Capitolo 8 Stima

Contenuto del capitolo

Proprietà degli stimatori

Correttezza: E(Stimatore) = parametro da stimare

Efficienza
Consistenza

Intervalli di confidenza

Per la media - per una proporzione

Come si determina l'ampiezza campionaria

Proprietà - correttezza

Esistono stimatori distorti?

La **varianza campionaria** se è calcolata come $\frac{\sum_i (X_i - \bar{X})^2}{n}$

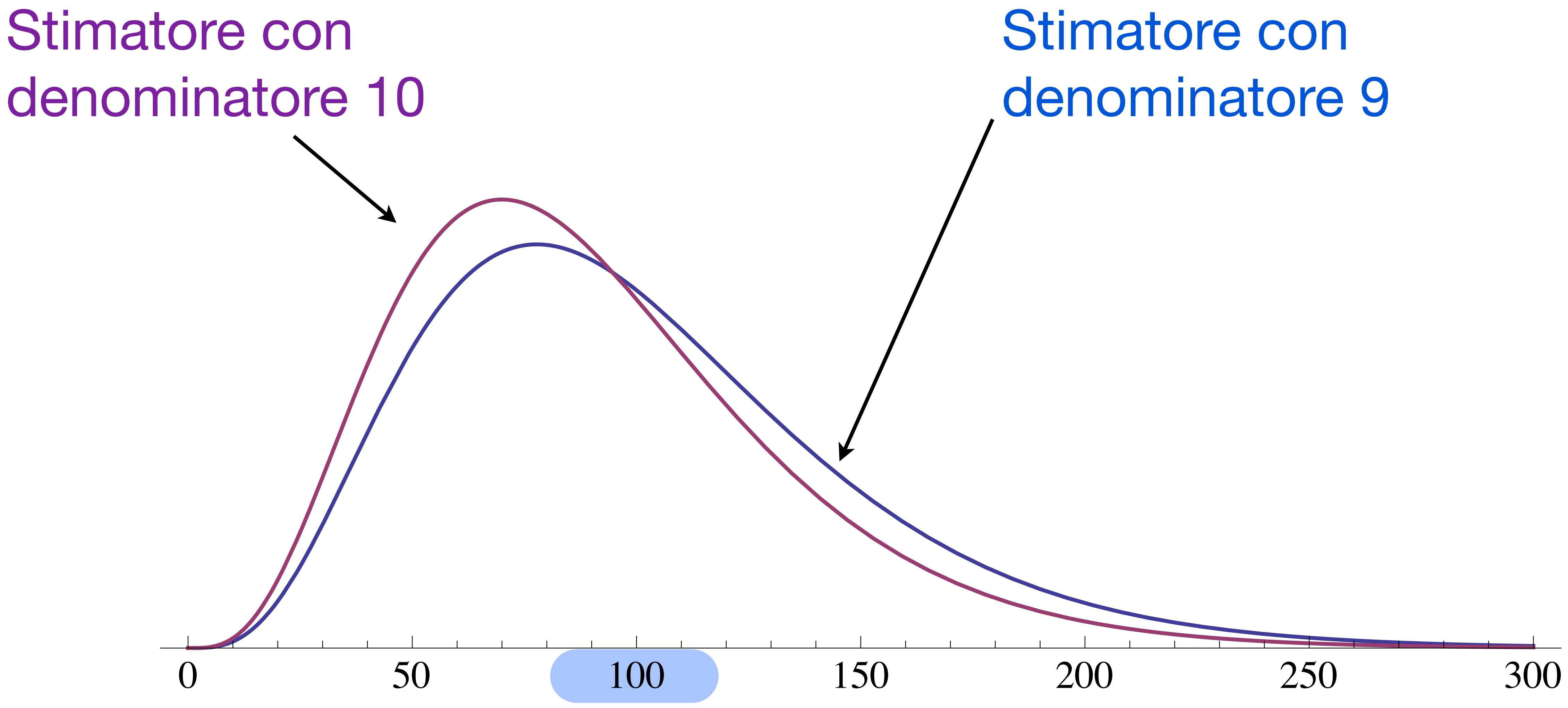
è uno stimatore **distorto** della varianza della popolazione

Invece se è calcolata come $S^2 = \frac{\sum_i (X_i - \bar{X})^2}{n-1}$

è uno stimatore **corretto** della varianza della popolazione

Esempio

Campione di n = 10 elementi da una normale con varianza 100



Efficienza

Dovendo scegliere tra due stimatori corretti si preferisce quello che ha una **varianza più piccola**.

Se la popolazione ha distribuzione normale con **media=mediana** μ sia la media campionaria $\bar{X}$ che la mediana campionaria M sono stimatori corretti di μ.

Ma la **media campionaria ha varianza minore della mediana**:

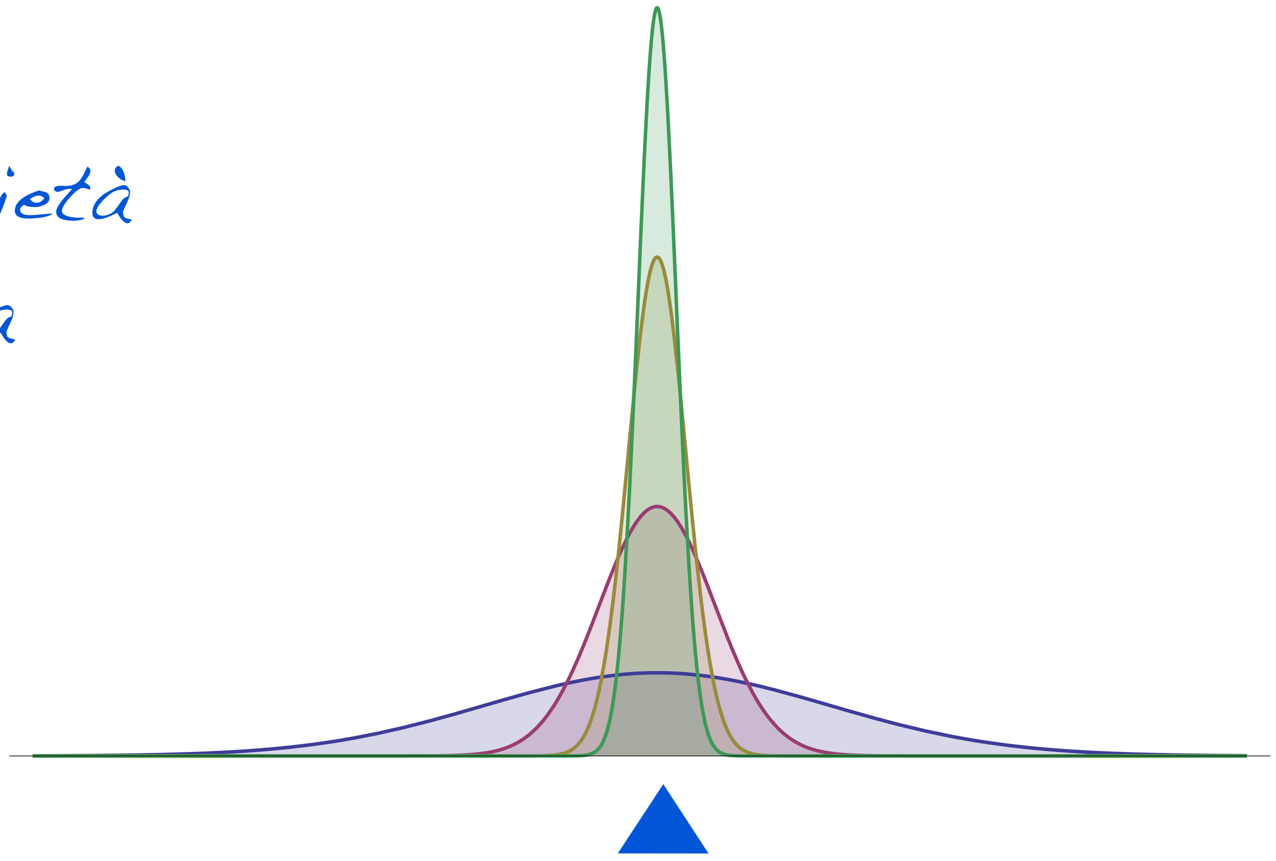
$$\mathrm{var}(\bar{X}) = \sigma^2/n \qquad \mathrm{var}(M) \approx 1.57 \cdot \sigma^2/n$$

La media è più **efficiente** della mediana:
per ottenere lo stesso ES della media con un campione di n = 100 casi con la mediana abbiamo bisogno di un campione di n = 157 casi.

Stimatore consistente

Uno stimatore si dice **consistente** se la sua distribuzione campionaria tende a concentrarsi sul parametro da stimare **all'aumentare della dimensione campionaria**.

