Cenni teoria dei campioni teorema limite centrale

SAPIENZA
UNIVERSITÀ DI ROMA

annarita.vestri@uniroma1.it

Il campione che viene estratto è soltanto uno dei numerosissimi campioni possibili (tutti di uguale numerosità n).

Al variare del campione varieranno anche le unità statistiche raccolte e di conseguenza gli indici sintetici (statistiche) ottenuti dall’elaborazione dei dati campionari.

Come esemplificazione facciamo riferimento alla media campionaria

Campioni diversi daranno luogo a medie generalmente diverse, ma alcune medie campionarie, pur provenendo da campioni diversi, saranno uguali fra loro.

L’insieme di tutte le possibili medie campionarie, ottenibili da tutti i campioni di uguale numerosità n, darà luogo ad una distribuzione di medie campionarie.

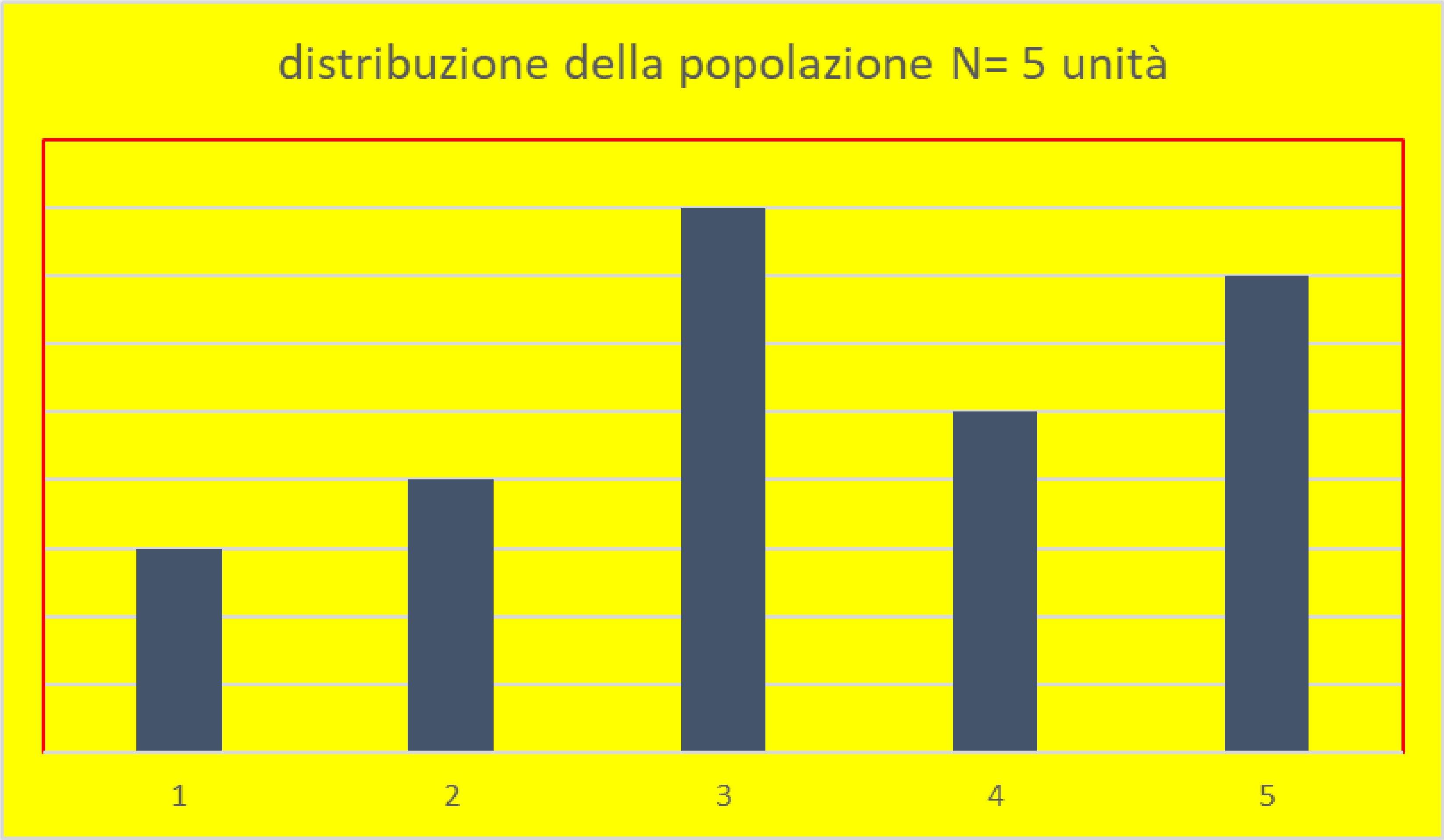
esempio

una popolazione composta da N unità,
$\{x\}_{N=5} = \{6, 8, 16, 14, 10\}$

