9 Stima per intervalli

9.1 Intervalli di confidenza

Anche se la precisione della stima aumenta con grandi campioni, non è vero che una stima puntuale, valore specifico di una statistica calcolata in corrispondenza di un dato campione, sia esattamente uguale al parametro della popolazione che si intende stimare. A volte è preferibile determinare un intervallo per il quale si ha un certo *livello di fiducia* o *confidenza* che il parametro vi appartenga.
Quindi una stima per intervallo è un intervallo costruito attorno allo stimatore puntuale (di cui si conosce la distribuzione di probabilità) in modo che sia nota e fissata a priori la probabilità che il parametro vero appartenga all'intervallo stesso.
Tale probabilità è detta livello di confidenza e viene indicato con $(1-\alpha)\%$ dove $\alpha \in (0,1)$ è la probabilità che il parametro si trovi al di fuori dell'intervallo di confidenza.
Nel caso di un campione casuale semplice di dimensione n, estratto da una popolazione normale $N(\mu, \sigma^2)$

si possono costruire intervalli di confidenza al livello $(1-\alpha)\%$ sia per la media μ (con σ nota o incognita) sia per la varianza σ^2 (con μ nota o incognita)

9.2 Intervallo di confidenza per la media μ quando la varianza σ^2 è nota

$X_i \sim N(\mu, \sigma^2)\,,\, i = 1, ..., n \Rightarrow \overline{X}_n \sim N(\mu, \frac{\sigma^2}{n})$
La variabile casuale

$$Z_n = \frac{\overline{X}_n - \mu}{\frac{\sigma}{\sqrt{n}}} \sim N(0,1)$$

è indipendente da μ e quindi è la statistica idonea per costruire l'intervallo per μ, essendo σ nota.
Si possono costruire intervalli bilaterali e unilaterali destro e sinistro.
Intervallo bilaterale

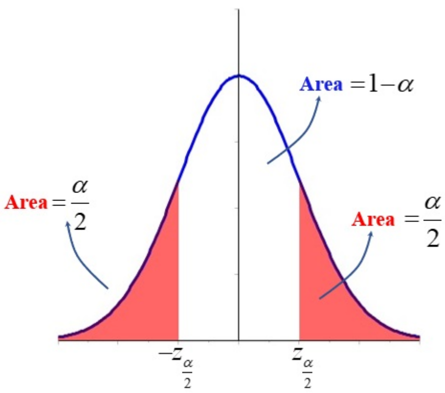


Figura 1: Intervallo bilaterale

$\forall \alpha \in (0,1)\ z_{\frac{\alpha}{2}} : P[Z_n > z_{\frac{\alpha}{2}}] = \frac{\alpha}{2}$
$\Rightarrow P[-z_{\frac{\alpha}{2}} < Z_n < z_{\frac{\alpha}{2}}] = 1 - \alpha$
Sostituendo $Z_n = \frac{\overline{X}_n - \mu}{\frac{\sigma}{\sqrt{n}}}$

$$P[-z_{\frac{\alpha}{2}} < \frac{\overline{X}_n - \mu}{\frac{\sigma}{\sqrt{n}}} < z_{\frac{\alpha}{2}}] = 1 - \alpha$$

ovvero

$$P[\overline{X}_n - z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} < \mu < \overline{X}_n + z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}] = 1 - \alpha$$

$$\rightarrow I = \left(\overline{X}_n - z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}, \overline{X}_n + z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \right)$$

intervallo di confidenza bilat. al livello $(1-\alpha)$ per μ

La lunghezza dell'intervallo bilaterale è

$$\ell_I = 2 z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$$

- se n aumenta, ℓ_I cala $\Rightarrow$ stima più precisa.
- se $(1-\alpha)$ aumenta, $z_{\frac{\alpha}{2}}$ aumenta, ℓ_I aumenta $\Rightarrow$ la stima è meno precisa.
- se σ aumenta, ℓ_I aumenta.

(n, α) si possono controllare, σ dipende dai dati.

