

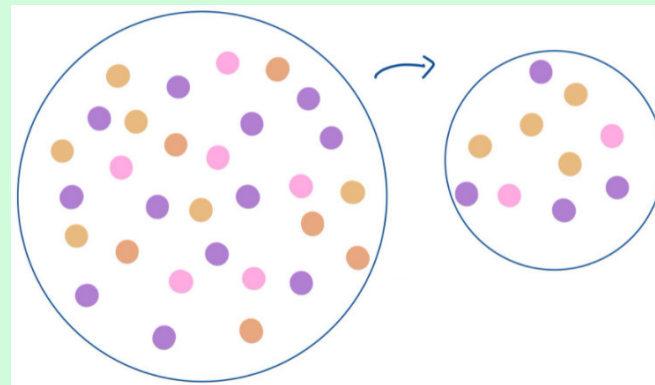
# Campionamento

Con il termine **popolazione** in statistica si intende un insieme finito o infinito di tutte le unità statistiche di cui si vuole indagare una certa caratteristica che le individua come omogenee.

Per determinare le caratteristiche fondamentali di una popolazione statistica non è necessario analizzare tutte le unità statistiche della popolazione d'interesse, ma è sufficiente una parte di esse



**CAMPIONE**



# Statistica inferenziale

**Statistica inferenziale** è volta all'induzione probabilistica delle caratteristiche ignote della popolazione.

Osservazioni svolte su un campione di unità rappresentative di tutta la popolazione, selezionate con date procedure, entro dati livelli di errore consentono di ottenere conclusioni che possono essere generalizzate all'intera popolazione.

Le 4 fasi principali sono:

- 1- estrazione di un campione della popolazione
- 2- calcolo delle statistiche campionarie, cioè dei valori di sintesi relativi ai dati del campione
- 3- stima dei parametri della popolazione in base ai risultati del campione
- 4- verifica dei risultati raggiunti

# Rappresentatività del campione

Il procedimento di **S. I.** conduce a risultati esatti solo se il campione è perfettamente **rappresentativo** della popolazione.

Significa che il campione dovrebbe rispecchiare e riprodurre le caratteristiche essenziali e la stessa distribuzione della popolazione.

La **rappresentatività** è garantita dalla **casualità** della selezione delle unità statistiche del campione.

# Campioni probabilistici

I campioni probabilistici sono caratterizzati dalla **casualità**, ciascuna unità della popolazione ha la stessa probabilità, diversa da zero, di essere estratta.

Consentono inferenza, ossia la **generalizzazione** dei risultati ottenuti alla popolazione intera, con scarti non significativi imputabili al caso.

Se estraiamo un campione di studenti a sorte tra quelli presenti in qualsiasi giorno in università non è un campione probabilistico:

- I non frequentanti hanno probabilità nulla di essere estratti
- Le matricole hanno una probabilità più alta di essere estratte
- Gli studenti fuori corso hanno una minore probabilità di essere estratti.



# Statistica campionaria

Per **parametro** della popolazione si intende quel valore numerico utilizzato come misura di una delle caratteristiche della popolazione di riferimento (ad es. la media, la varianza etc.).

Nell'inferenza statistica la **statistica campionaria** rappresenta lo stimatore puntuale del corrispondente parametro della popolazione.

La **stima puntuale** è il valore di uno stimatore puntuale utilizzato per stimare un parametro della popolazione.

# Statistica campionaria

	Parametro della popolazione	Statistica campionaria
<b>Media</b>	$\mu = \frac{\sum_{i=1}^N x_i}{N}$	$\bar{x} = \frac{\sum_{i=1}^n x_i}{n}$
<b>Varianza</b>	$\sigma^2 = \frac{\sum_{i=1}^N (x_i - \mu)^2}{N}$	$S^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n - 1}$
<b>Deviazione standard</b>	$\sigma = \sqrt{\frac{\sum_{i=1}^N (x_i - \mu)^2}{N}}$	$S = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n - 1}}$

La **statistica campionaria** è una caratteristica del campione come la media campionaria, la deviazione standard campionaria etc.

# Tipo di campionamento

**Campionamento casuale semplice:** è la procedura di scelta casuale più semplice. Ogni unità ha la stessa probabilità di far parte del campione.

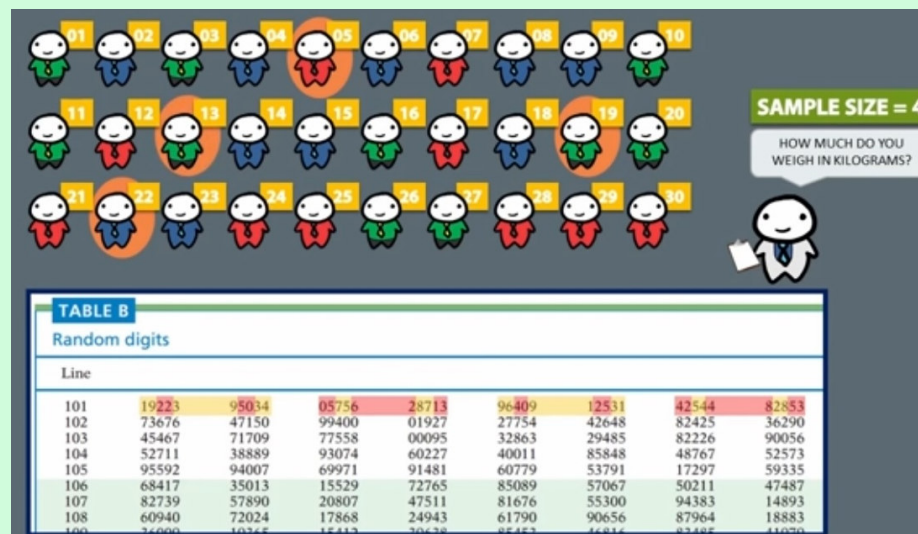
La casualità viene ottenuta estraendo numeri a partire da un elenco (detto "lista di campionamento") in cui sono presenti tutti gli individui della popolazione da studiare.

