Statistica

Capitolo 8

Problemi di Stima
su una Singola Popolazione

Obiettivi del Capitolo

Dopo aver completato il capitolo, sarete in grado di:

- Distinguere una stima puntuale da un intervallo di confidenza
- Costruire ed interpretare un intervallo di confidenza per la media di una singola popolazione usando sia Z che t
- Formare ed interpretare un intervallo di confidenza per la proporzione di una singola popolazione

Intervalli di Confidenza

Contenuto di questo capitolo

- Intervalli di confidenza per la Media della Popolazione, μ
  - quando la varianza della popolazione σ^2 è nota
  - quando la varianza della popolazione σ^2 non è nota
- Intervalli di confidenza per la Proporzione della Popolazione, p (grandi campioni)

Definizioni

- Uno stimatore di una parametro della popolazione è
  - una variabile aleatoria che dipende dalla informazione contenuta nel campione . . .
  - il cui valore fornisce un’approssimazione del valore sconosciuto del parametro

- Uno specifico valore della variabile aleatoria viene chiamaro stima puntuale

Stima Puntuale e per Intervallo

- Una stima puntuale è un unico valore,
- Un intervallo di confidenza fornisce ulteriori informazioni circa la variabilità

Stimatori Puntuali

Possiamo stimare il parametro della popolazione …		con una Statistica Campionaria (Stimatore Puntuale)
Media	μ	$\bar{X}$
Proporzione	p	$\hat{P}$

Non distorsione

- Uno stimatore puntale $\hat{\theta}$ viene definito stimatore non distorto per il parametro $\boldsymbol{\theta}$ se il suo valore atteso, o media, è $\boldsymbol{\theta}$,

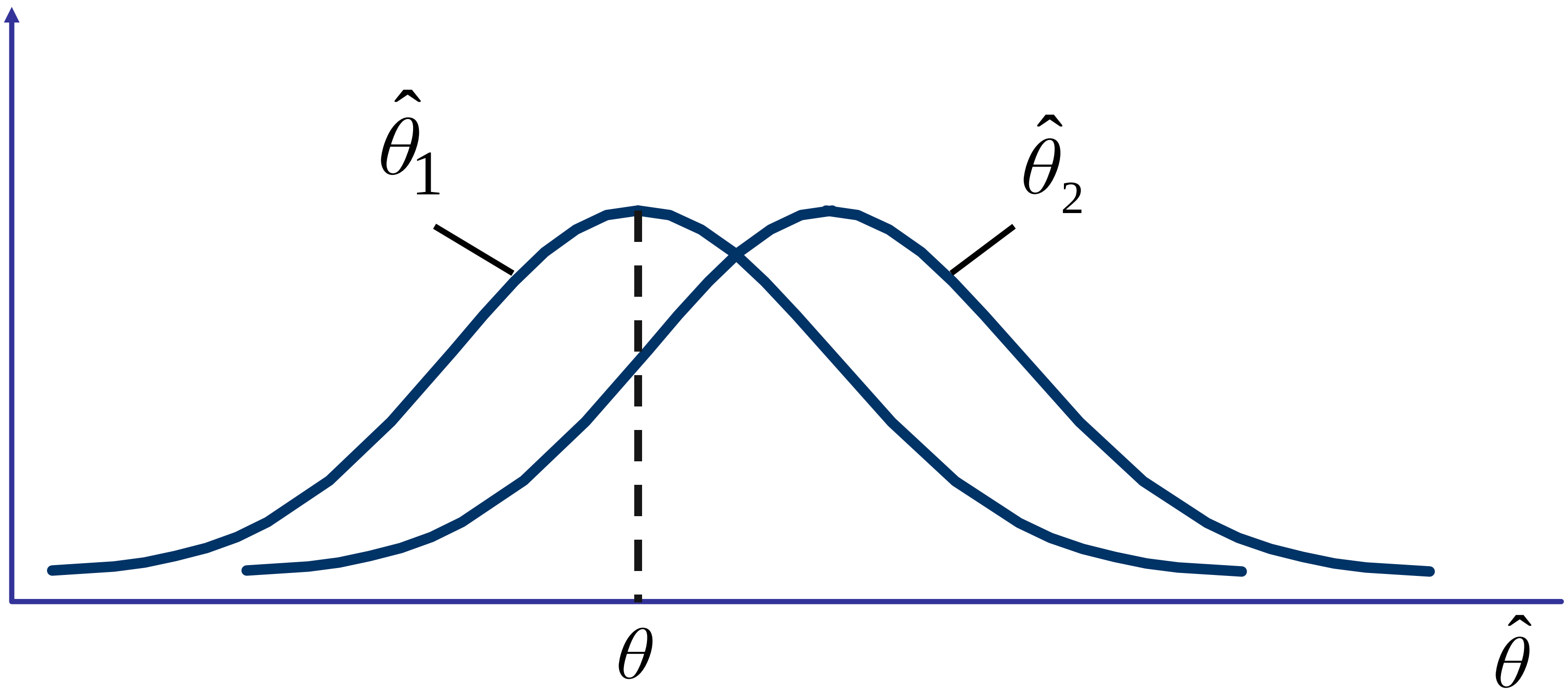
$$\boxed{\mathrm{E}(\hat{\theta}) = \theta}$$

- Esempi:
  - La media campionaria è uno stimatore non distorto per μ
  - La varianza campionaria è uno stimatore non distorto per σ^2
  - La proporzione campionaria è uno stimatore non distorto per *p*

Non distorsione

(continuazione)

- $\hat{\theta}_1$ è uno stimatore non distorto, $\hat{\theta}_2$ è distorto:



Distorsione

- Sia $\hat{\theta}$ uno stimatore per $\boldsymbol{\theta}$

- La distorsione di $\hat{\theta}$ è definita come la differenza tra la sua media e $\boldsymbol{\theta}$

$$\boxed{\mathrm{D}(\hat{\theta}) = \mathrm{E}(\hat{\theta}) - \theta}$$

- La distorsione di uno stimatore non distorto è 0

Non distorsione asintotica

- Sia $\hat{\theta}$ uno stimatore per $\boldsymbol{\theta}$

- $\hat{\theta}$ è uno stimatore asintoticamente non distorto per $\boldsymbol{\theta}$ se la differenza tra il valore atteso di $\hat{\theta}$ e $\boldsymbol{\theta}$ diminuisce al crescere dell’ampiezza del campione

- La non distorsione asintotica è una proprietà desiderata quando non si possono ottenere stimatori non distorti

