Metodologia Statistica Applicata in Ambito Biomedico e Clinico

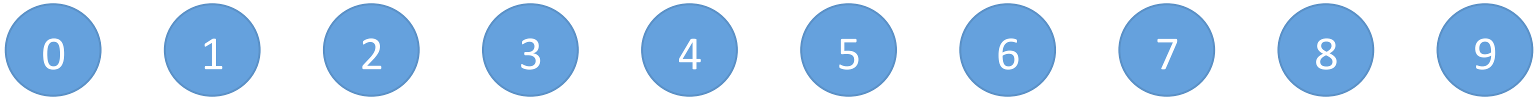
TERZA PARTE

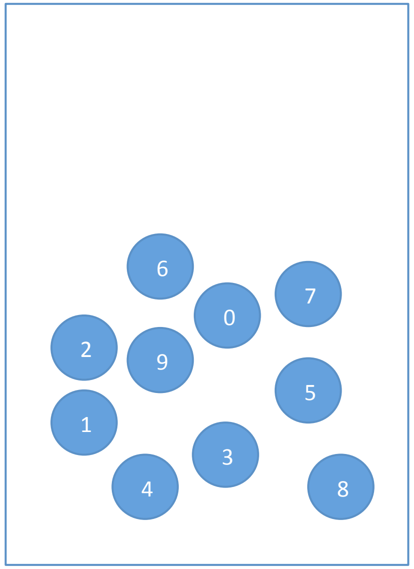
Mauro Gambaccini

Anno accademico 2018/19

DISTRIBUZIONI CAMPIONARIE

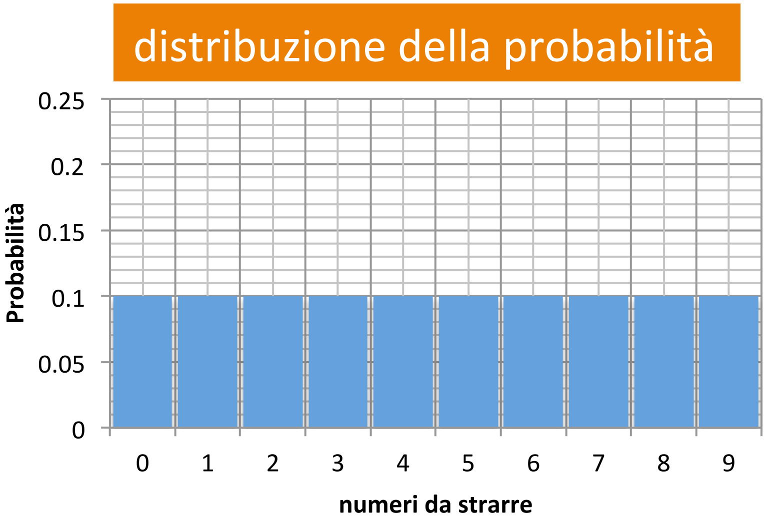
Prendiamo 10 palline uguali numerate da 0 a 9





Le mettiamo in un'urna ed estraiamo 100 volte rimettendo
Sempre nell'urna la pallina estratta

Per ogni numero la probabilità di essere estratto è
sempre la stessa p= 1/10 = 0.1



