

Campionamento e distribuzioni campionarie

Vedremo i concetti di:

- Popolazione e suoi parametri
- Campionamento da popolazioni finite
- Campionamento da popolazioni infinite
- Statistiche campionarie e loro distribuzioni
- Distribuzione della media campionaria

2 **PREMESSA**

- Ogni indagine/analisi statistica si riferisce ad una certa **popolazione** che può **finita** o **infinita** e che costituisce l'obiettivo d'interesse.

Es. in uno studio finalizzato a misurare il tasso di disoccupazione di un paese, la popolazione (finita) è costituita da tutti gli abitanti in età 18-60.



Si definisce la popolazione d'interesse



Si acquisiscono i dati rilevanti



Si analizzano i dati mediante appropriate tecniche statistiche

3

In alcuni casi i dati sono raccolti sulla intera Popolazione **P**, spesso ragioni di tempo, economicità e impossibilità pratica obbligano ad asservare solo una parte della popolazione, un ***campione***.

Il campione è estratto da P seguendo alcune regole probabilistiche (***il piano di campionamento***).

Dall'analisi dei dati campionari si ottengono tramite ***procedure inferenziali*** informazioni su caratteristiche rilevanti (***parametri***) riguardanti l'intera popolazione.

4 Popolazione finita e suoi parametri

- Una **Popolazione finita P** è un insieme di N unità su cui si può osservare un certo carattere X. (es: gli investimenti annui di tutte le aziende di un paese; il numero di figli di ogni famiglia italiana; i giorni di degenza di tutti i ricoverati in un certo ospedale.)
- I **parametri** della popolazione sono delle costanti che descrivono aspetti caratteristici della distribuzione del carattere nella popolazione stessa.

Esempio:

Media di P

$$\mu = \frac{1}{N} \sum_{i=1}^N x_i$$

Varianza di P

$$\sigma^2 = \frac{1}{N} \sum_{i=1}^N (x_i - \mu)^2 = \frac{1}{N} \sum_{i=1}^N x_i^2 - \mu^2$$

SIMBOLOGIA: attenzione

A SECONDA CHE CI RIFERIAMO ALLA POPOLAZIONE O AL CAMPIONE:

- **N** numero di unità della popolazione
- **n** numero di unità del campione
- μ e σ^2 la media e la varianza nella Popolazione
- $\bar{x}$ e s^2 la media e la varianza nel campione

6 Popolazione infinita e suoi parametri

- Una **Popolazione infinita** è composta da un numero molto elevato di unità, ossia da tutte le unità *potenzialmente osservabili* e non necessariamente già esistenti fisicamente (es. i circuiti elettrici che una azienda può produrre in un certo periodo; coloro i quali potrebbero ammalarsi di una certa malattia; etc).
- Il campione è composto dalla n unità tra quelle effettivamente osservabili (es. n circuiti tra quelli effettivamente prodotti).

7

Il carattere d'interesse può essere rappresentato da una variabile casuale con una certa distribuzione di probabilità. In questo caso si indicherà con "*popolazione X*" la variabile casuale X .

- Se la v.c. X è discreta con k valori distinti

$$x_1, \dots, x_k \quad \text{sarà } p(x_j) = P(X = x_j), j = 1, \dots, k$$

la sua distribuzione di probabilità

- Se la v.c. X è continua sarà $f(x)$ la funzione di densità che ne definisce la distribuzione di probabilità

8

Come nel caso di P finita un parametro è una statistica descrittiva che sintetizza qualche aspetto della distribuzione della v.c.

Media della popolazione (o valore atteso):

Discreta:

Continua:

$$\mu = E(X) = \sum_{j=1}^k x_j p(x_j)$$

$$\mu = E(X) = \int_{-\infty}^{+\infty} x f(x) dx$$

Varianza della popolazione:

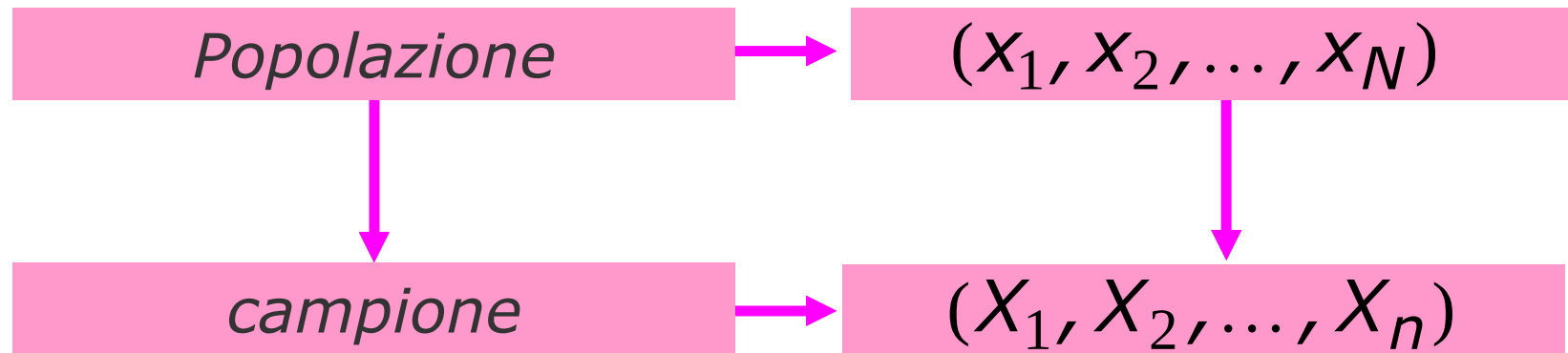
Discreta:

Continua:

$$\sigma^2 = \text{Var}(X) = E[X - E(X)]^2 = \sum_{j=1}^k x_j^2 p(x_j) - \mu^2$$

$$\sigma^2 = \text{Var}(X) = \int_{-\infty}^{+\infty} x^2 f(x) dx - \mu^2$$

9 Campionamento da Popolazioni finite (trattazione più intuitiva)



n = dimensione campionaria

n/N = frazione di campionamento o di sondaggio

10 **Campionamento da Popolazioni finite**

Se la **regola di selezione** del campione è di tipo probabilistico, l'estrazione del campione avviene in accordo con qualche specifica distribuzione di probabilità. Per una **selezione probabilistica** è necessario individuare:

- ✦ lo spazio campionario Ω , formato da tutti i possibili campioni estraibili con una medesima tecnica da una popolazione.
- ✦ la probabilità di ogni campione c in Ω di essere estratto

La coppia $\{\Omega, \text{probabilità dei campioni in } \Omega\}$ è detta **piano di campionamento o disegno campionario**.

Campioni casuali

I campioni possono essere estratti casualmente dalla popolazione:

- **con ripetizione**: una volta estratta un'unità viene rimessa dentro la popolazione e quindi potrebbe essere nuovamente estratta;
- **senza ripetizione**: una volta estratta un'unità questa viene messa da parte e quindi non può essere estratta più di una volta.

Due campioni **non ordinati** di uguale numerosità sono diversi tra loro se almeno un'unità del primo campione non è contenuta nel secondo campione. Nei campioni **ordinati** conta invece anche l'ordine con cui si presentano le diverse unità.

