Distribuzioni campionarie (deduzione): Perchè e come si usano i campioni?

A cura di
Francesco Fabi

ce3s
CENTRO STUDI
STATISTICI
E SOCIALI

Perche si usano I campioni?

- Se si vuole indagare un aspetto di una popolazione, che in pratica sarà un parametro, può essere non conveniente o non possibile analizzare l’intera popolazione.
- Per cui si estrae un sottoinsieme rappresentativo di essa.

Distribuzioni campionarie

- Una distribuzione campionaria è la distribuzione di tutti i possibili valori di una statistica, funzione dei dati da rilevare (media, mediana, deviazione standard….),
- per una fissata numerosità campionaria (numero di unità di osservazione)
- al variare del campione estratto dalla popolazione di interesse

Esempio

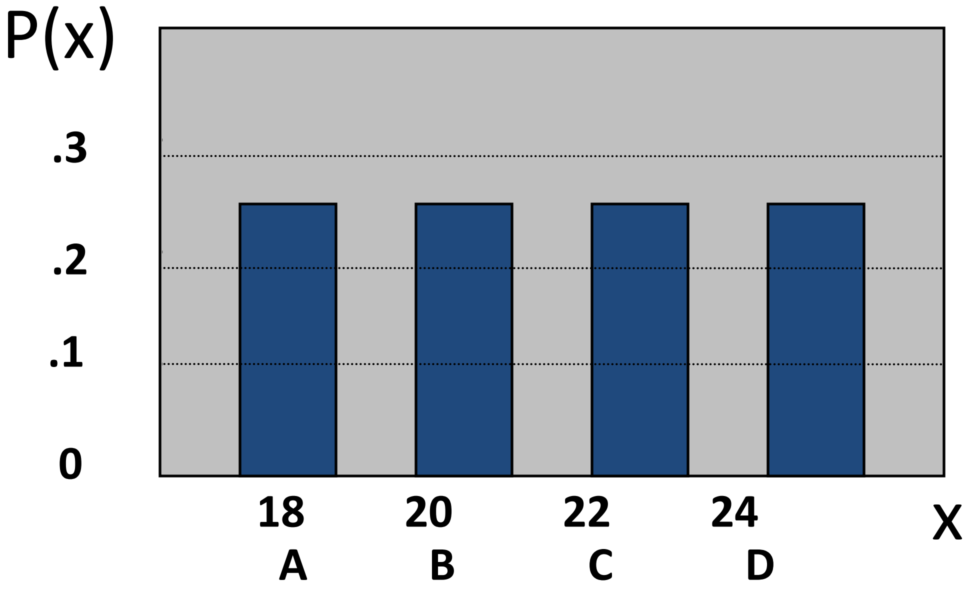
- **Consideriamo una popolazione …**
- Numerosità della popolazione N=4
- La variabile di interesse, X, è l'età dei soggetti
- Valori di X: 18, 20, 22, 24 (anni)

I parametri della popolazione

$$\mu = \frac{\sum X_i}{N}$$

$$= \frac{18 + 20 + 22 + 24}{4} = 21$$

$$\sigma = \sqrt{\frac{\sum (X_i - \mu)^2}{N}} = 2.236$$



Distribuzione uniforme di estrazione dei soggetti

Campionamento con ripetizione (n=2)

1st Obs	**2nd Observation**			
	18	20	22	24
18	18,18	18,20	18,22	18,24
20	20,18	20,20	20,22	20,24
22	22,18	22,20	22,22	22,24
24	24,18	24,20	24,22	24,24

16 campioni possibili (campionamento con reimmissione)