Il campionamento per randomizzazione semplice viene agevolmente applicato quando si dispone di una popolazione già numerata, preferibilmente composta di un numero non elevato di unità.

In questa procedura di selezione casuale si distinguono due modalità di estrazione dei campioni: con ripetizione (o bernoulliani) e senza ripetizione (o in blocco); a seconda che vi sia reimmissione o no delle unità estratte.

# Tavole dei numeri casuali

Esistono delle apposite tavole costituite da un insieme di numeri ricavati mediante algoritmi matematici, in modo che nel lungo andare ogni cifra, ogni coppia di cifre, ogni terna, ecc. abbia la stessa frequenza di ogni altra. Si individua casualmente un punto di partenza dal quale procedere ordinatamente per riga o per colonna.



**SAMPLE SIZE = 4**

HOW MUCH DO YOU WEIGH IN KILOGRAMS?

**TABLE B**  
Random digits

Line								
101	19223	95034	05756	24713	96409	12531	42544	82853
102	73676	47150	99400	01927	27754	42648	82425	36290
103	45467	71709	77558	00095	32863	29485	82226	90056
104	52711	38889	93074	60227	40011	85848	48767	52573
105	95592	94007	69971	91481	60779	53791	17297	59335
106	68417	35013	15529	72765	85089	57067	50211	47487
107	82739	57890	20807	47511	81676	55300	94383	14893
108	60940	72024	17868	24943	61790	90656	87964	18883



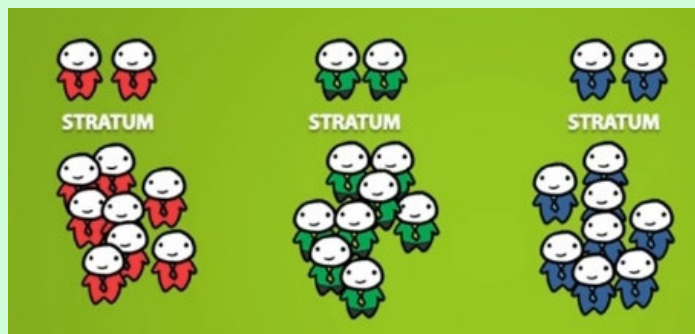
# Tipo di campionamento

**Campionamento stratificato:** è una procedura di campionamento con la quale inizialmente si procede nella suddivisione della popolazione in un numero determinato di strati o classi il più possibile omogenei al loro interno, rispetto al carattere indagato e successivamente nell'estrazione di un campione casuale semplice di numerosità prefissata da ciascuno strato.



**Campione estratto**

**Popolazione divisa in strati**



# Tipo di campionamento

**Campionamento stratificato:** la base per la creazione degli strati è a discrezione del soggetto che disegna il campione.

Es. area geografica, età, settore di attività economica.

Il campionamento stratificato funziona meglio quando la varianza tra gli elementi di ciascuno strato è relativamente piccola.

L'efficienza dipende, quindi, da quanto sono omogenei gli elementi all'interno di ciascuno strato.

Se gli strati sono omogenei, il campionamento stratificato fornisce risultati precisi quanto il campionamento casuale semplice, ma utilizzando una dimensione campionaria più piccola.

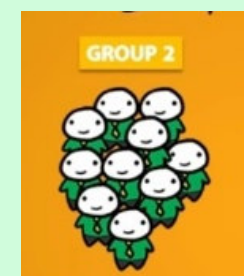
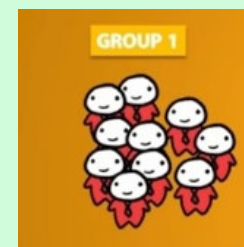
# Tipo di campionamento

**Campionamento multistadio:** è una procedura di campionamento composito che presuppone l'individuazione di una struttura gerarchica della popolazione, in cui le unità finali sono incluse in insiemi di livello via via più elevato.

Esempio

1. estrazione casuale di un campione di comuni (unità di primo stadio)

2. estrazione di un campione casuale di famiglie (unità di secondo stadio) da ciascuna lista anagrafica per ogni comune selezionato

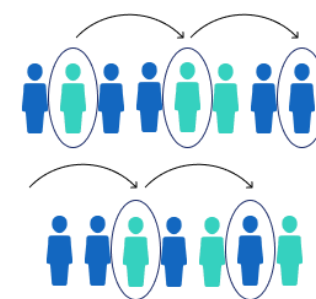


# Tipo di campionamento

**Campionamento sistematico:** differisce dal campionamento casuale semplice soprattutto dal punto di vista della tecnica di estrazione dei soggetti: si estraggono unità da una lista a intervalli regolari nella stessa. La prima unità va scelta casualmente poi si sceglie un'unità ogni intervallo  $k=N/n$ .

Si seleziona un punto di partenza casuale dalla popolazione e poi si procede selezionando le altre unità ad es. con una progressione aritmetica di ragione 15 fino all'esaurimento della lista.

Il campionamento sistematico è stato ideato per ridurre il lavoro sulle tavole dei numeri casuali ed è tutt'oggi ancora molto utilizzato.



# Campionamento non probabilistico

**Campionamento di convenienza:** è una tecnica di campionamento non probabilistico, in cui gli elementi sono selezionati in base alla convenienza.

