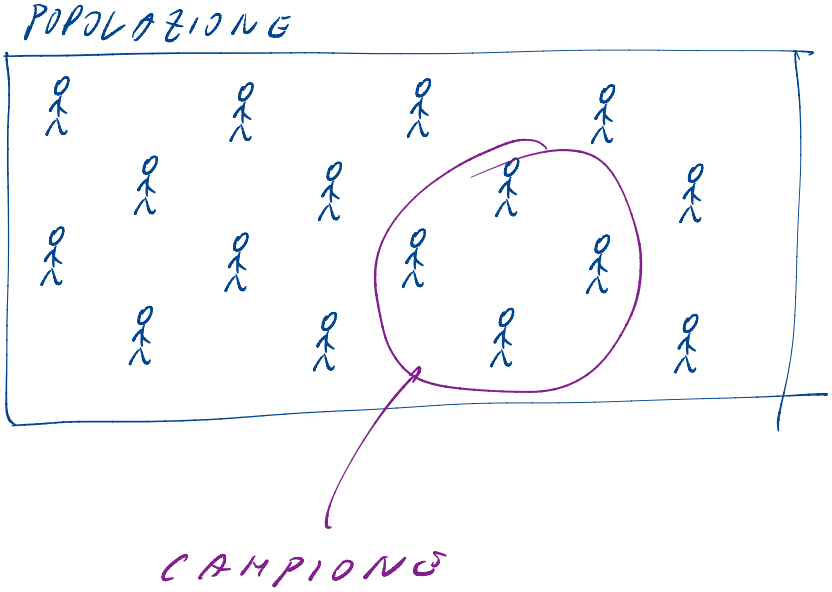
Probabilità

$$\text{probabilità di } A = \frac{\text{n. di ripetizioni in cui si verifica } A}{\text{n. di ripetizioni totali}}$$

A: evento

Stime e intervalli di confidenza



X v. stat. numerica

$$\mu(\text{pop.}) = \frac{\text{somma di tutti i val. di } X}{\text{numerosità della popol.}}$$

STIMA

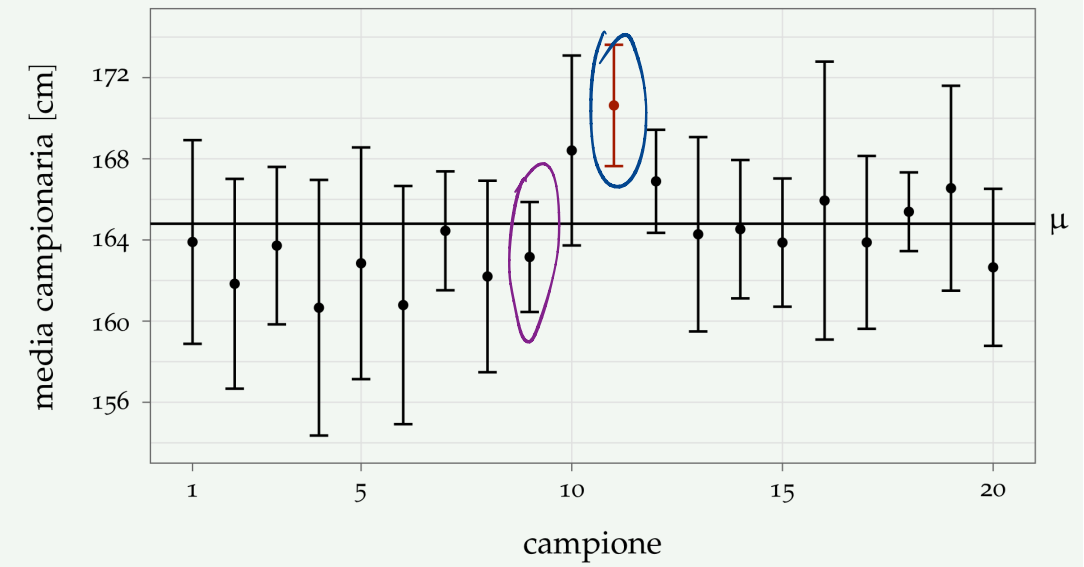
$$\hat{\mu} = \mu(\text{campione}) = \frac{x_1 + x_2 + \dots + x_n}{n}$$

n: numerosità del c.