16 medie campionarie

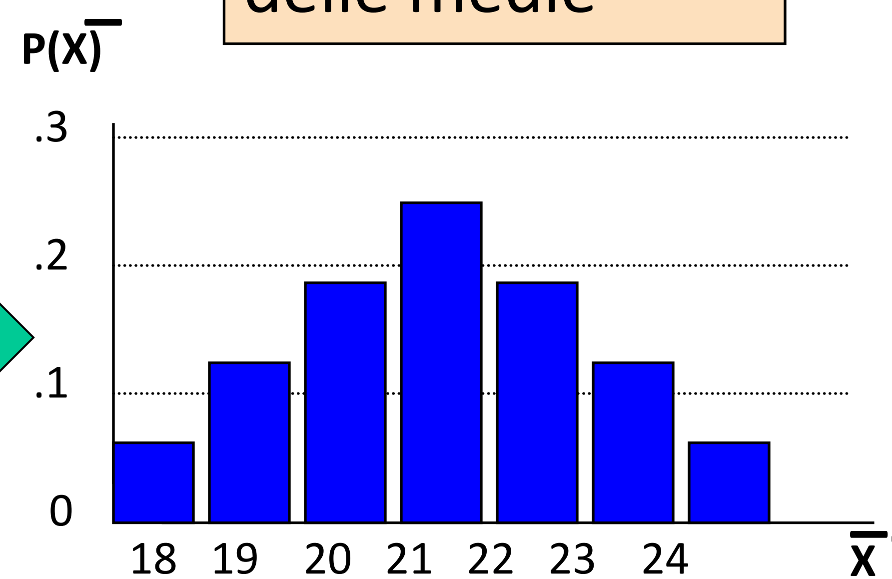
1st Obs	**2nd Observation**			
	18	**20**	**22**	**24**
18	18	19	20	21
20	19	20	21	22
22	20	21	22	23
24	21	22	23	24

Distribuzione campionaria delle medie

16 medie campionarie

1st Obs	2nd Observation			
	18	**20**	**22**	**24**
18	18	19	20	21
20	19	20	21	22
22	20	21	22	23
24	21	22	23	24

Distribuzione campionaria delle medie



Non più uniforme

Indici di posizione e di dispersione della distribuzione campionaria

$$\mu_{\bar{X}} = \frac{\sum \bar{X}_i}{N} = \frac{18 + 19 + 21 + \cdots + 24}{16} = 21$$