Es. le interviste volontarie per ricerche di mercato

Hanno il vantaggio che la selezione campionaria e la raccolta dei dati sono relativamente facili.

Tuttavia è impossibile valutare il campione in termini di rappresentatività della popolazione, per cui bisogna prestare molta attenzione nell'interpretazione dei risultati.

La **distorsione del campionamento (bias di selezione)** determina una differenza costante tra i risultati del campione e i risultati teorici dell'intera popolazione.

# Margine di errore

La stima del parametro è probabilistica, essa comporta, cioè, un errore dovuto all'impossibilità di determinare con esattezza il parametro.

Il margine d'errore nell'inferenza statistica è fisiologico trattandosi di stime.

L'effettiva estrazione di tutti i campioni possibili (di pari ampiezza) da una popolazione è l'unico modo per capire quanto le statistiche calcolate su un campione possano discostarsi dai dati ricavati dall'analisi di tutti gli altri campioni.

Per evitare tale operazione esistono degli strumenti in grado di valutare la robustezza delle stime effettuate.

# Distribuzione della media campionaria

La **media campionaria** è la variabile aleatoria  $X_n$  che descrive le medie di tutti i possibili campioni di ampiezza  $n$  che si possono estrarre dalla popolazione. Ha una propria distribuzione, dispersione etc.

Se si hanno a disposizione i valori medi di tutti i campioni possibili della stessa popolazione di riferimento si può generare la **distribuzione della media campionaria**, ossia quella distribuzione che raccoglie tutti i valori possibili che la media campionaria può assumere nei vari campioni estratti da una popolazione.

Conoscere i valori più frequenti della media campionaria vuol dire anche sapere quali sono i più probabili.

# Distribuzione della media campionaria

## Medie dei campioni estratti

935, 867, 743, 654, 194, 234, 236, 704, 560, 350, 353, 760, 705, 413, 520, 712, 340, 342, 813, 403, 204, 378,	$X_1 = \dots$
800, 788, 675, 456, 657, 702, 456, 412, 506, 675, 645, 430, 506, 348, 344, 514, 488, 605, 320, 500, 501, 653,	$X_2 = \dots$
621, 704, 706, 566, 534, 390, 400, 333, 475, 506, 721,	$X_3 = \dots$
711, 509, 430, 322, 421	$\cdot$
	$\cdot$
	$X_k = \dots$

Il valore medio della distribuzione della media campionaria è dato dalla media aritmetica delle medie di ogni campione.

Si sommano tutte le medie e si divide per il numero di osservazioni, ossia i campioni. Si ottiene così la media della distribuzione.



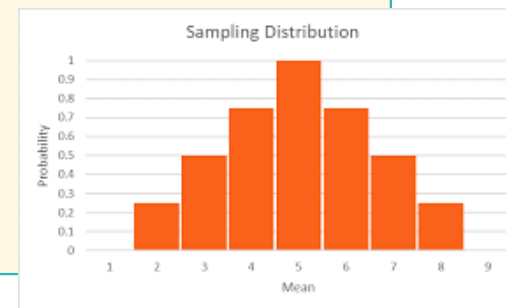
# Distribuzione campionaria

La **distribuzione campionaria** è una distribuzione di probabilità che consiste di tutti i valori possibili che una statistica campionaria può assumere e delle probabilità associate di ciascun valore.

Poiché campioni casuali differenti forniscono valori differenti per gli stimatori puntuali, questi ultimi sono considerati variabili casuali.

Tale distribuzione è generata teoricamente prendendo infiniti campioni di dimensione  $n$  e calcolando i valori della statistica per ogni campione.

Le più utilizzate sono le distribuzioni della media campionaria e della varianza campionaria.



# Esempio

Popolazione di 4 unità con peso medio 73,25kg.

Unità della popolazione	Peso (kg)
A	78
B	65
C	68
D	82

# Esempio

Consideriamo tutti i campioni di 2 unità estraibili e calcoliamo la media per ciascun campione:

Campioni	Media campionaria
AB	$(78+65/2)= 71,5$
AC	$(78+68/2)= 73$
AD	$(78+82/2)= 80$
BA	71,5
BC	66,5
BD	73,5
CA	73
CB	66,5
CD	75
DA	80
DB	73,5
DC	75

Media delle medie campionarie = 73,25 (nessuno dei campioni estratti presenta esattamente la media di 73,25).

La media delle medie campionarie è proprio uguale alla media della popolazione, quindi la media campionaria è uno stimatore non distorto della media della popolazione.

# Errore standard

Occorre considerare la volatilità della media campionaria. Per farlo bisogna ricorrere alla deviazione standard che in caso di distribuzioni campionarie si chiama **errore standard**.

L'errore standard (o scarto quadratico medio della media campionaria) indica lo scostamento medio delle singole medie campionarie rispetto al valore medio della popolazione, quindi la variabilità dalla media della popolazione. È una stima di quanto la media del campione si avvicini alla media della popolazione.

È una misura dell'errore che ci si può aspettare quando scegliamo un campione di una certa ampiezza dalla popolazione.

# Calcolo dell'errore standard

L'errore standard dei valori della distribuzione della media campionaria è uguale alla deviazione standard della popolazione divisa per la radice quadrata dell'ampiezza del campione:

$$SE(\bar{x}) = \frac{\sigma}{\sqrt{n}}$$

n = numerosità  
del campione

In questo caso  $6,97/\sqrt{2} = 4,93$

## Errore standard nel caso di $\sigma$ non nota

Nel caso in cui non sia nota la deviazione standard della popolazione si può utilizzare la deviazione standard del campione.

$$SE(\bar{x}) = \frac{s}{\sqrt{n}}$$



$$s = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n - 1}}$$

# Deviazione standard del campione

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

$$S = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n - 1}}$$

**Bessel's correction**

In un campione la dispersione intorno alla media è minore perché è composto da meno elementi rispetto alla popolazione.

