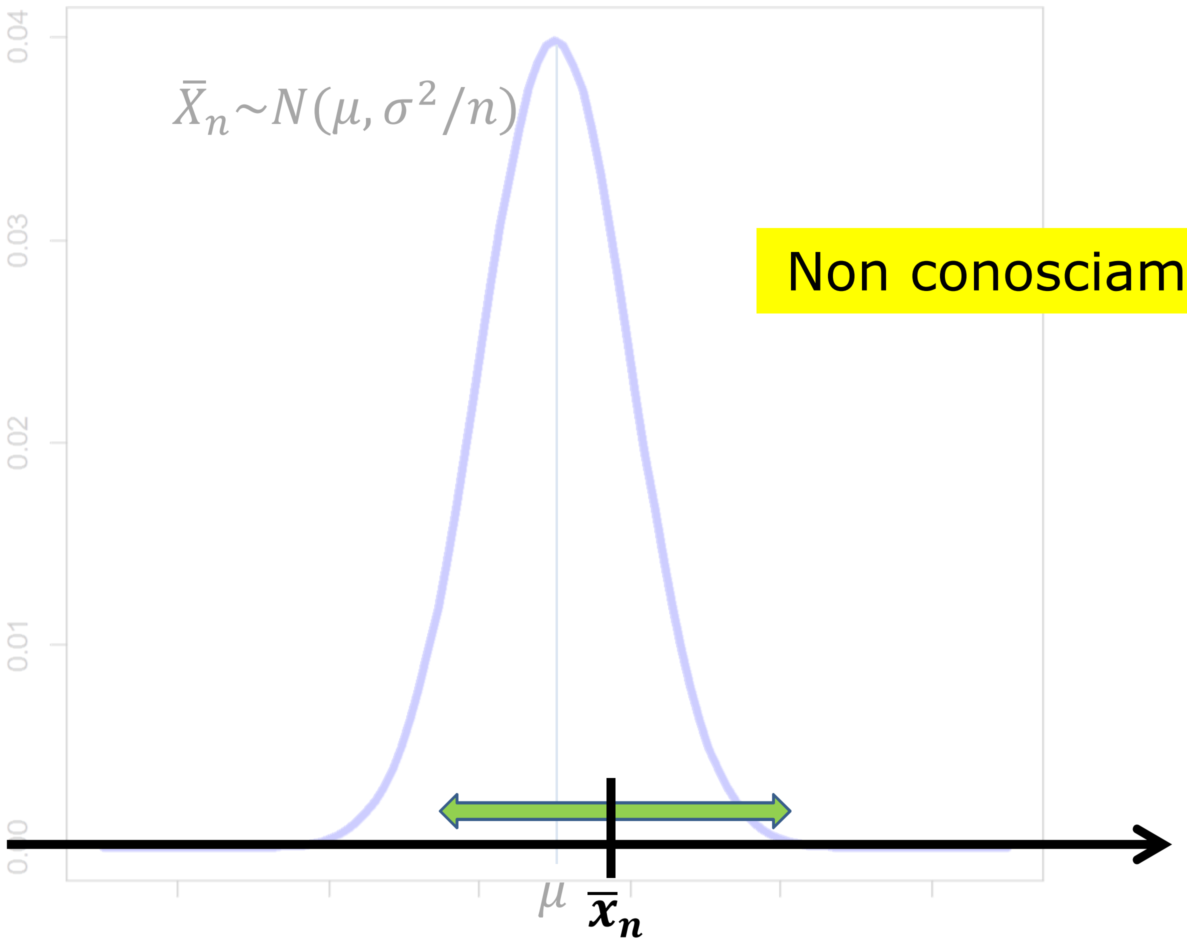
95 contengono il vero valore di μ

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



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.

