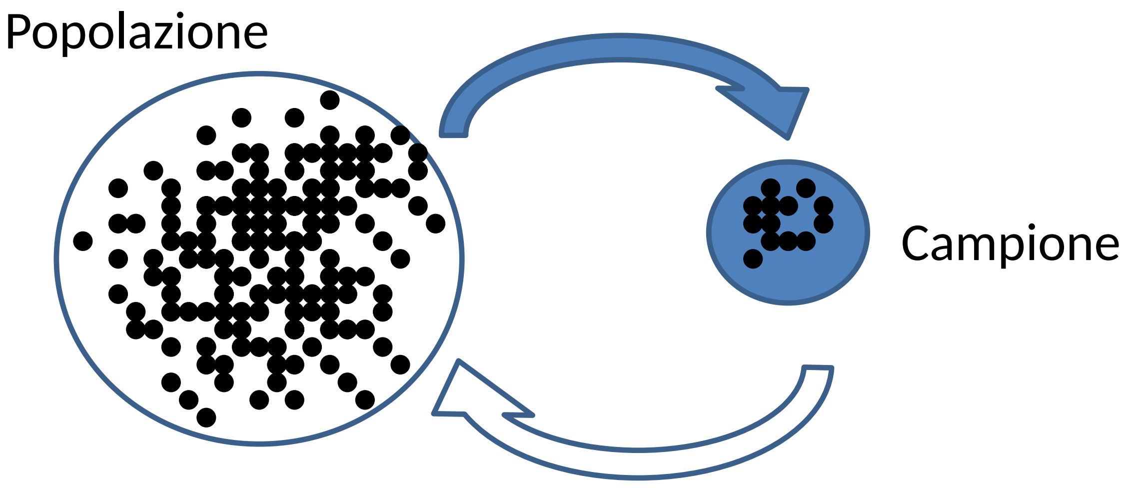
Last updated Aprile 4, 2023

Introduzione all'inferenza
Lezione 4

G. Bacaro

Statistica
CdL in Scienze e Tecnologie per l'Ambiente e la Natura
I anno, II semestre

Introduzione all'inferenza



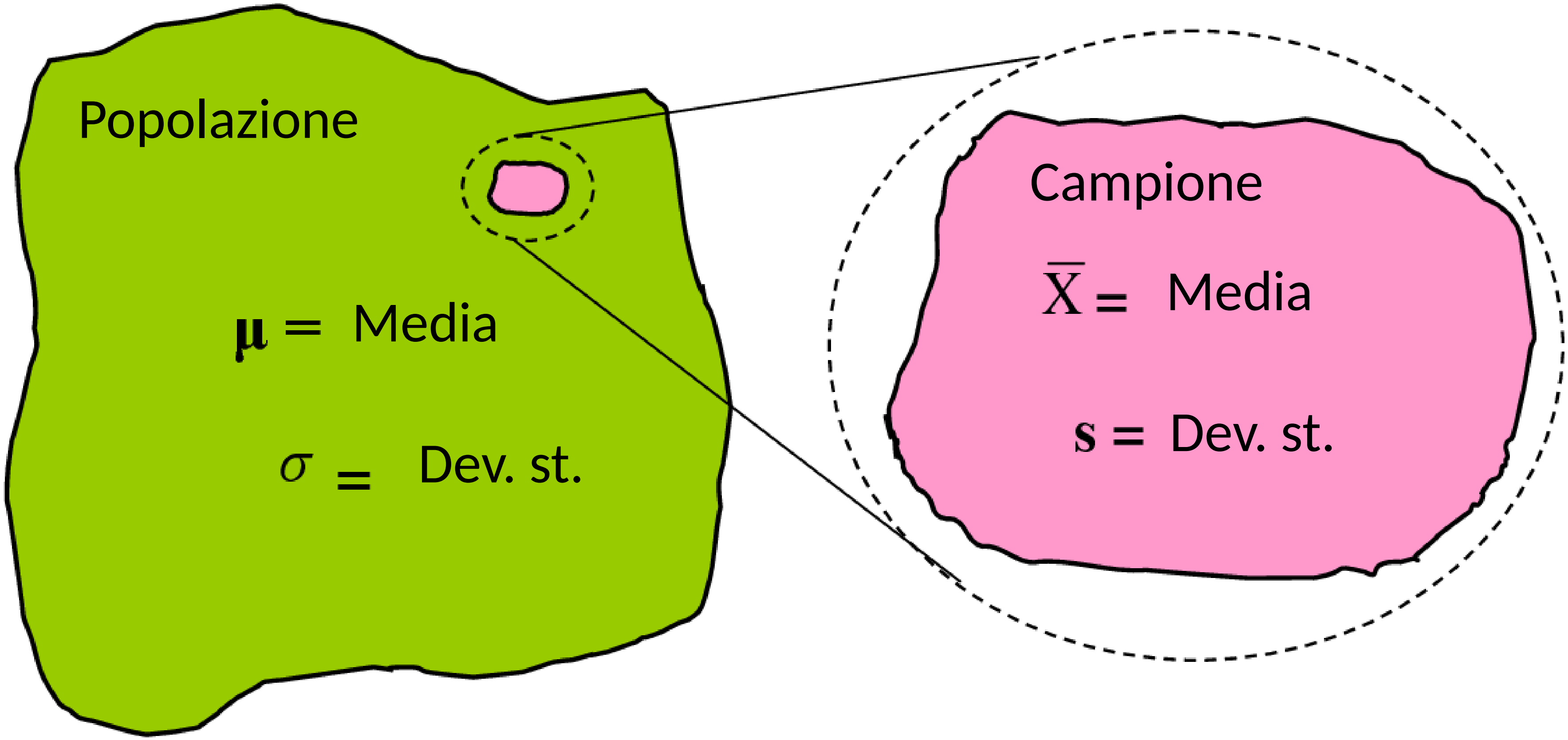
1. Distribuzione delle medie campionarie