Intervalli unilaterali

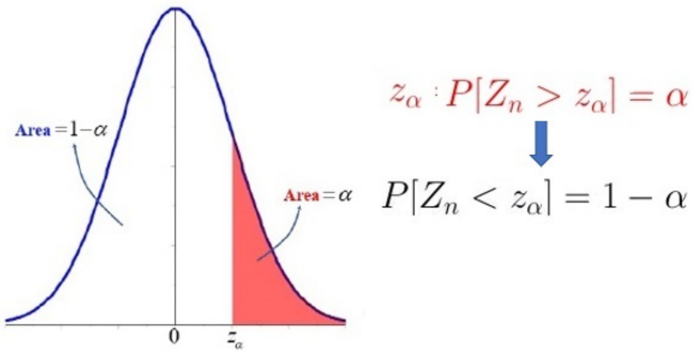


Figura 2: Intervallo unilaterale destro

$$\frac{\overline{X}_n-\mu}{\frac{\sigma}{\sqrt{n}}} < z_\alpha \Rightarrow \mu > \overline{X}_n - z_\alpha\frac{\sigma}{\sqrt{n}}$$

$$\rightarrow \left(\overline{X}_n - z_\alpha\frac{\sigma}{\sqrt{n}}, +\infty\right)$$

intervallo unilaterale destro per μ al livello $(1-\alpha)$.

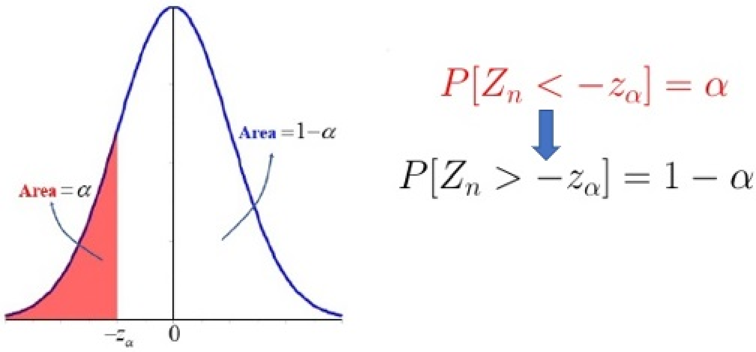


Figura 3: Intervallo unilaterale sinistro

$$\frac{\overline{X}_n-\mu}{\frac{\sigma}{\sqrt{n}}} > -z_\alpha \Rightarrow \mu < \overline{X}_n + z_\alpha\frac{\sigma}{\sqrt{n}}$$

$$\rightarrow \left(-\infty, \overline{X}_n + z_\alpha\frac{\sigma}{\sqrt{n}}\right)$$

intervallo unilaterale sinistro per μ al livello $(1-\alpha)$.

<u>Osservazione</u>

Se viene a mancare l'ipotesi che la distribuzione della popolazione sia normale, a condizione che n, la dimensione del campione, sia sufficientemente grande, in accordo con il Teorema del Limite Centrale ($n \geq 30$)

$$\frac{\overline{X}_n - \mu}{\frac{\sigma}{\sqrt{n}}} \dot{\sim} N(0,1)$$

e si possono costruire ugualmente gli intervalli di confidenza.

<u>Esempio</u>

Vogliamo calcolare

$$P[-1.96 < Z_n < 1.96] = \int_{-1.96}^{1.96} \varphi(z)dz =$$

$$2\int_{0}^{1.96} \varphi(z)dz = 2\left(\int_{-\infty}^{1.96} \varphi(z)dz - 0.5\right) = (\text{tabella})$$

$$= 2(0.975 - 0.5) = 2 \cdot 0.475 = 0.95$$

è equivalente a

$$P\left[\overline{X}_n - 1.96\frac{\sigma}{\sqrt{n}} < \mu < \overline{X}_n + 1.96\frac{\sigma}{\sqrt{n}}\right] = (1-\alpha) = 0.95$$

intervallo di confidenza bilaterale al 95% per μ.
Il 95% delle volte μ si troverà ad una distanza non superiore a $(1.96\frac{\sigma}{\sqrt{n}})$ dalla media campionaria dei dati.
Posso concludere che col 95% di confidenza la media μ della popolazione appartiene ad

$$I = \left(\overline{X}_n - 1.96\frac{\sigma}{\sqrt{n}}, \overline{X}_n + 1.96\frac{\sigma}{\sqrt{n}}\right)$$