Esercizio 1

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

Esercizio 1

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

$X_1, X_2, \ldots, X_n$ **i.i.d**, $X_i \sim N(\mu, 0.07^2)$, $\sigma = 0.07$, $n = 100$, $\alpha = 0.05$

IC(0.95): $\left(\bar{X}_n - z_{\frac{0.05}{2}} \times \sqrt{\frac{\sigma^2}{n}}, \bar{X}_n + z_{\frac{0.05}{2}} \times \sqrt{\frac{\sigma^2}{n}}\right)$

$$\bar{x}_{100} = 1 \Rightarrow \left(1 - z_{0.025} \times \sqrt{\frac{0.07^2}{100}}, 1 + z_{0.025} \times \sqrt{\frac{0.07^2}{100}}\right)$$

$\left(1 - 1.96 \times \frac{0.07}{10}, 1 + 1.96 \times \frac{0.07}{10}\right) = (0.986, 1.014)$ secondi

Esercizio 1

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.

b) IC(**0.99**) del tempo medio di reazione

$X_1, X_2, \dots, X_n$ **i.i.d**, $X_i \sim N(\mu, 0.07^2)$, $\sigma = 0.07$, $n = 100$, $\boldsymbol{\alpha = 0.01}$

IC(0.99): $$\left(\bar{X}_n - z_{\frac{0.01}{2}} \times \sqrt{\frac{\sigma^2}{n}}, \bar{X}_n + z_{\frac{0.01}{2}} \times \sqrt{\frac{\sigma^2}{n}}\right)$$

$$\bar{x}_{100} = 1 \Rightarrow \left(1 - \boldsymbol{z_{0.005}} \times \sqrt{\frac{0.07^2}{100}}, 1 + \boldsymbol{z_{0.005}} \times \sqrt{\frac{0.07^2}{100}}\right)$$

$$\left(1 - \mathbf{2.5758} \times \frac{0.07}{10}, 1 + \mathbf{2.5758} \times \frac{0.07}{10}\right) = (0.982, 1.018) \text{ secondi}$$

Esercizio 1

Tempi di reazione in un esame di psicologia di un campione di **150** individui: $\bar{x}_{150} = 1$ secondo. Si sa da studi precedenti che lo scarto quadratico medio (*dev. standard*) è $\sigma = 0.07$ secondi.

b) IC(**0.99**) del tempo medio di reazione

$X_1, X_2, \dots, X_n$ **i.i.d**, $X_i \sim N(\mu, 0.07^2)$, $\sigma = 0.07$, $\boldsymbol{n = 150}$, $\alpha = 0.01$

IC(0.99): $$\boxed{\left(\bar{X}_n - z_{\frac{0.01}{2}} \times \sqrt{\frac{\sigma^2}{n}}, \bar{X}_n + z_{\frac{0.01}{2}} \times \sqrt{\frac{\sigma^2}{n}}\right)}$$

$$\bar{x}_{150} = 1 \Rightarrow \boxed{\left(1 - z_{0.005} \times \sqrt{\frac{0.07^2}{\mathbf{150}}}, 1 + z_{0.005} \times \sqrt{\frac{0.07^2}{\mathbf{150}}}\right)}$$

$$\left(1 - 2.5758 \times \frac{0.07}{\sqrt{150}}, 1 + 2.5758 \times \frac{0.07}{\sqrt{150}}\right) = (0.985, 1.015) \text{ secondi}$$

Inferenza sulla **media di una popolazione**

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

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

$(x_1, x_2, \ldots, x_n)$: $\{34,19,6,12,24,12,20,27,33,29\}$ dati (campione casuale)