Pertanto, la dispersione statistica intorno al valore medio nel campione è naturalmente inferiore rispetto a quella dell'intera popolazione. Per attenuare questo effetto, nel caso dei campioni viene usata la formula della deviazione standard campionaria.

# Esempio

$x_i$	$x_i - \bar{x}$	$(x_i - \bar{x})^2$
11	0,74	0,55
8,5	-1,76	3,10
7,8	-2,46	6,05
7,4	-2,86	8,18
11,4	1,14	1,30
9	-1,26	1,59
9,5	-0,76	0,58
13,2	2,94	8,64
14,3	4,04	16,32
10,5	0,24	0,06
$\bar{x}$ 10,3		46,39

Campione di  
dimensione  
 $n = 10$

$$S = \left( \frac{46,39}{10-1} \right)^{(1/2)} = 2,27$$

Se si calcolasse la dispersione usando la formula della deviazione standard per l'intera popolazione la variabilità risulterebbe inferiore e pari a 2,15.



# Errore standard

L'entità dell'errore standard dipende da:

1- dimensione del campione.

Maggiore è l'ampiezza, minore è l'errore standard. Un campione più ampio che si avvicina alla numerosità della popolazione determina una stima più precisa perché si basa su un numero maggiore di osservazioni.

2- deviazione standard della popolazione. Maggiore è la volatilità della distribuzione dei dati nella popolazione, maggiore sarà l'errore standard.

# Errore standard

Se la deviazione standard della popolazione esprime quanto variano i singoli valori rispetto alla media, l'errore standard indica la variazione della media campionaria tra i diversi campioni.

La dispersione della media campionaria è inferiore rispetto a quella della popolazione, perché la media attenua l'effetto dei valori estremi che possono essere presenti nel campione.

Mentre le osservazioni nella popolazione assumono anche valori estremamente piccoli o estremamente grandi, la media campionaria è caratterizzata da una minore variabilità rispetto ai dati originali. Le medie campionarie saranno quindi caratterizzate, in generale, da valori meno dispersi rispetto a quelli che si osservano nella popolazione.

# Deviazione standard $\neq$ errore standard

La **deviazione standard** descrive la variabilità di una serie di misure effettuate su un campione o una popolazione.

L'**errore standard** descrive l'incertezza nella stima di un valore statistico (es. media, proporzione ecc.).

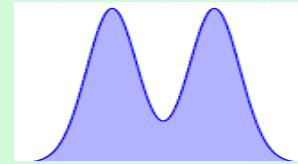
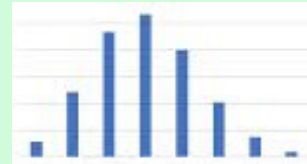
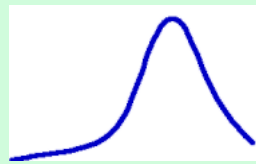
L'errore standard viene utilizzato per misurare l'accuratezza statistica di una stima. Es. l'errore standard fornisce l'accuratezza di una media campionaria misurando la variabilità da campione a campione delle medie campionarie. Descrive, quindi quanto sia precisa la media del campione come stima della media reale della popolazione.

# Teorema del limite centrale

Il **teorema del limite centrale** afferma che quando l'ampiezza del campione casuale diventa sufficientemente grande ( $n > 30$ ), la distribuzione delle medie campionarie può essere approssimata dalla distribuzione di probabilità normale con media  $\mu$  e deviazione standard  $\sigma/\sqrt{n}$ .

$$\bar{X} \sim N(\mu, \sigma/\sqrt{n})$$

Ciò è possibile indipendentemente dalla forma della distribuzione dei singoli valori della popolazione.



# Teorema del limite centrale

Se la popolazione d'interesse è distribuita come una normale

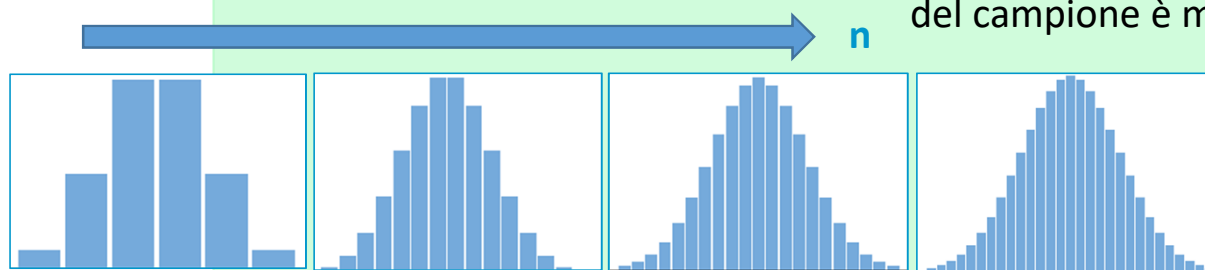
La distribuzione campionaria è distribuita come una normale indipendentemente dall'ampiezza del campione



$$N(\mu, \sigma / \sqrt{n})$$

Se la popolazione non è distribuita come una normale

La distribuzione campionaria è distribuita **approssimativamente** come una normale, se l'ampiezza del campione è maggiore di 30.