$(1-\alpha) = 0.95$
$\alpha = 0.05$, $\frac{\alpha}{2} = 0.025$
$1.96 = z_{\frac{\alpha}{2}} = z_{0.025}$

Voglio calcolare:

$$P[Z_n < 1.645] = \int_{-\infty}^{1.645} \varphi(z)dz = (\text{tabella}) = 0.95$$

è equivalente a

$$P\left[\overline{X}_n - 1.645\frac{\sigma}{\sqrt{n}} < \mu\right] = 0.95$$

intervallo di confidenza unilaterale destro al 95%, cioè

$$I = \left(\overline{X}_n - 1.645\frac{\sigma}{\sqrt{n}}, +\infty\right)$$

$(1-\alpha) = 0.95$
$\alpha = 0.05$
$z_\alpha = z_{0.05} = 1.645$
L'intervallo di confidenza unilaterale sinistro al 95% sarà:

$$I = \left(-\infty, \overline{X}_n + 1.645\frac{\sigma}{\sqrt{n}}\right)$$

I più comuni livelli di confidenza

$$1-\alpha = 90\% \Rightarrow \alpha = 0.1 \quad \begin{cases} z_\alpha = z_{0.1} = 1.28 \\ z_{\frac{\alpha}{2}} = z_{0.05} = 1.645 \end{cases}$$

$$1-\alpha = 95\% \Rightarrow \alpha = 0.05 \quad \begin{cases} z_\alpha = z_{0.05} = 1.645 \\ z_{\frac{\alpha}{2}} = z_{0.025} = 1.96 \end{cases}$$

$$1-\alpha = 99\% \Rightarrow \alpha = 0.01 \quad \begin{cases} z_\alpha = z_{0.01} = 2.326 \\ z_{\frac{\alpha}{2}} = z_{0.005} = 2.58 \end{cases}$$

<u>Osservazione</u>
Ci possono essere diversi intervalli di confidenza bilaterali che hanno lo stesso livello di confidenza.
$P[-1.96 < Z_n < 1.96] = 0.95$
ma anche
$P[-1.68 < Z_n < 2.70] = 0.95$

Problema: quale intervallo è migliore?
Si sceglie l'intervallo più corto.
Si può dimostrare che l'intervallo simmetrico rispetto ad $\overline{X}_n$ è quello che rende minima l'ampiezza dell'intervallo bilaterale I_{bil}.

Esempio

Sia dato un campione casuale di dimensione $n = 4$, estratto da una popolazione normale $N(\mu, \sigma = 3)$.
DATI: 1.2, 3.4, 0.6, 5.6
$\overline{x}_n = \overline{x}_4 = 2.7$ è la stima per μ.
L'intervallo bilaterale simmetrico al 95% per μ è:

$$\left(2.7 - 1.96 \cdot \frac{3}{2}, 2.7 + 1.96 \cdot \frac{3}{2}\right) = (-0.24, 5.64)$$

e $\ell_I = 5.88$
Se usiamo $P[-1.68 < Z_n < 2,70] = 0.95$ ricaviamo:

$$\left(2.7 - 2.7 \cdot \frac{3}{2}, 2.7 + 1.68 \cdot \frac{3}{2}\right) = (-1.35, 5.22)$$

e $\ell_I = 6.57$ è più ampio

Gli intervalli unilaterali in corrispondenza a $z_\alpha = z_{0.05} = 1.645$ saranno

$(0.23, +\infty)$ destro, $(-\infty, 5.16)$ sinistro.

In alcune situazioni si richiede che l'intervallo di confidenza bilaterale abbia una lunghezza fissata.

Problema: determinare la numerosità del campione che garantisca il risultato richiesto.

Da $\ell_I = 2z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \Rightarrow n = \left(2z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\ell_I}\right)^2$
(si approssima all'intero successivo).

<u>Esempio</u>
Dato un campione casuale estratto da una popolazione normale $N(\mu, \sigma = 0.2)$, calcolare n affinchè l'intervallo di confidenza bilaterale al 99% per μ abbia lunghezza $\ell_I = 0.1$.