12 Esempio disegno campionario

Popolazione composta da 4 grandi aziende.
($N=4$); Carattere="Fatturato annuo";

$$x_1 = 52, x_2 = 49, x_3 = 65, x_4 = 74$$

Spazio campionario Ω , costituito dai campioni *ordinati* di dimensione 3, estratti *senza ripetizione*.

$c_1 = 52 \ 49 \ 65$	$c_7 = 49 \ 65 \ 74$	$c_{13} = 65 \ 74 \ 52$	$c_{19} = 74 \ 52 \ 49$
$c_2 = 52 \ 65 \ 49$	$c_8 = 49 \ 74 \ 65$	$c_{14} = 65 \ 52 \ 74$	$c_{20} = 74 \ 49 \ 52$
$c_3 = 49 \ 52 \ 65$	$c_9 = 65 \ 49 \ 74$	$c_{15} = 74 \ 52 \ 65$	$c_{21} = 52 \ 74 \ 49$
$c_4 = 49 \ 65 \ 52$	$c_{10} = 65 \ 74 \ 49$	$c_{16} = 74 \ 65 \ 52$	$c_{22} = 52 \ 49 \ 74$
$c_5 = 65 \ 52 \ 49$	$c_{11} = 74 \ 49 \ 65$	$c_{17} = 52 \ 65 \ 74$	$c_{23} = 49 \ 52 \ 74$
$c_6 = 65 \ 49 \ 52$	$c_{12} = 74 \ 65 \ 49$	$c_{18} = 52 \ 74 \ 65$	$c_{24} = 49 \ 74 \ 52$

Ogni campione ha uguale probabilità di essere estratto, pari a $1/24$. Se non fosse stato considerato l'ordine di estrazione i campioni diversi sarebbero stati 4.

NOTA BENE

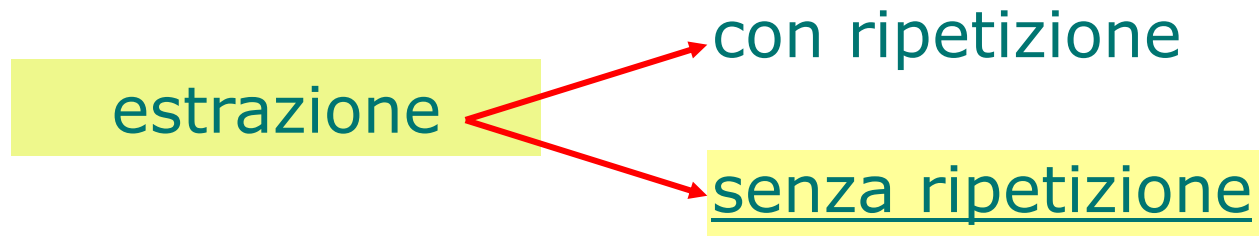
Svolgendo una indagine si può incorrere in diversi tipi di errori anche materiali

Nel caso di indagini campionarie si manifesta un diverso ed importante tipo di errore, l'**errore campionario**, derivante dal fatto che ogni tipo di conclusione riguardante la popolazione è basata solo sull'osservazione di una sua parte

Campionamento casuale semplice (P Finita)

Nel campionamento casuale semplice i campioni di uguale dimensione hanno tutti stessa probabilità di essere estratti.

- ✦ si devono conoscere le unità della popolazione;
- ✦ tutte le unità devono essere reperibili;
- ✦ si deve procedere all'estrazione casuale delle unità (*tavole dei numeri casuali o software per la generazione di numeri pseudo-casuali*).



Il campione casuale ottenuto con estrazioni senza ripetizione è composto da n variabili casuali $X_1, X_2, \dots, X_n$ che hanno marginalmente (individualmente) stessa distribuzione di probabilità ma non sono indipendenti.



La distribuzione di probabilità della generica X_i è uguale a quella del carattere X nella popolazione.

15 Numero di campioni:

	Estrazione con ripetizione	Estrazione senza ripetizione
Campioni ordinati	N^n	$N!/(N-n)!$
Campioni non ordinati	$(N+n-1)!/n!(N-1)!$	$N!/(N-n)!$

Nel campionamento senza ripetizione e campioni non ordinati nell'esempio Ω è formato solo da 4 campioni per esempio

c6=65 49 52, c12=74 65 49, c18=52 74 65, c24=49 74 52

Inoltre nell'esempio ogni v.c. X_i ha la stessa distribuzione:

X_i	$P(X_i)$
49	0.25
52	0.25
65	0.25
74	0.25

16

Estraendo senza ripetizione le variabili che compongono il campione non sono indipendenti. Per es. per $c_1=52,49,65$ si ha:

$$P(X_1=52, X_2=49, X_3=65) = P(X_1=52)P(X_2=49/X_1=52)P(X_3=65/X_1=52, X_2=49) =$$

$$= \frac{1}{4} \cdot \frac{1}{3} \cdot \frac{1}{2} = \frac{1}{24} \neq \frac{1}{4} \cdot \frac{1}{4} \cdot \frac{1}{4} = \frac{1}{64}$$

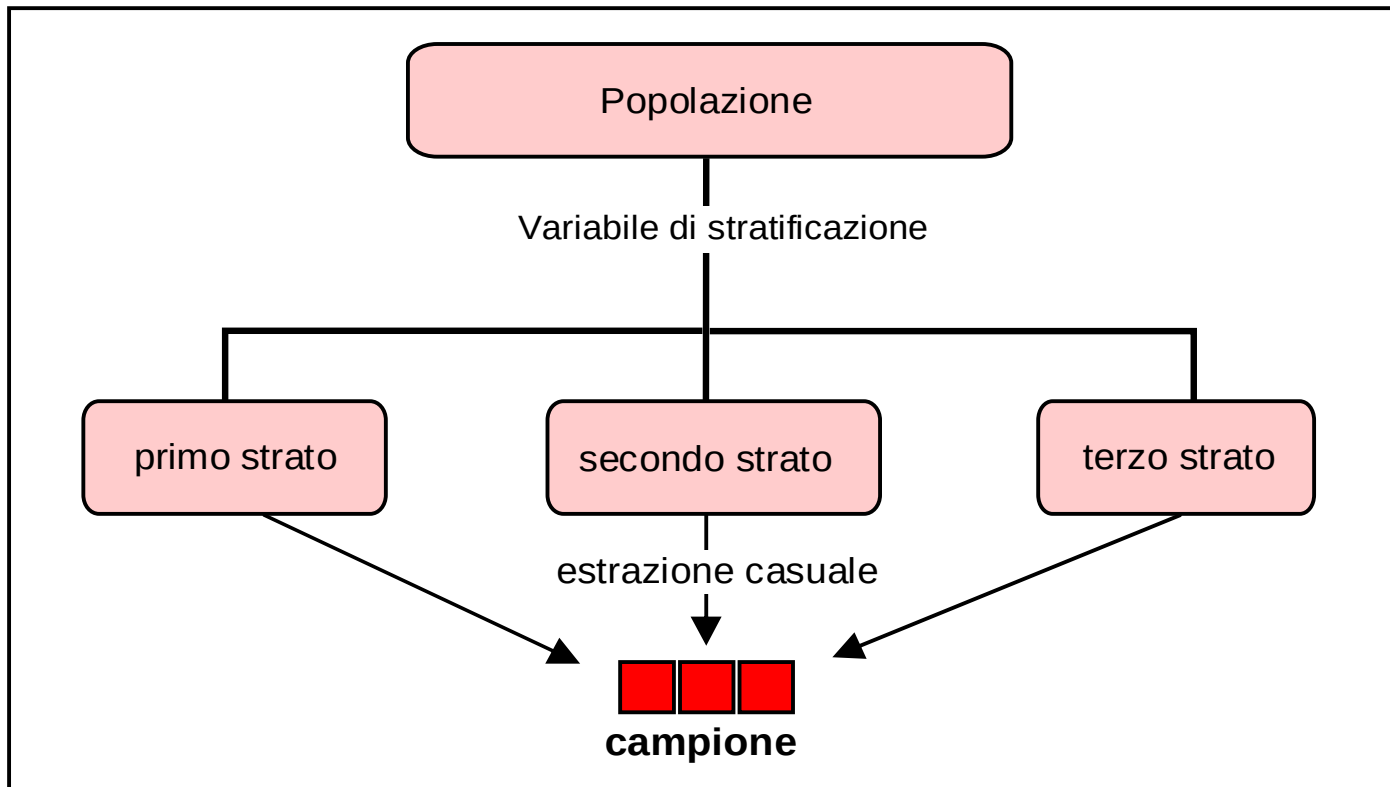
Considerando campioni ordinati ed estrazione con ripetizione, il n° di campioni possibili è $N^n=4^3=64$. Poiché essi sono equiprobabili si ricava:

