Esempio 2a

Consideriamo la distribuzione dei livelli di colesterolo della popolazione maschile di ipertesi fumatori. Sappiamo che è una distribuzione con media μ ignota e deviazione $\sigma = 56\,\mathrm{mg/dL}$. Supponiamo di selezionare un campione casuale di dimensione $n = 49$. La media di tale campione è $\bar{x} = 217\,\mathrm{mg/dL}$.

- Qual è intervallo di confidenza per μ con confidenza del 95%?

$$\begin{aligned}
0.95 &= P\Big(\bar{X} - \varepsilon \leq \mu \leq \bar{X} + \varepsilon\Big) \\
&= P\Big(\mu - \varepsilon \leq \bar{X} \leq \mu + \varepsilon\Big) \\
&= P\Big(-\frac{\varepsilon}{\sigma/\sqrt{n}} \leq Z \leq \frac{\varepsilon}{\sigma/\sqrt{n}}\Big) \\
&= P\Big(-\frac{\varepsilon}{8} \leq Z \leq \frac{\varepsilon}{8}\Big).
\end{aligned}$$

Dalla tavola sappiamo che $0.95 = P\big(-1.96 \leq Z \leq 1.96\big)$ quindi $\varepsilon = 8 \cdot 1.96 = 15.68$. L'intervallo di confidenza è $[202, 233]$.

Esempio 2b

Consideriamo la distribuzione dei livelli di colesterolo della popolazione maschile di ipertesi fumatori. Sappiamo che è una distribuzione con media μ ignota e deviazione $\sigma = 46\,\mathrm{mg/dL}$.

- Che dimensione deve avere il campione per ottenere un intervallo di confidenza per μ di ampiezza $20\,\mathrm{mg/dL}$ con confidenza del 99%?

$$\begin{aligned} 0.99 &= P\Big(-\frac{\varepsilon}{\sigma/\sqrt{n}} \leq Z \leq \frac{\varepsilon}{\sigma/\sqrt{n}}\Big) \\ &= P\Big(-\frac{10}{46/\sqrt{n}} \leq Z \leq \frac{10}{46/\sqrt{n}}\Big) \end{aligned}$$

Dalla tavola sappiamo che $0.99 = P\big(-2.58 \leq Z \leq 2.58\big)$ quindi

$$2.58 = \frac{10}{46/\sqrt{n}}$$

Quindi $n \geq 141$.

Esempio 3

I bambini sani hanno un livello noto di emoglobina μ_s. I bambini intossicati da piombo hanno un livello di emoglobina generalmente più basso.

Vogliamo stabilire se una popolazione di bambini è da ritenersi intossicata. Da un campione di dimensione $n = 74$ vogliamo stimare il livello medio di emoglobina μ supponendo nota la deviazione standard $\sigma = 0.85\,\mathrm{g/dL}$. La media campionaria è $\bar{x} = 10.6\,\mathrm{g/dL}$.

- In base a questo campione calcolare l'intervallo di confidenza **unilaterale** al 95% per μ.

$$\begin{aligned} 0.95 &= P\big(\mu \le \bar{X} + \varepsilon\big) \\ &= P\Big(Z \le \frac{\varepsilon}{\sigma/\sqrt{n}}\Big) \\ &= P\Big(Z \le \frac{\varepsilon}{0.85/\sqrt{74}}\Big) \end{aligned}$$

dalle tavole otteniamo che $\varepsilon = 1.65 \cdot 0.85/\sqrt{74} = 0.1625$.

Varianza ignota

La statistica

$$\sqrt{\frac{1}{n}\sum_{i=1}^{n}(X_i - \mu)^2}$$

È uno stimatore non distorto della deviazione standard σ.

Purtroppo generalmente μ non è nota.

Possiamo sostituire μ con $\bar{X}$ ma dobbiamo correggere il fattore di normalizzazione:

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

Questo è uno stimatore non distorto della deviazione standard $E(S) = \sigma$

Si chiama **deviazione standard campionaria**.

Distribuzione t_n di Student

Ulteriore complicazione: la variabile aleatoria

$$\frac{\bar{X} - \mu}{S/\sqrt{n}}$$

Non è normale!