$$\sim N(\mu, \sigma / \sqrt{n})$$

# Teorema del limite centrale

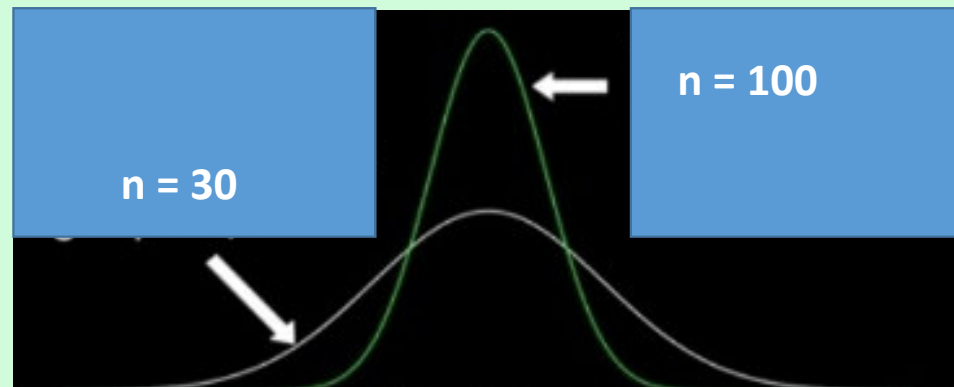
Sia data una popolazione avente media  $\mu$  e varianza  $\sigma^2$ , e da essa si estraggano campioni casuali di ampiezza  $n$ , indicando con  $\bar{x}$  la media campionaria, la variabile

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

è una variabile aleatoria la cui distribuzione tende alla distribuzione normale standardizzata per  $n \longrightarrow \infty$

# Relazione tra la dimensione del campione e la distribuzione campionaria di $\bar{x}$

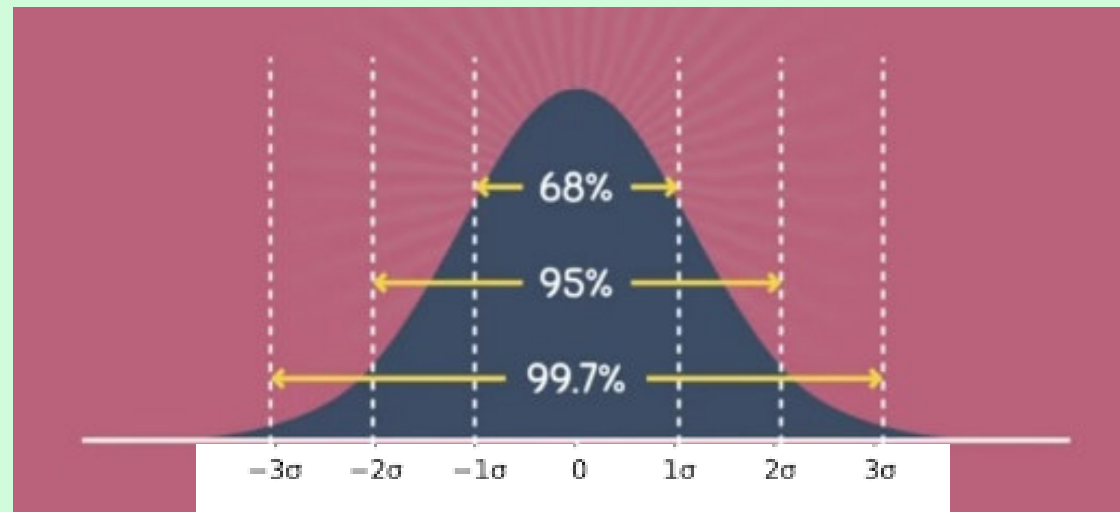
Poiché la distribuzione campionaria con ad es.  $n = 100$  ha un errore standard più piccolo, i valori di  $\bar{x}$  hanno una minore variabilità e tendono ad essere più vicini alla media della popolazione rispetto ad  $\bar{x}$  con  $n = 30$ .



# Teorema del limite centrale

Sapendo che la distribuzione della media campionaria è normale possiamo utilizzare la **regola empirica** della distribuzione normale.

Si può calcolare una probabilità associata a un campione.





# Esempio

Una popolazione ha una media di 220 e una deviazione standard di 60. Viene selezionato un campione di 100 unità e la media campionaria viene utilizzata per stimare la media della popolazione.

- 1- Qual è il valore atteso di  $\bar{x}$ ?
- 2- Qual è la deviazione standard di  $\bar{x}$ ?
- 3- Mostrare la distribuzione campionaria di  $\bar{x}$ .
4. Cosa mostra la distribuzione campionaria di  $\bar{x}$ ?

# Soluzione

1- Qual è il valore atteso di  $\bar{x}$ ?

220

2- Qual è la deviazione standard di  $\bar{x}$ ?

=  $60/\sqrt{100} = 6$

3- Mostrare la distribuzione campionaria di  $\bar{x}$ .

Approssimativamente una normale con  $E(\bar{x}) = 220$  e  $\sigma_{\bar{x}} = 6$

4. Cosa mostra la distribuzione campionaria di  $\bar{x}$ ?

La distribuzione di probabilità di  $\bar{x}$ .

# Esempio

Assumiamo che la deviazione standard sia  $\sigma = 15$ .  
Calcolare l'errore standard della media per campioni di numerosità 50, 100, 150, 200.

Cosa si può dire sulla dimensione dell'errore standard della media all'aumentare della dimensione del campione?

# Soluzione

- 2,12
- 1,5
- 1,22
- 1,06

$\sigma_x$  diminuisce al crescere di n.

Un campione più grande fornisce una maggiore probabilità che la media campionaria sia compresa entro una certa distanza dalla media della popolazione.

# Esempio

Una popolazione ha una media di 220 e una deviazione standard di 60. Viene selezionato un campione di 100 unità e la media campionaria viene utilizzata per stimare la media della popolazione.