$1 - \alpha = 0.99, \Rightarrow \alpha = 0.01, \Rightarrow z_{\frac{\alpha}{2}} = z_{0.005} = 2.58.$
$\ell_I = 2 \cdot z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}} = 2 \cdot 2.58 \cdot \frac{0.2}{\sqrt{n}} \Rightarrow$
$\Rightarrow n = (5.16 \cdot 0.2)^2 \simeq 106.5 \Rightarrow n = 107$

9.3 Intervallo di confidenza per la media μ quando la varianza σ^2 non è nota

Non è ora più possibile usare la variabile casuale $Z_n = \frac{\overline{X}_n - \mu}{\sigma/\sqrt{n}}$ poichè contiene l'incognita σ e quindi non è una statistica.
Bisogna usare la statistica

$$T = \frac{\overline{X}_n - \mu}{\frac{S}{\sqrt{n}}} \sim t_{n-1} \quad \text{(Teorema 5)}$$

In analogia al caso precedente si costituiscono gli intervalli di confidenza bilaterali e unilaterali destro e sinistro al livello $(1-\alpha)\%$, $\forall \alpha \in (0,1)$ sostituendo alla deviazione standard σ della popolazione la deviazione campionaria standard S e ai percentili della normale standard Z i percentili della t di Student al grado di libertà $(n-1)$.

$$I = \left(\overline{X}_n - t_{\frac{\alpha}{2},n-1} \frac{S}{\sqrt{n}}, \overline{X}_n + t_{\frac{\alpha}{2},n-1} \frac{S}{\sqrt{n}} \right)$$ intervallo bilaterale per μ

$$\left(\overline{X}_n - t_{\alpha,n-1} \frac{S}{\sqrt{n}}, +\infty \right)$$ intervallo unilat. dx per μ

$$\left(-\infty, \overline{X}_n + t_{\alpha,n-1}\frac{S}{\sqrt{n}}\right)$$ intervallo unilat. sx per μ

Quando σ non è nota, l'intervallo costruito attraverso la t di Student è più ampio poichè le code della t sono più pesanti.

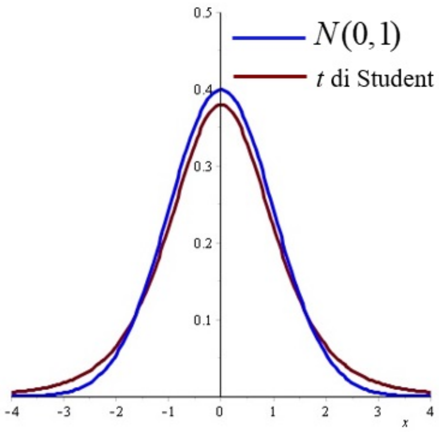


Figura 4: Normale standard vs t di Student

Esempio

Sia dato un campione casuale di dimensione $n = 4$ estratto da una popolazione normale $N(\mu, \sigma^2)$. Determinare un intervallo di confidenza bilaterale al 95% per μ.

DATI: 1.2, 3.4, 0.6, 5.6
$\overline{x}_n = \overline{x}_4 = 2.7$
$s^2 = \frac{1}{3}\left[(1.2)^2 + (3.4)^2 + (0.6)^2 + (5.6)^2 - 4 \cdot (2.7)^2\right]$

$\downarrow$
$s \simeq 2.277$
$1 - \alpha = 0.95 \Rightarrow \alpha = 0.05 \Rightarrow \frac{\alpha}{2} = 0.025$
$t_{\frac{\alpha}{2},n-1} = t_{0.025,3} = (\text{tabella}) = 3.182$
$\ell_I = 2t_{\frac{\alpha}{2},n-1}\frac{s}{\sqrt{n}} \Rightarrow \ell_I = 2 \cdot 3.182 \cdot \frac{2.277}{2}$
$\Rightarrow \ell_I = 7.25$ (più ampio rispetto a 5.88 calcolato quando σ è nota)
Osservazione
Nella pratica si costruiscono gli intervalli di confidenza per μ quando σ non è nota per mezzo della t di Student solo per piccoli campioni ($n \leq 30$). Per grandi campioni si usa invece la normale standard Z_n sostituendo alla σ incognita la S calcolata dai dati.