varianza campionaria

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

stimatore di σ^2

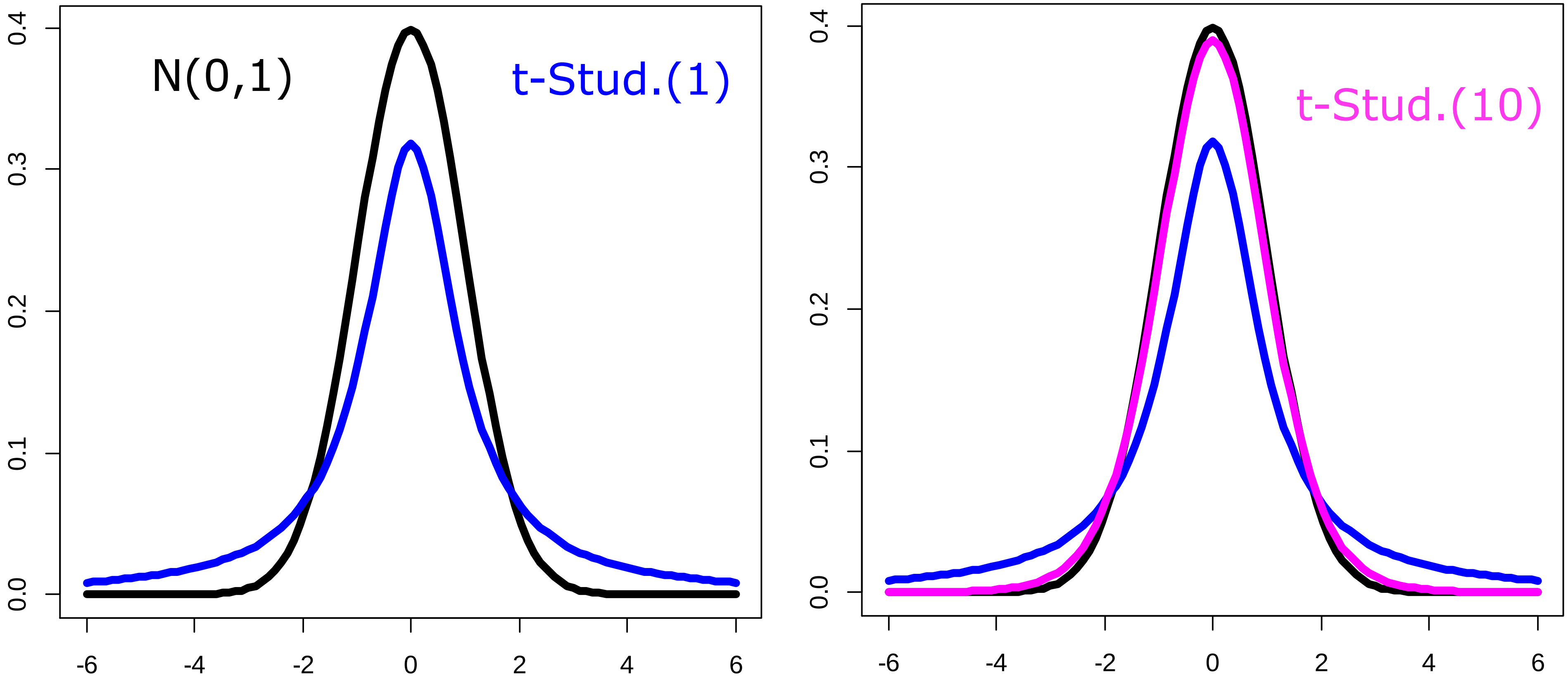
$$\frac{\bar{X}_n - \mu}{\sqrt{\frac{S_n{}^2}{n}}} \sim t(n-1)$$

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

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

σ^2 nota

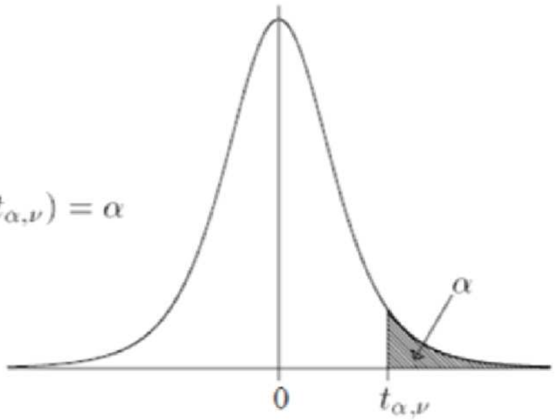
Inferenza sulla **media di una popolazione**