Se la variaile $\bar{X}$ è normale allora

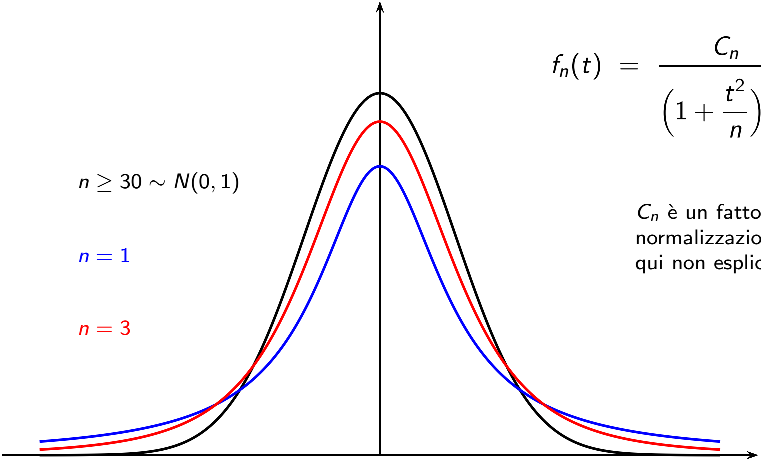
$$T = \frac{\bar{X} - \mu}{S/\sqrt{n}}$$

Si chiama distribuzione t_{n-1} di Student (con $n-1$ gradi di libertà).

Se n è abbastanza grande t_{n-1} è ben approssimata dalla disribuzione normale standard.

Le distribuzioni t di Student.

Anche qui una famiglia di distribuzioni (n è un qualsiasi intero positivo)



$$f_n(t) \;=\; \frac{C_n}{\left(1 + \dfrac{t^2}{n}\right)^{\frac{n+1}{2}}}$$

C_n è un fattore di normalizzazione che qui non esplicitiamo.

Esempio

Consideriamo un campione di $n = 10$ selezionato fra la popolazione di neonati cui viene somministrato un medicinale contenente alluminio. Assumiamo che i livelli di alluminio plasmatico abbiano distribuzione normale con media μ e deviazione standard σ ignote.

Sappiamo che la media del campione è $\bar{x} = 37.2\,\mu$g/l e la deviazione standard campionaria è $s = 7.13$ μg/l.

Vogliamo l'intervallo di confidenza al 95% per μ.

$$\begin{aligned} P(\bar{X} - \varepsilon \leq \mu \leq \bar{X} + \varepsilon) &= 0.95 \\ P(-\frac{\varepsilon}{s/\sqrt{n}} \leq T_{n-1} \leq \frac{\varepsilon}{s/\sqrt{n}}) &= 0.95 \\ P(-2.262 \leq T_9 \leq 2.262) &= 0.95 \end{aligned}$$

$$\frac{\varepsilon}{s/\sqrt{n}} = \frac{\varepsilon}{7.13/\sqrt{10}} = 2.262 \qquad \Rightarrow \qquad \varepsilon = 5.1$$

L'intervallo di confidenza è $[37.2 - 5.1,\ 37.2 + 5.1] = [32.1,\ 42.3]$.

Esempio

Consideriamo un campione di $n = 10$ selezionato fra la popolazione di neonati cui viene somministrato un medicinale contenente alluminio. Assumiamo che i livelli di alluminio plasmatico abbiano distribuzione normale con media μ e deviazione standard σ ignote.

Sappiamo che la media del campione è $\bar{x} = 37.2\,\mu$g/l e la deviazione standard campionaria è $s = 7.13\ \mu$g/l.

Vogliamo l'intervallo di confidenza al 99% per μ.

$$
\begin{aligned}
P(\bar{X} - \varepsilon \leq \mu \leq \bar{X} + \varepsilon) &= 0.99 \\
P\left(-\frac{\varepsilon}{s/\sqrt{n}} \leq T_{n-1} \leq \frac{\varepsilon}{s/\sqrt{n}}\right) &= 0.99 \\
P(-3.250 \leq T_9 \leq 3.250) &= 0.99
\end{aligned}
$$

$$
\frac{\varepsilon}{s/\sqrt{n}} = \frac{\varepsilon}{7.13/\sqrt{10}} = 3.250 \qquad \Rightarrow \qquad \varepsilon = 7.3
$$

L'intervallo di confidenza è $[37.2 - 7.3,\ 37.2 + 7.3] = [29.9,\ 44.5]$.

Esempio

Consideriamo un campione di $n = 10$ selezionato fra la popolazione di neonati cui viene somministrato un medicinale contenente alluminio. Assumiamo che i livelli di alluminio plasmatico abbiano distribuzione normale con media μ e deviazione standard σ ignote.

Sappiamo che la media del campione è $\bar{x} = 37.2\,\mu$g/l e la deviazione standard campionaria è $s = 7.13\ \mu$g/l.

Vogliamo l'intervallo di confidenza unilaterale 95% per μ.

$$\begin{aligned} P(\mu \leq \bar{X} + \varepsilon) &= 0.95 \\ P(T_{n-1} \leq \frac{\varepsilon}{s/\sqrt{n}}) &= 0.95 \\ P(T_9 \leq 1.833) &= 0.95 \end{aligned}$$

$$\frac{\varepsilon}{s/\sqrt{n}} = \frac{\varepsilon}{7.13/\sqrt{10}} = 1.833 \qquad \Rightarrow \qquad \varepsilon = 4.1$$

L'intervallo di confidenza è $[0,\ 37.2 + 4.1] = [0,\ 41.3]$.