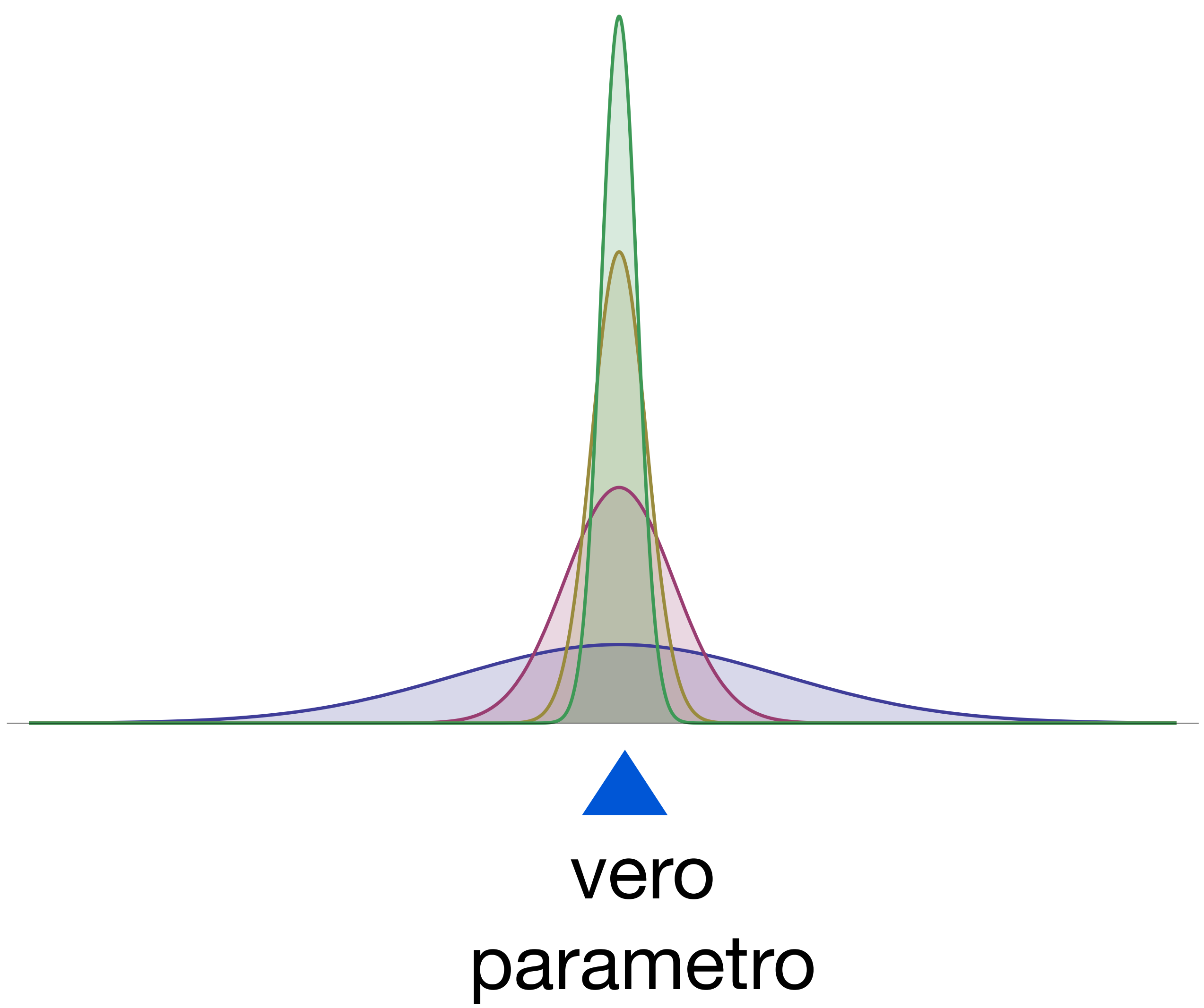
È una proprietà asintotica



vero parametro

Stimatore consistente

Due condizioni sono **sufficienti**. Al crescere di n
1) la media della distribuzione campionaria dello stimatore tende al parametro da stimare
2) la varianza della distribuzione campionaria dello stimatore tende a zero.



Stimatore consistente

Due condizioni sono **sufficienti**. Al crescere di n
1) la media della distribuzione campionaria dello stimatore tende al parametro da stimare
2) la varianza della distribuzione campionaria dello stimatore tende a zero.

La media campionaria e consistente
La proporzione campionaria è consistente
La varianza campionaria è consistente

Stima per intervallo

Quantificazione dell'incertezza

Stima,   Errore standard

La stima è
un valore puntuale

Lo stimatore è una variabile aleatoria, l'ES è la dev. st.

Intervallo di stima,   Intervallo di confidenza

Un intervallo di stima
è una **coppia di valori** (a,b)
che definiscono gli estremi

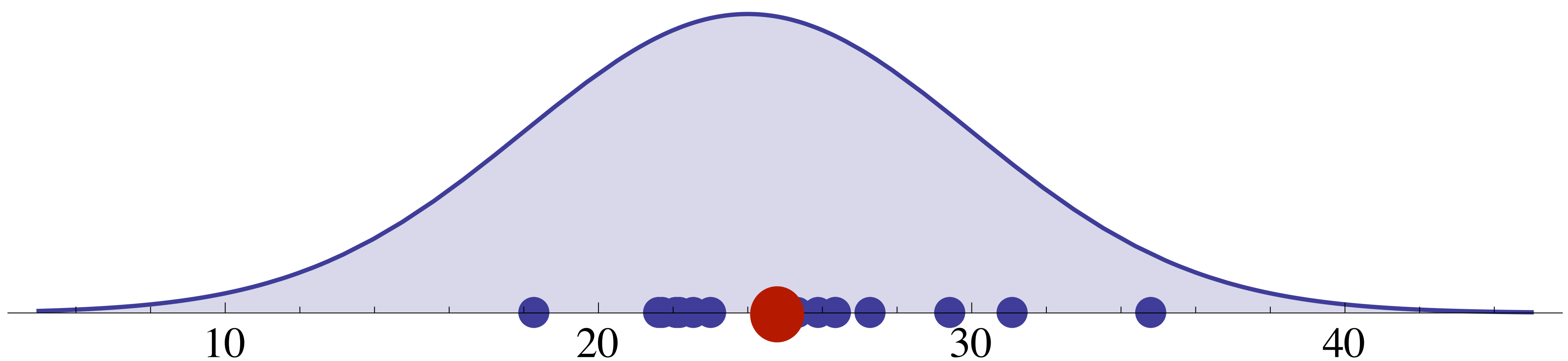
Un **intervallo di confidenza**
è una coppia
di variabile aleatorie che
con probabilità data contiene
il parametro

Esempio

Il tempo trascorso dai clienti in
un negozio è Normale $N(\mu =?,\ \sigma = 6 \text{ min})$

Si estrae un campione casuale di 16 clienti

La media campionaria è $\bar{x} = 25 \text{ min}$



$\bar{x} = 25$

Esempio

L'errore standard della stima è ES = $\sigma/\sqrt{16} = 6/4 = 1.5 \text{ min}$

Un intervallo di stima per la media è

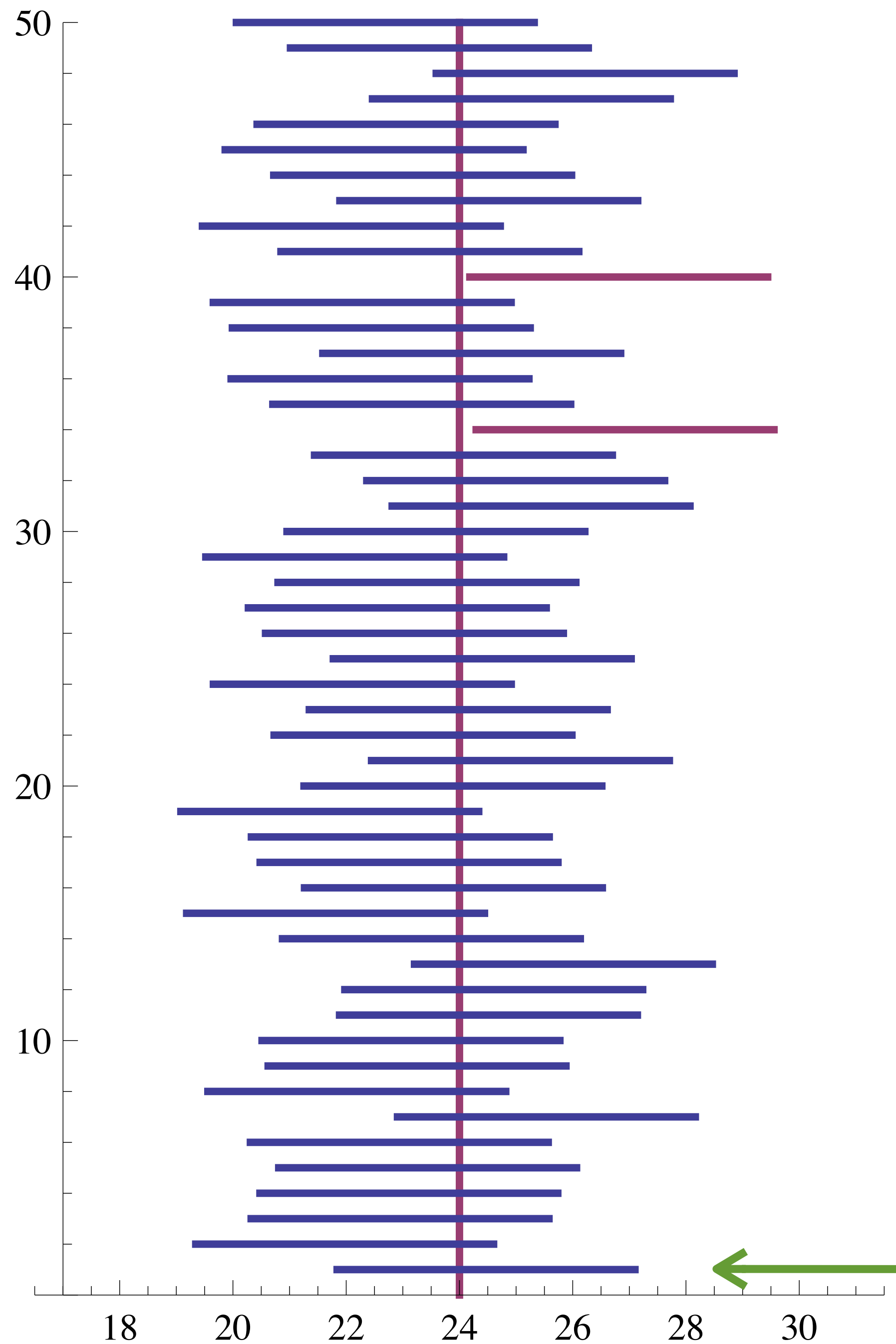
$$[\bar{x} - 1.96\text{ES},\ \bar{x} + 1.96\text{ES}]$$

