Introduzione agli intervalli di confidenza

A cura di

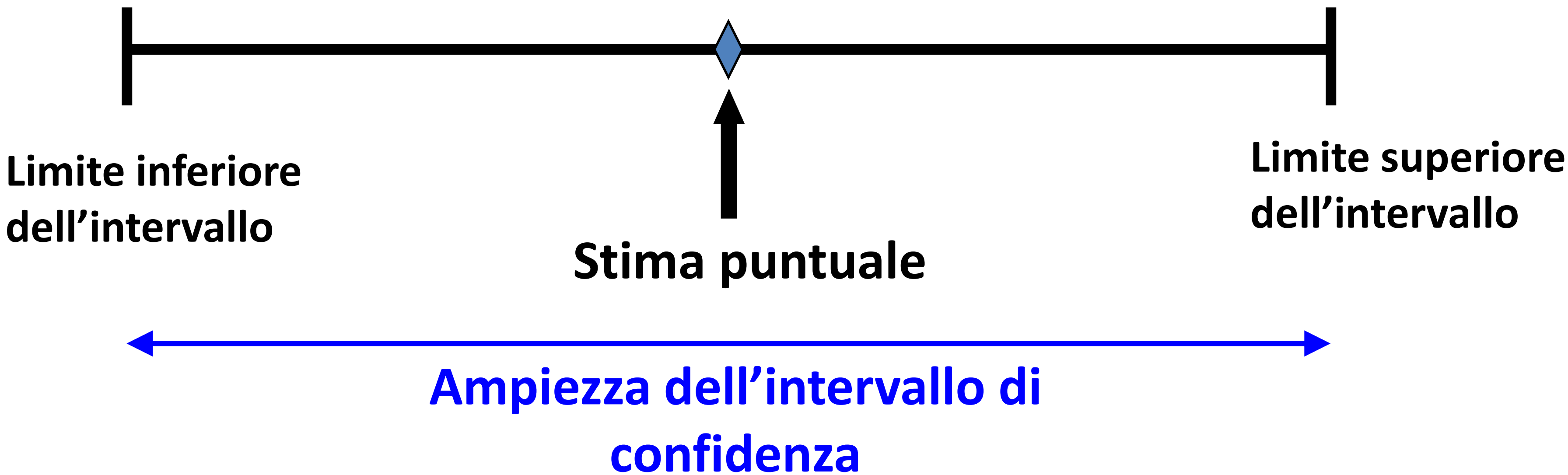
Francesco Fabi

ce3s

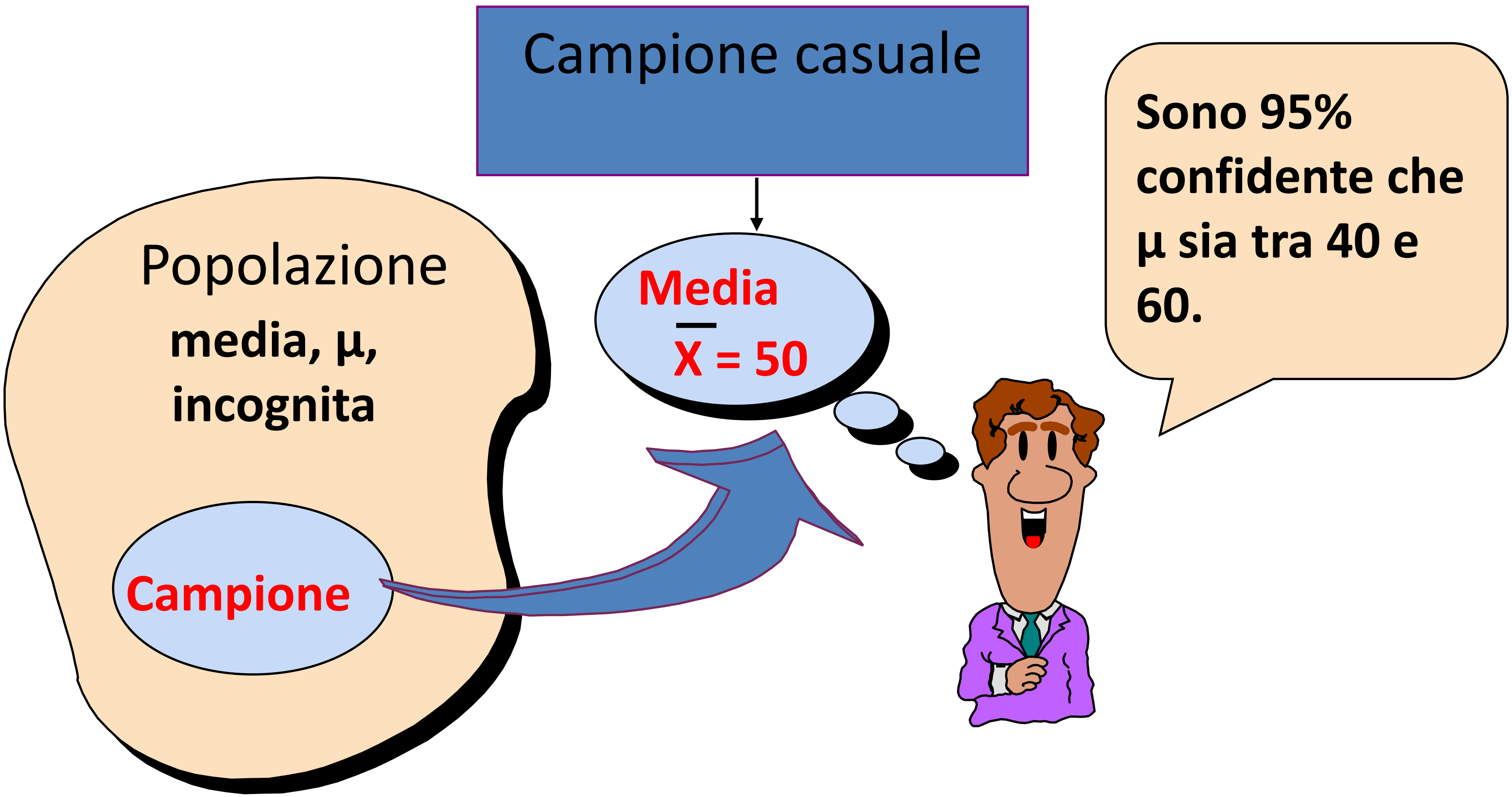
CENTRO STUDI STATISTICI E SOCIALI

Stime puntuali e di intervallo

- Una stima puntuale è un numero,
- Un intervallo di confidenza fornisce anche informazioni sulla variabilità