9.4 Intervallo di confidenza per la varianza σ^2 quando la media μ non è nota

Una stima per σ^2 può essere determinata usando la statistica

$$V = (n-1)\frac{S^2}{\sigma^2} \sim \chi^2_{n-1}, \quad \text{(Teorema 4)}$$

Intervallo bilaterale

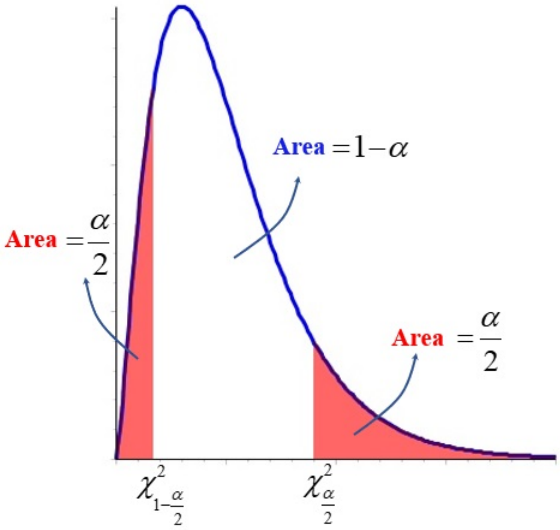


Figura 5: Intervallo bilaterale

$$P\left[\chi^2_{1-\frac{\alpha}{2},n-1} < V < \chi^2_{\frac{\alpha}{2},n-1}\right] = 1-\alpha$$

sostituendo V si ha

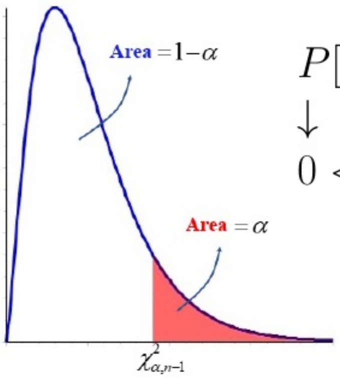
$$\chi^2_{1-\frac{\alpha}{2},n-1} < \frac{(n-1)S^2}{\sigma^2} < \chi^2_{\frac{\alpha}{2},n-1}$$

cioè

$$\frac{(n-1)S^2}{\chi^2_{\frac{\alpha}{2},n-1}} < \sigma^2 < \frac{(n-1)S^2}{\chi^2_{1-\frac{\alpha}{2},n-1}}$$

$$\Rightarrow \left(\frac{(n-1)S^2}{\chi^2_{\frac{\alpha}{2},n-1}}, \frac{(n-1)S^2}{\chi^2_{1-\frac{\alpha}{2},n-1}}\right)$$ intervallo bilaterale per σ^2 quando μ non è nota

Intervalli unilaterali

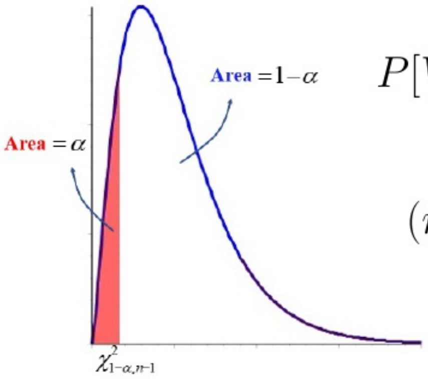


$$P[0 < V < \chi^2_{\alpha,n-1}] = 1-\alpha$$
$$\downarrow$$
$$0 < (n-1)\frac{S^2}{\sigma^2} < \chi^2_{\alpha,n-1}$$

Figura 6: Intervallo unilaterale destro

$$\Rightarrow \sigma^2 > (n-1)\frac{S^2}{\chi^2_{\alpha,n-1}}$$

$$\Rightarrow \left(\frac{(n-1)S^2}{\chi^2_{\alpha,n-1}}, +\infty\right)$$ intervallo unilaterale destro per σ^2



$$P[V > \chi^2_{1-\alpha,n-1}] = 1-\alpha$$
$$\downarrow$$
$$(n-1)\frac{S^2}{\sigma^2} > \chi^2_{1-\alpha,n-1}$$

Figura 7: Intervallo unilaterale sinistro

$$\Rightarrow \sigma^2 < (n-1)\frac{S^2}{\chi^2_{1-\alpha,n-1}}$$

$$\Rightarrow \left(0, \frac{(n-1)S^2}{\chi^2_{1-\alpha,n-1}}\right)$$ intervallo unilaterale sinistro per σ^2

Esempio

Dato un campione casuale di dimensione $n = 10$, di rondelle di spessore ridotto, estratto da una popolazione normale $N(\mu, \sigma^2)$, stimare σ^2 con μ non nota.