Estremi: 25 - 1.96 * 1.5 = **22.06**; 25 + 1.96 * 1.5 = **27.94**

Nel campionamento ripetuto questi intervalli comprendono la vera media nel 95% dei casi

$$P(\bar{X} - 1.96\text{ES} < \mu < \bar{X} + 1.96\text{ES}) = 0.95$$

Esempio

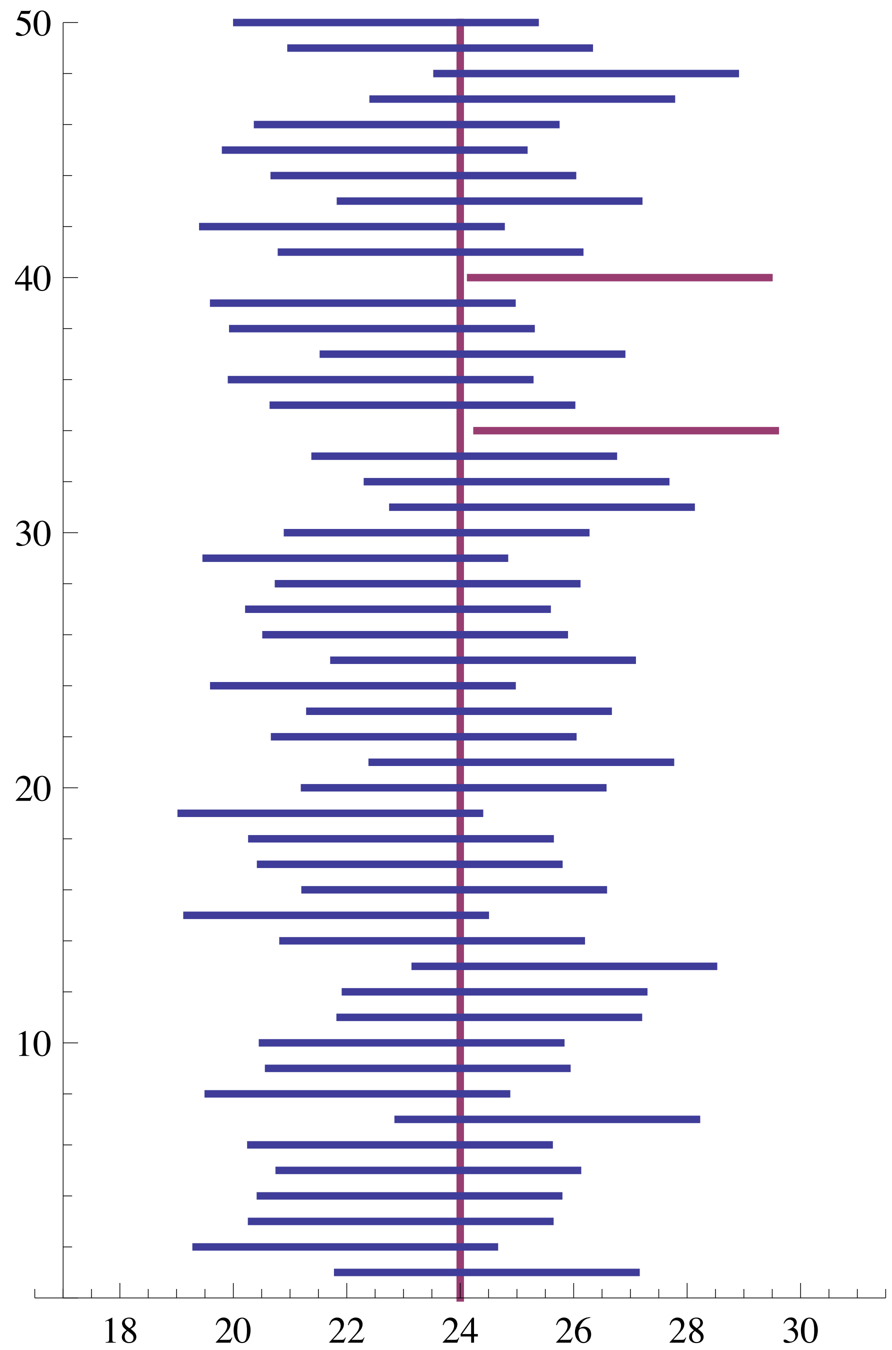


il mio intervallo è generato
da una procedura che nel 95% dei casi
copre la vera media

il mio intervallo

Vera media

Perché?