2. Il teorema del limite centrale

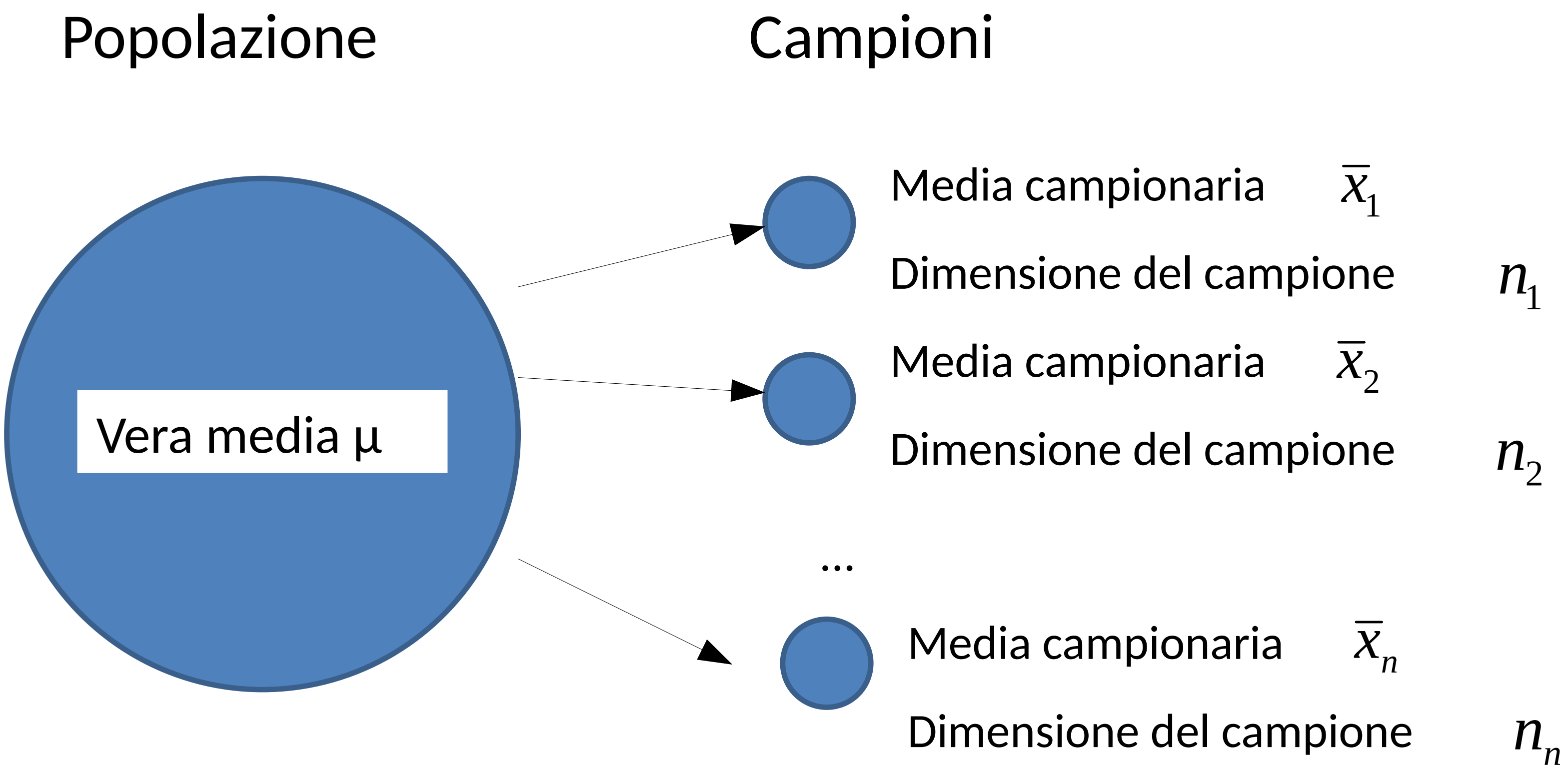
3. Attendibilità della stima campionaria: errore standard della media e limiti fiduciali