$$\sigma_{\bar{X}} = \sqrt{\frac{\sum (\bar{X}_i - \mu_{\bar{X}})^2}{N - 1}}$$

$$= \sqrt{\frac{(18 - 21)^2 + (19 - 21)^2 + \cdots + (24 - 21)^2}{15}} = 1.69$$

Confronti

Popolazione

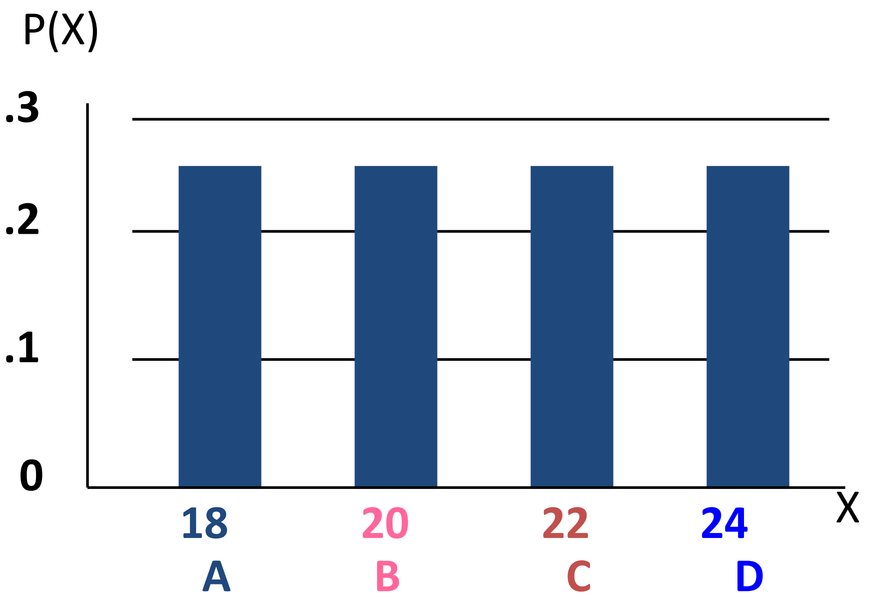
N = 4

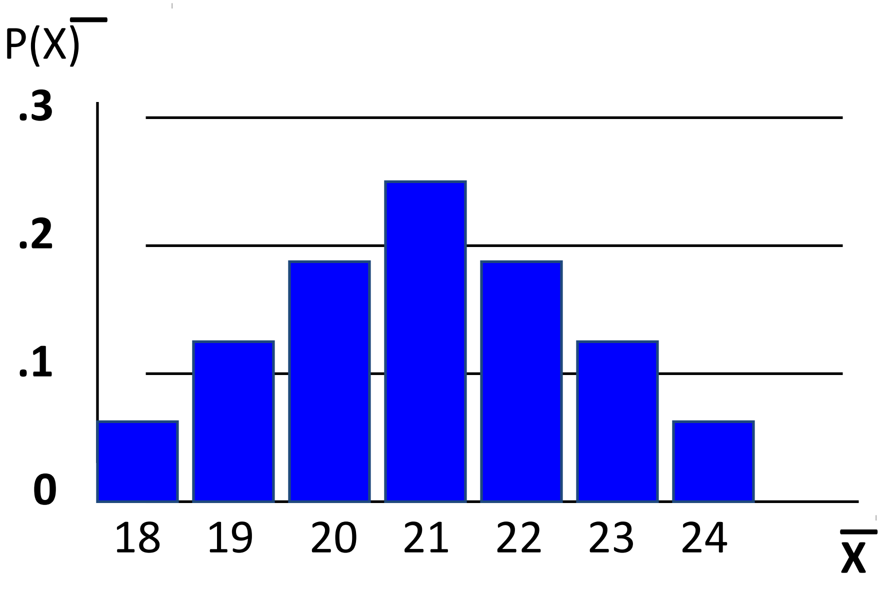
$$\mu = 21 \qquad \sigma = 2.236$$

Distribuzione delle medie campionarie

n = 2

$$\mu_{\overline{X}} = 21 \qquad \sigma_{\overline{X}} = 1.69$$





Errore standard della media

- Diversi campioni con stessa numerosità campionaria danno luogo a diverse medie campionarie.
- Una misura della variabilità della media tra campioni diversi è l'errore standard della media dato da:

$$\boxed{\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}}$$

- L'errore standard della media diminuisce al crescere della numerosità campionaria.

Se la popolazione è normale

- Se la popolazione è normale, con media μ e scarto standard σ, la distribuzione campionaria della media sarà ancora normale con:

e

$$\boxed{\mu_{\bar{X}} = \mu}$$

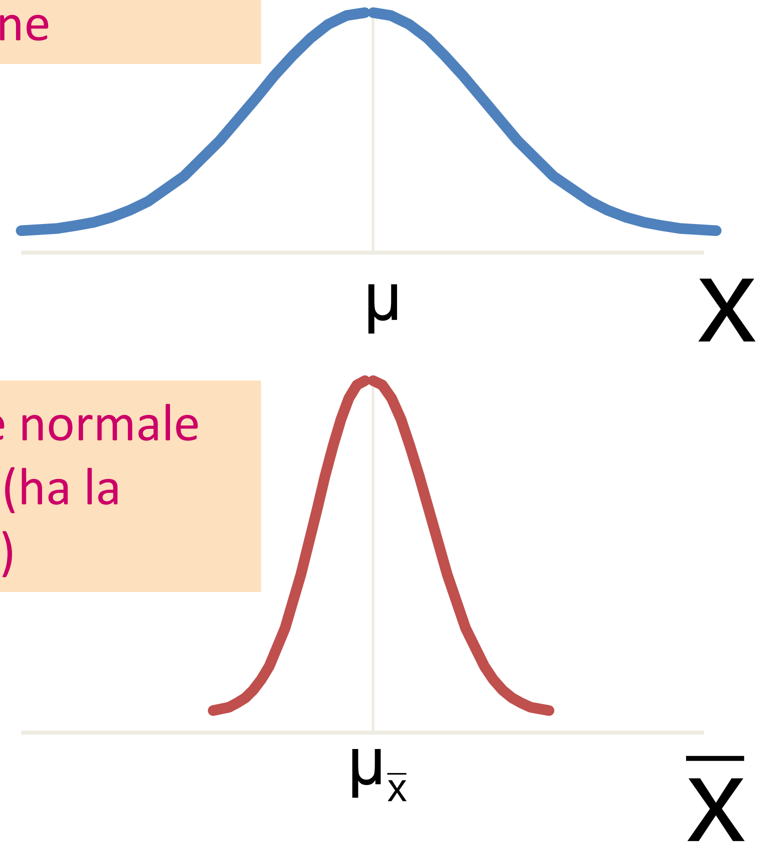
$$\boxed{\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}}$$