$$P(X_1=52, X_2=49, X_3=65) = P(X_1=52)P(X_2=49)P(X_3=65) =$$

$$= \frac{1}{4} \cdot \frac{1}{4} \cdot \frac{1}{4} = \frac{1}{64}$$

Ossia le tre v.c. X_1, X_2, X_3 sono indipendenti

17 Campionamento casuale stratificato –P Finita



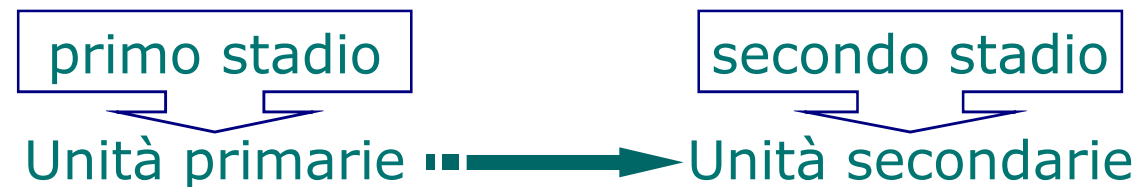
Nel campionamento casuale stratificato la popolazione viene suddivisa in **strati**. Da ogni strato vengono poi estratti, tramite un campionamento casuale semplice, le unità da inserire nel campione. *Esempio strati: Regioni; età; sesso.*

- Migliori stime se gli strati sono ben scelti
- Stime non solo per P ma anche per i singoli strati

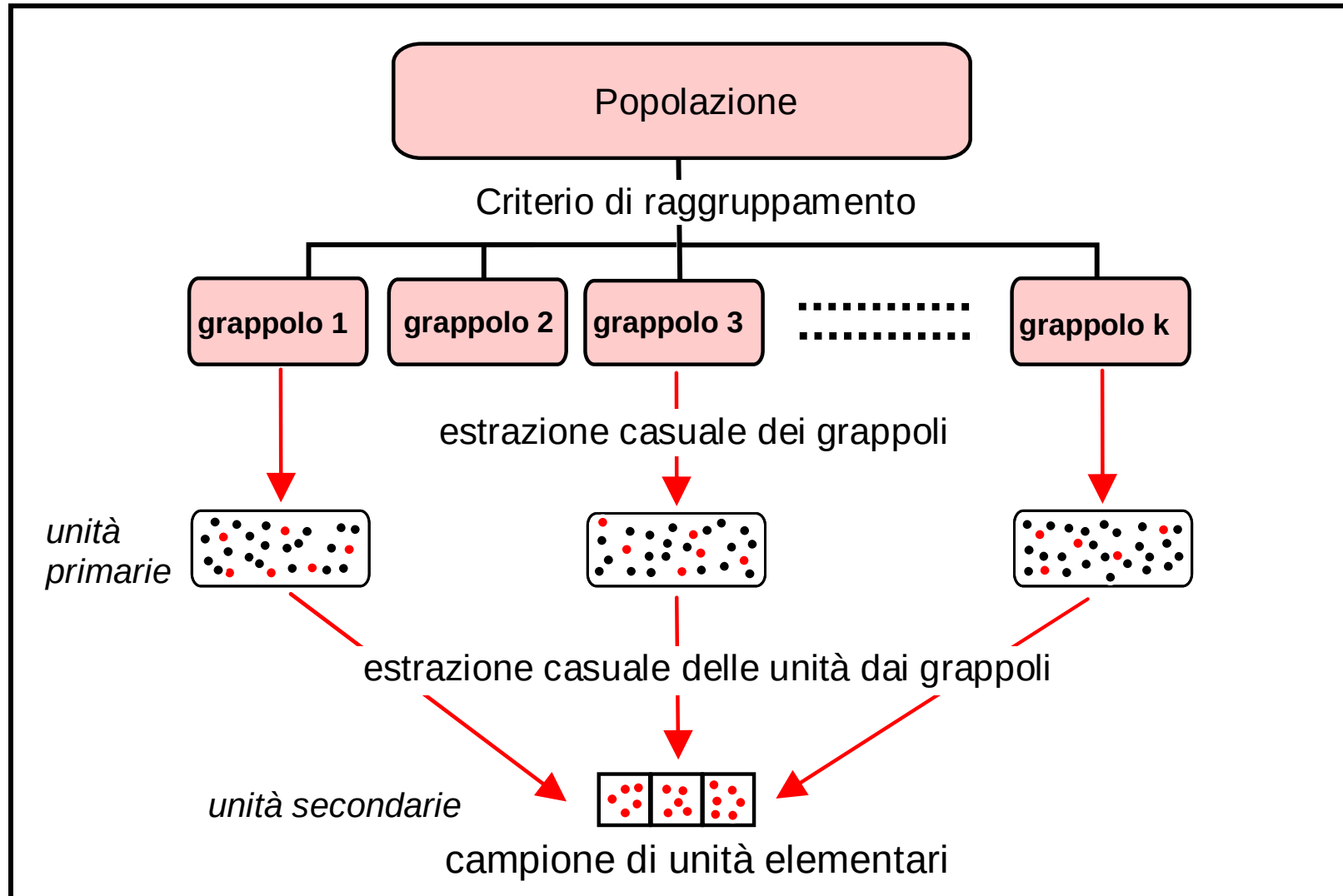
18 Campionamento casuale a grappoli e a stadi – P Finita

Nel campionamento **casuale a grappoli** la popolazione viene suddivisa in sottoinsiemi detti **grappoli**. Si selezionano, con un'estrazione casuale senza ripetizione, un certo numero di grappoli e si prendono come unità campionarie tutte le unità appartenenti ai grappoli estratti.