Introduzione all’inferenza

Campionamento



Distribuzione delle medie campionarie



Ci sono infiniti campioni che possono essere estratti dalla popolazione

Esempio estrazione di 100 campioni casuali di 5 studenti

Popolazione degli studenti

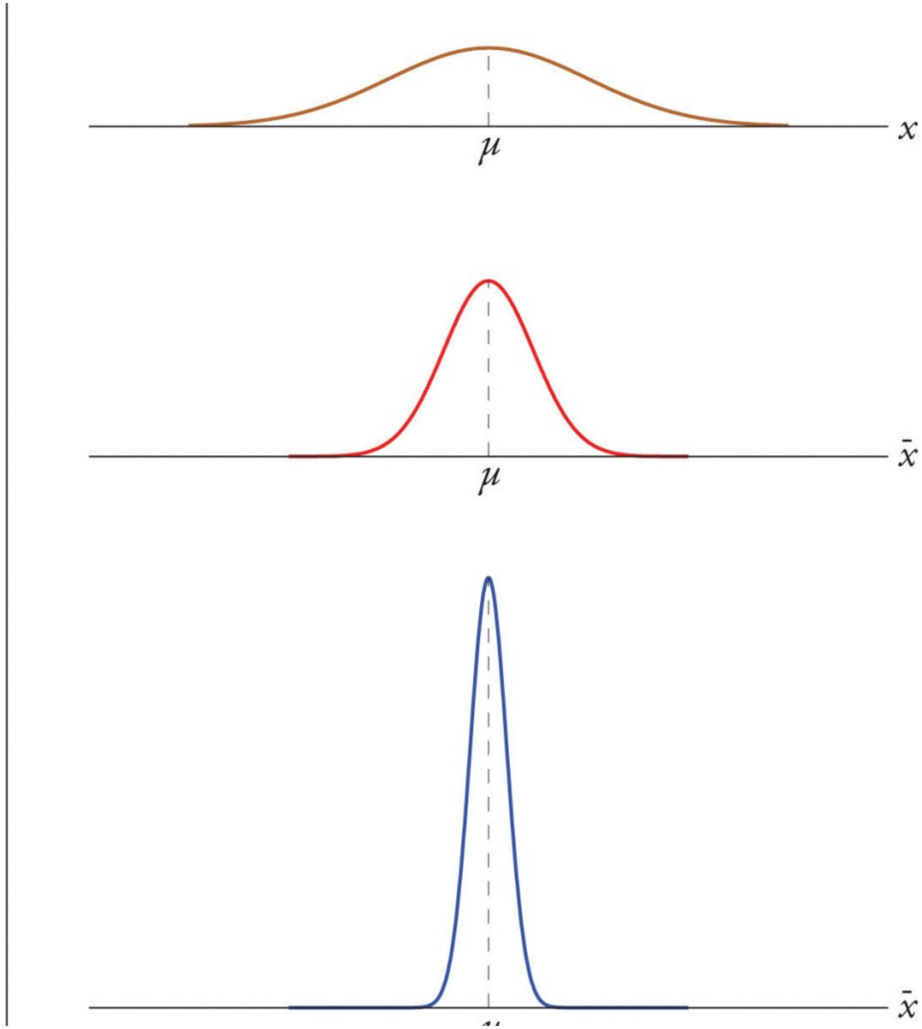
Altezza media dei 5 per ogni campione

Come si distribuiscono le medie dei 100 campioni?

Distribuzione della popolazione
x

Distribuzione delle medie campionarie con n=5

Distribuzione delle medie campionarie con n=30



Le medie di campioni estratti da una popolazione normale si distribuiscono esse stesse secondo una distribuzione normale attorno alla media della popolazione

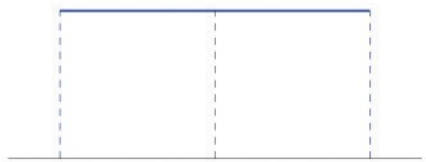
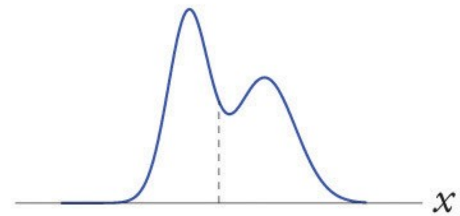
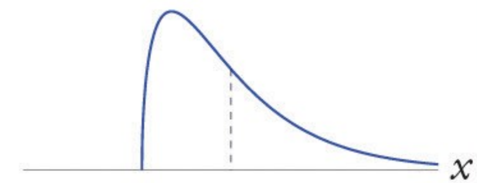
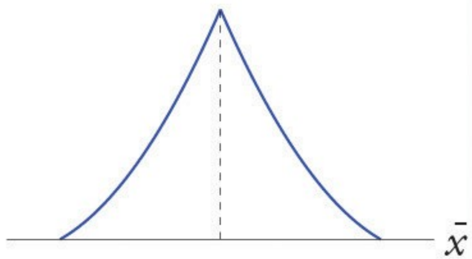
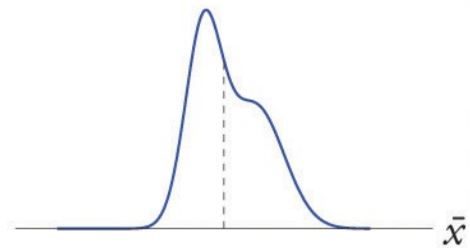
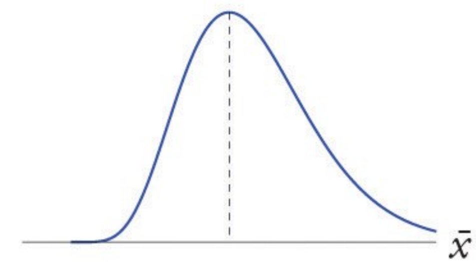
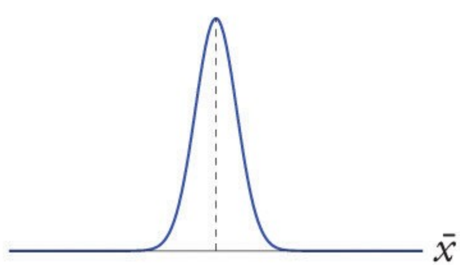
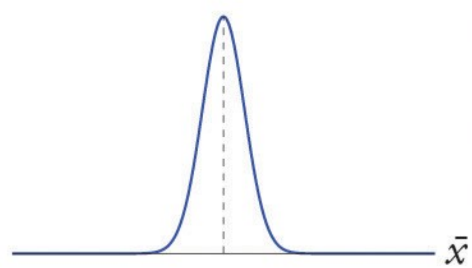
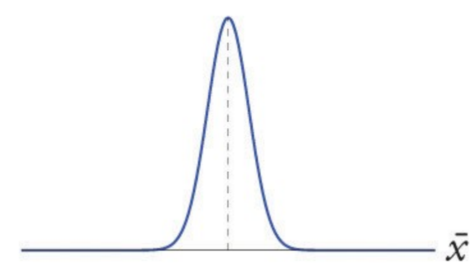
La dimensione del campione non altera questa proprietà ma la variabilità di questa distribuzione

Teorema del limite centrale

Cosa succede se la popolazione di partenza non è normale?