	x_i	n_i	$(x_i - \mu)$	$(x_i - \mu)^2$
	6	1	-4	16
	8	1	-6	4
	10	1	0	0
	16	1	6	4
	14	1	4	16
Σ	50	5	0	40

μ= 50/5=10

σ^2= 40/5=8


distribuzione della popolazione N= 5 unità
1
2
3
4
5

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

(6 , 6)	(6 , 8)	(6 , 10)	(6 , 12)	(6 ,14)
(8 , 6)	(8 , 8)	(8 , 10)	(8 , 12)	(8 ,14)
(10, 6)	(10, 8)	(10, 10)	(10 12)	(10,14)
(12, 6)	(12, 8)	(12 10)	(12, 12)	(12,14)
(14,6)	(14,8)	(14,10)	(14,12)	(14,14)

… e le medie campionarie sono …

6	7	8	9	10
7	8	9	10	11
8	9	10	11	12
9	10	11	12	13
10	11	12	13	14

campionamento
con ripetizione

DISTRIBUZIONE DELLE MEDIE CAMPIONARIE

x_i	n_i	freq rel	$x_i n_i$	dev
6	1	1/25	6	16
7	2	2/25	14	18
8	3	3/25	24	12
9	4	4/25	36	4
10	5	1/25	50	0
11	4	4/25	44	4
12	3	3/25	36	12
13	2	2/25	26	18
14	1	1/25	14	16
	25	1	250	100

Dalla distribuzione di frequenza delle medie campionarie si calcola:

la media aritmetica della media campionaria.