DATI: spessori (in mm)

0.123, 0.133, 0.124, 0.125, 0.126

0.128, 0.120, 0.124, 0.130, 0.126

$$\Rightarrow \overline{x}_n = \overline{x}_{10} = 0.125$$

$$s^2 = \frac{1}{9}\left[(0.123^2 + 0.133^2 + \ldots + 0.126^2) - 10 \cdot 0.125^2\right] \simeq 1.366 \cdot 10^{-5}$$

Vogliamo determinare un intervallo di confidenza bilaterale per σ^2 al 90%.

$$1 - \alpha = 0.90,\ \alpha = 0.1,\ \Rightarrow \frac{\alpha}{2} = 0.05,\ 1 - \frac{\alpha}{2} = 0.95$$

$\chi^2_{\frac{\alpha}{2},n-1} = \chi^2_{0.05,9} = 16.919$

$\chi^2_{1-\frac{\alpha}{2},n-1} = \chi^2_{0.95,9} = 3.325$

quindi

$\frac{(n-1)s^2}{\chi^2_{\frac{\alpha}{2},n-1}} < \sigma^2 < \frac{(n-1)s^2}{\chi^2_{1-\frac{\alpha}{2},n-1}}$

diventa

$\frac{9 \cdot 1.366 \cdot 10^{-5}}{16.919} < \sigma^2 < \frac{9 \cdot 1.366 \cdot 10^{-5}}{3.325}$

$\Rightarrow (7.26 \cdot 10^{-6}, 36.97 \cdot 10^{-6})$ è l'intervallo bilaterale per σ^2.

9.5 Intervallo di confidenza per la varianza σ^2 quando la media μ è nota

In questo caso si può usare la statistica

$$U = \sum_{i=1}^{n} \frac{(X_i - \mu)^2}{\sigma^2} \sim \chi^2_n, \quad \text{(Teorema 3)}$$

e ragionare come nel caso 9.4, sostituendo alla statistica V la statistica U (il grado di libertà è n).

$$\left(\frac{\sum_{i=1}^{n}(X_i-\mu)^2}{\chi^2_{\frac{\alpha}{2},n}}, \frac{\sum_{i=1}^{n}(X_i-\mu)^2}{\chi^2_{1-\frac{\alpha}{2},n}}\right)$$ intervallo bilaterale per σ^2 quando μ è nota

$$\left(\frac{\sum_{i=1}^{n}(X_i-\mu)^2}{\chi^2_{\alpha,n}}, +\infty\right)$$ intervallo unilaterale destro per σ^2

$$\left(0, \frac{\sum_{i=1}^{n}(X_i-\mu)^2}{\chi^2_{1-\alpha,n}}\right)$$ intervallo unilaterale sinistro per σ^2

SCHEMA

stima per μ:

- se σ è nota $\Rightarrow N(0,1)$
- se σ è incognita $\Rightarrow t_{n-1}$ solo per piccoli campioni $n \leq 30$
  - $\Rightarrow N(0,1)$ per $n \geq 30$ sostituendo σ con s

stima per σ^2:

- se μ è nota $\Rightarrow \chi^2_n$
- se μ è incognita $\Rightarrow \chi^2_{n-1}$

e la numerosità del campione non conta

Esempi

1 Stimare il numero medio di battiti cardiaci al minuto di una popolazione normale $N(\mu, \sigma = 10)$ considerando un campione casuale, di dimensione $n = 49$, avente $\overline{x}_{49} = 90$.
Determinare gli intervalli di confidenza bilaterali per μ al 90%, 95%, 99%.