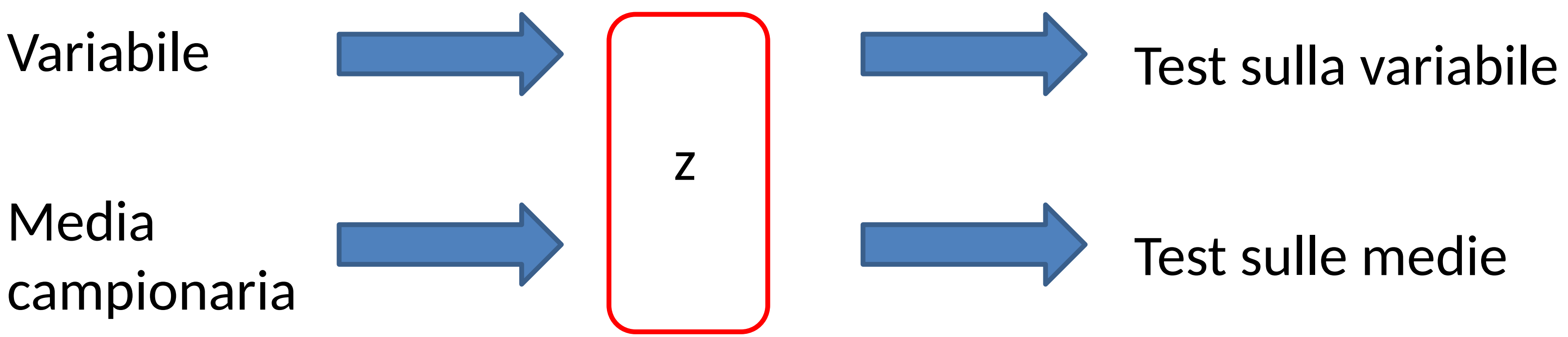
Il **teorema del limite centrale** afferma che, al crescere della dimensione del campione, **medie di misurazioni casuali** ricavate da una popolazione tendono a distribuirsi secondo **una distribuzione normale**, a prescindere dalla distribuzione di partenza

Teorema del limite centrale

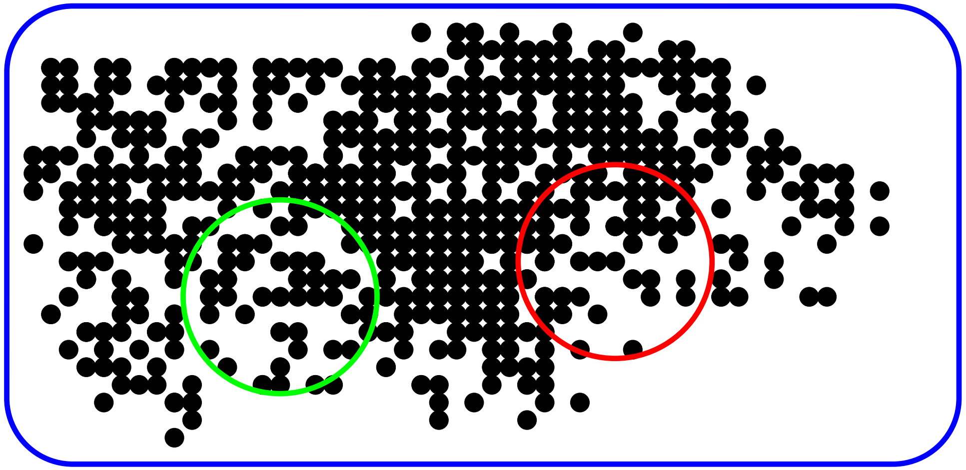
Distribuzione popolazione di partenza			
Distribuzione delle medie con n=5			
Distribuzione delle medie con n=30			