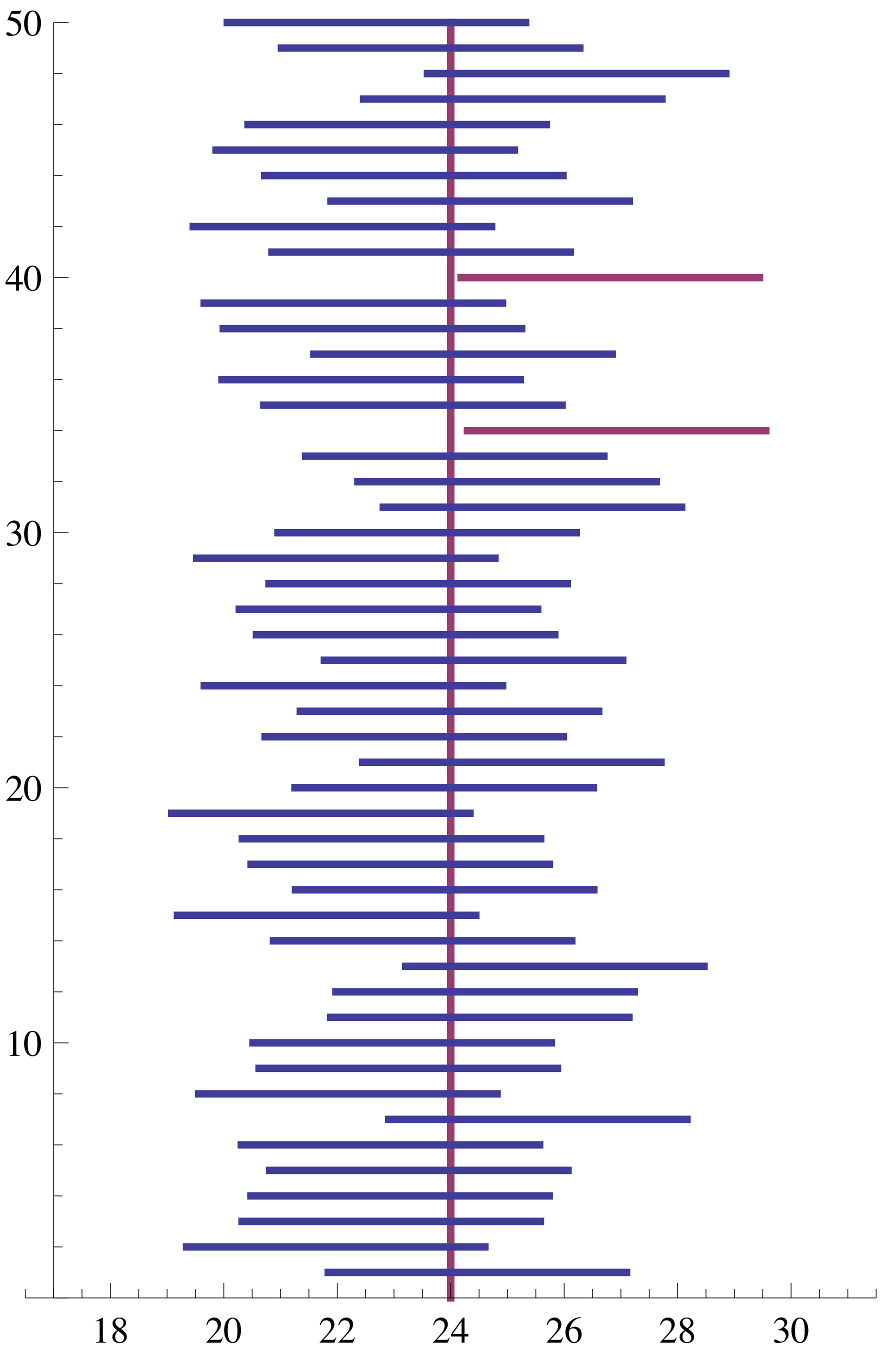
Vera media

$$P\left(-1.96 < \frac{\bar{X}-\mu}{\mathrm{ES}} < 1.96\right) = 0.95$$

è equivalente a

$$P(\bar{X} - 1.96\,\mathrm{ES} < \mu < \bar{X} + 1.96\,\mathrm{ES}) = 0.95$$

Interpretazione

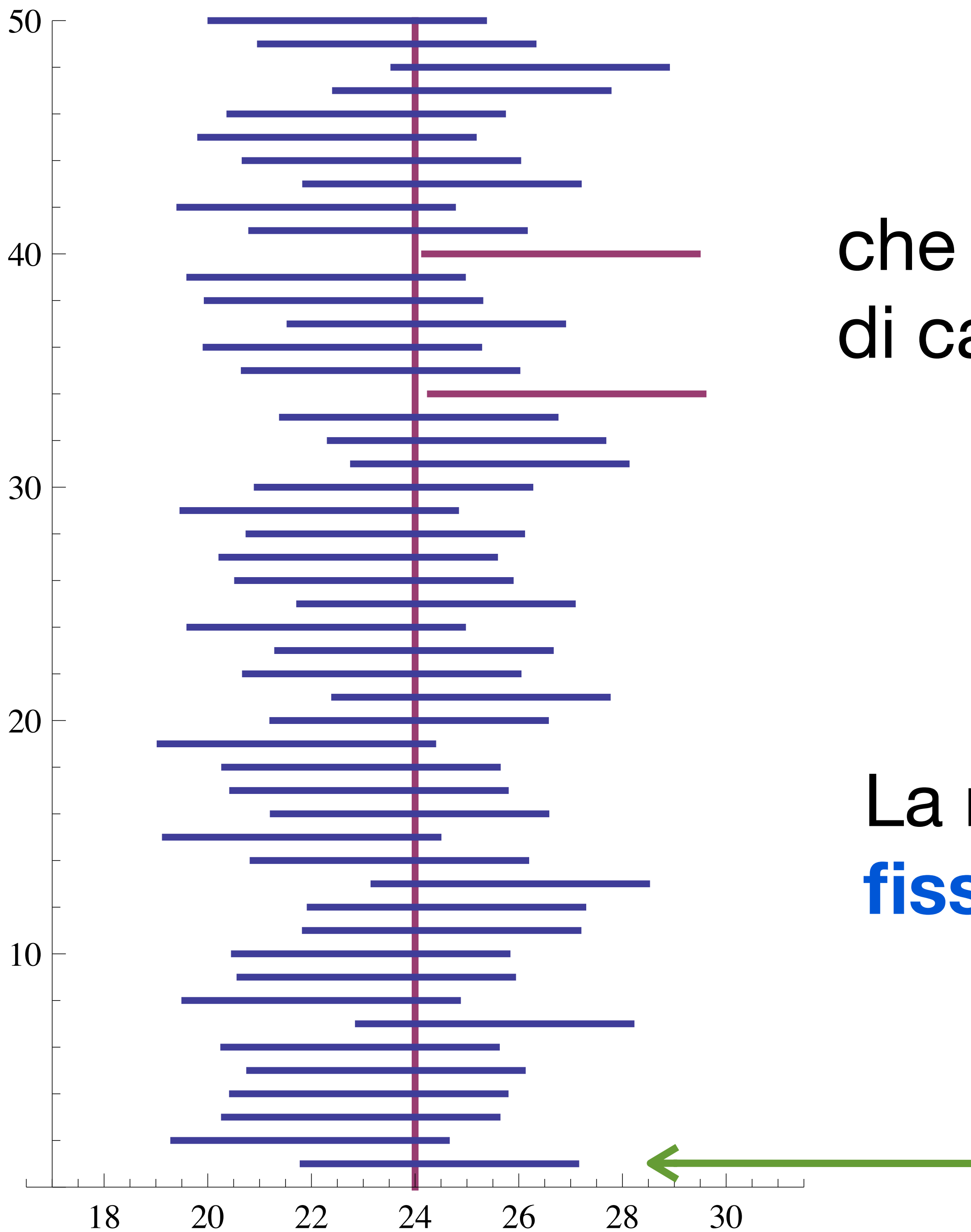


Vera media

Nel lungo andare gli intervalli costruiti come la media più o meno 1.96 ES comprendono la media incognita nel 95% dei casi

$$P(\bar{X} - 1.96\,\mathrm{ES} < \mu < \bar{X} + 1.96\,\mathrm{ES}) = 0.95$$

Interpretazione sbagliata



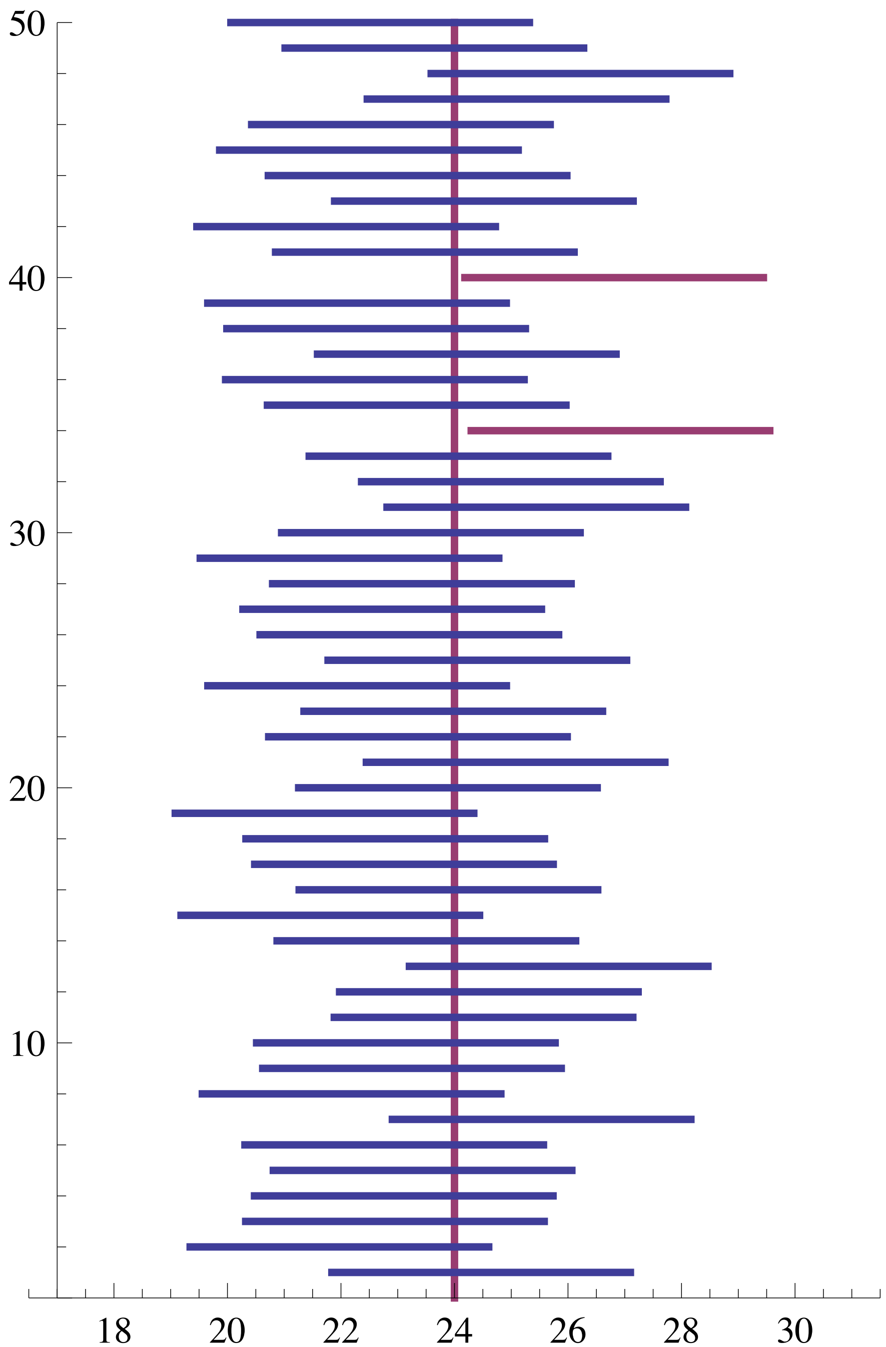
ATTENZIONE! NON È VERO

che la media incognita ha probabilità 0.95 di cadere nel **mio** intervallo (**22.06**, **27.94**)

$$\cancel{P(22.06 < \mu < 27.94) = 0.95}$$

La media della popolazione è una quantità **fissa** e non una variabile aleatoria

Intervallo di confidenza



Vera media

Siccome la **procedura** di costruzione degli intervalli nel lungo andare comprende la media incognita nel 95% dei casi

ho **fiducia** che il mio intervallo che è stato ottenuto da questa procedura sia vincente