Proprietà

$$\boxed{\mu_{\bar{x}} = \mu}$$

Distribuzione normale della popolazione

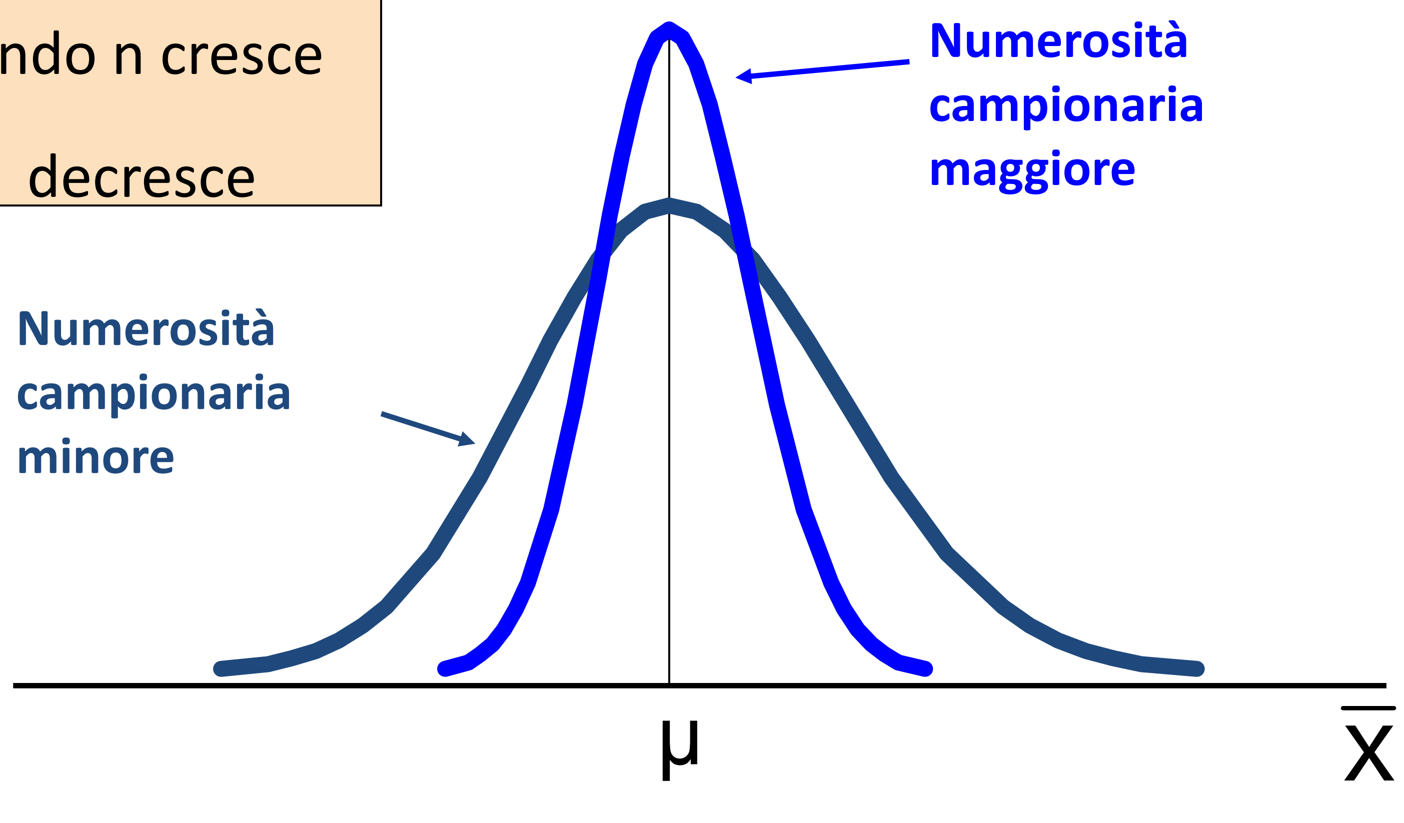
Distribuzione normale campionaria (ha la stessa media)