Teorema del limite centrale

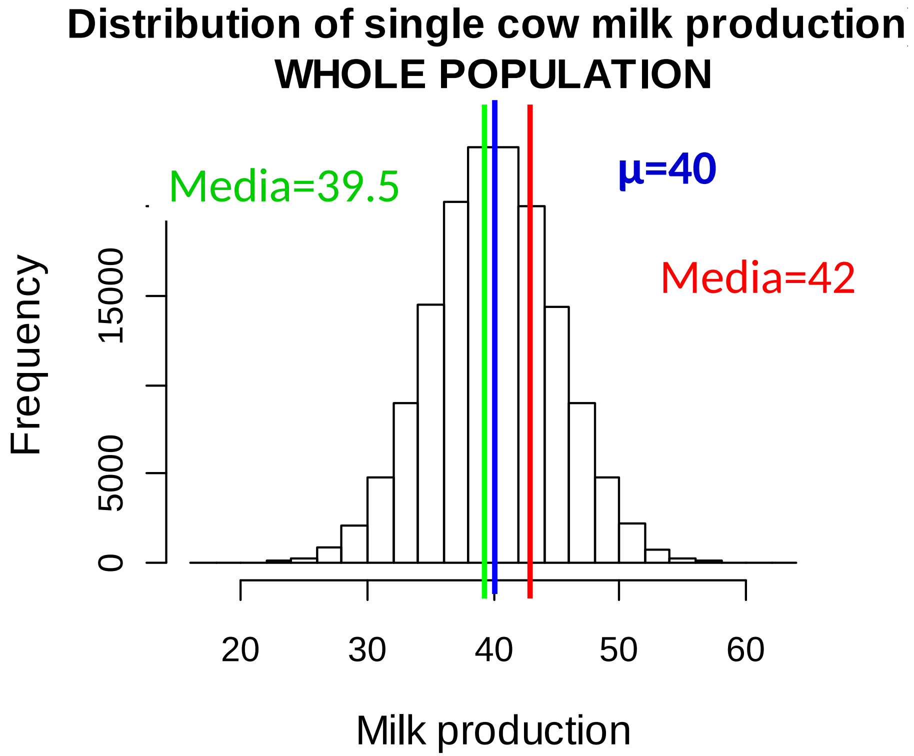
Se le medie seguono una distribuzione normale posso usare le proprietà di Z per testare ipotesi sulle medie e non più sui valori della variabile!!!



Attendibilità della stima



Di solito lavoriamo solo con uno degli infiniti campioni



Poiché lavoriamo con i campioni abbiamo sempre un certo grado di incertezza nella stima della media

Come possiamo misurare l'attendibilità della stima

Distribuzione delle medie: Errore standard

Errore standard della media (SE)

$$SE = \frac{\sigma}{\sqrt{n}}$$

è la deviazione standard della distribuzione delle medie!!!

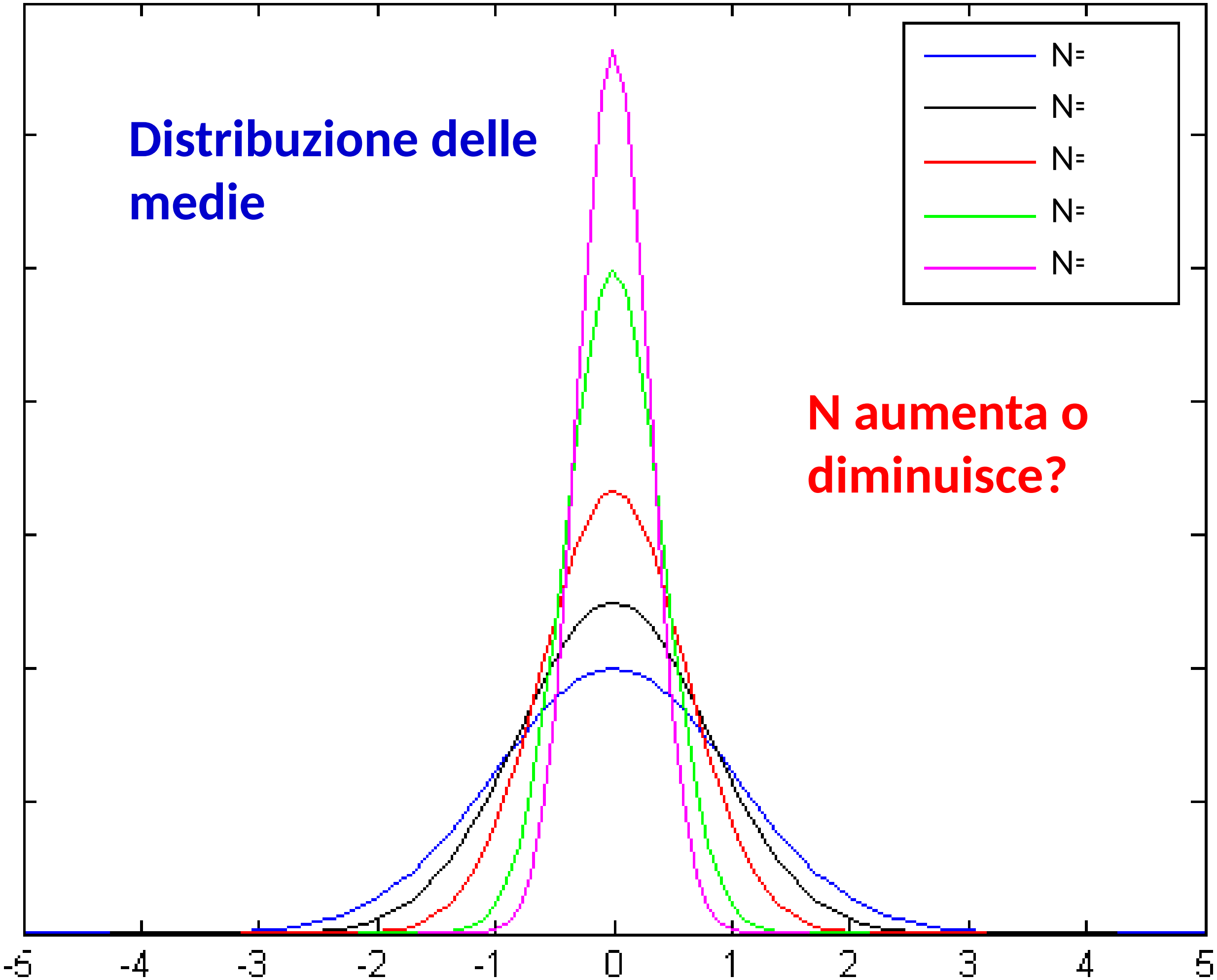
Ci serve la deviazione standard popolazione