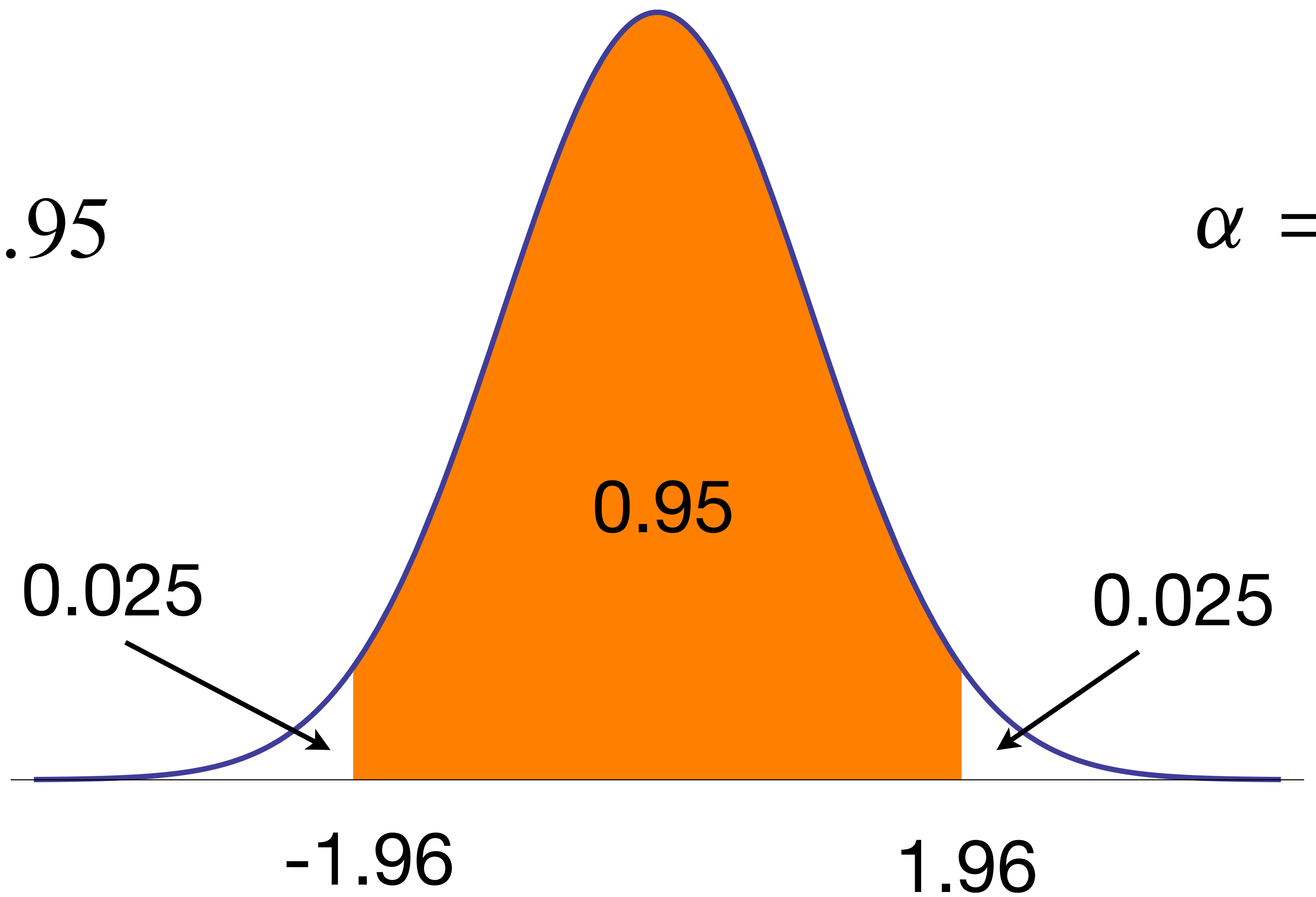
Per questo l'intervallo si chiama **intervallo di confidenza** e 0.95 si dice **livello di confidenza**

Livelli di confidenza

Livello di confidenza	80%	90%	95%	99%	99.9%
$z_{\alpha/2}$	1.28	1.64	**1.96**	2.58	3.29

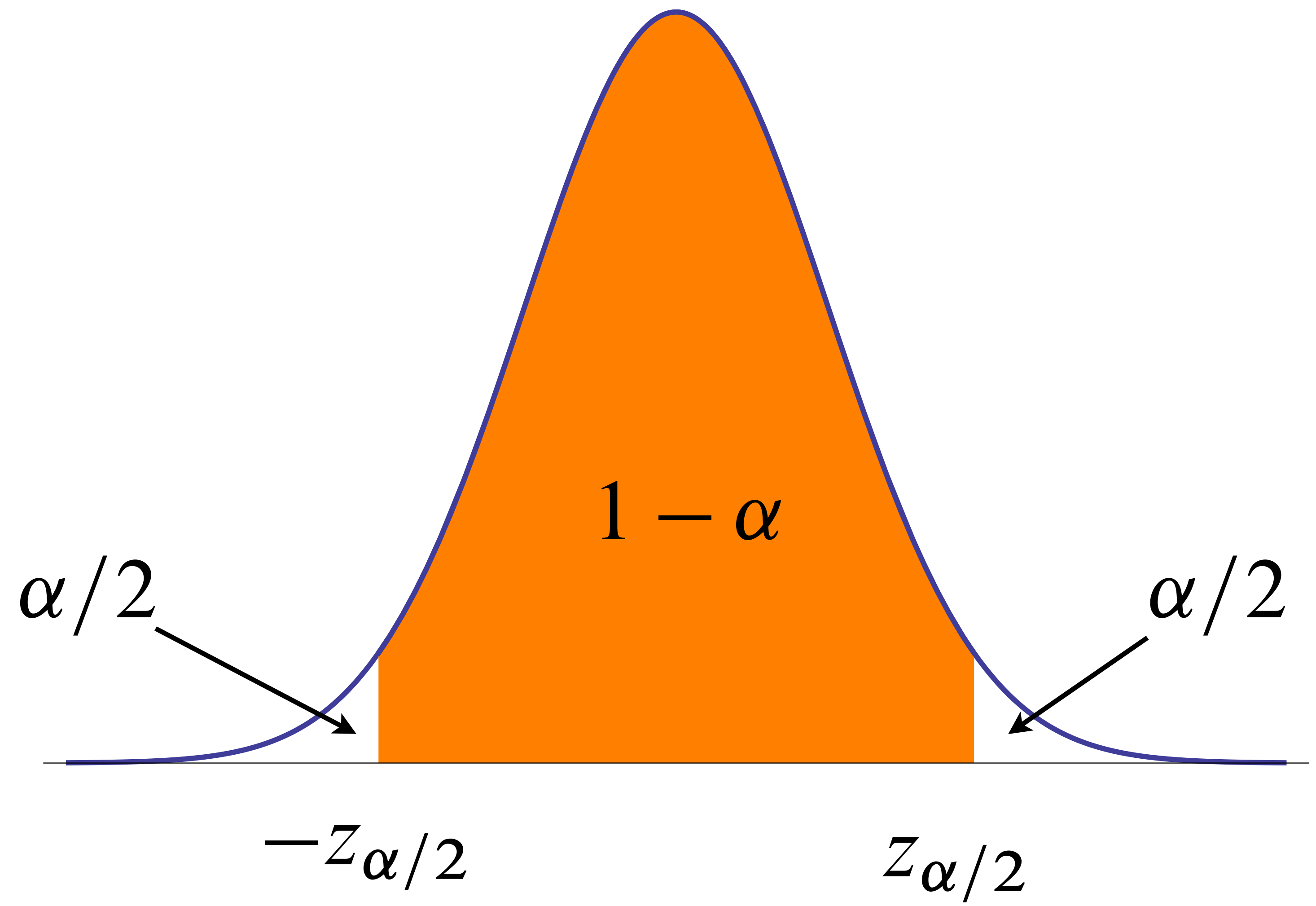
$$\mathrm{LC} = 0.95$$

$$\alpha = 0.05$$



Livelli di confidenza

$1-\alpha$	80%	90%	95%	99%	99.9%
$z_{\alpha/2}$	1.28	1.64	**1.96**	2.58	3.29



Margine di errore

In un intervallo di confidenza la semiampiezza

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

il libro la chiama **margine di errore**

$$\bar{X} - \text{ME} \qquad \bar{X} \qquad \bar{X} + \text{ME}$$

IC per μ (X~Normale, σ^2 non nota)

Quando si costruisce un IC per la media μ la deviazione std σ non è di diretto interesse, ma è comunque un ingrediente necessario perché entra nell'espressione dell'IC

Nella maggior parte delle applicazioni la deviazione std σ **non è nota** e quindi per poter determinare l'IC per μ occorre rimpiazzare σ con una sua stima

Stima della varianza

Lo **stimatore corretto** della varianza della popolazione σ^2 è la varianza campionaria, quella con il divisore n-1

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

gradi di libertà

Sostituendo, lo **stimatore dell'errore standard** è

$$\text{ES stimato} = \frac{S}{\sqrt{n}}$$

Cosa sono i gradi di libertà

Gli scarti dalla media campionaria

$$X_1 - \bar{X}, X_2 - \bar{X}, \ldots, X_n - \bar{X}$$

Non sono indipendenti fra loro perché l**a loro somma deve essere zero**.

Per esempio se ci sono tre osservazioni: 12, 1, ? con **media = 5**

Gli scarti dalla media sono 7, -4, ?

L'ultimo scarto per forza è -3 !

Sono libero di scegliere solo n-1 scarti dalla media, l'ultimo no.