Media =250/25=10 = μ

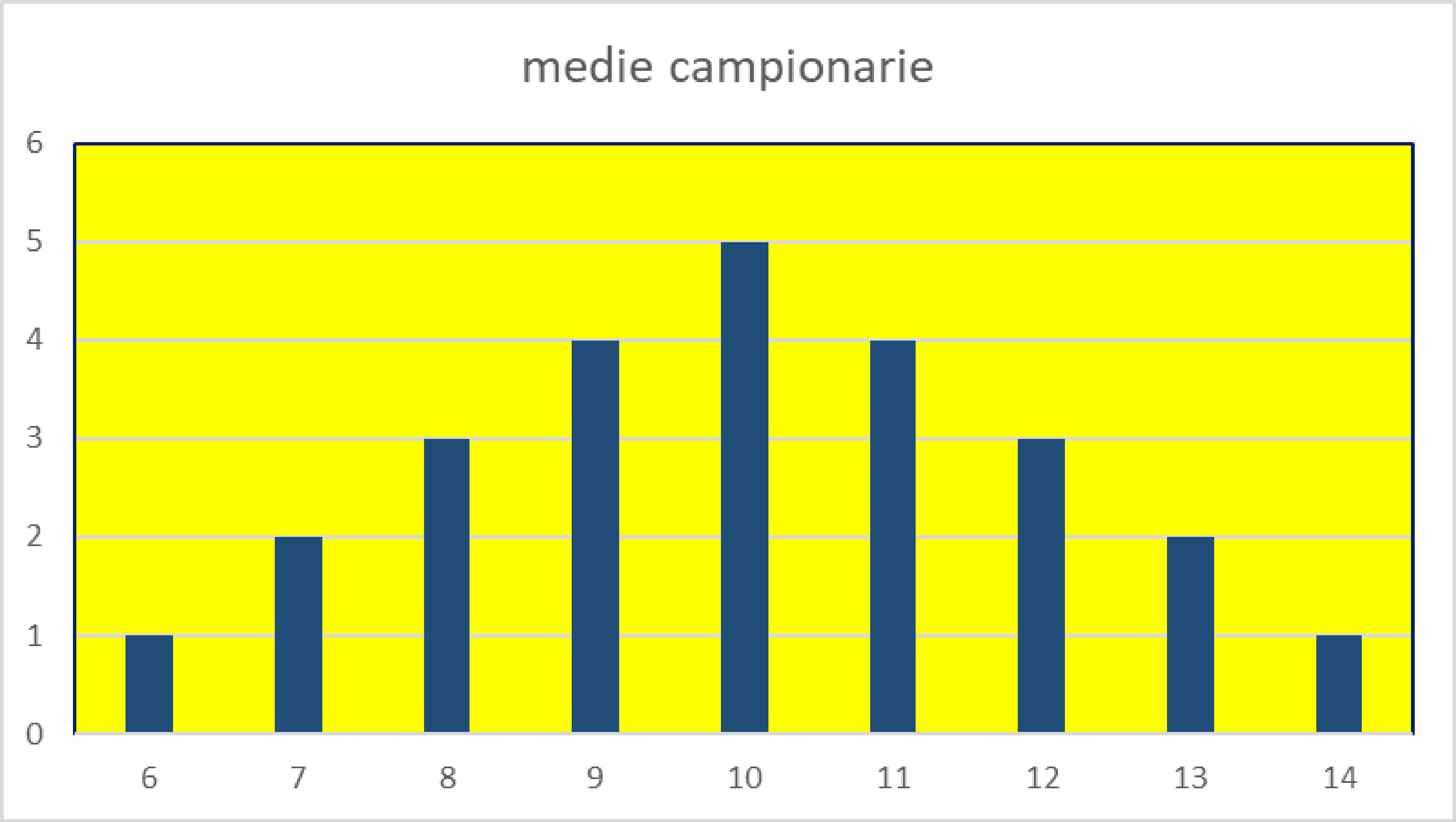
Solo 1 dei 25 possibili campioni ha media coincidente con la vera media della popolazione.

la varianza della media campionaria

Var(x) = 100/25=4

Var(x) = σ^2/n=8/2=4

La varianza della distribuzione di campionamento è uguale alla varianza della popolazione divisa per la dimensione del campione


medie campionarie
6
5
4
3
2
1
0
6
7
8
9
10
11
12
13
14

Si può dimostrare che:

se nella popolazione la variabile considerata si distribuisce normalmente, anche la **distribuzione di medie campionarie seguirà la curva normale**

- Se nella popolazione la variabile considerata non si distribuisce normalmente, la distribuzione di medie campionarie…
- per grandi campioni ($n > 30$) tenderà alla distribuzione normale
- per piccoli campioni ($n < 30$) non seguirà la distribuzione normale

Disporre di una distribuzione campionaria che segua la curva normale è fondamentale per l'inferenza statistica in quanto ci permette di stimare i valori caratteristici di una popolazione partendo da quelli di un suo campione.

Pertanto, dovendo studiare la media campionaria di una variabile non normale, o di una variabile della quale non si conosce il tipo di distribuzione, sarebbe opportuno, se possibile, aumentare la numerosità campionaria almeno fino a 30 unità.