Dimensioni del campione

Errrore standard della media (SE) con σ nota

$$SE = \frac{\sigma}{\sqrt{n}}$$

Poiché lavoriamo con i campioni abbiamo sempre un certo grado di incertezza nella stima della media

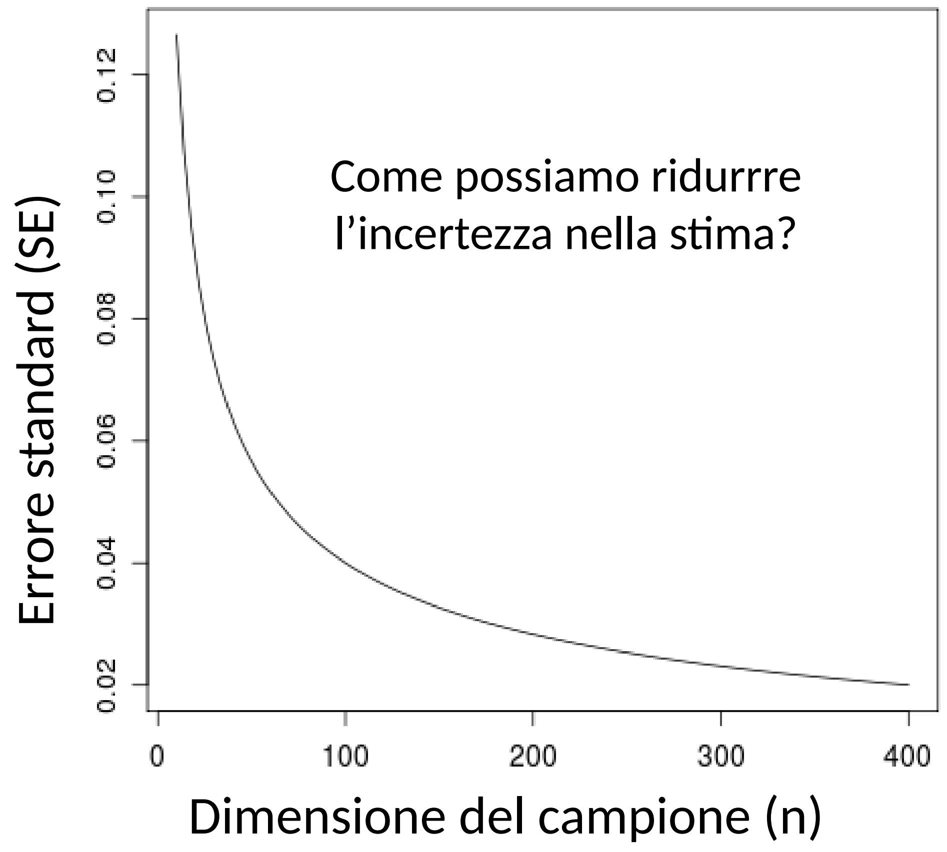


Distribuzione delle medie

N aumenta o diminuisce?

Intervalli di confidenza (o fiduciali) con σ nota

$$SE = \frac{\sigma}{\sqrt{n}}$$



Intervalli di confidenza (o fiduciali) con σ nota

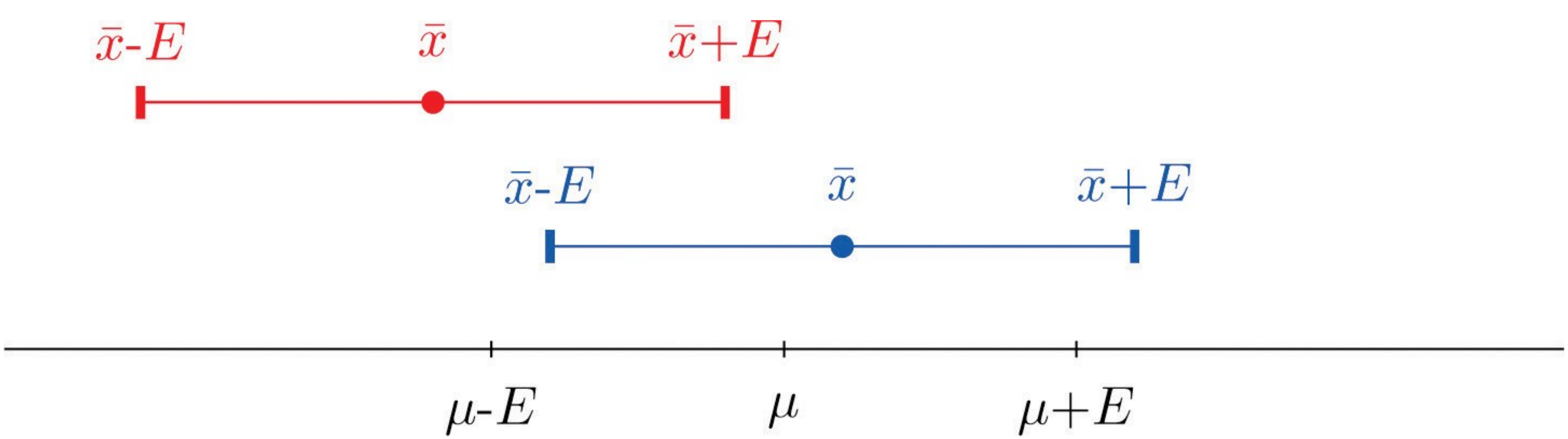
Si può definire un intervallo di valori di x attorno alla media campionaria entro cui il parametro può essere incluso e determinare una probabilità associata (e.g. 95%)