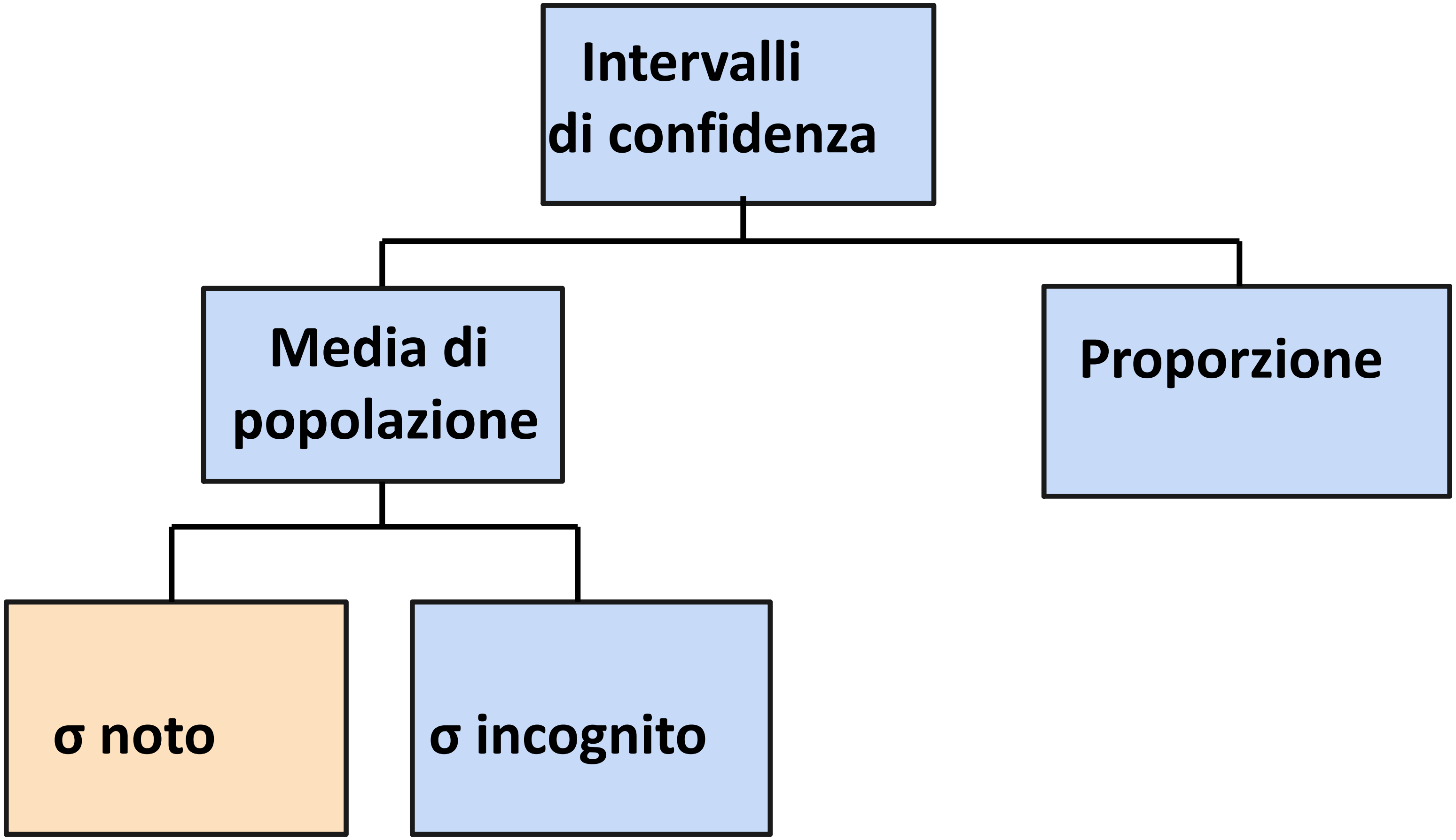
Processo di stima



Livello di confidenza, (1-α)

- Supponiamo di fissare il livello = 95%
- Si scrive $(1 - \alpha) = 0.95$
- Una interpretazione:
  - Se si effettuano molti esperimenti nelle stesse condizioni, il 95% di tutti gli intervalli di confidenza costruiti conterrà il vero parametro incognito
- Un singolo intervallo però può o meno contenere il parametro e non è possibile saperlo

Intervalli di Confidenza



Formula Generale

- La formula generale per il calcolo di ogni intervallo di confidenza è:

Stima puntuale $\pm$ (valore critico)(errore standard)

Intervalli di confidenza per μ (σ noto)

- Assunzioni
  - Lo scarto standard della popolazione σ è noto
  - La popolazione è distribuita normalmente
  - In caso contrario, il campione è grande
- Calcolo dell’intervallo:

$$\boxed{\bar{X} \pm Z \frac{\sigma}{\sqrt{n}}}$$

dove $\bar{X}$ è la stima puntuale della media

Z è il valore critico della distribuzione normale standardizzata con $\alpha/2$ in ogni coda

$\sigma/\sqrt{n}$ è l’errore standard

Livelli di confidenza usuali

- Livelli comunemente usati 90%, 95%, e 99%

Livello di confidenza	$1-\alpha$	*z*
80%	**0.80**	**1.28**
90%	**0.90**	**1.645**
95%	**0.95**	**1.96**
98%	**0.98**	**2.33**
99%	**0.99**	**2.58**
99.8%	**0.998**	**3.08**
99.9%	**0.999**	**3.27**