- 1- Qual è la probabilità che  $\bar{x}$  sia compresa entro  $\pm 6$  rispetto alla media della popolazione?
- 2- Qual è la probabilità che  $\bar{x}$  sia compresa entro  $\pm 12$  rispetto alla media della popolazione?

# Soluzione

Essendo il campione  $> 30$  possiamo utilizzare la tavola delle probabilità normale standardizzata:

$$z = \frac{x - \mu}{\sigma} \quad \rightarrow \quad z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}}$$

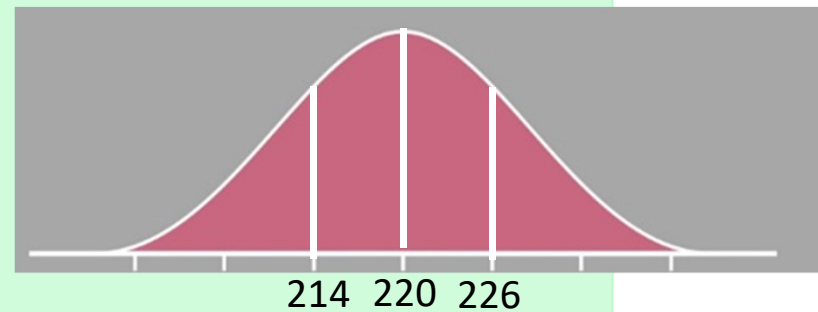
$$\sigma_x = 60 / \sqrt{100} = 6$$

Cerco l'intervallo  $(214 \leq \bar{x} \leq 226)$

$$\text{Per } x \ 214 = (214 - 220) / 6 = -1$$

$$\text{Per } x \ 226 = (226 - 220) / 6 = 1$$

$$(-1 \leq P \leq 1)$$



# Soluzione

$$P \leq 1 = 0,5 + 0,3413 = 0,8413$$

$$P \leq -1 = 1 - 0,8413 = 0,1587$$

$$(214 \leq x \leq 226) = 0,8413 - 0,1587 = 0,6826$$

z	0	1	2	3	4	5	6	7	8	9
0,0	0,0000	0,0040	0,0080	0,0120	0,0160	0,0199	0,0239	0,0279	0,0319	0,0359
0,1	0,0398	0,0438	0,0478	0,0517	0,0557	0,0596	0,0636	0,0675	0,0714	0,0754
0,2	0,0793	0,0832	0,0871	0,0910	0,0948	0,0987	0,1026	0,1064	0,1103	0,1141
0,3	0,1179	0,1217	0,1255	0,1293	0,1331	0,1368	0,1406	0,1443	0,1480	0,1517
0,4	0,1554	0,1591	0,1628	0,1664	0,1700	0,1736	0,1772	0,1809	0,1844	0,1879
0,5	0,1915	0,1950	0,1985	0,2019	0,2054	0,2088	0,2123	0,2157	0,2190	0,2224
0,6	0,2258	0,2291	0,2324	0,2357	0,2389	0,2422	0,2454	0,2486	0,2518	0,2549
0,7	0,2580	0,2612	0,2642	0,2673	0,2704	0,2734	0,2764	0,2794	0,2823	0,2852
0,8	0,2881	0,2910	0,2939	0,2967	0,2996	0,3023	0,3051	0,3078	0,3106	0,3133
0,9	0,3159	0,3186	0,3212	0,3238	0,3264	0,3289	0,3315	0,3340	0,3365	0,3389
1,0	0,3413	0,3438	0,3461	0,3485	0,3508	0,3531	0,3554	0,3577	0,3599	0,3621
1,1	0,3643	0,3665	0,3686	0,3708	0,3729	0,3749	0,3770	0,3790	0,3810	0,3830
1,2	0,3849	0,3869	0,3888	0,3907	0,3925	0,3944	0,3962	0,3980	0,3997	0,4015
1,3	0,4032	0,4049	0,4066	0,4082	0,4099	0,4115	0,4131	0,4147	0,4162	0,4177
1,4	0,4192	0,4207	0,4222	0,4236	0,4251	0,4265	0,4279	0,4292	0,4306	0,4319
1,5	0,4332	0,4345	0,4357	0,4370	0,4382	0,4394	0,4406	0,4418	0,4429	0,4441
1,6	0,4452	0,4463	0,4474	0,4484	0,4495	0,4505	0,4515	0,4525	0,4535	0,4545
1,7	0,4554	0,4564	0,4573	0,4582	0,4591	0,4599	0,4608	0,4616	0,4625	0,4633
1,8	0,4641	0,4649	0,4656	0,4664	0,4671	0,4678	0,4686	0,4693	0,4699	0,4706
1,9	0,4713	0,4719	0,4726	0,4732	0,4738	0,4744	0,4750	0,4756	0,4761	0,4767
2,0	0,4772	0,4778	0,4783	0,4788	0,4793	0,4798	0,4803	0,4808	0,4812	0,4811
2,1	0,4821	0,4826	0,4830	0,4834	0,4838	0,4842	0,4846	0,4850	0,4854	0,4857
2,2	0,4861	0,4864	0,4868	0,4871	0,4875	0,4878	0,4881	0,4884	0,4887	0,4890
2,3	0,4893	0,4896	0,4898	0,4901	0,4904	0,4906	0,4909	0,4911	0,4913	0,4916
2,4	0,4918	0,4920	0,4922	0,4925	0,4927	0,4929	0,4931	0,4932	0,4934	0,4936
2,5	0,4938	0,4940	0,4941	0,4943	0,4945	0,4946	0,4948	0,4949	0,4951	0,4952
2,6	0,4953	0,4955	0,4956	0,4957	0,4959	0,4960	0,4961	0,4962	0,4963	0,4964
2,7	0,4965	0,4966	0,4967	0,4968	0,4969	0,4970	0,4971	0,4972	0,4973	0,4974
2,8	0,4974	0,4975	0,4976	0,4977	0,4977	0,4978	0,4979	0,4979	0,4980	0,4981
2,9	0,4981	0,4982	0,4982	0,4983	0,4984	0,4984	0,4985	0,4985	0,4986	0,4986