Nell’Esempio abbiamo visto come sebbene la variabile X nella popolazione sia chiaramente non normale, la distribuzione di medie campionarie tenda comunque a normalizzare.

Si può dimostrare che

$$M(\bar{x}_i) = \mu$$

La media delle medie campionarie coincide con μ (la media della popolazione)

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

La SD delle medie campionarie è uguale alla SD della popolazione divisa la radice quadrata di n

Questo è l’errore standard, descrive l’incertezza della stima della media cioè quanto le stime sono variabili intorno al valore da stimare.

- Nel caso frequente in cui σ sia ignoto, si sostituisce con la sua corrispondente stima campionaria s
- Si potranno avere pertanto 2 possibili ES: uno noto ed uno stimato

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

ES noto

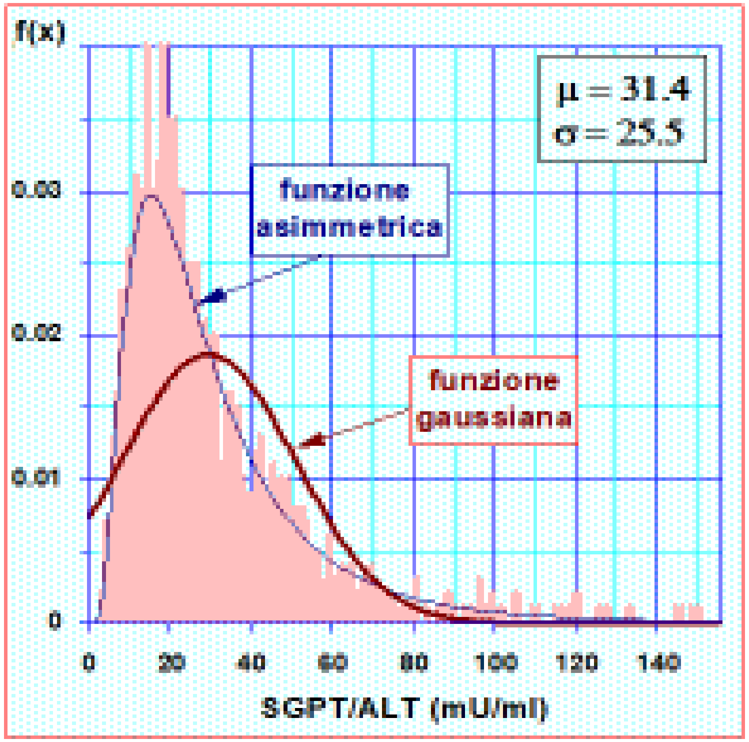
$$\hat{ES} = \hat{\sigma}(\bar{x}_i) = \frac{s}{\sqrt{n}}$$

ES stimato

Concludendo, posto che la distribuzione di medie campionarie sia normale, avremo al centro di tale distribuzione il valore μ (in genere ignoto) con la variabilità espressa dall'ES

DISTRIBUZIONE DELLE MEDIE CAMPIONARIE

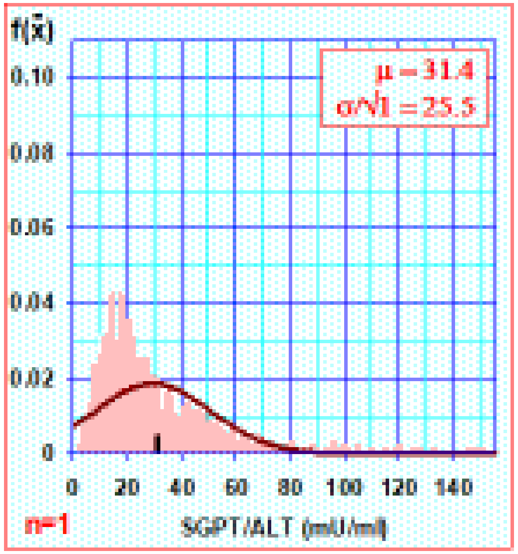
Dalla popolazione generale si sono estratti a caso 1000 soggetti maschi adulti, e su ciascuno di essi si è determinato il livello ematico di ALT (alanina amino-transferasi). È noto che, nella popolazione generale, la **distribuzione** di ALT è **fortemente asimmetrica positiva** per la presenza di individui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