Calcolo dell'intervallo per la media

- Supponiamo di avere un campione di grande numerosità
- L'intervallo di confidenza al 95% si ricava considerando che, per la variabile normale standardizzata esattamente il 95% delle osservazioni è tra
- – 1,96 e + 1,96
- Al variare del campione allora osserveremo circa il 95% delle volte un valore di Z=$\dfrac{\bar{X}-\mu}{\sigma/\sqrt{n}}$ compreso tra – 1,96 e + 1,96.

Intervallo di confidenza al 95% per la media

- Passando alla misure non standardizzate, possiamo scrivere:

$$P(\mu - 1{,}96\sigma/\sqrt{n} \le \bar{X} \le \mu + 1{,}96\sigma/\sqrt{n}) = 0{,}95$$

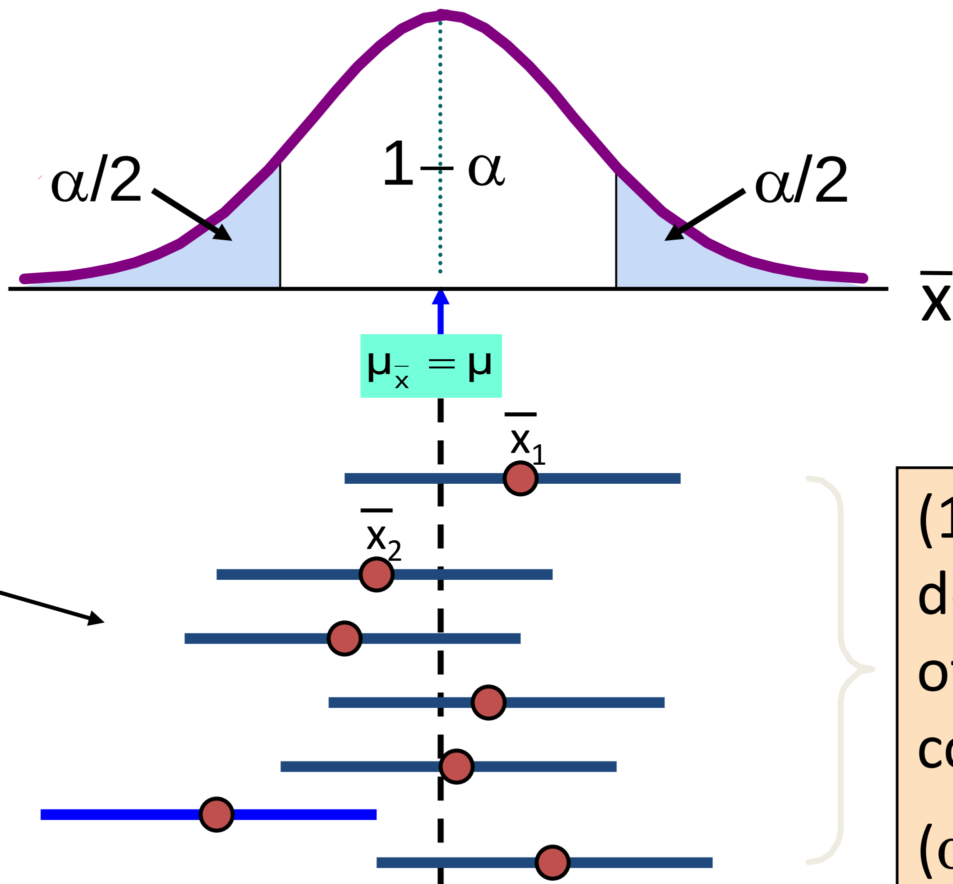
Esplicitiamo mettendo al centro la media incognita:

$$P(\bar{X} - 1{,}96\sigma/\sqrt{n} \le \mu \le \bar{X} + 1{,}96\sigma/\sqrt{n}) = 0{,}95$$

- Se non conosciamo σ, lo sostituiamo con s, che è la sua stima puntuale, purché la numerosità campionaria sia superiore a 30.
- L'intervallo è centrato su $\bar{X}$ e ha semiampiezza 1,96s/√(n).