Proprietà

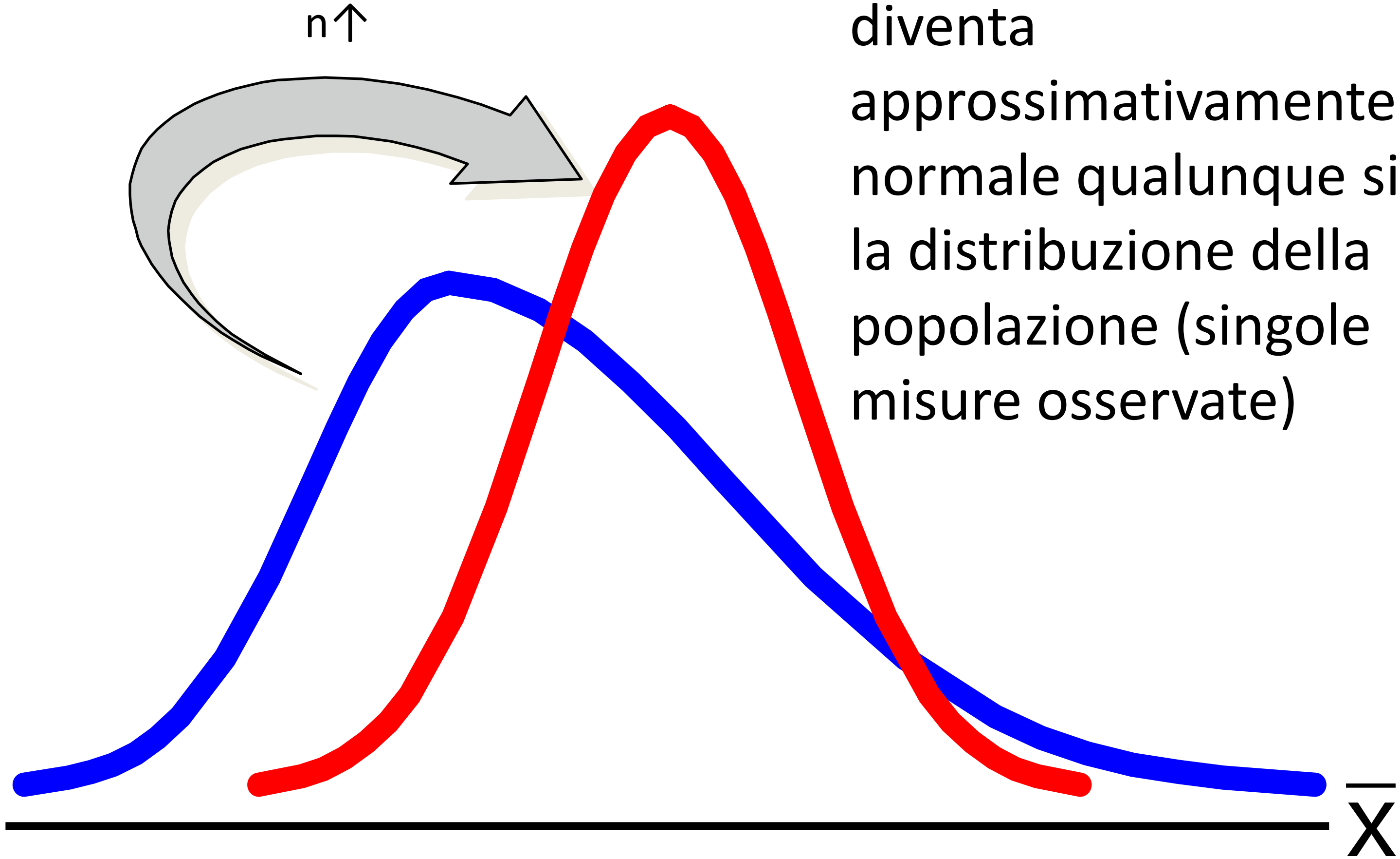
Quando n cresce

$\sigma_{\bar{x}}$ decresce



Il Teorema Centrale

Quando la numerosità campionaria cresce…



la distribuzione della media campionaria diventa approssimativamente normale qualunque sia la distribuzione della popolazione (singole misure osservate)

Popolazione non normale

- Si può applicare il teorema centrale:
  - Anche se la popolazione non è normale,
  - …le medie saranno approssimativamente normali se la numerosità campionaria è abbastanza elevata.
  - La distribuzione campionaria ha parametri:

$$\boxed{\mu_{\bar{x}} = \mu}$$