IC per μ (X~Normale, σ^2 non nota)

Alla base dell'intervallo di confidenza per la media di una normale con varianza nota c'è la relazione

$$P\left(-1.96 < \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} < 1.96\right) = 0.95$$

Normale Standard

Vale per qualsiasi n

IC per μ (X~Normale, σ^2 non nota)

Non si può usare per l'intervallo di confidenza per la media di una normale con varianza ignota

$$P\left(-1.96 < \frac{\bar{X}-\mu}{S/\sqrt{n}} < 1.96\right) \neq 0.95$$

Non è Normale Standard
ma t di Student

IC per μ (X~Normale, σ^2 <u>non</u> nota)

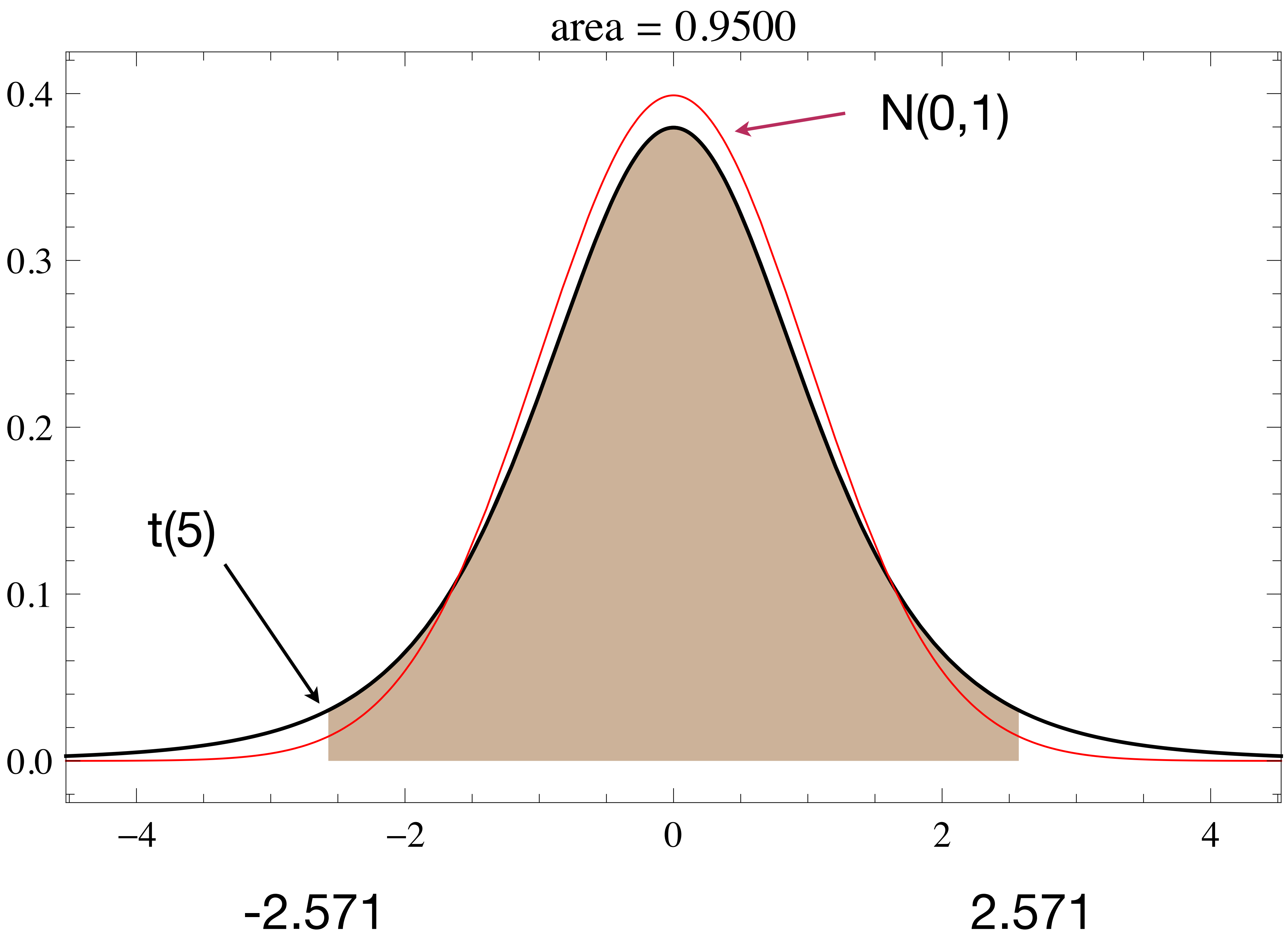
La distribuzione della media standardizzata non è normale ma **t di Student**. Per esempio per un campione di dimensione 6 si ottiene

$$P\left(-2.571 < \frac{\bar{X}-\mu}{S/\sqrt{n}} < 2.571\right) = 0.95$$

È una t di Student con 5 gradi di libertà

C'è una distribuzione diversa
per ogni dimensione campionaria n

Distribuzione della t di Student



La distribuzione t di Student

La t di Student è una famiglia parametrica di v.a. continue che assomigliano alla Normale Standard

Il parametro della famiglia è un numero intero **detto gradi di libertà**

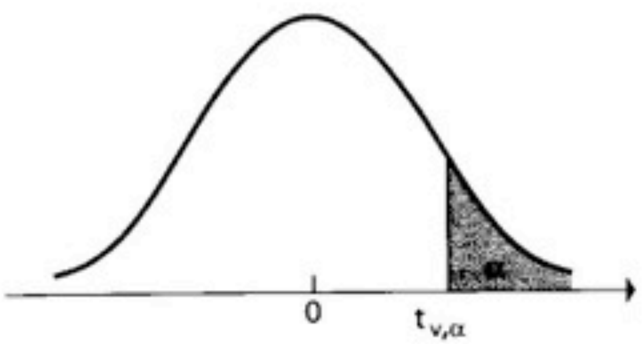
Ogni t ha una distribuzione simmetrica con media 0, varianza appena maggiore di 1 e code più spesse della Normale standard

La t di Student è sostanzialmente diversa dalla Normale standard quando il numero di gdl è piccolo (meno di 20)

Al crescere dei gradi di libertà **converge alla Normale standard**

Tavole della t di Student

Tavola 2 Distribuzione t di Student.



Fornisce i valori di t data la probabilità

In corrispondenza alla variabile aleatoria t di Student con v gradi di libertà la tavola contiene, per determinati valori di α, i valori di $t_{v,\alpha}$ tali che $P(t_v > t_{v,\alpha}) = \alpha$ (ovvero il quantile di ordine $1 - \alpha$). Ad esempio, la probabilità che la variabile aleatoria t di Student con 10 gradi di libertà superi 1.372 è pari a 0.10.

ν	α				
	0.100	**0.050**	**0.025**	**0.010**	**0.005**
1	3.078	6.314	12.706	31.821	63.657
2	1.886	2.920	4.303	6.965	9.925
3	1.638	2.353	3.182	4.541	5.841
4	1.533	2.132	2.776	3.747	4.604
5	1.476	2.015	2.571	3.365	4.032
6	1.440	1.943	2.447	3.143	3.707
7	1.415	1.895	2.365	2.998	3.499
8	1.397	1.860	2.306	2.896	3.355
9	1.383	1.833	2.262	2.821	3.250
10	1.372	1.812	2.228	2.764	3.169
11	1.363	1.796	2.201	2.718	3.106
12	1.356	1.782	2.179	2.681	3.055
13	1.350	1.771	2.160	2.650	3.012
14	1.345	1.761	2.145	2.624	2.977

ν	α				
	0.100	**0.050**	**0.025**	**0.010**	**0.005**
1	3.078	6.314	12.706	31.821	63.657
2	1.886	2.920	4.303	6.965	9.925
3	1.638	2.353	3.182	4.541	5.841
4	1.533	2.132	2.776	3.747	4.604
5	1.476	2.015	2.571	3.365	4.032
6	1.440	1.943	2.447	3.143	3.707
7	1.415	1.895	2.365	2.998	3.499
8	1.397	1.860	2.306	2.896	3.355
9	1.383	1.833	2.262	2.821	3.250
10	1.372	1.812	2.228	2.764	3.169
11	1.363	1.796	2.201	2.718	3.106
12	1.356	1.782	2.179	2.681	3.055
13	1.350	1.771	2.160	2.650	3.012
14	1.345	1.761	2.145	2.624	2.977
15	1.341	1.753	2.131	2.602	2.947
16	1.337	1.746	2.120	2.583	2.921
17	1.333	1.740	2.110	2.567	2.898
18	1.330	1.734	2.101	2.552	2.878
19	1.328	1.729	2.093	2.539	2.861
20	1.325	1.725	2.086	2.528	2.845
21	1.323	1.721	2.080	2.518	2.831
22	1.321	1.717	2.074	2.508	2.819
23	1.319	1.714	2.069	2.500	2.807
24	1.318	1.711	2.064	2.492	2.797
25	1.316	1.708	2.060	2.485	2.787
26	1.315	1.706	2.056	2.479	2.779
27	1.314	1.703	2.052	2.473	2.771
28	1.313	1.701	2.048	2.467	2.763
29	1.311	1.699	2.045	2.462	2.756
30	1.310	1.697	2.042	2.457	2.750
40	1.303	1.684	2.021	2.423	2.704
60	1.296	1.671	2.000	2.390	2.660
∞	1.282	1.645	1.960	2.326	2.576

Se i gdl sono infiniti dà il valore della N(0,1)

IC per la media con la t di Student

Quando σ è nota si usa il valore tabulato z della Normale standard

Quando σ non è nota si usa il valore tabulato della t di Student con n–1 gradi di libertà.

Quando σ non è nota ma n è grande (> 100) si usa il valore tabulato z della Normale standard perchè è lo stesso.