Confronto tra densità Gaussiana e densità t-Student

Tavola 2: Valori critici della Distribuzione t

$$P(T \geq t_{\alpha,\nu}) = \alpha$$



$t(v)_\alpha$

ν	0.20	0.10	0.05	0.025	α 0.01	0.005	0.001	0.0005	0.0001
1	1.3764	3.0777	6.3137	12.7062	31.8210	63.6559	318.2888	636.5776	3185.2722
2	1.0607	1.8856	2.9200	4.3027	6.9645	9.9250	22.3285	31.5998	70.7060
3	0.9785	1.6377	2.3534	3.1824	4.5407	5.8408	10.2143	12.9244	22.2027
4	0.9410	1.5332	2.1318	2.7765	3.7469	4.6041	7.1729	8.6101	13.0385
5	0.9195	1.4759	2.0150	2.5706	3.3649	4.0321	5.8935	6.8685	9.6764
6	0.9057	1.4398	1.9432	2.4469	3.1427	3.7074	5.2075	5.9587	8.0233
7	0.8960	1.4149	1.8946	2.3646	2.9979	3.4995	4.7853	5.4081	7.0641
8	0.8889	1.3968	1.8595	2.3060	2.8965	3.3554	4.5008	5.0414	6.4424
9	0.8834	1.3830	1.8331	2.2622	2.8214	3.2498	4.2969	4.7809	6.0094
10	0.8791	1.3722	1.8125	2.2281	2.7638	3.1693	4.1437	4.5868	5.6939
11	0.8755	1.3634	1.7959	2.2010	2.7181	3.1058	4.0248	4.4369	5.4529
12	0.8726	1.3562	1.7823	2.1788	2.6810	3.0545	3.9296	4.3178	5.2631
13	0.8702	1.3502	1.7709	2.1604	2.6503	3.0123	3.8520	4.2209	5.1106
14	0.8681	1.3450	1.7613	2.1448	2.6245	2.9768	3.7874	4.1403	4.9849
15	0.8662	1.3406	1.7531	2.1315	2.6025	2.9467	3.7329	4.0728	4.8801
16	0.8647	1.3368	1.7459	2.1199	2.5835	2.9208	3.6861	4.0149	4.7905
17	0.8633	1.3334	1.7396	2.1098	2.5669	2.8982	3.6458	3.9651	4.7148
18	0.8620	1.3304	1.7341	2.1009	2.5524	2.8784	3.6105	3.9217	4.6485

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**

$(x_1, x_2, \dots, x_n)$: $\{34,19,6,12,24,12,20,27,33,29\}$ dati (campione casuale)

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

$\bar{x}_{10} = 21.6\ \mu g/m^3$

$${s_{10}}^2 = \frac{1}{9}\sum_{i=1}^{10}(x_i - 21.6)^2 = 90.04$$

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**

$(x_1, x_2, \dots, x_n)$: $\{34,19,6,12,24,12,20,27,33,29\}$ dati (campione casuale)

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

$$\left.\begin{aligned} &\bar{x}_{10} = 21.6\ \mu g/m^3 \\ &{s_{10}}^2 = \frac{1}{9}\sum_{i=1}^{10}(x_i - 21.6)^2 = 90.04 \end{aligned}\right\}$$

$\alpha = 0.05 \Longrightarrow t(9)_{\mathbf{0.025}}$

$$21.6 \mp 2.2622 \times \sqrt{\frac{90.04}{10}}$$

$$IC(0.95):\ (14.81, 28.39)$$

Esercizio (8 pg. 234, circa…)

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 (8 pg. 234, circa…)

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, $X_i \sim N(\mu, \sigma^2)$ (*) $\alpha = 0.05$

$$\left(\bar{x}_n - t(n-1)_{\alpha/2}\sqrt{\frac{{s_n}^2}{n}}, \bar{x}_n + t(n-1)_{\alpha/2}\sqrt{\frac{{s_n}^2}{n}}\right)$$

$$\left(11.2 - 2.0086\sqrt{\frac{10.7^2}{50}}, 11.2 + 2.0086\sqrt{\frac{10.7^2}{50}}\right) = (8.2, 14.2) \text{ cm}$$

Esercizio (8 pg. 234, circa…)

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.