$$\boxed{\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}}$$

Popolazione non normale

Proprietà della distribuzione campionaria:

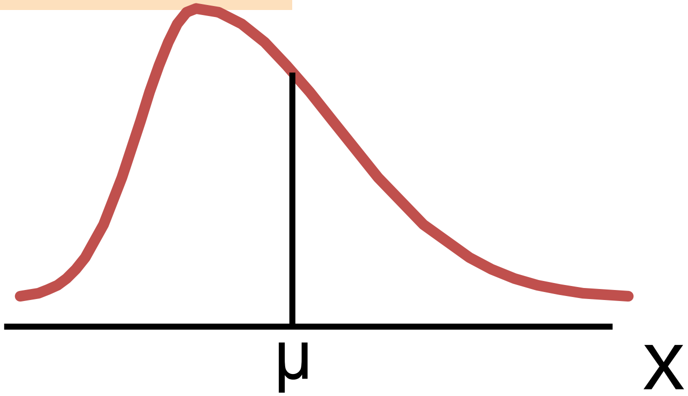
media

$$\mu_{\bar{x}} = \mu$$

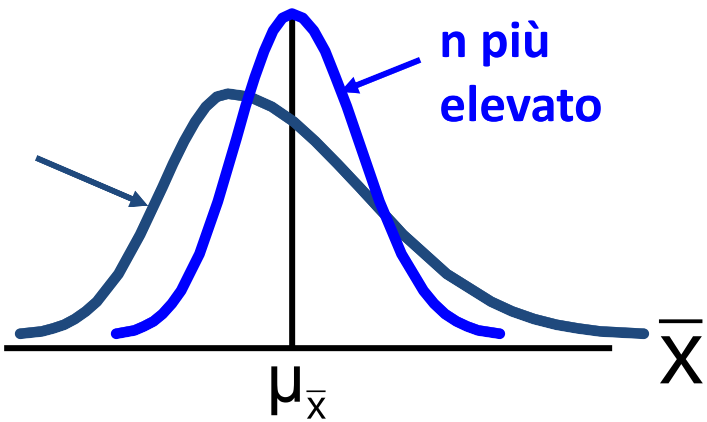
Scarto standard

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

Distribuzione della X



Distribuzione di $\bar{X}$



Quanto deve essere grande n?

- Nella maggior parte delle situazioni è sufficiente che n sia maggiore di 30
- Per distribuzioni simmetriche, $n > 15$
- Per popolazioni normali la distribuzione della media è sempre normale