Efficienza

Supponiamo esistano diversi stimatori non distorti per θ

- Lo stimatore più efficiente per θ è lo stimatore non distorto con la varianza più piccola
- Siano $\hat{\theta}_1$ e $\hat{\theta}_2$ due stimatori non distorti per θ, basati sullo stesso numero di osservazioni campionarie. Allora,
  - $\hat{\theta}_1$ è più efficiente di $\hat{\theta}_2$ se $\boxed{\mathrm{Var}(\hat{\theta}_1) < \mathrm{Var}(\hat{\theta}_2)}$
  - L'efficienza relativa di $\hat{\theta}_1$ rispetto a $\hat{\theta}_2$ è il rapporto tra le loro varianze:

$$\boxed{\text{Efficienza Relativa} = \frac{\mathrm{Var}(\hat{\theta}_2)}{\mathrm{Var}(\hat{\theta}_1)}}$$

Intervalli di Confidenza

- Quanta incertezza è associata ad una stima puntuale di un parametro della popolazione?
- Rispetto ad una stima puntuale, una stima per intervallo fornisce maggiori informazioni sulla caratteristica della popolazione oggetto di studio
- Tali stime per intervallo sono chiamate intervalli di confidenza

Intervallo di Confidenza

- Un intervallo fornisce una serie di valori:
  - Prende in considerazione la variazione nel valore della statistica da campione a campione
  - Basato sulle osservazioni del campione
  - Fornisce informazioni sulla vicinanza al parametro sconosciuto della popolazione
  - Espresso in termini di livello di confidenza
    - Non è possibile essere confidenti al 100%

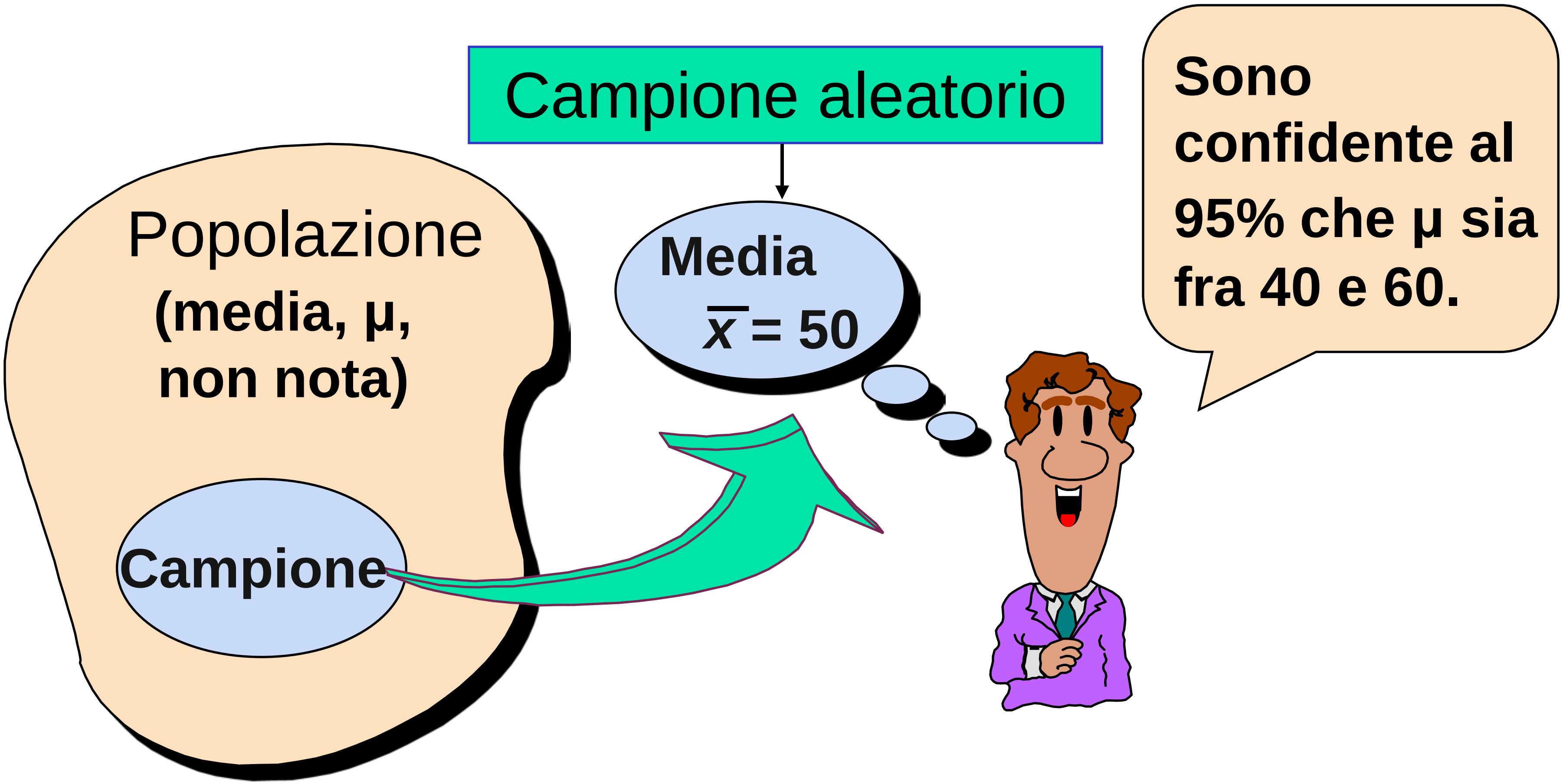
Intervallo di Confidenza e Livello di Confidenza

- Se $P(A < \theta < B) = 1 - \alpha$ allora l'intervallo da A a B viene definito intervallo di confidenza a livello $100(1 - \alpha)\%$ per θ.

- La quantità $(1 - \alpha)$ viene definita livello di confidenza dell'intervallo (α compreso tra 0 e 1)

  - Se si estraggono ripetutamente più campioni dalla popolazione, il vero valore del parametro θ sarà contenuto nel $100(1 - \alpha)\%$ degli intervalli determinati in questo modo.
  - L'intervallo di confidenza calcolato in questo modo viene indicato come $A < \theta < B$ con livello di confidenza $100(1 - \alpha)\%$

Processo di Stima



Livello di Confidenza, (1-α)

(continuazione)

- Assumiamo il livello di confidenza = 95%, anche scritto come (1 - α) = 0.95
- Interpretazione basata sulla frequenza relativa:
  - Se si estraggono ripetutamente più campioni, il 95% degli intervalli di confidenza che possono essere costruiti conterrà il vero valore sconosciuto del parametro
- Uno specifico intervallo conterrà o non conterrà il vero valore del parametro
  - La probabilità non è coinvolta in uno specifico intervallo