$X_1, X_2, \ldots, X_n$ i.i.d, $X_i \sim N(\mu, \sigma^2)$ (*) $\alpha = 0.05$

$$\left(\bar{x}_n - t(n-1)_{\alpha/2}\sqrt{\frac{{s_n}^2}{n}}, \bar{x}_n + t(n-1)_{\alpha/2}\sqrt{\frac{{s_n}^2}{n}}\right)$$

$$\left(11.2 - 2.0086\sqrt{\frac{10.7^2}{50}}, 11.2 + 2.0086\sqrt{\frac{10.7^2}{50}}\right) = (0 < 8.2, 14.2) \text{ cm}$$

Inferenza sulla **proporzione**

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

$X_i \sim b(p)$ 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

Inferenza sulla **proporzione**

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

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

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

Inferenza sulla **proporzione**

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

$X_i \sim b(p)$ modello

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

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

$n = 1000$
Sì: 33.8%
No: 37.0%
Indeciso: 29.2%

$$IC(0.95) - S\grave{\imath}:\ 0.338 \mp 1.96 \times \sqrt{\frac{0.338(1-0.338)}{1000}}$$

$(0.309, 0.367)$

$$IC(0.95) - No:\ 0.37 \mp 1.96 \times \sqrt{\frac{0.370(1-0.370)}{1000}}$$

$(0.340, 0.400)$

Inferenza sulla **proporzione**

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

$X_i \sim b(p)$ modello

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

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

$n = 250$
Sì: 33.8%
No: 37.0%
Indeciso: 29.2%

$$IC(0.95) - S\text{ì}:\ 0.338 \mp 1.96 \times \sqrt{\frac{0.338(1-0.338)}{250}}$$