- $(1-\alpha) = 0.90 \Rightarrow \frac{\alpha}{2} = 0.05$, $z_{\frac{\alpha}{2}} = z_{0.05} = 1.645$

$90 - 1.645 \cdot \frac{10}{7} < \mu < 90 + 1.645 \cdot \frac{10}{7}$
$\downarrow$
$\mu \in (87.65; 92.35)$
$\ell_I = 4.70$

- $(1-\alpha) = 0.95 \Rightarrow \frac{\alpha}{2} = 0.025$, $z_{\frac{\alpha}{2}} = z_{0.025} = 1.96$
$\ell_I = 5.60$

- $(1-\alpha) = 0.99 \Rightarrow \frac{\alpha}{2} = 0.005$, $z_{\frac{\alpha}{2}} = z_{0.005} = 2.58$
$\ell_I = 7.36$

Se $(1-\alpha)$ aumenta, ℓ_I aumenta.

2 Da una popolazione si estrae un campione di 16 oggetti di cui si misura il peso. Stimare il peso medio μ della popolazione, sapendo che $\overline{x}_{16} = 3.42$ grammi e $s = 0.68$ grammi, con un livello di confidenza del 99%.
Poichè eseguo delle misure, anche se il testo non lo specifica, la distribuzione della popolazione è una normale $N(\mu, \sigma^2)$. Bisogna stimare μ con σ^2 incognita.

$n = 16$ campione piccolo (< 30)
$\nu = n - 1 = 15$ gradi di libertà della t di Student
$1 - \alpha = 0.99$, $\frac{\alpha}{2} = 0.005$

$t_{\frac{\alpha}{2},n-1} = t_{0.005,15} = 2.947$

$3.42 - 2.947 \cdot \frac{0.68}{4} < \mu < 3.42 + 2.947 \cdot \frac{0.68}{4}$
$\Rightarrow \mu \in (2.91, 3.93)$

3 Dato un campione casuale di 16 studenti scelti da una classe quinta di una scuola superiore, se ne misura l'altezza. È nota la varianza campionaria $s^2 = 37.09 \text{cm}^2$.
Determinare un intervallo di confidenza bilaterale al

95% per la varianza σ^2 della popolazione costituita da tutti gli studenti delle classi quinte.

Poichè faccio delle misure, la densità della popolazione è una normale $N(\mu, \sigma^2)$.
Per costruire un intervallo di confidenza per σ^2 quando μ non è nota uso la statistica V e quindi χ^2_{n-1}.

$1-\alpha = 0.95$, $\alpha = 0.05$, $\frac{\alpha}{2} = 0.025$, $1-\frac{\alpha}{2} = 0.975$

$\nu = n-1 = 15$ gradi di libertà della χ^2

$\chi^2_{\frac{\alpha}{2},n-1} = \chi^2_{0.025,15} = 27.488$

$\chi^2_{1-\frac{\alpha}{2},n-1} = \chi^2_{0.975,15} = 6.262$

$\frac{15 \cdot 37.09}{27.488} < \sigma^2 < \frac{15 \cdot 37.09}{6.262}$

$\Rightarrow \sigma^2 \in (20.23; 88.84)$

4 Le misure di un campione casuale di 200 sferette da cuscinetto prodotte da un macchinario, in una settimana, forniscono i seguenti dati:
$\overline{x}_{200} = 0.824\text{cm}$

$s = 0.042$cm
Determinare un intervallo di confidenza bilaterale per la media μ della popolazione di sferette al 95%.

La funzione di densità della popolazione è una normale $N(\mu, \sigma^2)$.
La σ^2 non è nota $\rightarrow$ dovrei applicare la statistica T, ma la tabella della t di Student arriva fino al grado di libertà $\nu = 50$ e nel nostro caso $\nu = n-1 = 199$!!.
Poichè i grafici della t e della Z sono simili, si può usare la $N(0, 1)$ sostituendo a σ incognita il valore campionario $s = 0.042$.

$1 - \alpha = 0.95$, $\alpha = 0.05$, $\frac{\alpha}{2} = 0.025$, $z_{0.025} = 1.96$

$0.824 - 1.96 \cdot \frac{0.042}{\sqrt{200}} < \mu < 0.824 + 1.96 \cdot \frac{0.042}{\sqrt{200}}$

$\Rightarrow \mu \in (0.818; 0.83)$