$$\bar{x} \pm \overset{E}{\boxed{z_\alpha \times \frac{s}{\sqrt{n}}}}$$

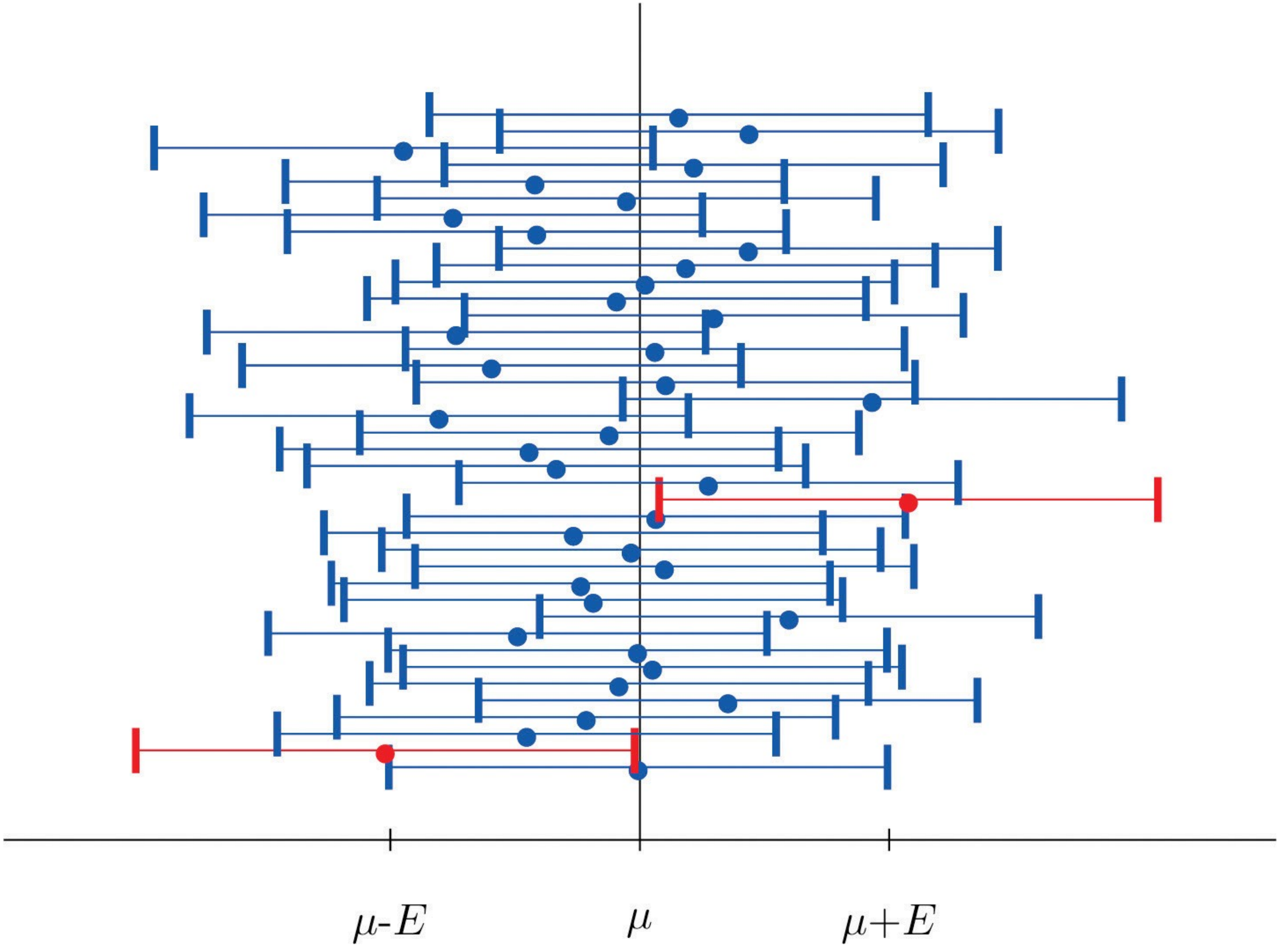
$z_{50\%}$= 0.647 contiene il 50%

$z_{95\%}$=1.960 contiene il 95%

$z_{99\%}$=2.576 contiene il 99%



Intervalli di confidenza (o fiduciali) con σ nota



Intervalli di confidenza (o fiduciali) con σ ignota

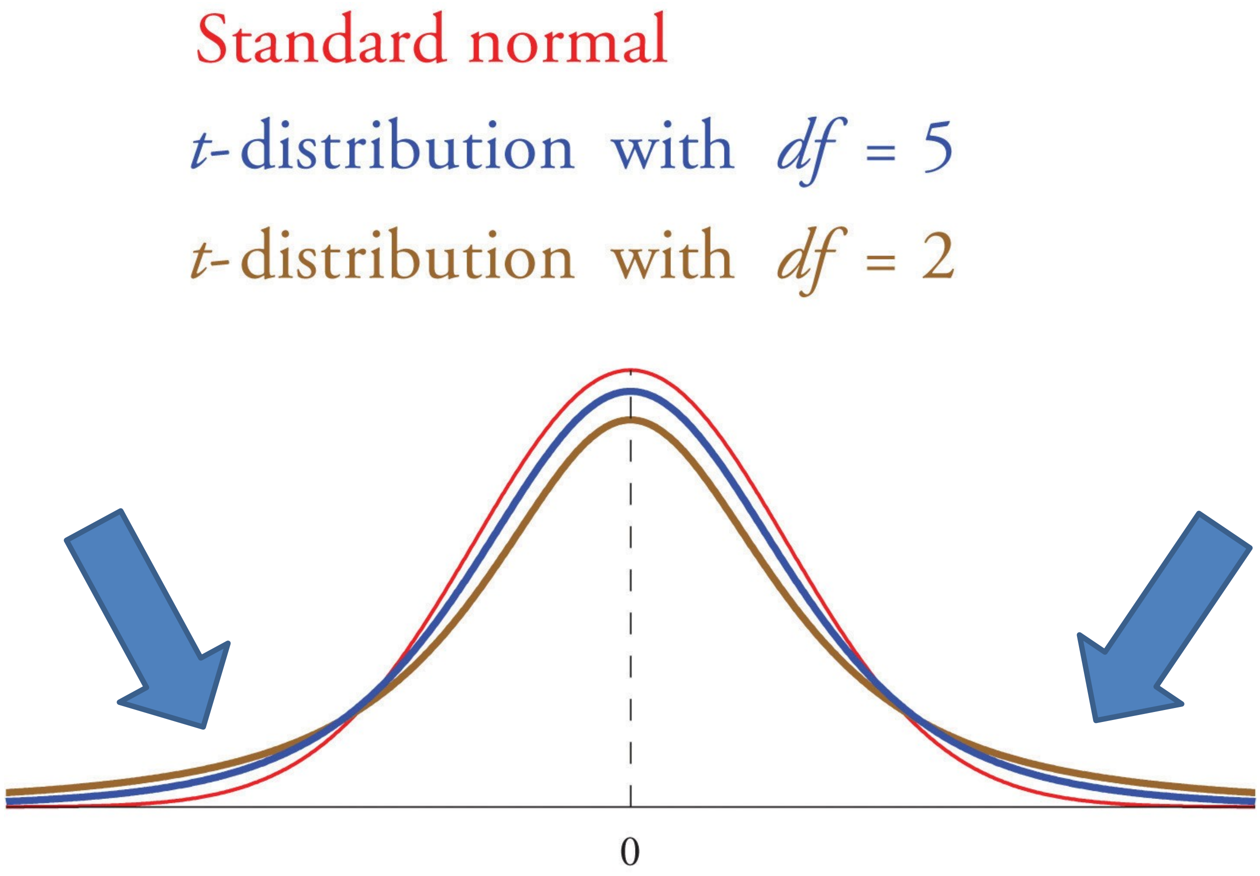
Per campioni piccoli con (e.g. $n<30$) non siamo sicuri che la distribuzione delle medie sia normale

Aumenta il grado di incertezza nella stima