$(0.279, 0.397)$

$$IC(0.95) - No:\ 0.37 \mp 1.96 \times \sqrt{\frac{0.370(1-0.370)}{250}}$$

$(0.310, 0.430)$

Inferenza sulla **proporzione**

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

$X_i \sim b(p)$ modello

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

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

$\boldsymbol{n = 2500}$

Sì: 33.8%

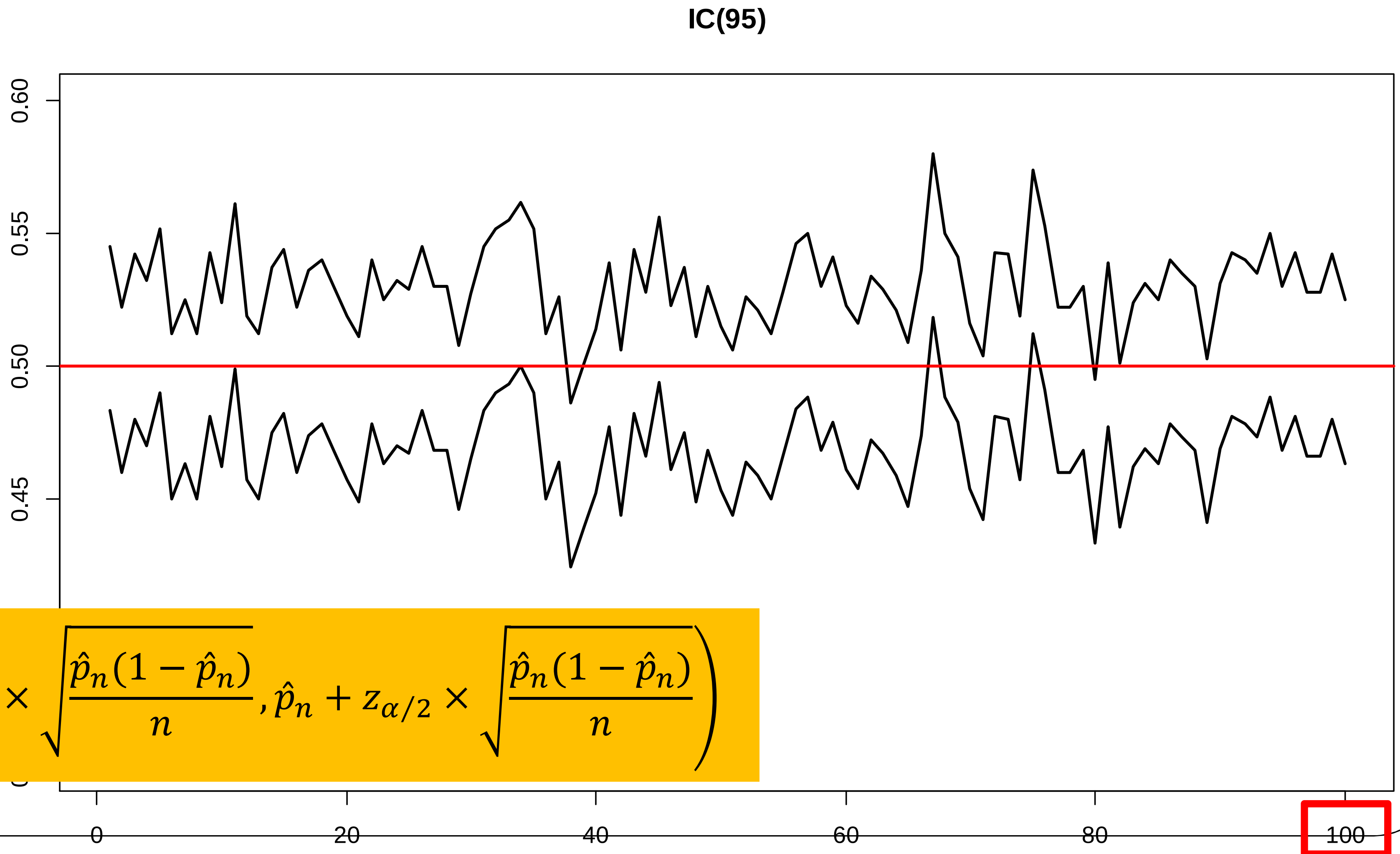
No: 37.0%

Indeciso: 29.2%

di compito…

Intervalli di confidenza

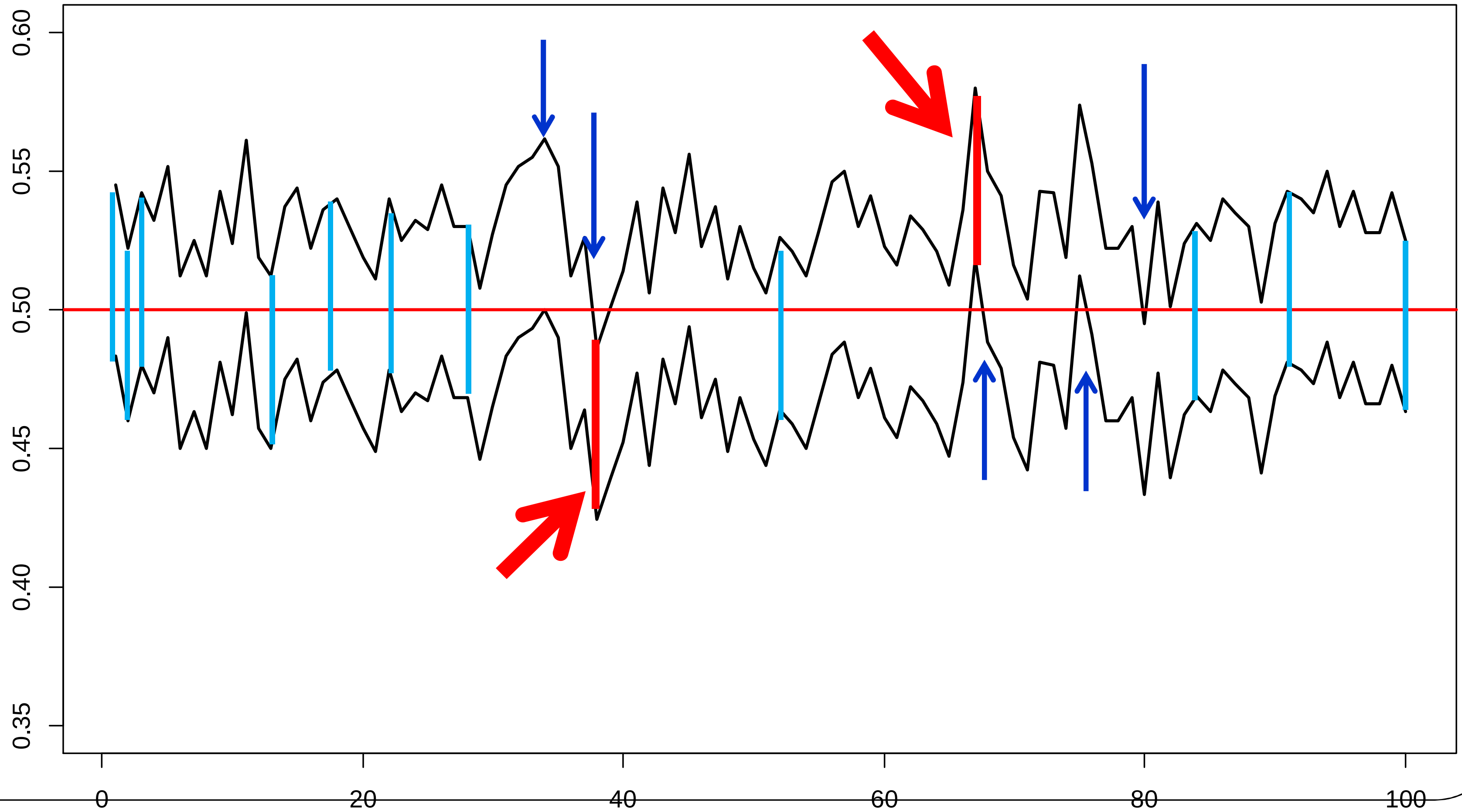
Immaginiamo di avere 100 campioni casuali, ciascuno di 1000 individui, da una popolazione in cui il 50% “vota sì” ($p = 0.5$): per ciascun campione costruisco l’intervallo di confidenza del 95%