"μ è nell'intervallo $\hat{\mu} \pm \Delta$ con il 95% di probabilità"

$\hat{\mu} \pm \Delta$: intervallo di confidenza

95%: livello di confidenza



Int. di conf. della media

X v. stat. numerica.

Teorema centrale del limite

Per n grande la media campionaria (campioni con numerosità n) si distribuisce in maniera gaussiana con

media = μ e d. st. = $\frac{\sigma}{\sqrt{n}}$

$\hat{\mu}$ è in $\mu \pm \frac{\sigma}{\sqrt{n}}$ col ~68% di p.

$\hat{\mu}$ è in $\mu \pm 2\frac{\sigma}{\sqrt{n}}$ col ~95% di p.

μ è in $\hat{\mu} \pm 2\frac{\sigma}{\sqrt{n}}$ col ~95% di p.

$\hat{\sigma} = \sigma(\text{campione})$

$\hat{\mu} = \mu(\text{campione})$ $\quad \hat{\sigma} = \sigma(\text{campione})$

$\widehat{SEM} = \frac{\hat{\sigma}}{\sqrt{n}}$ ← "err. standard della media"

$\hat{\mu} \pm 1.96\ \widehat{SEM}$ al 95%

$\hat{\mu} \pm 2.58\ \widehat{SEM}$ al 99%

Metodo per campioni piccoli

$$\hat{SEM} = \frac{S(\text{campione})}{\sqrt{n}}$$

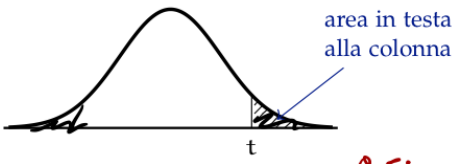
$$\hat{\eta} \pm t_{\frac{\alpha}{2}, n-1} \cdot \hat{SEM} \quad \text{al livello } 1-\alpha$$

$1-\alpha$ → 95%

$\frac{\alpha}{2}$ → 2.5% = 0.025

§C.2 La t di Student

Per ciascun numero di gradi di libertà (riga) e ciascuna probabilità (colonna), la tavola riporta il valore t tale che la t di Student supera t con quella probabilità (area della coda destra). Nella notazione del testo, la colonna $\frac{\alpha}{2}$ e la riga con ν gradi di libertà danno la soglia $t_{\alpha/2,\nu}$ (per il test sulla media di un campione di n individui, $\nu = n - 1$). L'ultima riga, ∞, è la gaussiana standard.



95% 99%

gradi di libertà	0.25	0.10	0.05	0.025	0.01	0.005
1	1.00	3.08	6.31	12.71	31.82	63.66
2	0.82	1.89	2.92	4.30	6.96	9.92
3	0.76	1.64	2.35	3.18	4.54	5.84
4	0.74	1.53	2.13	2.78	3.75	4.60
5	0.73	1.48	2.02	2.57	3.36	4.03
6	0.72	1.44	1.94	2.45	3.14	3.71
7	0.71	1.41	1.89	2.36	3.00	3.50
8	0.71	1.40	1.86	2.31	2.90	3.36
9	0.70	1.38	1.83	2.26	2.82	3.25
10	0.70	1.37	1.81	2.23	2.76	3.17
11	0.70	1.36	1.80	2.20	2.72	3.11
12	0.70	1.36	1.78	2.18	2.68	3.05
13	0.69	1.35	1.77	2.16	2.65	3.01

$n = 10$

Esercizio
Delle 36 studentesse che hanno partecipato alla raccolta delle stature (capitolo 17) ne sono state estratte a caso 10, con le stature seguenti, in centimetri.