Influenza della distribuzione *t* di Student

La t di Student ha code più lunghe della Normale standard

Quindi il quantile della t è più grande (= spostato verso destra) rispetto a quello della Normale standard

L'IC per μ è più ampio quando σ non è nota in quanto il valore t è più grande di z

È dovuto all'**incertezza maggiore** causata dalla necessità di stimare σ

n	gdl	*t*	*z*	*t/z*
3	2	4.30	1.96	2.20
5	4	2.78	1.96	1.42
10	9	2.26	1.96	1.15
20	19	2.09	1.96	1.07
30	29	2.05	1.96	1.04
40	39	2.02	1.96	1.03
50	49	2.01	1.96	1.03
75	74	1.99	1.96	1.02
100	99	1.98	1.96	1.01
120	119	1.98	1.96	1.01
200	199	1.97	1.96	1.01
500	499	1.96	1.96	1.00

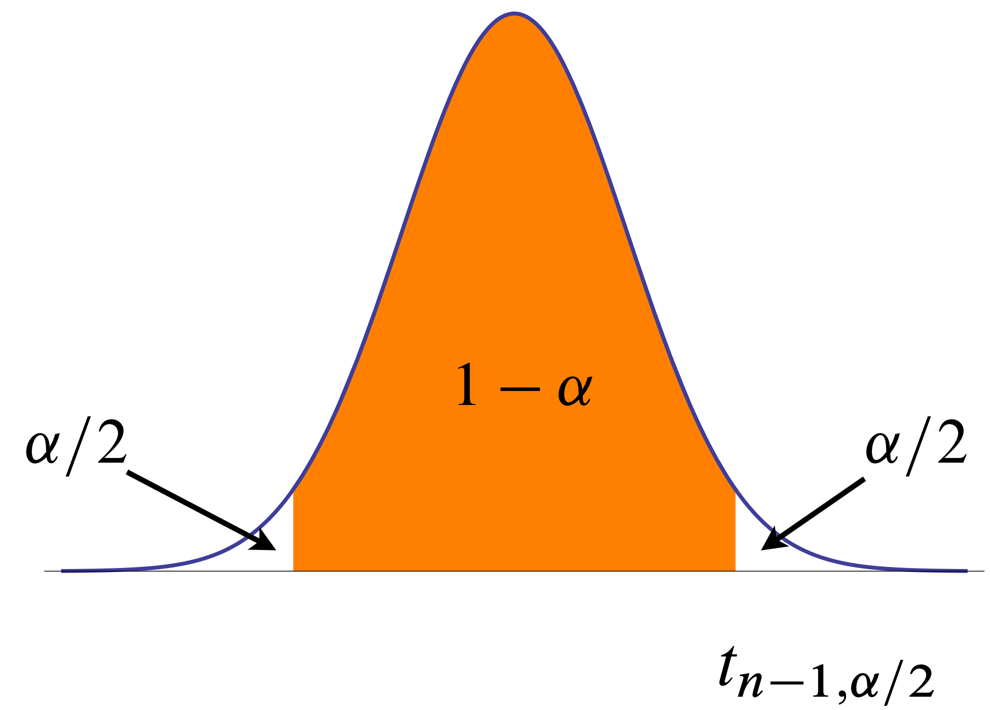
Esempio: valori critici che lasciano sulla coda destra $\alpha=0.025$

IC per μ (X~Normale, σ^2 non nota)

Se la popolazione è $X \sim N(\mu,\sigma^2)$ con μ e σ^2 entrambi ignoti e con un campione casuale di dimensione n, l'intervallo di confidenza **con livello di confidenza 1−α** è

$$\bar{X} \pm t_{n-1,\alpha/2} \frac{S}{\sqrt{n}}$$

Il valore $t_{n-1,\alpha/2}$ è il valore che, nella distribuzione t di Student con n−1 gdl, lascia a destra un probabilità α/2



Esempio

Consumo di carburante (miglia/gallone) in un campione casuale di 24 autocarri

15.5 21.0 18.5 19.3 19.7 16.9 20.2 14.5
16.5 19.2 18.7 18.2 18.0 17.5 18.5 20.5
18.6 19.1 19.8 18.0 19.8 18.2 20.3 21.8

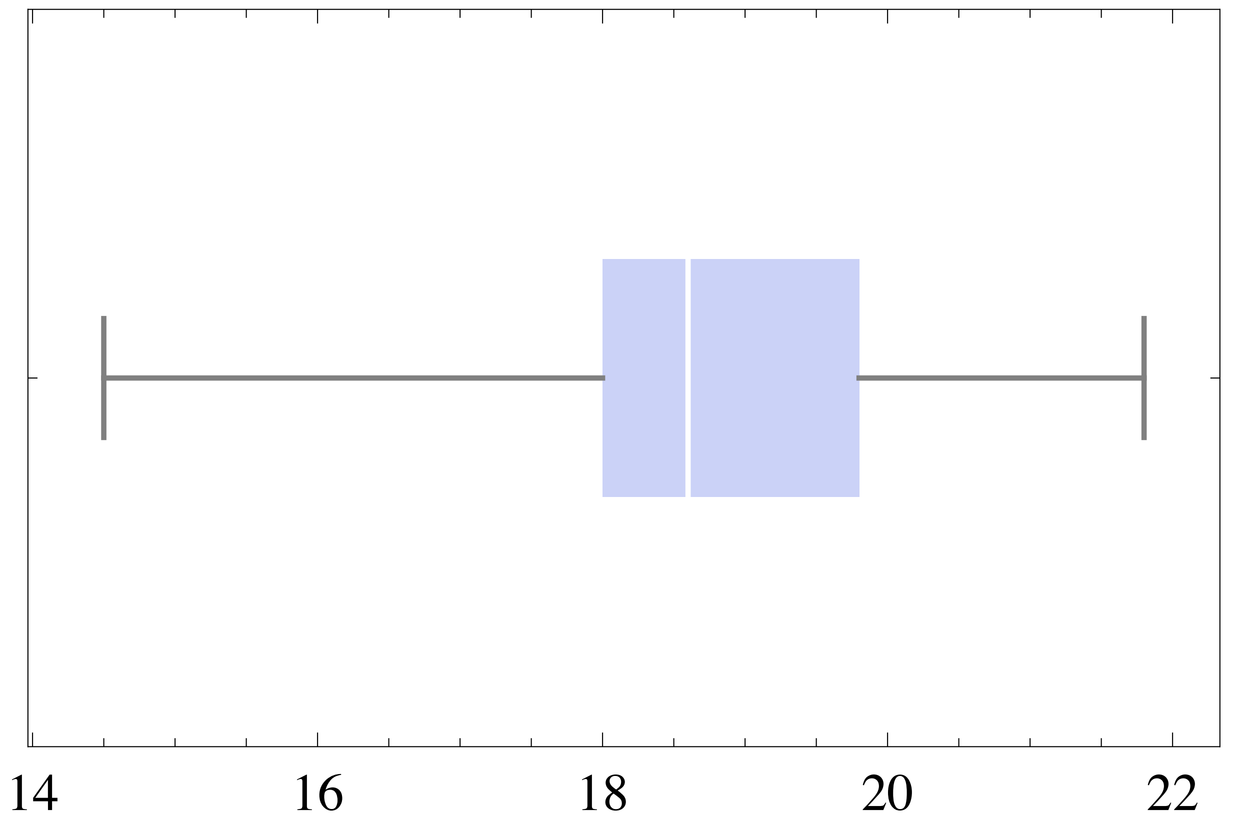
media = 18.68

s = 1.69526 *(divisore n-1)*

ES = $s/\sqrt{n}$

= 1.69526 / 4.89

= **0.346**



Esempio

media = 18.68 **s = 1.69526** **ES = 1.69526**

Intervallo di confidenza al **livello** 0.95 $\alpha/2 = 0.025$

Gradi di libertà n - 1 = 23

$$18.68 \pm t_{23,0.025} \times 0.346$$

ν	α				
	0.100	**0.050**	**0.025**	**0.010**	**0.005**
1	3.078	6.314	12.706	31.821	63.657
2	1.886	2.920	4.303	6.965	9.925
3	1.638	2.353	3.182	4.541	5.841
4	1.533	2.132	2.776	3.747	4.604
5	1.476	2.015	2.571	3.365	4.032
6	1.440	1.943	2.447	3.143	3.707
7	1.415	1.895	2.365	2.998	3.499
8	1.397	1.860	2.306	2.896	3.355
9	1.383	1.833	2.262	2.821	3.250
10	1.372	1.812	2.228	2.764	3.169
11	1.363	1.796	2.201	2.718	3.106
12	1.356	1.782	2.179	2.681	3.055
13	1.350	1.771	2.160	2.650	3.012
14	1.345	1.761	2.145	2.624	2.977
15	1.341	1.753	2.131	2.602	2.947
16	1.337	1.746	2.120	2.583	2.921
17	1.333	1.740	2.110	2.567	2.898
18	1.330	1.734	2.101	2.552	2.878
19	1.328	1.729	2.093	2.539	2.861
20	1.325	1.725	2.086	2.528	2.845
21	1.323	1.721	2.080	2.518	2.831
22	1.321	1.717	2.074	2.508	2.819
23	1.319	1.714	2.069	2.500	2.807
24	1.318	1.711	2.064	2.492	2.797
25	1.316	1.708	2.060	2.485	2.787
26	1.315	1.706	2.056	2.479	2.779
27	1.314	1.703	2.052	2.473	2.771
28	1.313	1.701	2.048	2.467	2.763
29	1.311	1.699	2.045	2.462	2.756
30	1.310	1.697	2.042	2.457	2.750
40	1.303	1.684	2.021	2.423	2.704
60	1.296	1.671	2.000	2.390	2.660
∞	1.282	1.645	1.960	2.326	2.576