Nel campionamento **casuale a due stadi** la popolazione viene suddivisa in un certo numero di **grappoli**. Al primo stadio si estrae senza ripetizione un certo numero di grappoli. Da ciascuno di questi si estrae con ripetizione (secondo stadio) un certo numero di unità.



Campionamento casuale a grappoli e a stadi – Popolazione Finita



20 Campionamento da popolazioni infinite

In una popolazione infinita, la n -pla di variabili casuali:

$$(X_1, X_2, \dots, X_n)$$

che compongono il campione casuale di dimensione n presenta le seguenti proprietà (i.i.d.):

✦ $X_1, X_2, \dots, X_n$ sono variabili casuali **indipendenti** (non ha senso distinguere tra estrazione con e senza ripetizione) .

✦ ogni v.c. X_i possiede la stessa distribuzione di probabilità della popolazione X .

Nelle popolazioni finite in cui la dimensione campionaria è molto più piccola della numerosità della popolazione, si può applicare la teoria del campionamento da popolazioni infinite poiché n/N è trascurabile.

21 **Statistiche campionarie e distribuzioni campionarie**

Una **statistica campionaria** è una funzione a valori reali delle osservazioni campionarie: $T = t(X_1, X_2, \dots, X_n)$

Esempi di statistiche campionarie:

media campionaria: $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$

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

La statistica campionaria è una variabile casuale a cui è associata una distribuzione di probabilità detta **distribuzione campionaria**.

Statistiche campionarie e distribuzioni campionarie - Esempio

Consideriamo una popolazione finita composta dalle seguenti 5 unità: $X_1 = 8$ $X_2 = 4$ $X_3 = 2$ $X_4 = 11$ $X_5 = 6$

Risulta che $\mu=6.2$ (e $\sigma^2=9.76$).

Si consideri l'estrazione senza ripetizione di campioni di dimensione $n=2$ e per ognuno di essi si calcoli la statistica *media campionaria*.

X_1	8	8	8	8	4	4	4	4	2	2	2	2	11	11	11	11	6	6	6	6
X_2	4	2	11	6	8	2	11	6	8	4	11	6	8	4	2	6	8	4	2	11
$\bar{X}$	6	5	9,5	7	6	3	7,5	5	5	3	6,5	4	9,5	7,5	6,5	8,5	7	5	4	8,5

Ogni campione ha probabilità 0,05 di essere estratto e quindi la distribuzione di probabilità della media campionaria è data da:

$\bar{X}$	3	4	5	6	6,5	7	7,5	8,5	9,5
$P(\bar{X})$	0,1	0,1	0,2	0,1	0,1	0,1	0,1	0,1	0,1

23 Distribuzione della media campionaria nelle popolazioni infinite

Sia X la v.c. e siano μ e σ^2 media e varianza di P .

➡ $\mu = E(X)$ e $\sigma^2 = V(X)$

Sia $X_1, X_2, \dots, X_n$ un campione casuale di dimensione n estratto dalla popolazione X . Sarà costituito da n v.c. i.i.d. ciascuna con $\mu = E(X_i)$ e $\sigma^2 = V(X_i)$.



Proprietà della media campionaria:

✦ il valore atteso $E(\bar{X}) = \mu$

✦ la varianza

$$Var(\bar{X}) = \sigma^2/n$$

Nel precedente esempio

$$E(\bar{X}) = \sum_{i=1}^9 \bar{x}_i p_i = 6.2 = \mu$$

I precedenti risultati valgono quale che sia la distribuzione della popolazione X e quale che sia la dimensione n campionaria.

Ma possiamo dire ancora di più (sulla distribuzione campionaria di $\bar{X}$) nel caso in cui X segua alcune distribuzioni.

In particolare:

$$\text{Se } X \sim N(\mu; \sigma^2) \quad \text{allora} \quad \bar{X} \sim N\left(\mu; \frac{\sigma^2}{n}\right)$$

N.B. della popolazione e la sua distribuzione della media campionaria è meno variabile di quella di variabilità cresce all'aumentare della dimensione campionaria

ALTRI IMPORTANTI RISULTATI

Inoltre:

Se $X \sim Be(\pi)$ allora

$$P(\bar{X} = x) = \binom{n}{nx} \pi^{nx} (1-\pi)^{n-nx}, \quad x = 0, \frac{1}{n}, \frac{2}{n}, \dots, \frac{n-1}{n}, 1$$

con media π e varianza $\pi(1-\pi)/n$

Ossia

$$\bar{X} \sim \frac{1}{n} Bin(n, \pi)$$

IN GENERALE

Tuttavia quale che sia la distribuzione nella popolazione X **per il Teorema del Limite Centrale**

$$\lim_{n \rightarrow \infty} P\left(\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \leq z\right) = P(Z \leq z)$$

dove Z è una v.c. Normale standardizzata

27 Distribuzione della media campionaria nelle popolazioni finite

Si consideri una popolazione finita dalla quale viene estratto senza ripetizione un campione casuale. In questo caso:

- ✦ il valore atteso $E(\bar{X}) = \mu$
- ✦ la varianza $Var(\bar{X}) = \left(\frac{N - n}{N - 1} \right) \frac{\sigma^2}{n}$
- ✦ Dove $(N - n)/(N - 1)$ è detto fattore di correzione per popolazioni finite.
- ✦ Se n è sufficientemente ampio ma molto più piccolo di N , allora la distribuzione di $\bar{X}$ può essere approssimata a una Normale con media μ e varianza $\left(\frac{N - n}{N - 1} \right) \frac{\sigma^2}{n}$

28 Caratteristiche della distribuzione della media campionaria

Sia $\bar{X}$ la media campionaria di un campione di n osservazioni provenienti da una **P** con **μ e σ^2 media e varianza**. La distribuzione di $\bar{X}$

1. ha media $E(\bar{X}) = \mu$
2. ha deviazione standard $Var(\bar{X}) = \sigma^2/n$
3. Se l'ampiezza del campione n non è sufficientemente piccola rispetto alla dimensione della popolazione N allora:

$$Var(\bar{X}) = \left(\frac{N - n}{N - 1} \right) \frac{\sigma^2}{n}$$

1. Se la distribuzione di P è normale allora la distribuzione di $\bar{X}$ è normale e la variabile:

$$Z = \frac{\bar{X} - \mu}{\sigma / \sqrt{n}}$$

è una normale standardizzata

Vediamo ora nel caso del
Campionamento casuale semplice
mediante dati reali una "verifica"
del
Teorema del limite centrale

Riassumendo

- Per studiare le proprietà (parametri: μ , σ) di una popolazione si fa ricorso ad un campione estratto dalla popolazione.
- Abbiamo in particolare illustrato una serie di risultati riguardanti la media della Popolazione e la media campionaria
- Ma... fino a che punto le "stime", ricavate dal campione, possono essere riferite alla popolazione ?
- A tal fine, si consideri una popolazione composta da **N** unità, $\{x\}_{N=5} = \{2, 3, 6, 8, 11\}$

e da questa popolazione si estraggano tutti i possibili campioni di una data ampiezza **n**.

una popolazione composta da N unità,
 $\{x\}_{N=5} = \{2, 3, 6, 8, 11\}$

	x_i	f_i	$(x_i - \mu)$	$(x_i - \mu)^2$
	2	1	-4	16
	3	1	-3	9
	6	1	0	0
	8	1	2	4
	11	1	5	25
Σ	30	5	0	54

$$\mu = \frac{30}{5} = 6$$

$$\sigma^2 = \frac{54}{5} = 10,8$$

**Tutti i possibili campioni di ampiezza (n=2) ,
(estraibili dalla popolazione) ,sono ...**

(2 , 2)	(2 , 3)	(2 , 6)	(2 , 8)	(2 , 11)
(3 , 2)	(3 , 3)	(3 , 6)	(3 , 8)	(3 , 11)
(6 , 2)	(6 , 3)	(6 , 6)	(6 , 8)	(6 , 11)
(8 , 2)	(8 , 3)	(8 , 6)	(8 , 8)	(8 , 11)
(11,2)	(11,3)	(11,6)	(11,8)	(11,11)

... e le medie campionarie sono ...