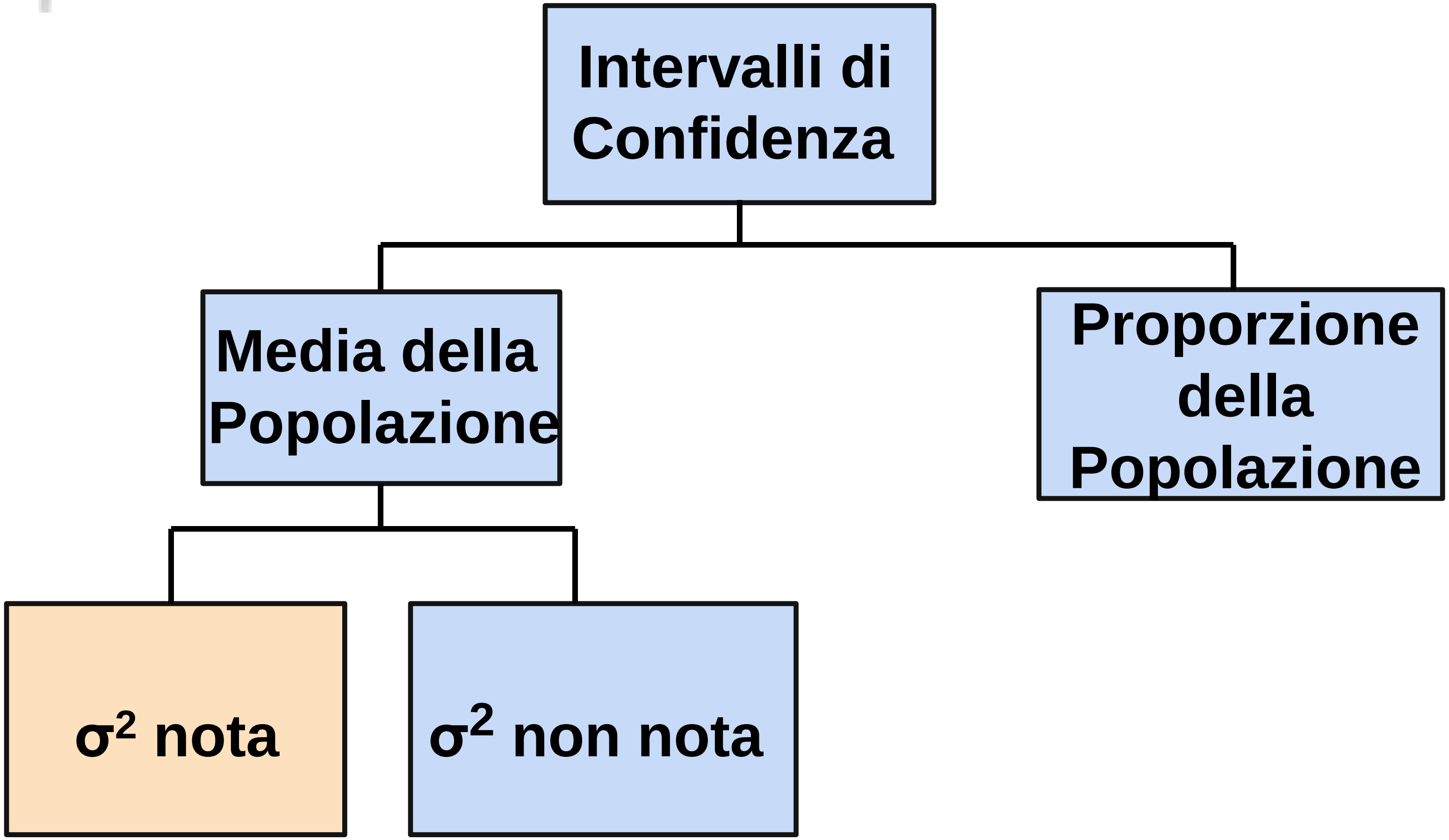
Formula Generale

- La formula generale per tutti gli intervalli di confidenza è:

Stima Puntuale ± (Fattore di Affidabilità)(Errore Standard dello Stimatore)

- Il valore del fattore di affidabilità dipende dal livello di confidenza desiderato

Intervalli di Confidenza



Intervallo di Confidenza per μ (σ^2 nota)

- Assunzioni
  - Varianza della popolazione σ^2 è nota
  - Popolazione ha distribuzione normale
  - Se la popolazione non ha distribuzione normale, si devono utilizzare grandi campioni

- Intervallo di Confidenza:

$$\boxed{\bar{x} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} < \mu < \bar{x} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}}$$

(dove $z_{\alpha/2}$ è il valore critico della distribuzione normale standard che lascia nella coda destra una probabilità pari ad $\alpha/2$)

Margine di Errore

- L'intervallo di confidenza

$$\bar{x} - z_{\alpha/2} \frac{\sigma}{\sqrt{n}} < \mu < \bar{x} + z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$$