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

Intervalli di confidenza

Le linee verticali azzurre sono alcuni dei 100 intervalli di confidenza del 95% corrispondenti: l'interpretazione frequentista dell'IC è che se posso ripetere l'esperimento tante volte, sempre nelle stesse condizioni, il (1-α)% degli intervalli campionari ottenuti contiene il «vero valore» del parametro.



Inferenza sulla **media di una popolazione**

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

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

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

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

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

Grazie al TLC vale anche per modello *non* gaussiano

Inferenza sulla **media di una popolazione**

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

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

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

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

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

$$\boxed{n \geq 30 \Rightarrow t(n-1) \approx N(0,1)}$$

Grazie al TLC vale anche per modello *non* gaussiano

(*) Esercizio (8 pg. 234, circa…)

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, $X_i \sim N(\cancel{\mu,\sigma^2})$ $\qquad \alpha = 0.05$

$$\left(\bar{x}_n - t(n-1)_{\alpha/2}\sqrt{\frac{{s_n}^2}{n}}, \bar{x}_n + t(n-1)_{\alpha/2}\sqrt{\frac{{s_n}^2}{n}}\right)$$

$$\left(11.2 - 2.0086\sqrt{\frac{10.7^2}{50}}, 11.2 + 2.0086\sqrt{\frac{10.7^2}{50}}\right) = (8.2, 14.2) \text{ cm}$$

Intervalli di confidenza

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

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

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

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

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

Esercizio (3 pg. 246)

In una indagine della Gallup(*), ai soggetti 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 o "non hanno espresso un'opinione".

a) quale percentuale di intervistati ha detto sì?
b) IC(95%) per la stima dei detentori di un'arma in casa

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

Esercizio (3 pg. 246)

In una indagine della Gallup, ai soggetti 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 o "non hanno espresso un'opinione".

a) quale percentuale di intervistati ha detto sì?
b) IC(95%) per la stima dei detentori di un'arma in casa

$$\hat{p}_n = \frac{413}{413+646} = 0.400$$

$$\left(\hat{p}_n - z_{\alpha/2}\sqrt{\frac{\hat{p}_n(1-\hat{p}_n)}{n}}, \hat{p}_n + z_{\alpha/2}\sqrt{\frac{\hat{p}_n(1-\hat{p}_n)}{n}}\right) = 0.400 \mp 1.96\sqrt{\frac{0.4 \times 0.6}{(413+646)}}$$

$$\Rightarrow \ \mathbf{(0.370, 0.429)}$$

Esercizio (3 pg. 246)

In una indagine della Gallup, ai soggetti 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 o "non hanno espresso un'opinione".