2.0	2.5	4.0	5.0	6.5
2.5	3.0	4.5	5.5	7.0
4.0	4.5	6.0	7.0	8.5
5.0	5.5	7.0	8.0	9.5
6.5	7.0	8.5	9.5	11.0

**campionamento
con ripetizione**

DISTRIBUZIONE DELLE MEDIE CAMPIONARIE

media
campionaria

$\bar{x}_i$	f_i	$\bar{x}_i \cdot f_i$	$(\bar{x}_i - \bar{\bar{x}})^2 \cdot f_i$
2.0	1	2	16
2.5	2	5	24.5
3.0	1	3	9
4.0	2	8	8
4.5	2	9	4.5
5.0	2	10	2
5.5	2	11	0.5
6.0	1	6	0
6.5	2	13	0.5
7.0	4	28	4
8.0	1	8	4
8.5	2	17	12.5
9.5	2	19	24.5
11.0	1	11	25
Σ	25	150	135

Dalla distribuzione di frequenza delle **medie campionarie** si calcola:

a) la media

b) la dispersione :

la **media aritmetica** della media campionaria.

$$Media(\bar{x}) = \frac{150}{25} = 6$$

$$E[\bar{x}] = \mu$$

la **varianza** della media campionaria

$$Var(\bar{x}) = \frac{135}{25} = 5.4$$

$$Var(\bar{x}) = \frac{\sigma^2}{n}$$

media
campionaria

$\bar{x}_i$	f_i	$(\bar{x}_i - \bar{\bar{x}}) \times f_i$	$(\bar{x}_i - \bar{\bar{x}})^2 \times f_i$
2.0	1	-4.0	16
2.5	2	-3.5	24.5
3.0	1	-3.0	9
4.0	2	-2.0	8
4.5	2	-1.5	4.5
5.0	2	-1.0	2
5.5	2	-0.5	0.5
6.0	1	0.0	0
6.5	2	0.5	0.5
7.0	4	1.0	4
8.0	1	2.0	4
8.5	2	2.5	12.5
9.5	2	3.5	24.5
11.0	1	5.0	25
TOTALE	25	150	135

Si noti che solo 1 dei 25 possibili campioni ha media coincidente con la vera media della popolazione.

I restanti campioni hanno media che si discosta da 6.0 per più o meno 5 unità (da 2 a 11);

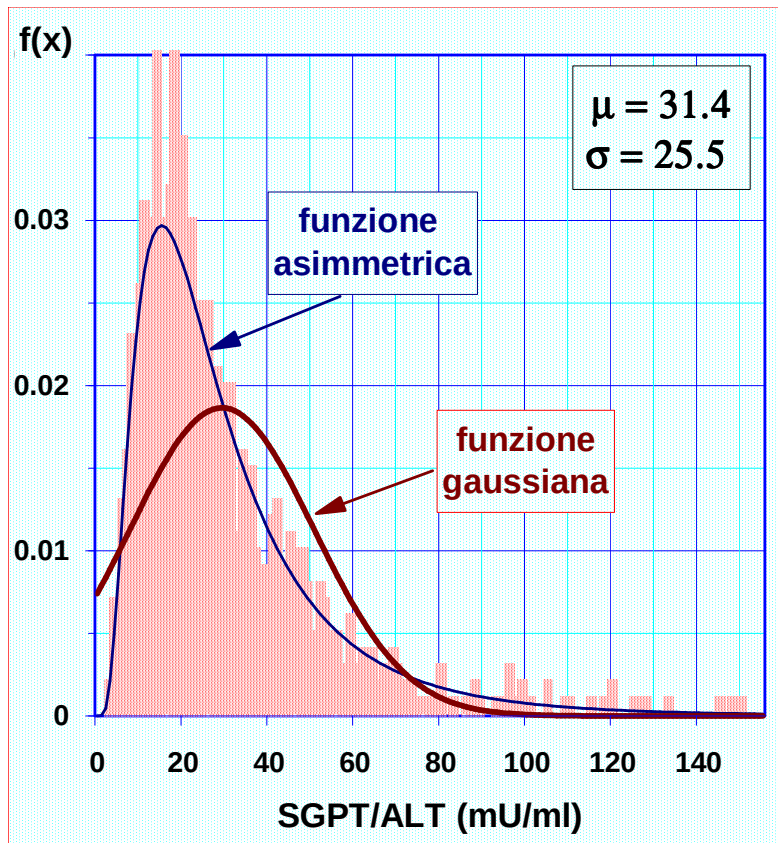
35 LA MEDIA DELLE MEDIE CAMPIONARIE

L'operazione consiste nel calcolo della media e della varianza delle MEDIE CAMPIONARIE, cioè delle medie ricavate su tutti i possibili campioni di ampiezza finita estraibili dalla popolazione.

In una popolazione con media μ e deviazione standard σ , abbiamo visto che la distribuzione delle medie campionarie di dimensione n ha tre proprietà:

1. La media della distribuzione di campionamento delle medie è uguale alla media della popolazione.
2. La deviazione standard della distribuzione di campionamento delle medie campionarie è detta "errore standard" ed è uguale $\sigma/\sqrt{n}$.
3. La forma della distribuzione delle medie campionarie è approssimativamente normale, indipendentemente dalla forma della distribuzione e per n sufficientemente grande.