può anche essere scritto come $\bar{x} \pm \text{ME}$

dove ME è noto come margine di errore

$$\text{ME} = z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$$

- L'ampiezza dell'intervallo, *w*, è uguale al doppio del margine di errore

Ridurre il Margine di Errore

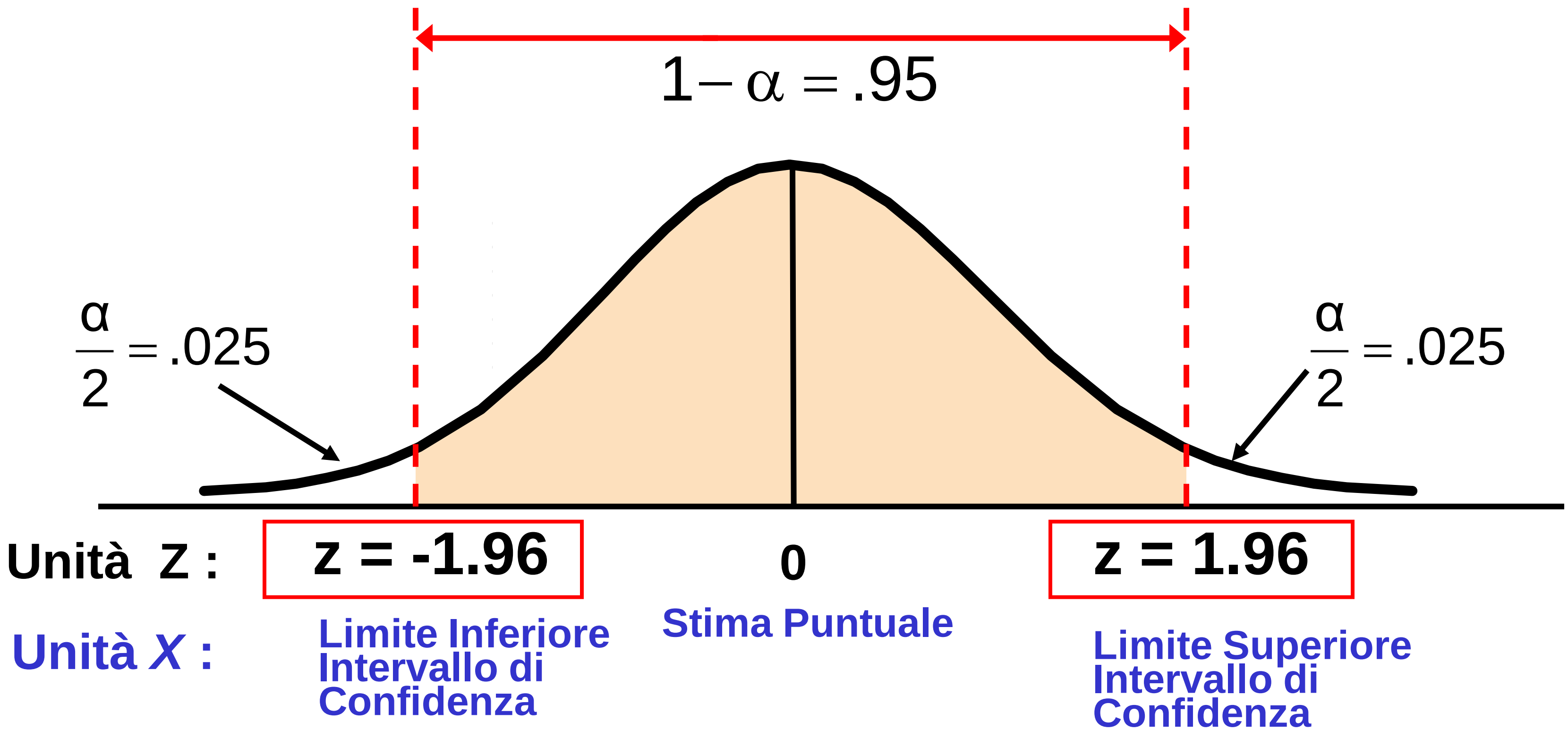
$$\text{ME} = z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$$

Il margine di errore può essere ridotto se

- La deviazione standard della popolazione può essere ridotta ($\sigma\downarrow$)
- La dimensione del campione aumenta ($n\uparrow$)
- Il livello di confidenza diminuisce, $(1 - \alpha)\downarrow$

Trovare il Fattore di Affidabilità, $z_{\alpha/2}$

- Consideriamo un intervallo di confidenza al 95%:



- Nella tavola della distribuzione normale standard si trova $z_{.025} = \pm 1.96$

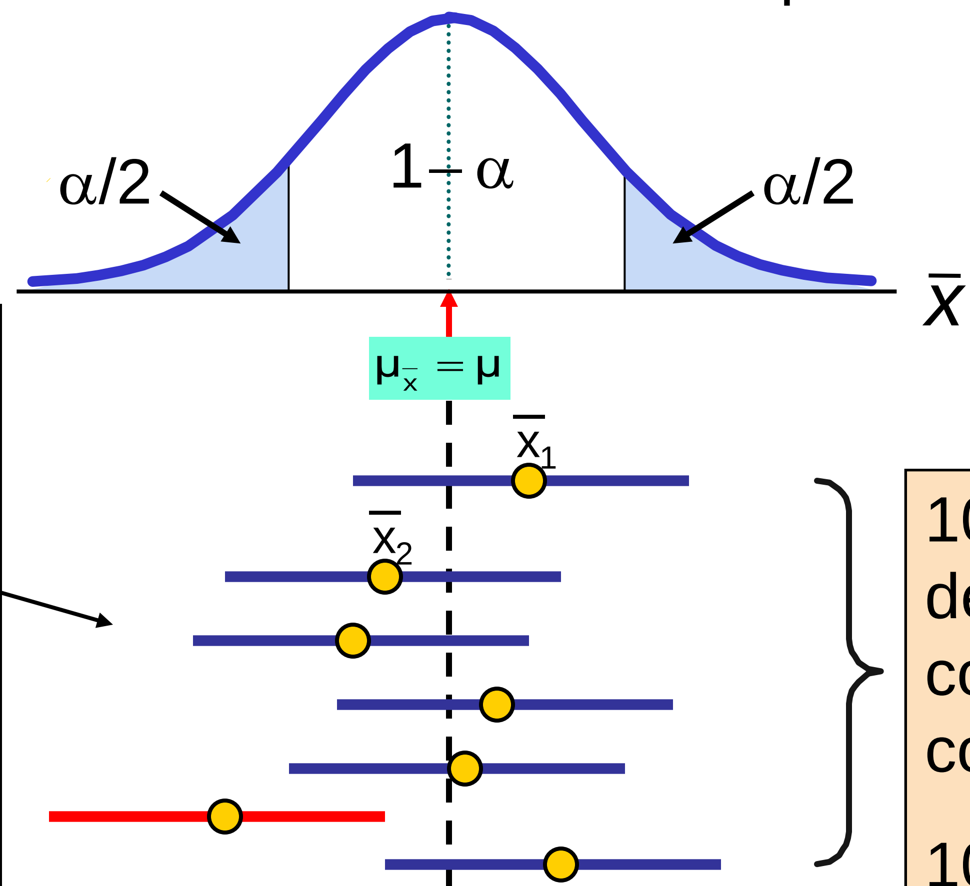
Livelli Comuni di Confidenza

- I livelli di confidenza comunemente usati sono 90%, 95% e 99%

Livello di Confidenza	*Coefficiente di Confidenza,* $1-\alpha$	*Valore* $Z_{\alpha/2}$
80%	**.80**	**1.28**
90%	**.90**	**1.645**
95%	**.95**	**1.96**
98%	**.98**	**2.33**
99%	**.99**	**2.575**
99.8%	**.998**	**3.08**
99.9%	**.999**	**3.27**

Intervalli e Livello di Confidenza

Distribuzione della media campionaria



Gli intervalli vanno da

$$\bar{x} - z\frac{\sigma}{\sqrt{n}}$$

a

$$\bar{x} + z\frac{\sigma}{\sqrt{n}}$$

Intervalli di Confidenza

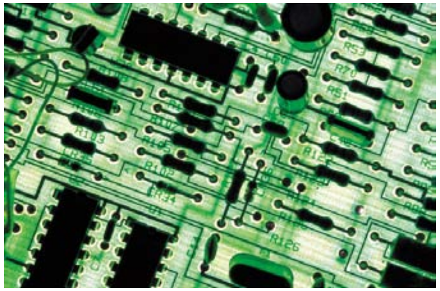
100(1-α)% degli intervalli costruiti contengono μ;

100(α)% non lo contengono.