# Soluzione

$$P \leq 2 = 0,5 + 0,4772 = 0,9772$$

$$P \leq -2 = 1 - 0,9772 = 0,0228$$

$$(208 \leq x \leq 232) = 0,9772 - 0,0228 = 0,9544$$

z	0	1	2	3	4	5	6	7	8	9
0,0	0,0000	0,0040	0,0080	0,0120	0,0160	0,0199	0,0239	0,0279	0,0319	0,0359
0,1	0,0398	0,0438	0,0478	0,0517	0,0557	0,0596	0,0636	0,0675	0,0714	0,0754
0,2	0,0793	0,0832	0,0871	0,0910	0,0948	0,0987	0,1026	0,1064	0,1103	0,1141
0,3	0,1179	0,1217	0,1255	0,1293	0,1331	0,1368	0,1406	0,1443	0,1480	0,1517
0,4	0,1554	0,1591	0,1628	0,1664	0,1700	0,1736	0,1772	0,1809	0,1844	0,1879
0,5	0,1915	0,1950	0,1985	0,2019	0,2054	0,2088	0,2123	0,2157	0,2190	0,2224
0,6	0,2258	0,2291	0,2324	0,2357	0,2389	0,2422	0,2454	0,2486	0,2518	0,2549
0,7	0,2580	0,2612	0,2642	0,2673	0,2704	0,2734	0,2764	0,2794	0,2823	0,2852
0,8	0,2881	0,2910	0,2939	0,2967	0,2996	0,3023	0,3051	0,3078	0,3106	0,3133
0,9	0,3159	0,3186	0,3212	0,3238	0,3264	0,3289	0,3315	0,3340	0,3365	0,3389
1,0	0,3413	0,3438	0,3461	0,3485	0,3508	0,3531	0,3554	0,3577	0,3599	0,3621
1,1	0,3643	0,3665	0,3686	0,3708	0,3729	0,3749	0,3770	0,3790	0,3810	0,3830
1,2	0,3849	0,3869	0,3888	0,3907	0,3925	0,3944	0,3962	0,3980	0,3997	0,4015
1,3	0,4032	0,4049	0,4066	0,4082	0,4099	0,4115	0,4131	0,4147	0,4162	0,4177
1,4	0,4192	0,4207	0,4222	0,4236	0,4251	0,4265	0,4279	0,4292	0,4306	0,4319
1,5	0,4332	0,4345	0,4357	0,4370	0,4382	0,4394	0,4406	0,4418	0,4429	0,4441
1,6	0,4452	0,4463	0,4474	0,4484	0,4495	0,4505	0,4515	0,4525	0,4535	0,4545
1,7	0,4554	0,4564	0,4573	0,4582	0,4591	0,4599	0,4608	0,4616	0,4625	0,4633
1,8	0,4641	0,4649	0,4656	0,4664	0,4671	0,4678	0,4686	0,4693	0,4699	0,4706
1,9	0,4713	0,4719	0,4726	0,4732	0,4738	0,4744	0,4750	0,4756	0,4761	0,4767
2,0	0,4772	0,4778	0,4783	0,4788	0,4793	0,4798	0,4803	0,4808	0,4812	0,4811
2,1	0,4821	0,4826	0,4830	0,4834	0,4838	0,4842	0,4846	0,4850	0,4854	0,4857
2,2	0,4861	0,4864	0,4868	0,4871	0,4875	0,4878	0,4881	0,4884	0,4887	0,4890
2,3	0,4893	0,4896	0,4898	0,4901	0,4904	0,4906	0,4909	0,4911	0,4913	0,4916
2,4	0,4918	0,4920	0,4922	0,4925	0,4927	0,4929	0,4931	0,4932	0,4934	0,4936
2,5	0,4938	0,4940	0,4941	0,4943	0,4945	0,4946	0,4948	0,4949	0,4951	0,4952
2,6	0,4953	0,4955	0,4956	0,4957	0,4959	0,4960	0,4961	0,4962	0,4963	0,4964
2,7	0,4965	0,4966	0,4967	0,4968	0,4969	0,4970	0,4971	0,4972	0,4973	0,4974
2,8	0,4974	0,4975	0,4976	0,4977	0,4977	0,4978	0,4979	0,4979	0,4980	0,4981
2,9	0,4981	0,4982	0,4982	0,4983	0,4984	0,4984	0,4985	0,4985	0,4986	0,4986



# Distribuzione campionaria di $\bar{x}$

La distribuzione campionaria di  $x$  può essere utilizzata per fornire informazioni probabilistiche sulla differenza tra la media campionaria e la media della popolazione.

Esempio in un'azienda lavorano 2.500 dipendenti, il cui stipendio medio è di 21.800€ mensili. Il direttore del personale ritiene che la media campionaria possa essere utilizzata se compresa entro i 400€ dalla media della popolazione.

La domanda in termini probabilistici è: qual è la probabilità che la media campionaria calcolata utilizzando un campione casuale semplice di 30 dipendenti sia compresa entro i 400€ dalla media della popolazione. La deviazione standard della popolazione è 2.000.