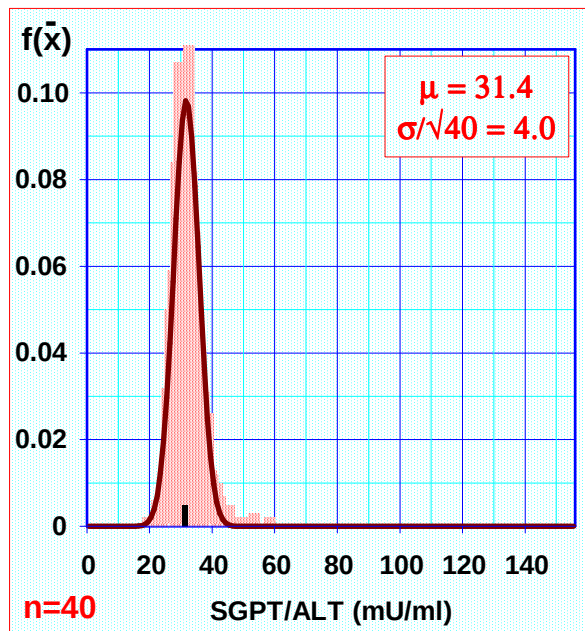
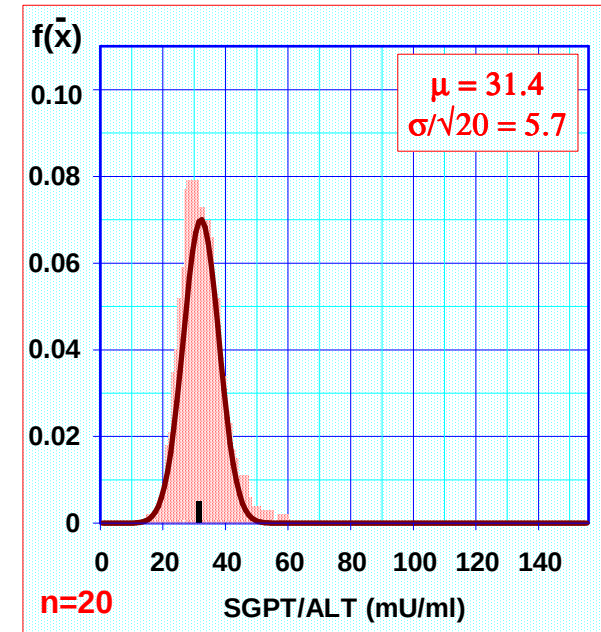
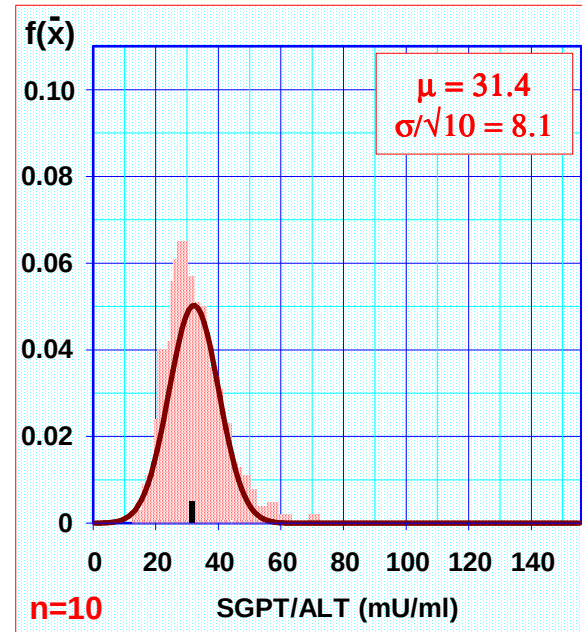
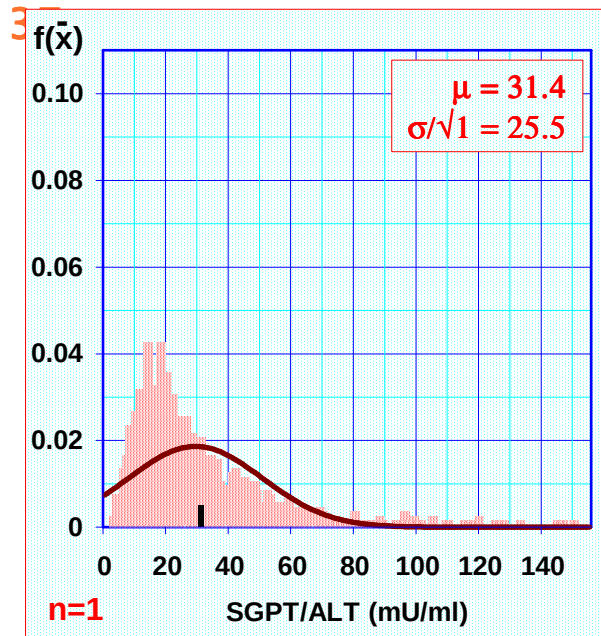
Dalla popolazione generale si sono estratti a caso 1000 soggetti adulti, e su ciascuno di essi si è determinato il livello ematico di ALT (alanina amino-transferasi). È noto che, nella popolazione generale, la **distribuzione** di ALT è **fortemente asimmetrica positiva** per la presenza di individui con danni epatici causati da alcol, farmaci e virus.



Dalla stessa popolazione sono poi estratti:

- **1000** campioni di dimensione **n = 10**
- **1000** campioni di dimensione **n = 20**
- **1000** campioni di dimensione **n = 40**

di ciascun campione si è calcolata la media.



All'aumentare della dimensione del campione la distribuzione delle medie campionarie non solo riduce la sua dispersione ma tende ad assumere la forma della funzione di Gauss.

TEOREMA DEL LIMITE CENTRALE

si calcolano la media e la varianza delle 1000 medie campionarie per ciascun valore di n si trova che:

n	$\bar{\bar{X}}$	$E(\bar{X})$	$S_{\bar{X}}^2$	$S_{\bar{X}}$	$\frac{\sigma}{\sqrt{n}}$
1	29.1	31.4	457.70	21.4	25.5
10	31.7	31.4	62.95	7.9	8.1
20	31.7	31.4	32.39	5.7	5.7
40	31.3	31.4	16.45	4.1	4.0

Quanto sopra **mostrato** con un esempio è **dimostrato** dal « teorema del limite centrale » :

Dato un campione di dimensione n , tratto da una **variabile casuale qualunque** (x) con media μ e varianza σ^2 ,

la **variabile casuale media campionaria** $\bar{\bar{X}}$

approssima, al crescere di n , la **distribuzione gaussiana** con valore atteso pari μ e varianza pari a σ^2/n :

Il teorema del limite centrale definisce l'errore standard della media

dato $\{x_1, x_2, \dots, x_n\}$ tratto da $x \sim (\mu, \sigma^2)$ si ha che

$$(\bar{x}) \sim N(\mu, \sigma^2/n) \text{ per } n \rightarrow \infty$$

Il parametro $\frac{\sigma}{\sqrt{n}}$

è detto **errore standard della media** ed è "**indice dell'imprecisione**" con la quale la media campionaria

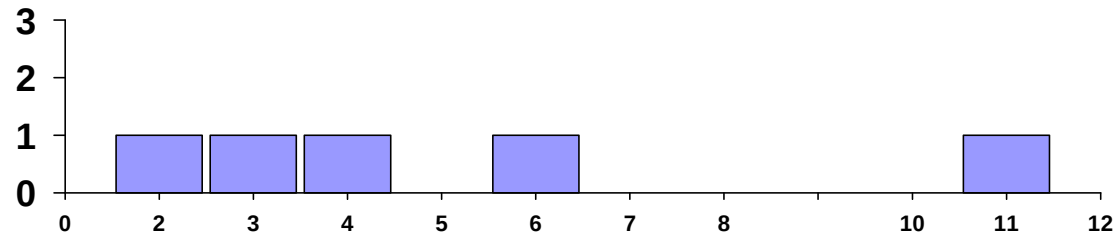
$(\bar{x})$ **stima** la media (μ) della popolazione