Esempio

- Un campione di 11 circuiti, estratti da una popolazione molto numerosa, con distribuzione normale, ha una resistenza media di 2.2 ohm. Sappiamo da test precedenti che la deviazione standard della popolazione è 0.35 ohm.

- Determinare un intervallo di confidenza al 95% per la vera resistenza media della popolazione.



Esempio

(continuazione)

- Un campione di 11 circuiti, estratti da una popolazione molto numerosa, con distribuzione normale, ha una resistenza media di 2.2 ohm. Sappiamo da test precedenti che la deviazione standard della popolazione è 0.35 ohm.

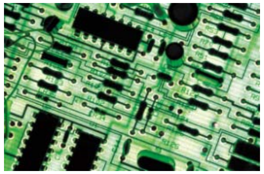
- Soluzione:

$$\bar{x} \pm z \frac{\sigma}{\sqrt{n}}$$

$$= 2.20 \pm 1.96\,(.35/\sqrt{11})$$

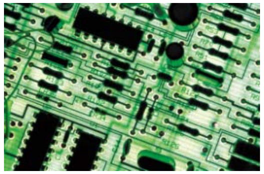
$$= 2.20 \pm .2068$$

$$\boxed{1.9932 < \mu < 2.4068}$$

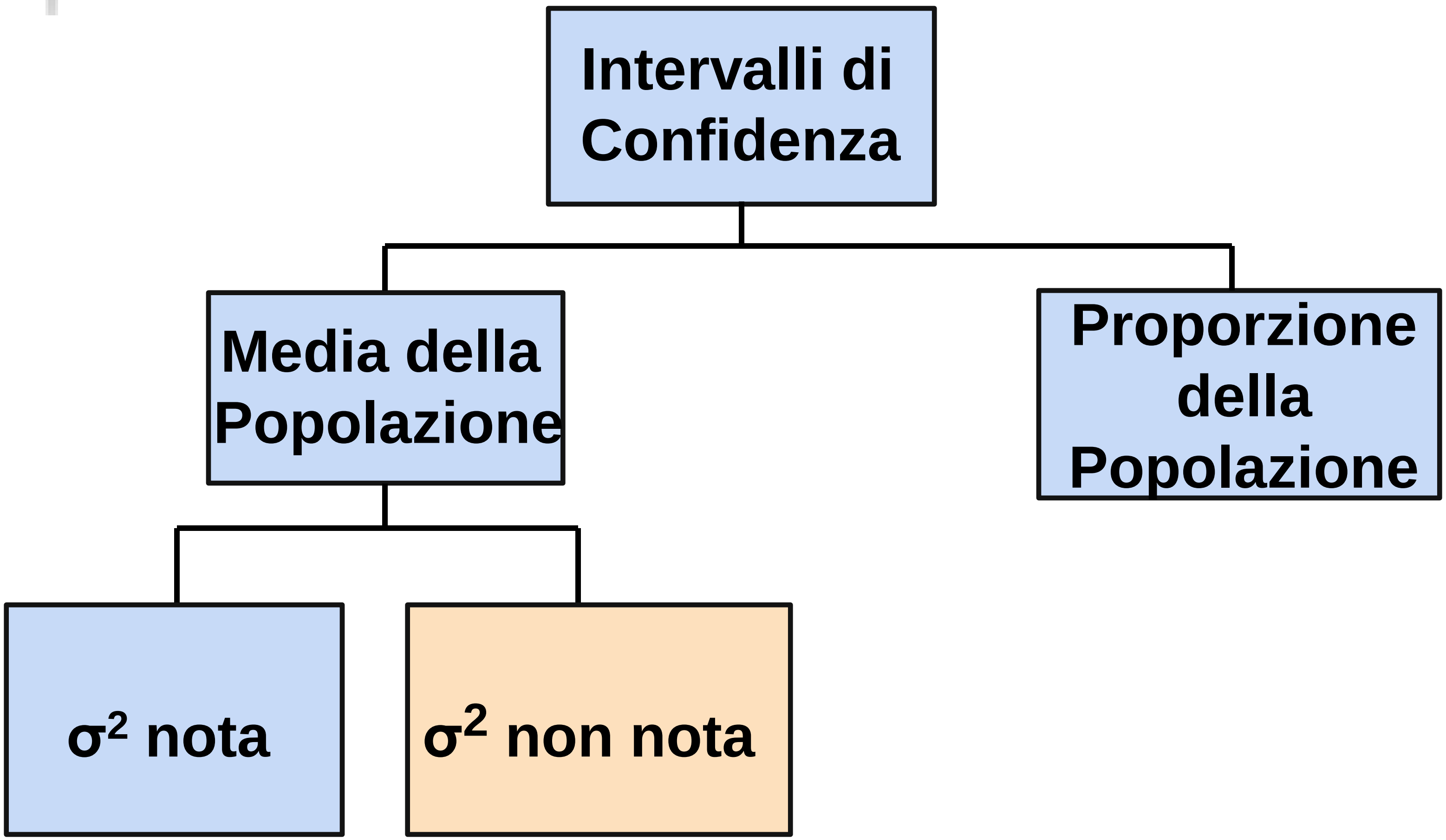


Interpretazione

- Siamo confidenti a livello 95% che la vera resistenza media sia compresa tra 1.9932 e 2.4068 ohm
- Sebbene la vera media possa essere o meno inclusa in questo intervallo, il 95% degli intervalli costruiti con questo metodo conterrà il vero valore medio



Intervalli di Confidenza



Distribuzione t di Student

- Consideriamo un campione aleatorio di *n* osservazioni
  - con media $\bar{X}$ e deviazione standard S
  - da una popolazione con distribuzione normale di media μ

- Allora la variabile

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

ha una distribuzione t di Student con (*n* - 1) gradi di libertà

Intervallo di Confidenza per μ (σ^2 non nota)

- Se la deviazione standard della popolazione, σ, non è nota, possiamo sostituirla con la deviazione standard campionaria, S
- Questo introduce ulteriore incertezza, siccome S varia da campione a campione
- Quindi usiamo la distribuzione t invece della distribuzione normale

Intervallo di Confidenza per μ (σ^2 non nota)

(continuazione)