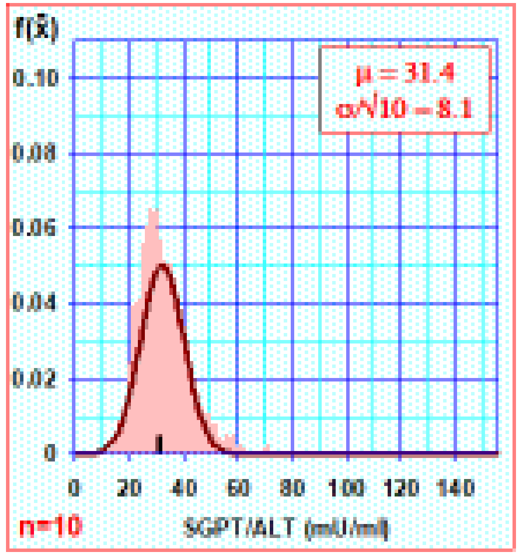


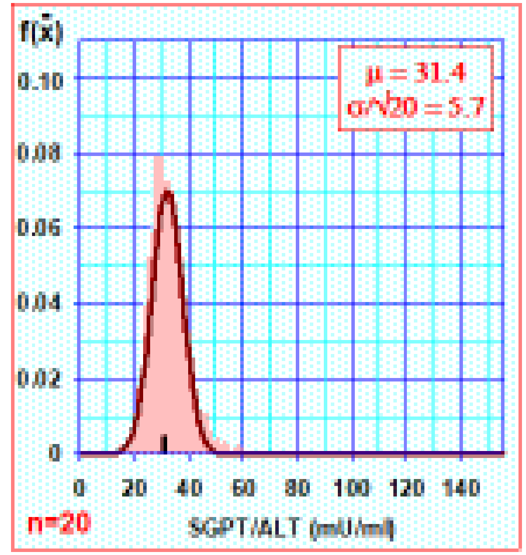
Dalla stessa popolazione sono poi estratti:

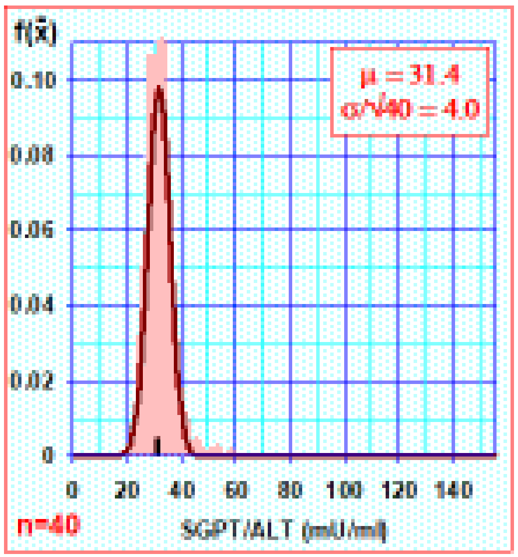
- **1000** campioni di dimensione **n = 10**
- **1000** campioni di dimensione **n = 60**
- **1000** campioni di dimensione **n = 40**

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	$\bar{x}$	$E(x)$	$S_{\bar{x}}$	$\frac{\sigma}{\sqrt{n}}$
1	29.1	31.4	21.4	25.5
10	31.7	31.4	7.9	8.1
20	31.7	31.4	5.7	5.7
40	31.3	31.4	4.1	4.0

Quanto sopra mostrato con un esempio è dimostrato da un teorema detto « teorema del limite centrale » il cui enunciato può essere così espresso:

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 $\bar{x}$

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