# Esempio

Si considera un intervallo di probabilità  $21.400€ \leq 21.800€ \leq 22.200€$

1- Calcolo l'errore standard =  $2000 / \sqrt{30} = 365,1$

$$z = (21.400 - 21.800) / 365,1 = 1,09$$

$$z = (22.200 - 21.800) / 365,1 = - 1,09$$

$$= 0,4713 + 0,5 = 0,9713$$

$$= 1 - 0,9713 = 0,0287$$

$$= 0,9713 - 0,0287 = 0,9426$$

Un campione casuale semplice di 30 dipendenti ha una probabilità del 94% di fornire una media campionaria che si trovi entro la soglia consentita entro 400€.

z	0	1	2	3	4	5	6	7	8	9
0,0	0,0000	0,0040	0,0080	0,0120	0,0160	0,0199	0,0239	0,0279	0,0319	0,0359
0,1	0,0398	0,0438	0,0478	0,0517	0,0557	0,0596	0,0636	0,0675	0,0714	0,0754
0,2	0,0793	0,0832	0,0871	0,0910	0,0948	0,0987	0,1026	0,1064	0,1103	0,1141
0,3	0,1179	0,1217	0,1255	0,1293	0,1331	0,1368	0,1406	0,1443	0,1480	0,1517
0,4	0,1554	0,1591	0,1628	0,1664	0,1700	0,1736	0,1772	0,1809	0,1844	0,1879
0,5	0,1915	0,1950	0,1985	0,2019	0,2054	0,2088	0,2123	0,2157	0,2190	0,2224
0,6	0,2258	0,2291	0,2324	0,2357	0,2389	0,2422	0,2454	0,2486	0,2518	0,2549
0,7	0,2580	0,2612	0,2642	0,2673	0,2704	0,2734	0,2764	0,2794	0,2823	0,2852
0,8	0,2881	0,2910	0,2939	0,2967	0,2996	0,3023	0,3051	0,3078	0,3106	0,3133
0,9	0,3159	0,3186	0,3212	0,3238	0,3264	0,3289	0,3315	0,3340	0,3365	0,3389
1,0	0,3413	0,3438	0,3461	0,3485	0,3508	0,3531	0,3554	0,3577	0,3599	0,3621
1,1	0,3643	0,3665	0,3686	0,3708	0,3729	0,3749	0,3770	0,3790	0,3810	0,3830
1,2	0,3849	0,3869	0,3888	0,3907	0,3925	0,3944	0,3962	0,3980	0,3997	0,4015
1,3	0,4032	0,4049	0,4066	0,4082	0,4099	0,4115	0,4131	0,4147	0,4162	0,4177
1,4	0,4192	0,4207	0,4222	0,4236	0,4251	0,4265	0,4279	0,4292	0,4306	0,4319
1,5	0,4332	0,4345	0,4357	0,4370	0,4382	0,4394	0,4406	0,4418	0,4429	0,4441
1,6	0,4452	0,4463	0,4474	0,4484	0,4495	0,4505	0,4515	0,4525	0,4535	0,4545
1,7	0,4554	0,4564	0,4573	0,4582	0,4591	0,4599	0,4608	0,4616	0,4625	0,4633
1,8	0,4641	0,4649	0,4656	0,4664	0,4671	0,4678	0,4686	0,4693	0,4699	0,4706
1,9	0,4713	0,4719	0,4726	0,4732	0,4738	0,4744	0,4750	0,4756	0,4761	0,4767

# Esercizio

I risultati di un'indagine indicano che l'errore standard della media è 20. La deviazione standard della popolazione è 500.

- 1- Quanto è grande il campione utilizzato per questa indagine?
- 2- Qual è la probabilità che la stima puntuale sia compresa entro  $\pm 25$  dalla media della popolazione?

# Soluzione

1- Quanto è grande il campione utilizzato per questa indagine?

$$20 = 500/x$$

$$20x = 500 \quad \rightarrow \quad 500/20 = 25 \quad \rightarrow \quad n = 625$$

2- Qual è la probabilità che la stima puntuale sia compresa entro  $\pm 25$  dalla media della popolazione?

$$\pm 25/20 = \pm 1,25$$

$$(-1,25 \leq P \leq 1,25) = 0,8944 - 0,1059 = 0,7885$$

# Domande multiple choice

**La variabilità del campione indica?**

- Quanto il valore della statistica campionaria possa variare da campione a campione
- La dispersione di una serie di misure effettuate su un campione
- Quanto il valore della statistica campionaria possa variare a seconda della numerosità del campione estratto
- Non si può applicare il concetto di variabilità ai campioni

**Nel campionamento casuale stratificato il campione è composto da tanti campioni casuali semplici quanti sono gli strati?**

SI

NO

**In riferimento al bias di selezione quali delle seguenti affermazioni non è corretta?**

- Compromette la validità interna di uno studio clinico randomizzato
- E' un errore sistematico
- E' un errore casuale

**Quali delle seguenti affermazione è vera (sono possibili più alternative)?**

- Nel rispetto del requisito di simultaneità, il censimento viene condotto nell'arco di un periodo di tempo limitato
- Dati più dettagliati possano essere ottenuti attraverso il ricorso a tecniche campionarie
- L'indagine sulle forze di lavoro utilizza un campione non probabilistico a 2 stadi
- Nella quantificazione dei costi gioca un ruolo fondamentale anche la tecnica di rilevazione adottata (es tipo di intervista)

**L'estrazione di un campione da un elenco telefonico è un campione probabilistico?**

SI

NO

**Con riferimento ai campioni probabilistici quale delle seguenti affermazioni è falsa?**

- Le unità statistiche sono estratte a sorte
- Eventuali scarti devono essere non significativi ed imputabili al caso
- Le unità statistiche da includere nel campione sono scelte in modo ragionato
- La scelta del campione non risulta influenzata dal ricercatore stesso