- Assunzioni
  - Deviazione stardard della popolazione non è nota
  - Popolazione ha distribuzione normale
  - Se la popolazione non ha distribuzione normale, si devono utilizzare grandi campioni
- Uso della Distribuzione t di Student
- Intervallo di Confidenza:

$$\boxed{\bar{x} - t_{n\text{-}1,\alpha/2}\frac{s}{\sqrt{n}} < \mu < \bar{x} + t_{n\text{-}1,\alpha/2}\frac{s}{\sqrt{n}}}$$

dove $t_{n\text{-}1,\alpha/2}$ è il valore critico della distribuzione t con n -1 g.d.l. che lascia nella coda destra una probabilità pari ad $\alpha/2$:

$$\boxed{P(t_{n-1} > t_{n-1,\alpha/2}) = \alpha/2}$$

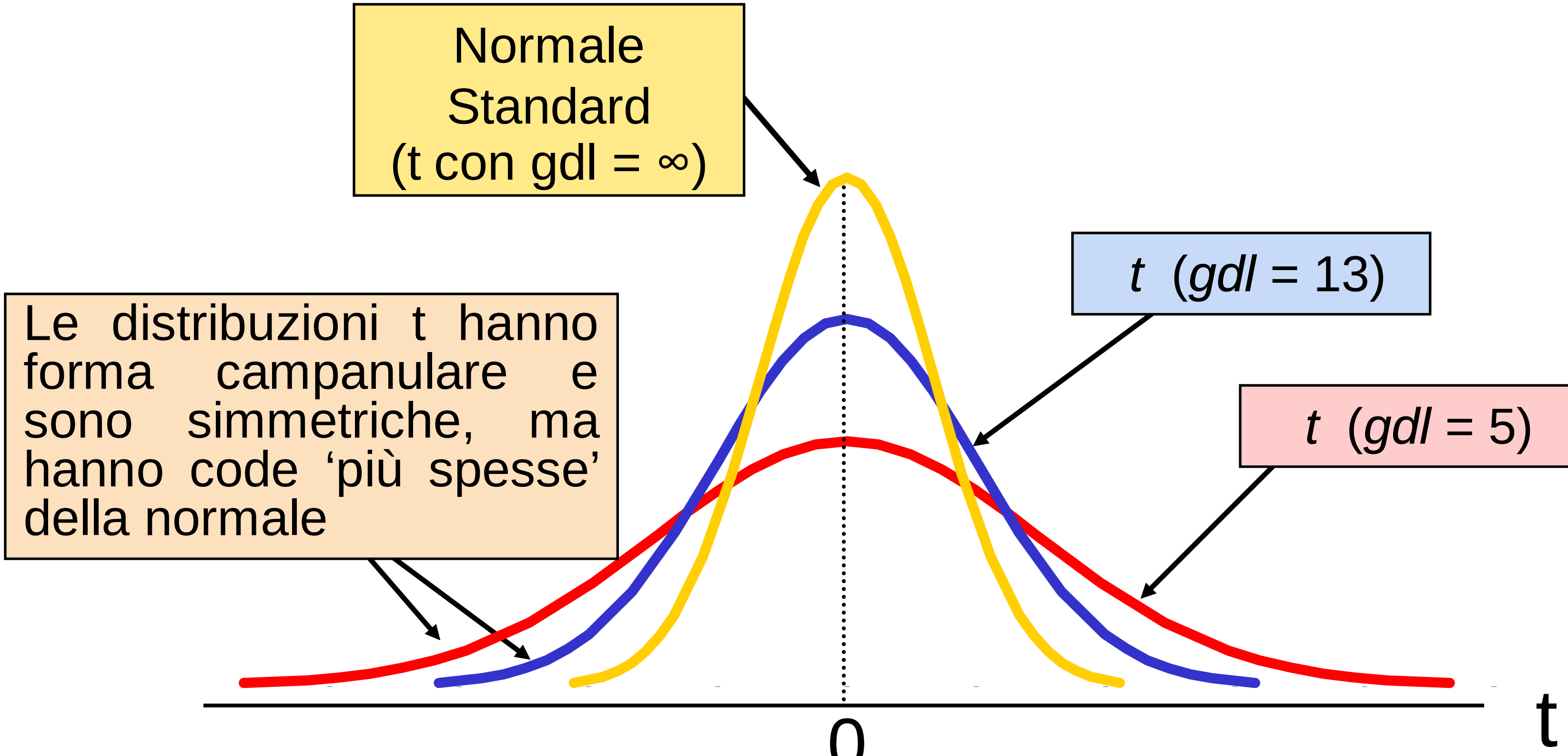
Distribuzione t di Student

- La t è una famiglia di distribuzioni
- Il valore di t dipende dai gradi di libertà (g.d.l.)
  - Numero di osservazioni che sono libere di variare dopo che la media è stata calcolata

$$\boxed{\text{g.d.l.} = n - 1}$$

Distribuzione t di Student

Notare: $t \longrightarrow Z$ quando n aumenta

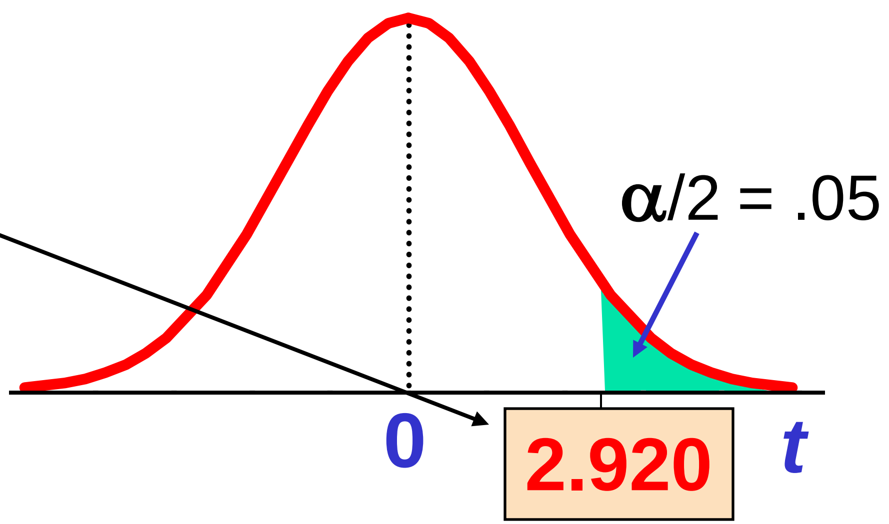


Tavole della distribuzione t di Student

	Area Coda Destra		
gdl	.10	**.05**	.025
1	3.078	6.314	12.706
2	1.886	**2.920**	4.303
3	1.638	2.353	3.182