Intervalli e livelli di confidenza

Distribuzione campionaria della media



$$\bar{X} - Z\frac{\sigma}{\sqrt{n}}$$

$$\bar{X} + Z\frac{\sigma}{\sqrt{n}}$$

Intervalli di confidenza

$(1-\alpha)$x100% degli intervalli ottenuti contiene μ;

(α)x100% non lo contiene

Intervallo di confidenza per μ (σ incognito)

- Se lo scarto standard della popolazione σ è incognito, possiamo sostituire lo scarto standard campionario, S
- Questo introduce altra incertezza, perché S varia da campione a campione
- Dobbiamo allora utilizzare la distribuzione t invece della normale, a meno che il campione non permetta di utilizzare il teorema centrale

Intervallo di confidenza per μ(σ incognito)

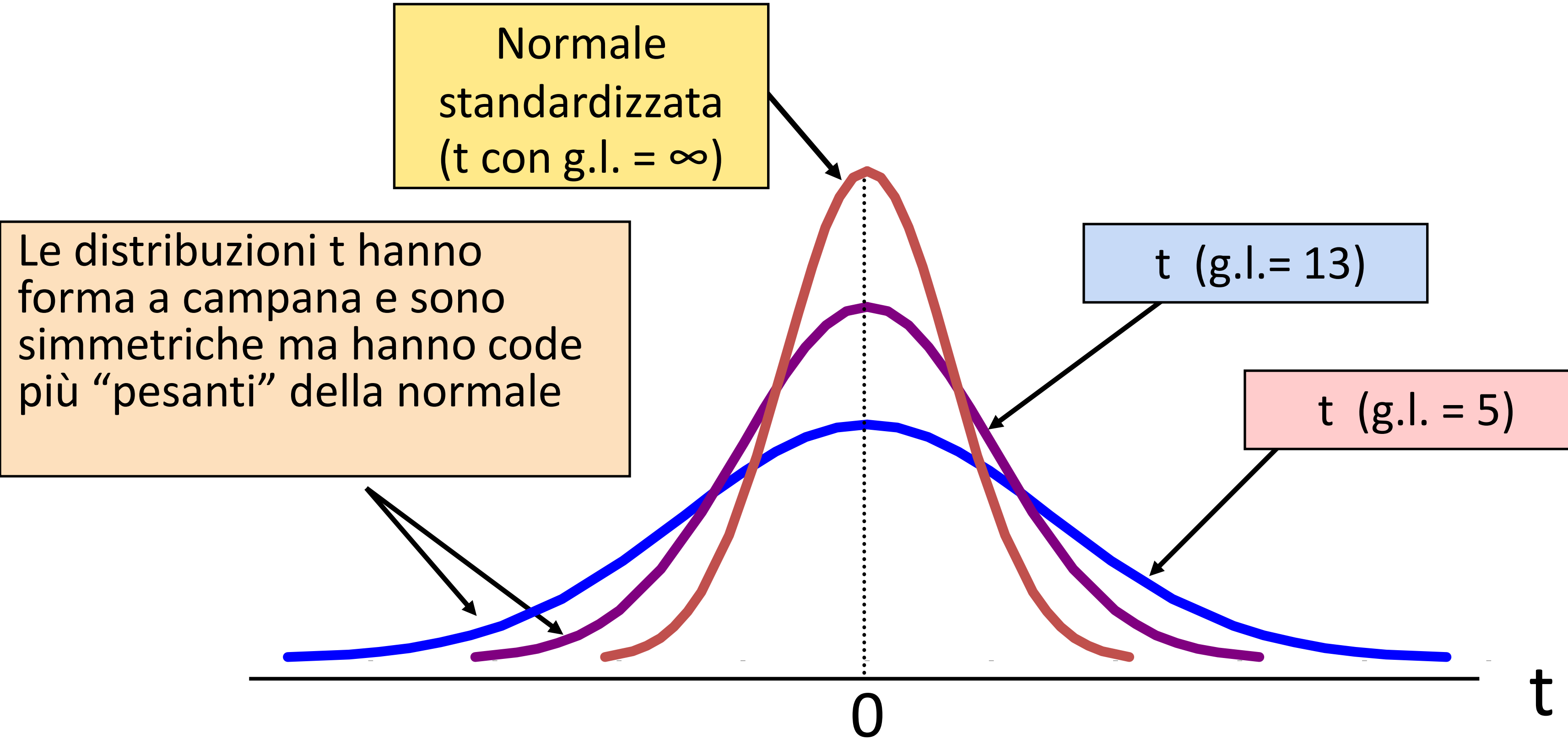
- Assunzioni
  - Lo scarto standard della popolazione è incognito
  - Se la popolazione non è normale e il campione non è grande.
  - Utilizziamo la distribuzione t di Student
- Intervallo di confidenza:

$$\boxed{\bar{X} \pm t_{n-1} \frac{S}{\sqrt{n}}}$$

(t è il valore critico della distribuzione t con $\alpha/2$ in ogni coda)

La distribuzione t di Student

Nota: $t \longrightarrow Z$ quando n cresce



Intervallo di confidenza per la proporzione π

- Una stima di intervallo per la proporzione (π) può essere calcolata aggiungendo un margine di incertezza alla frequenza campionaria (p) che è la stima puntuale.

Intervallo di confidenza per la proporzione π

- Se la numerosità campionaria è elevata la distribuzione campionaria della frequenza è approssimativamente normale con scarto standard

$$\boxed{\sigma_p = \sqrt{\frac{\pi(1-\pi)}{n}}}$$

- Che possiamo stimare a partire dai dati con:

$$\boxed{\sqrt{\frac{p(1-p)}{n}}}$$

Intervallo di Confidenza

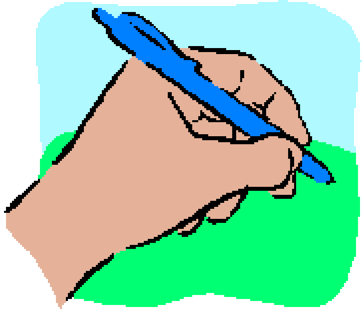
- L'intervallo di confidenza si calcola allora:

$$\boxed{p \pm Z\sqrt{\frac{p(1-p)}{n}}}$$

- dove
  - Z è il valore critico della distribuzione normale standardizzata al livello fissato
  - p è la frequenza campionaria
  - n è la numerosità campionaria

Esempio

- Un campione casuale di 100 individui ha mostrato che 25 sono mancini,
- Determinare un intervallo di confidenza al 95% per la vera proporzione di mancini nella popolazione



Esempio

$$p \pm Z\sqrt{p(1-p)/n}$$

$$= 25/100 \pm 1.96\sqrt{0.25(0.75)/100}$$

$$= 0.25 \pm 1.96\,(0.0433)$$

$$\boxed{0.1651 \le \pi \le 0.3349}$$

Interpretazione

- Siamo al 95% confidenti che la vera percentuale di mancini nella popolazione sia compresa tra

16.51% e 33.49%.

- Sebbene l'intervallo 0.1651 - 0.3349 possa contenere o meno la vera proporzione, il 95% degli intervalli costruiti in questo modo a partire da 100 osservazioni conterrà la v proporzione.

Come si determina la numerosità campionaria

Determinazione di n

Media

Errore di campionamento (margine di errore)

$$\bar{X} \pm Z\frac{\sigma}{\sqrt{n}}$$

$$e = Z\frac{\sigma}{\sqrt{n}}$$

Come si determina la numerosità campionaria

Determinazione di n

Media

$$e = Z\frac{\sigma}{\sqrt{n}}$$

Risolviamo rispetto a n

$$\boxed{n = \frac{Z^2\,\sigma^2}{e^2}}$$