167 178 168 165 170 160 159 170 165 160

Stima la statura media delle 36 studentesse con un intervallo di confidenza al 95%.

$$\hat{\mu} = \frac{167 + 178 + \dots}{10} = 166.2$$

$$\sigma^2(\text{campione}) = \frac{167^2 + 178^2 + \dots}{10} - \hat{\mu}^2 = \dots$$

$$\sigma(\text{campione}) = 5.51$$

$$s(\text{camp.}) = \sigma(\text{camp.}) \cdot \sqrt{\frac{n}{n-1}} = 5.81 \qquad n = 10$$

$$\widehat{SEM} = \frac{s(\text{camp.})}{\sqrt{n}} = 1.84$$

$$\hat{\mu} \pm 2.26 \cdot \widehat{SEM} = 166.2 \pm 4.2$$

RIFERITA ALLA CASUALITÀ DEL CAMPIONE

"col 95% di probabilità

μ è compresa fra 162.0 e 170.4"

Stima delle proporzioni

$$\hat{p} = \frac{\text{n. di casi}}{n} = \mu_x(\text{campione})$$

vale 1 per gli ind. positivi
0 ——— negativi

$$\sigma_x^2(\text{campione}) = \mu_x(\text{campione}) - \mu_x^2(\text{campione})$$

$$\hat{\mu}_x$$

$$= \hat{\mu}_x(1 - \hat{\mu}_x)$$

$$\hat{\sigma}_x = \sqrt{\hat{\mu}_x(1 - \hat{\mu}_x)}$$

$$\widehat{SE}\hat{n} = \sqrt{\frac{\hat{\mu}_x(1 - \hat{\mu}_x)}{n}}$$

$$\hat{p} \pm 1.96 \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \quad \text{al } 95\%$$

Campione 1000 persone

250 fumatori

$$\hat{p} = \frac{250}{1000} = 0.25 = 25\%$$

$$0.25 \pm 1.96 \sqrt{\frac{0.25\,(1-0.25)}{1000}} \quad \text{al } 95\%$$

$$= (25.0 \pm 2.7)\%$$

"$22.3\% \leq p \leq 27.7\%$ col 95% di prob."

Esercizio 104 (p. 143)**.** Su un campione di compresse di un lotto si stima il contenuto medio di principio attivo, con l'intervallo di confidenza al 95% da 249 a 253 mg. Quali delle affermazioni seguenti sono corrette?

1. Il 95% delle compresse del lotto contiene fra 249 e 253 mg di principio attivo.
2. Intervalli costruiti nello stesso modo, a partire da campioni diversi, contengono il contenuto medio vero del lotto nel 95% dei casi.
3. Se si ripetesse il campionamento, il 95% delle nuove medie campionarie cadrebbe fra 249 e 253 mg.
4. Con un campione più grande l'intervallo sarebbe, di regola, più stretto.

Soluzione.

1. L'affermazione è falsa, perché l'intervallo che abbiamo calcolato riguarda la *media* del lotto e non le singole compresse, che sono molto più disperse: la loro deviazione standard è $\sqrt{n}$ volte l'errore standard della media.
2. È corretta, ed è proprio il significato del livello di confidenza — a cambiare per caso è il campione che scegliamo, e con lui l'intervallo.
3. Anche questa è falsa: il livello di confidenza del 95% ci garantisce semplicemente che il 95% degli intervalli di confidenza contengono la media vera, ma, sul singolo intervallo — in questo caso (251 ± 2) mg — non ci offre alcuna garanzia. Inoltre, se ripetessimo il campionamento, le nuove medie campionarie si distribuirebbero attorno alla media vera μ, non attorno alla nostra stima $\hat{\mu} = 251$ mg, che è a sua volta spostata da μ di una quantità casuale che non conosciamo.
4. È corretta, perché l'errore standard della media è $\sigma/\sqrt{n}$ e diminuisce al crescere di n.

L'errore del primo punto — confondere la media con i singoli individui — è di gran lunga il più grossolano e il più comune.