40

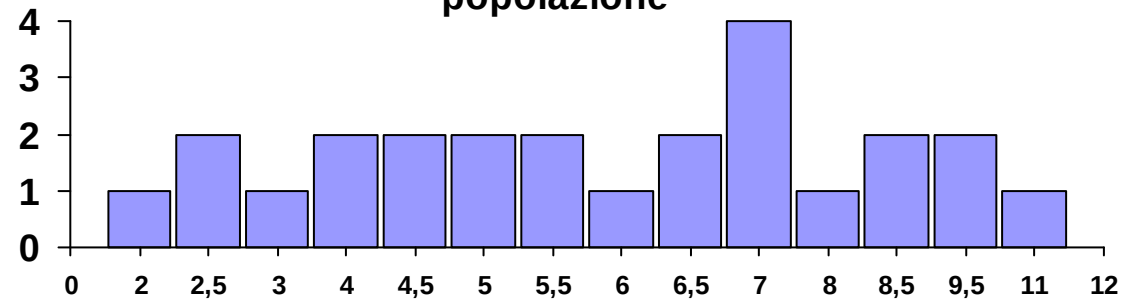
ESEMPIO $N=5$ $n=2$

$$\{X\}_{N=5} = \{X: 2, 3, 5, 7, 11\}$$

Distribuzione della variabile X nella popolazione



Distribuzione della v.c. media campionaria nella popolazione



Ed ora ci chiediamo "quali sono tutti i possibili campioni d'ampiezza ($n=3$) che possono venire estratti da tale popolazione?"

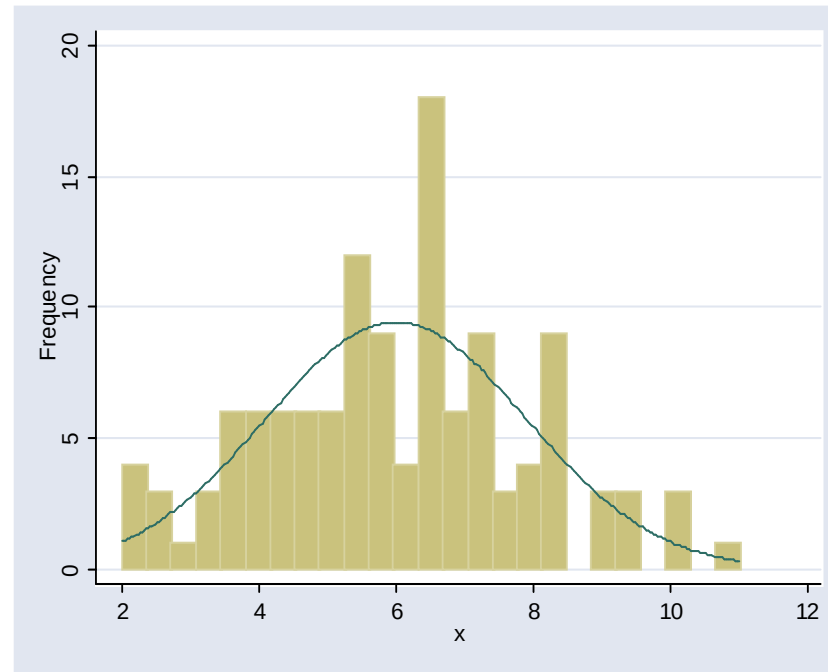
41 possibili campioni d'ampiezza (n=3)

(2, 2, 2)	(3, 2, 2)	(6, 2, 2)	(8, 2, 2)	(11, 2, 2)
(2, 2, 3)	(3, 2, 3)	(6, 2, 3)	(8, 2, 3)	(11, 2, 3)
(2, 2, 6)	(3, 2, 6)	(6, 2, 6)	(8, 2, 6)	(11, 2, 6)
(2, 2, 8)	(3, 2, 8)	(6, 2, 8)	(8, 2, 8)	(11, 2, 8)
(2, 2,11)	(3, 2,11)	(6, 2,11)	(8, 2,11)	(11, 2,11)
(2, 3, 2)	(3, 3, 2)	(6, 3, 2)	(8, 3, 2)	(11, 3, 2)
(2, 3, 3)	(3, 3, 3)	(6, 3, 3)	(8, 3, 3)	(11, 3, 3)
(2, 3, 6)	(3, 3, 6)	(6, 3, 6)	(8, 3, 6)	(11, 3, 6)
(2, 3, 8)	(3, 3, 8)	(6, 3, 8)	(8, 3, 8)	(11, 3, 8)
(2, 3,11)	(3, 3,11)	(6, 3,11)	(8, 3,11)	(11, 3,11)
(2, 6, 2)	(3, 6, 2)	(6, 6, 2)	(8, 6, 2)	(11, 6, 2)
(2, 6, 3)	(3, 6, 3)	(6, 6, 3)	(8, 6, 3)	(11, 6, 3)
(2, 6, 6)	(3, 6, 6)	(6, 6, 6)	(8, 6, 6)	(11, 6, 6)
(2, 6, 8)	(3, 6, 8)	(6, 6, 8)	(8, 6, 8)	(11, 6, 8)
(2, 6,11)	(3, 6,11)	(6, 6,11)	(8, 6,11)	(11, 6,11)
(2, 8, 2)	(3, 8, 2)	(6, 8, 2)	(8, 8, 2)	(11, 8, 2)
(2, 8, 3)	(3, 8, 3)	(6, 8, 3)	(8, 8, 3)	(11, 8, 3)
(2, 8, 6)	(3, 8, 6)	(6, 8, 6)	(8, 8, 6)	(11, 8, 6)
(2, 8, 8)	(3, 8, 8)	(6, 8, 8)	(8, 8, 8)	(11, 8, 8)
(2, 8,11)	(3, 8,11)	(6, 8,11)	(8, 8,11)	(11, 8,11)
(2,11, 2)	(3,11, 2)	(6,11, 2)	(8,11, 2)	(11,11, 2)
(2,11, 3)	(3,11, 3)	(6,11, 3)	(8,11, 3)	(11,11, 3)
(2,11, 6)	(3,11, 6)	(6,11, 6)	(8,11, 6)	(11,11, 6)
(2,11, 8)	(3,11, 8)	(6,11, 8)	(8,11, 8)	(11,11, 8)
(2,11,11)	(3,11,11)	(6,11,11)	(8,11,11)	(11,11,11)