Esempio

media = 18.68 **s = 1.69526** **ES = 1.69526**

Intervallo di confidenza al **livello** 0.95 $\alpha/2 = 0.025$

Gradi di libertà n - 1 = 23

$$18.68 \pm t_{23,0.025} \times 0.346$$

18.68 - 2.069 * 0.346 = **17.96** 18.68 + 2.069 * 0.346 = **19.39**

Esempio

media = 18.68 **s = 1.69526** **ES = 1.69526**

Intervallo di confidenza al **livello** 0.90 $\alpha/2 = 0.05$

Gradi di libertà n - 1 = 23

$$18.68 \pm t_{23,0.05} \times 0.346$$

ν	α				
	0.100	**0.050**	**0.025**	**0.010**	**0.005**
1	3.078	6.314	12.706	31.821	63.657
2	1.886	2.920	4.303	6.965	9.925
3	1.638	2.353	3.182	4.541	5.841
4	1.533	2.132	2.776	3.747	4.604
5	1.476	2.015	2.571	3.365	4.032
6	1.440	1.943	2.447	3.143	3.707
7	1.415	1.895	2.365	2.998	3.499
8	1.397	1.860	2.306	2.896	3.355
9	1.383	1.833	2.262	2.821	3.250
10	1.372	1.812	2.228	2.764	3.169
11	1.363	1.796	2.201	2.718	3.106
12	1.356	1.782	2.179	2.681	3.055
13	1.350	1.771	2.160	2.650	3.012
14	1.345	1.761	2.145	2.624	2.977
15	1.341	1.753	2.131	2.602	2.947
16	1.337	1.746	2.120	2.583	2.921
17	1.333	1.740	2.110	2.567	2.898
18	1.330	1.734	2.101	2.552	2.878
19	1.328	1.729	2.093	2.539	2.861
20	1.325	1.725	2.086	2.528	2.845
21	1.323	1.721	2.080	2.518	2.831
22	1.321	1.717	2.074	2.508	2.819
23	1.319	1.714	2.069	2.500	2.807
24	1.318	1.711	2.064	2.492	2.797
25	1.316	1.708	2.060	2.485	2.787
26	1.315	1.706	2.056	2.479	2.779
27	1.314	1.703	2.052	2.473	2.771
28	1.313	1.701	2.048	2.467	2.763
29	1.311	1.699	2.045	2.462	2.756
30	1.310	1.697	2.042	2.457	2.750
40	1.303	1.684	2.021	2.423	2.704
60	1.296	1.671	2.000	2.390	2.660
∞	1.282	1.645	1.960	2.326	2.576

Esempio

media = 18.68 **s = 1.69526** **ES = 1.69526**

Intervallo di confidenza al **livello** 0.90 $\alpha/2 = 0.05$

Gradi di libertà n - 1 = 23

$$18.68 \pm t_{23,0.05} \times 0.346$$

18.68 - 1.714 * 0.346 = **18.1** 18.68 + 1.714 * 0.346 = **19.3**

IC per la proporzione p

*Si possono determinare IC **approssimati** per p usando la tecnica basata sulla stima più o meno un multiplo dell'errore standard*

Sappiamo che $\hat{P} = X/n$ è uno stimatore corretto di *p*

con errore standard

$$\mathrm{ES} = \sqrt{\frac{p(1-p)}{n}}$$

Poiché non conosciamo *p* non possiamo sapere esattamente ES, ma possiamo stimarlo con

$$\widehat{\mathrm{ES}} = \sqrt{\frac{\hat{P}(1-\hat{P})}{n}}$$

IC per la proporzione p

Si possono determinare IC ***approssimati*** *per p usando la tecnica basata sulla stima più o meno un multiplo dell'errore standard*

Allora, se il campione è "grande" , l'IC al 95% per p è

$$\hat{P} \pm 1.96 \sqrt{\frac{\hat{P}(1-\hat{P})}{n}}$$

È un intervallo **approssimato** e vale se $n\hat{P}(1-\hat{P}) > 9$

Nota che è costruito come quello per la media di una normale con varianza nota

Giustificazione

Se $\quad n\hat{P}(1-\hat{P}) > 9$

la distribuzione della proporzione campionaria **è bene approssimata da una normale**

$$\frac{\hat{P}-p}{\widehat{\mathrm{ES}}} \approx N(0,1) \qquad \text{dove} \qquad \widehat{\mathrm{ES}} = \sqrt{\frac{\hat{P}(1-\hat{P})}{n}}$$

Quindi

$$P\left(-1.96 < \frac{\hat{P}-p}{\widehat{\mathrm{ES}}} < 1.96\right) \approx 0.95$$

Giustificazione

E invertendo le diseguaglianze si ha che

$$P\left(-1.96 < \frac{\hat{P} - p}{\widehat{\mathrm{ES}}} < 1.96\right) \approx 0.95$$

è equivalente a

$$P\left(\hat{P} - 1.96\,\widehat{\mathrm{ES}} < p < \hat{P} + 1.96\,\widehat{\mathrm{ES}}\right) \approx 0.95$$

Esempio

In una indagine di mercato 90 persone su 225 intervistate (40%) ricordano la pubblicità di un certo prodotto.

Se le risposte delle 225 persone intervistate sono un campione casuale, trovare l'IC al livello 95% della proporzione nella popolazione di persone che ricordano la pubblicità.

Esempio

In una indagine di mercato 90 persone su 225 intervistate (40%) ricordano la pubblicità di un certo prodotto.

n = 225; stima di p = 0.4

$$\widehat{\mathrm{ES}} = \sqrt{\frac{0.4 \cdot 0.6}{225}} = 0.03266$$

$$\begin{aligned}
\hat{p} \pm z_{\alpha/2} \frac{\sqrt{\hat{p}(1-\hat{p})}}{\sqrt{n}} &= 0.4 \pm 1.96 \frac{\sqrt{0.4(1-0.4)}}{\sqrt{225}} \\
&= 0.4 \pm 1.96 \frac{0.4899}{15} \\
&= 0.4 \pm 0.0640 \\
&= [0.3360, 0.4640] \quad \textit{ovvero} \quad [33.60\%, 46.40\%]
\end{aligned}$$

Scelta della dimensione del campione

La dimensione campionaria influenza l'ampiezza dell'IC

In fase di progettazione di un'indagine statistica la dimensione del campione è un fattore da determinare

Uno dei criteri per fissarla è quello di determinarla **in modo da ottenere un margine di errore** prefissato (= semiampezza dell'IC)

Nell'esempio precedente si è ottenuto un IC di 0.4 ± 0.064: dunque ME = 0.064, mentre la lunghezza totale è $0.064 \times 2 = 0.128$

Scelta dell'ampiezza campionaria

Tipo di IC	***ME***
IC per μ, σ nota	$z_{\alpha/2} \dfrac{\sigma}{\sqrt{n}}$
IC per μ, σ ignota	$t_{n-1,\alpha/2} \dfrac{s}{\sqrt{n}}$
IC per p	$z_{\alpha/2} \dfrac{\sqrt{\hat{p}(1-\hat{p})}}{\sqrt{n}}$

Si fissa ME e quindi si imposta un'equazione e si risolve per n

Scelta dell’ampiezza campionaria

Nel caso dell’IC per μ con σ ignota, occorre ipotizzare la dev.std. campionaria s. Questa talvolta è stimata con un’indagine preliminare, detta **indagine pilota**.

Nel caso dell’IC per p, n occorre ipotizzare la proporzione campionaria (con il caso peggiore $p = 0.5$) o stimarla con un’indagine pilota.

Scelta dell'ampiezza campionaria

Fissato un livello di confidenza (1–α) e scelto il margine di errore, il valore di n si ottiene risolvendo l'equazione

$$\text{ME} = \text{valore tabulato} \times \frac{\text{dev.stand}}{\sqrt{n}}$$

da cui si ottiene

$$n = \text{valore tabulato}^2 \times \frac{\text{dev.stand}^2}{\text{ME}^2}$$

Esempi

Popolazione: peso (g): distribuito normalmente

Deviazione standard: 15 g

Livello di confidenza: 95%

Margine di errore: ±3 grammi

Dimensione campionaria:

$$n = 1.96^2 \times \frac{15^2}{3^2} = 96.04 \rightarrow 97$$

Esempio

Popolazione: soddisfazione del cliente, distribuita come una Bernoulli

Stima di p: 0.75

Quindi deviazione standard: $\sqrt{\hat{p}\hat{q}} = \sqrt{0.75 \cdot 0.25} =$ 0.4330

Livello di confidenza 95%

Margine di errore: ±5 punti percentuali = ±0.05

Dimensione campionaria

$$n = 1.96^2 \times \frac{0.4330^2}{0.05^2} = 288.12 \rightarrow 289$$

Esempio (segue)

Popolazione: soddisfazione del cliente, distribuita come una Bernoulli

Valore di p incognito

Scelgo il caso peggiore p = 0.5 quindi deviazione std: $\sqrt{pq} = 0.5$

Livello di confidenza 95%

Margine di errore: ±5 punti percentuali = ±0.05

Dimensione campionaria

$$n = 1.96^2 \frac{0.5^2}{0.05^2} = 384$$