Il corpo della tavola contiene i valori t, non le probabilità

Sia: $n = 3$
gdl = $n - 1 = 2$
$\alpha = .10$
$\alpha/2 = .05$



Valori della distribuzione t

Confronto con i valori di Z

Livello di Confidenza	t (10 gdl)	t (20 gdl)	t (30 gdl)	Z
.80	1.372	1.325	1.310	1.282
.90	1.812	1.725	1.697	1.645
.95	2.228	2.086	2.042	1.960
.99	3.169	2.845	2.750	2.576

Notare: $t \rightarrow Z$ quando *n* cresce

Esempio

Un campione casuale di n = 25 ha $\bar{x}$ = 50 e s = 8. Determinare un intervallo di confidenza al 95% per μ

- g.d.l. = n – 1 = 24, allora $t_{n-1,\alpha/2} = t_{24,.025} = 2.064$

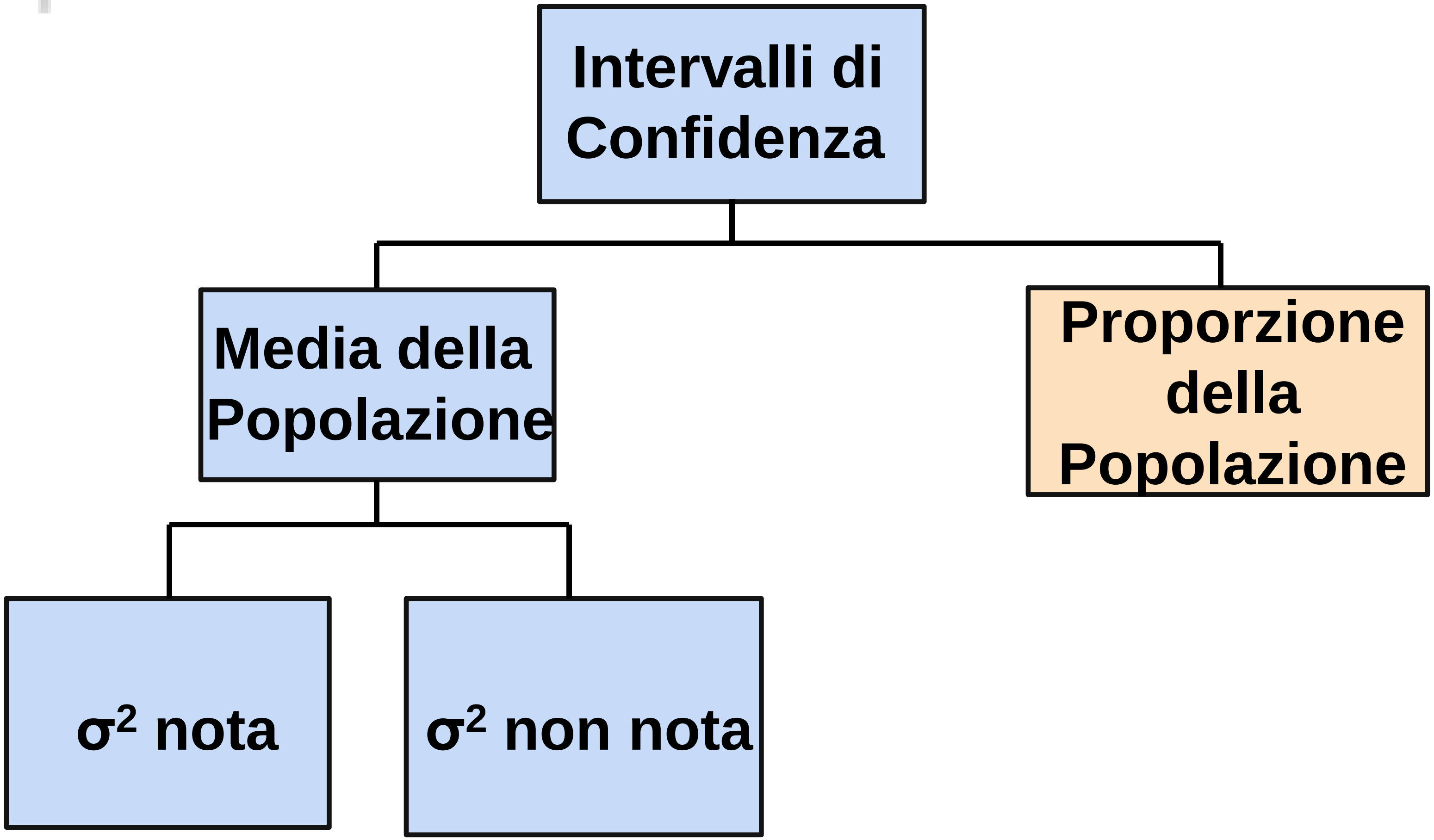
L'intervallo di confidenza è

$$\bar{x} - t_{n\text{-}1,\alpha/2} \frac{s}{\sqrt{n}} < \mu < \bar{x} + t_{n\text{-}1,\alpha/2} \frac{s}{\sqrt{n}}$$

$$50 - (2.064)\frac{8}{\sqrt{25}} < \mu < 50 + (2.064)\frac{8}{\sqrt{25}}$$

$$\boxed{46.698 < \mu < 53.302}$$

Intervalli di Confidenza



Intervallo di Confidenza per la Proporzione della Popolazione, p

- Una stima per intervallo per la proporzione della popolazione (p) può essere calcolata aggiungendo un fattore che tiene conto dell'incertezza della proporzione campionaria $(\hat{P})$

Intervallo di Confidenza per Proporzione della Popolazione, p

(continuazione)

- Ricordiamo che, se il campione è grande, la distribuzione della proporzione campionaria è approssimativamente normale, con deviazione standard

$$\sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}}$$

- La deviazione standard verrà stimata con i dati campionari:

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

Limiti dell'Intervallo di Confidenza

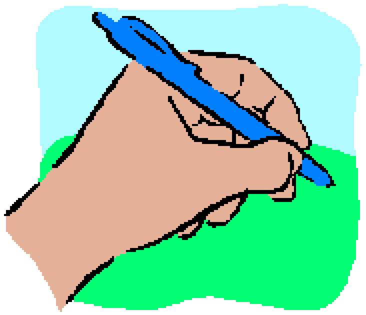
- I limiti superiore ed inferiore dell'intervallo di confidenza per la proporzione della popolazione sono calcolati usando la seguente formula:

$$\hat{p} - z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \; < \; p \; < \; \hat{p} + z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

- dove
  - $z_{\alpha/2}$ è il valore della distribuzione normale corrispondente al livello di confidenza desiderato
  - $\hat{p}$ è la proporzione campionaria
  - n è la dimensione del campione