La distribuzione di frequenza delle medie campionarie per (n=3)

x	Freq.	Percent	Cum.
2	1	0.80	0.80
2.3333	3	2.40	3.20
2.6666	3	2.40	5.60
3	1	0.80	6.40
3.3333	3	2.40	8.80
3.6666	6	4.80	13.60
4	6	4.80	18.40
4.3333	6	4.80	23.20
4.6666	6	4.80	28.00
5	6	4.80	32.80
5.3333	12	9.60	42.40
5.6666	9	7.20	49.60
6	4	3.20	52.80
6.3333	9	7.20	60.00
6.6666	9	7.20	67.20
7	6	4.80	72.00
7.3333	9	7.20	79.20
7.6666	3	2.40	81.60
8	4	3.20	84.80
8.3333	9	7.20	92.00
9	3	2.40	94.40
9.3333	3	2.40	96.80
10	3	2.40	99.20
11	1	0.80	100.00
	125	100.00	

$$\{x\}_{N=5} = \{2, 3, 5, 7, 11\} \quad n=3 \quad \text{pop}=5^3$$



$$\text{Somma } (x) = 750$$

$$\text{Devianza di } x = 450$$

$$\text{Media}(\bar{x}) = \frac{750}{125} = 6$$

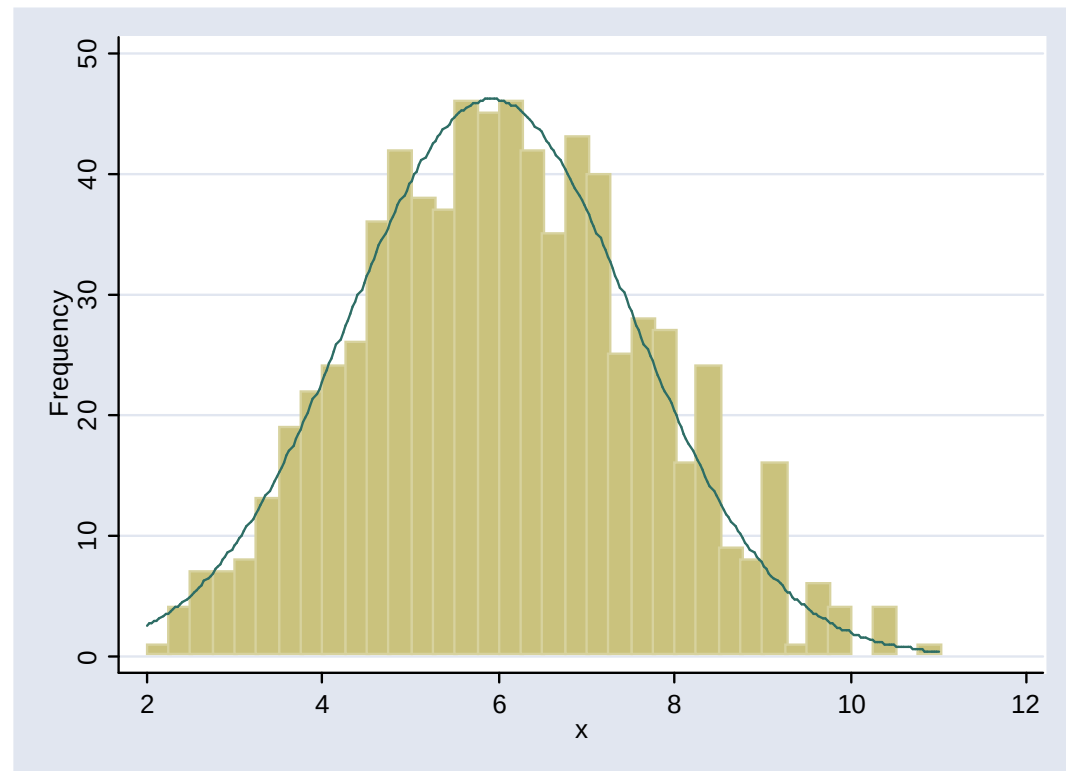
$$\text{Var}(\bar{x}) = \frac{\sigma^2}{n} = \frac{450}{125} = \frac{10,8}{3} = 3.6$$

43 x	Freq.	Percent	Cum.
2	1	0.13	0.13
2.25	4	0.53	0.67
2.5	7	0.93	1.60
2.75	7	0.93	2.53
3	8	1.07	3.60
3.25	13	1.73	5.33
3.5	19	2.53	7.87
3.75	22	2.93	10.80
4	24	3.20	14.00
4.25	26	3.47	17.47
4.5	36	4.80	22.27
4.75	42	5.60	27.87
5	38	5.07	32.93
5.25	37	4.93	37.87
5.5	46	6.13	44.00
5.75	45	6.00	50.00
6	46	6.13	56.13
6.25	42	5.60	61.73
6.5	35	4.67	66.40
6.75	43	5.73	72.13
7	40	5.33	77.47
7.25	25	3.33	80.80
7.5	28	3.73	84.53
7.75	27	3.60	88.13
8	16	2.13	90.27
8.25	24	3.20	93.47
8.5	9	1.20	94.67
8.75	8	1.07	95.73
9	16	2.13	97.87
9.25	1	0.13	98.00
9.5	6	0.80	98.80
9.75	4	0.53	99.33
10.25	4	0.53	99.87
11	1	0.13	100.00
Total	750	100.00	

La distribuzione di frequenza
delle medie campionarie per (n=4) :

$$\{x\}_{N=5} = \{2, 3, 5, 7, 11\}$$

$$n=4 \quad \text{pop}=5^4$$



“STIME” CAMPIONARIE PUNTUALI

Supponiamo che x sia una variabile d'interesse in una vasta popolazione reale.

Consideriamo la **media campionaria** di x in un qualsiasi campione casuale di dimensione n .

Due caratteristiche appaiono intuitivamente chiare:

1. Più la variabile casuale x , in quella popolazione, è mutevole altrettanto lo sono le **medie campionarie** dei distinti campioni casuali di dimensione n .
2. La **variabilità della media campionaria** di x dipende da n : più grandi sono i campioni casuali tanto più ci si aspetta che **i valori delle medie campionarie** siano vicini tra di loro.

CAMPIONAMENTO SENZA RIPETIZIONE

di campioni d'ampiezza (n=2)

I possibili campioni d'ampiezza (n=2) sono

(2 , 3)	(2 , 6)	(2 , 8)	(2 ,11)
	(3 , 6)	(3 , 8)	(3 ,11)
		(6 , 8)	(6 ,11)
			(8 ,11)

Le medie corrispondenti sono ...

2.5	4.0	5.0	6.5
	4.5	5.5	7.0
		7.0	8.5
			9.5

$$\bar{\bar{x}} = (\Sigma \bar{x})/10 = 6.0$$

CAMPIONAMENTO SENZA RIPETIZIONE

di campioni d'ampiezza (n=2)

I componenti della devianza sono ...

$(2.5-6.0)^2$	$(4.0-6.0)^2$	$(5.0-6.0)^2$	$(6.5-6.0)^2$
	$(4.5-6.0)^2$	$(5.5-6.0)^2$	$(7.0-6.0)^2$
		$(7.0-6.0)^2$	$(8.5-6.0)^2$
			$(9.5-6.0)^2$

Gli elementi della devianza sono ...

$(-3.5)^2$	$(-2.0)^2$	$(-1.0)^2$	$(0.5)^2$
	$(-1.5)^2$	$(-0.5)^2$	$(1.0)^2$
		$(1.0)^2$	$(2.5)^2$
			$(3.5)^2$

$$\Sigma(\bar{x} - \bar{\bar{x}})^2 / 10 = 4.5$$

47 CAMPIONAMENTO SENZA RIPETIZIONE

La varianza delle medie campionaria $\Sigma(\bar{x} - \bar{\bar{x}})^2 / 10 = 4.5$

da cui l'errore standard è $\sigma_{\bar{x}} = 2.01$

Questo illustra il fatto che $\sigma_{\bar{x}}^2 = \frac{\sigma^2}{n} \cdot \left(\frac{N - n}{N - 1} \right)$

Che nel nostro esempio è $\sigma_{\bar{x}}^2 = \frac{10,8}{2} \cdot \left(\frac{5 - 2}{5 - 1} \right)$