Si utilizza una nuova distribuzione (t-Student) che ci permette di stimare intervalli di confidenza corretti

Per n elevati t-Student e z convergono

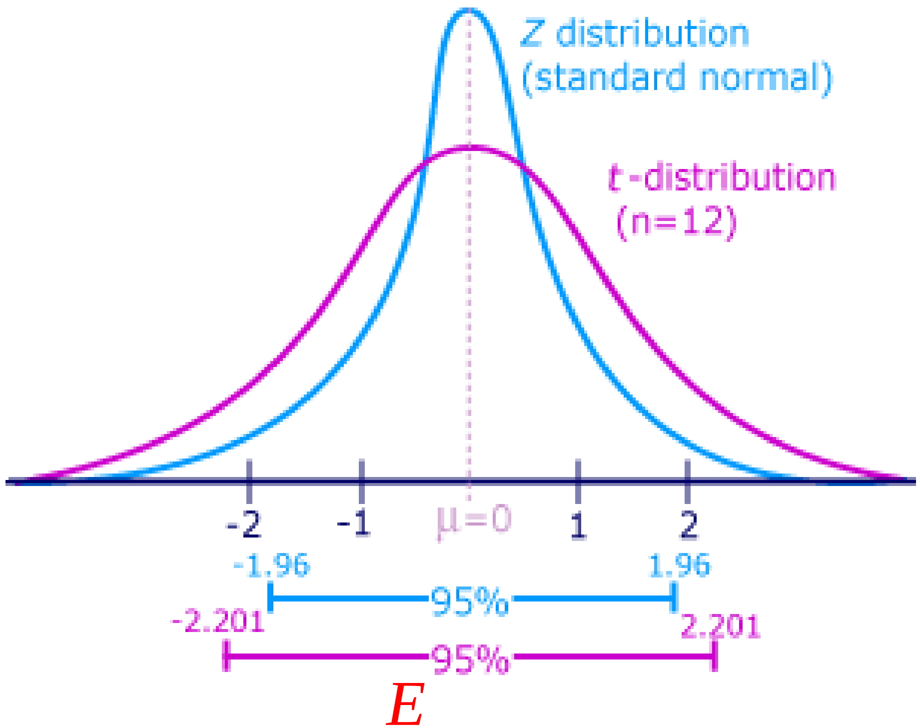
Intervalli di confidenza (o fiduciali) con σ ignota



t dipende dalla dimensione del campione z no!

Gli intervalli saranno maggiori o minori con t rispetto a z?

Intervalli di confidenza (o fiduciali) con σ ignota



$$\bar{x} \pm \boxed{t_{\alpha[n-1]} \times \frac{s}{\sqrt{n}}}$$

t dipende da n e dalla probabilità associata all'intervallo!