Esempio

- In un campione casuale di 100 individui verifichiamo che 25 sono mancini.
- Determinare un intervallo di confidenza al 95% per la vera proporzione di mancini



Esempio

(continuazione)

- In un campione casuale di 100 individui verifichiamo che 25 sono mancini. Determinare un intervallo di confidenza al 95% per la vera proporzione di mancini.

$$\hat{p} - z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \;<\; p \;<\; \hat{p} + z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

$$\frac{25}{100} - 1.96\sqrt{\frac{.25(.75)}{100}} \;<\; p \;<\; \frac{25}{100} + 1.96\sqrt{\frac{.25(.75)}{100}}$$

$$\boxed{0.1651 \;<\; p \;<\; 0.3349}$$

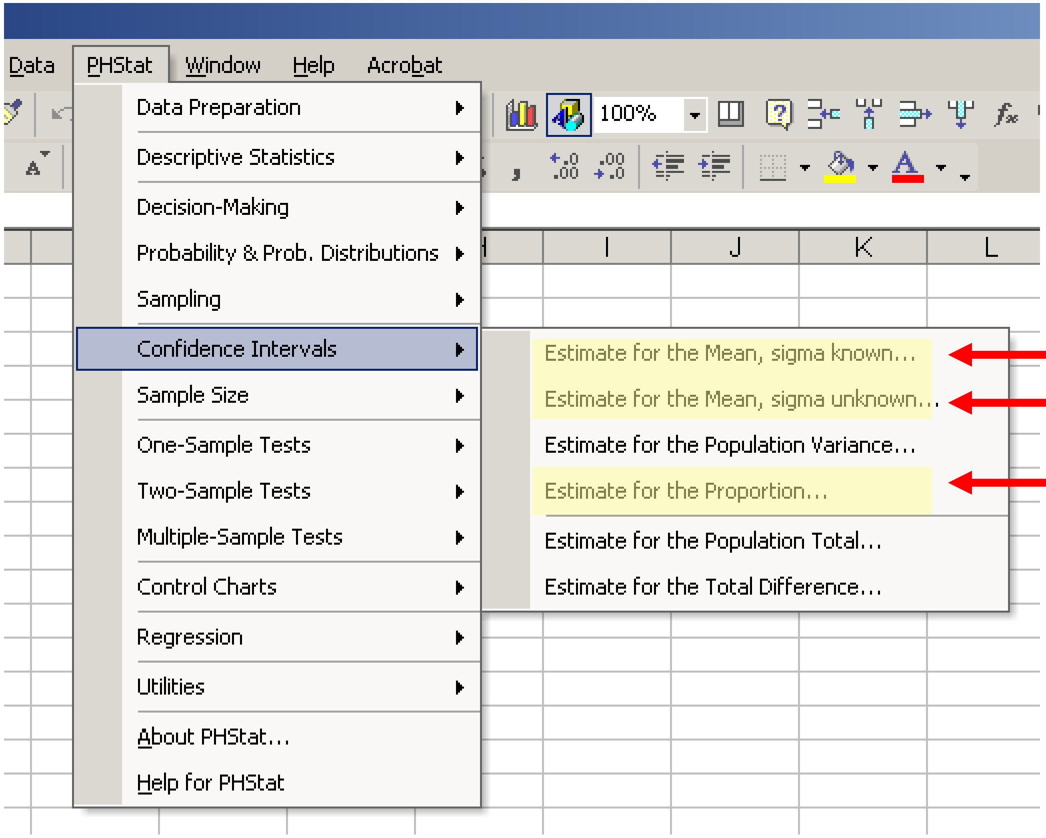
Interpretazione

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

  16.51% e 33.49%

- Sebbene l'intervallo da 0.1651 a 0.3349 possa contenere o non contenere la vera proporzione, 95% degli intervalli costruiti con questo metodo, con campioni di dimensione 100, conterranno il vero valore della proporzione di mancini.

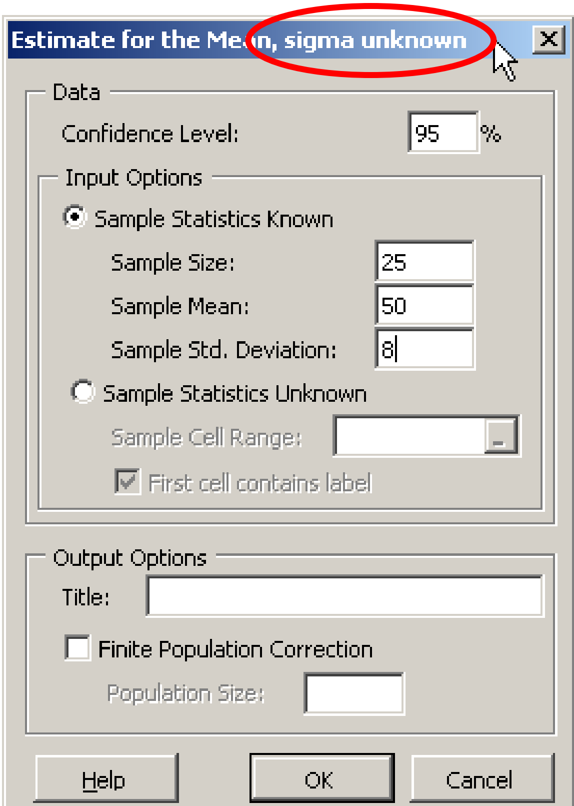
Intervalli di Confidenza con PHStat



opzioni

Uso di PHStat
(per μ, σ non noti)

Un campione casuale di $n = 25$ ha $\bar{x} = 50$ e $s = 8$.
Formare un intervallo di confidenza al 95% per μ



	A	B
1	**Confidence Interval Estimate for the Mean**	
2		
3	**Data**	
4	**Sample Standard Deviation**	**8**
5	**Sample Mean**	**50**
6	**Sample Size**	**25**
7	**Confidence Level**	**95%**
8		
9	Intermediate Calculations	
10	Standard Error of the Mean	1.6
11	Degrees of Freedom	24
12	*t* Value	2.063898137
13	Interval Half Width	3.302237019
14		
15	**Confidence Interval**	
16	**Interval Lower Limit**	**46.70**
17	**Interval Upper Limit**	**53.30**

Riepilogo del Capitolo

- Introdotto il concetto di intervallo di confidenza
- Discusso le stime puntuali
- Sviluppato il concetto di intervallo di confidenza
- Determinato l'intervallo di confidenza per la media (σ^2 nota)
- Introdotta la distribuzione t di Student
- Determinato l'intervallo di confidenza per la media (σ^2 non nota)
- Determinato l'intervallo di confidenza per la proporzione