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?"

$$\hat{p}_n = \frac{413}{413+646} = 0.400$$

$$\left(\hat{p}_n - z_{\alpha/2}\sqrt{\frac{\hat{p}_n(1-\hat{p}_n)}{n}}, \hat{p}_n + z_{\alpha/2}\sqrt{\frac{\hat{p}_n(1-\hat{p}_n)}{n}}\right) = 0.400 \mp 1.96\sqrt{\frac{0.4\times 0.6}{(413+646)}}$$

$$\Rightarrow \ (\mathbf{0.370, 0.429})$$

Esercizio (3 pg. 246)

In una indagine della Gallup, ai soggetti 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 o "non hanno espresso un'opinione".

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?"

$$\hat{p}_n = \frac{413}{413+646} = 0.400 \quad \Longleftrightarrow \quad \boxed{\textbf{CERTO!!}}$$

$$\left(\hat{p}_n - z_{\alpha/2}\sqrt{\frac{\hat{p}_n(1-\hat{p}_n)}{n}}, \hat{p}_n + z_{\alpha/2}\sqrt{\frac{\hat{p}_n(1-\hat{p}_n)}{n}}\right) = 0.400 \mp 1.96\sqrt{\frac{0.4\times 0.6}{(413+646)}}$$

$$\Rightarrow \quad (\mathbf{0.370, 0.429})$$

Esercizio (3 pg. 246)

In una indagine della Gallup, ai soggetti 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 o "non hanno espresso un'opinione".

c) "Possiamo concludere tranquillamente che **meno del 50% della popolazione** ha ~~risposto sì quando è stato loro chiesto se avevano~~ un'arma in casa?"

Sì, con un liv. di conf. del 95%

$$\hat{p}_n = \frac{413}{413+646} = 0.400$$

$$\left(\hat{p}_n - z_{\alpha/2}\sqrt{\frac{\hat{p}_n(1-\hat{p}_n)}{n}}, \hat{p}_n + z_{\alpha/2}\sqrt{\frac{\hat{p}_n(1-\hat{p}_n)}{n}}\right) = 0.400 \mp 1.96\sqrt{\frac{0.4\times 0.6}{(413+646)}}$$

$$\Rightarrow\ (\mathbf{0.370}, \mathbf{0.429} < \mathbf{0.5})$$

Esercizio (3 pg. 246)

In una indagine della Gallup, ai soggetti 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 o "non hanno espresso un'opinione".

d) Qual è una risposta ragionevole ("sensibile"...) alla critica che la Gallup non può fornire risultati attendibili perchè il campione è costituito da soli 1059 adulti, selezionati da una popolazione enorme di oltre 200 milioni di adulti?

$$\left(\hat{p}_n - z_{\alpha/2}\sqrt{\frac{\hat{p}_n(1-\hat{p}_n)}{n}}, \hat{p}_n + z_{\alpha/2}\sqrt{\frac{\hat{p}_n(1-\hat{p}_n)}{n}}\right) = 0.400 \mp 1.96\sqrt{\frac{0.4 \times 0.6}{(413+646)}}$$

$$\Rightarrow \; (\mathbf{0.370}, \mathbf{0.429} < \mathbf{0.5})$$

Esercizio (3 pg. 246)

In una indagine della Gallup, ai soggetti 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 o "non hanno espresso un'opinione".

d) Qual è una risposta ragionevole ("sensibile"...) alla critica che la Gallup non può fornire risultati attendibili perchè il campione è costituito da soli 1059 adulti, selezionati da una popolazione enorme di oltre 200 milioni di adulti?

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

Esercizio (3 pg. 246)

In una indagine della Gallup, ai soggetti 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 o "non hanno espresso un'opinione".

d) Qual è una risposta ragionevole ("sensibile"...) alla critica che la Gallup non può fornire risultati attendibili perchè il campione è costituito da soli 1059 adulti, selezionati da una popolazione enorme di oltre 200 milioni di adulti?

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

Qui il margine di errore è ancora il 3%.