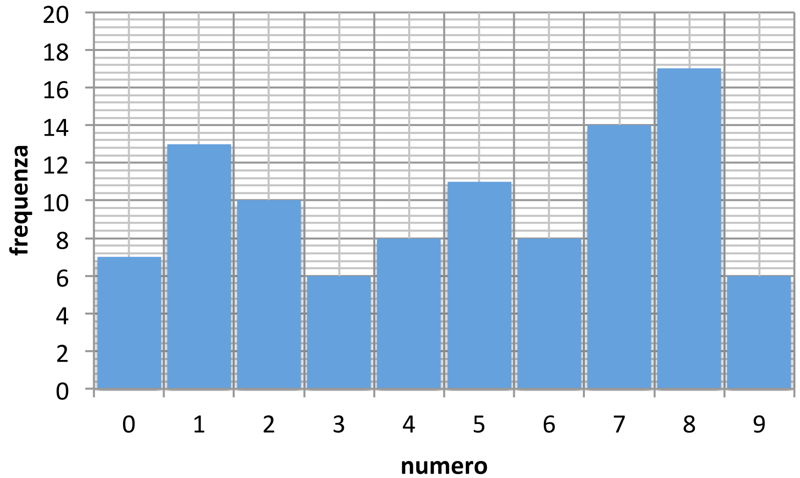
DISTRIBUZIONI CAMPIONARIE

CENTO numeri estratti casualmente tra 0 e 9									
9	1	0	7	5	6	9	5	8	8
1	0	5	7	6	5	0	2	1	2
1	8	8	8	5	2	4	8	3	1
6	5	5	7	4	1	7	3	3	3
2	8	1	8	5	8	4	0	1	9
2	1	6	9	4	4	7	6	1	7
1	9	7	9	7	2	7	7	0	8
1	6	3	8	0	5	7	4	8	6
7	0	2	8	8	7	2	5	4	1
8	6	8	3	5	8	2	7	2	4

I 100 numeri estratti sono riportati nella tabella a sinistra e come potete vedere essi non vengono estratti lo stesso numero di volte anche se a priori hanno la stessa probabilità di essere estratti

DISTRIBUZIONE DELLA FREQUENZA DEI NUMERI ESTRATTI	
numero	frequenza
0	7
1	13
2	10
3	6
4	8
5	11
6	8
7	14
8	17
9	6

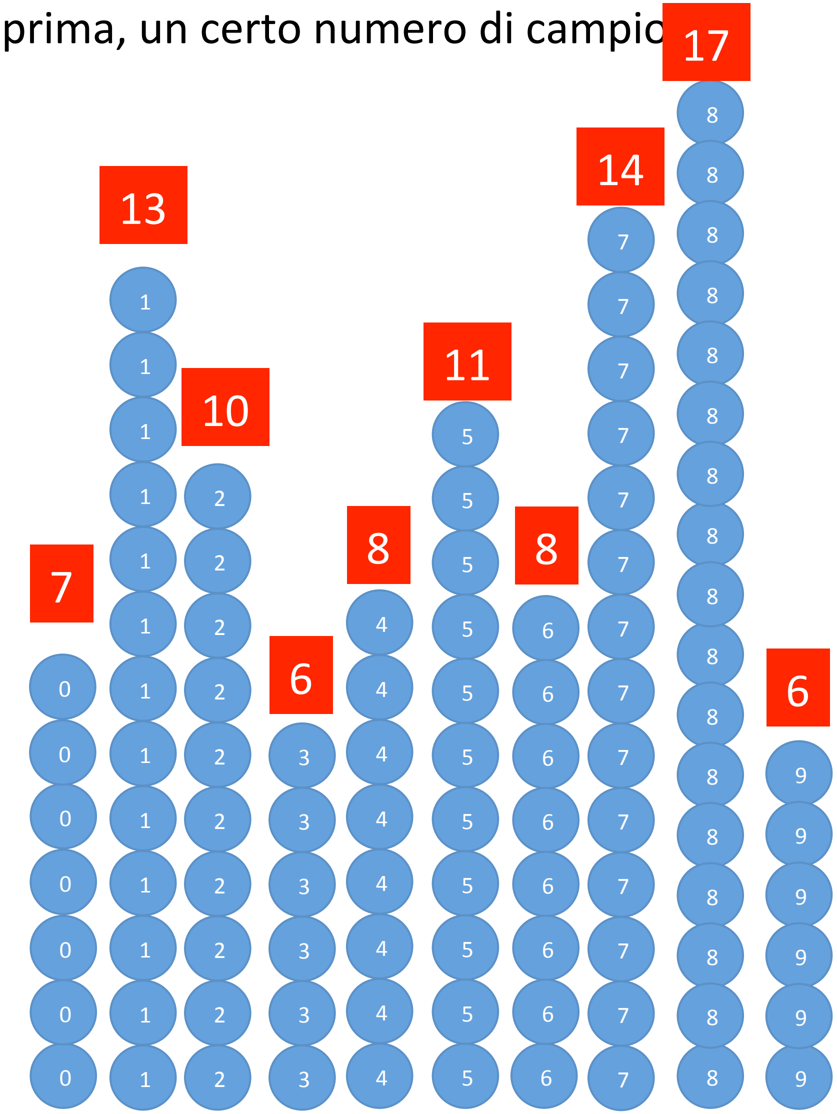
DISTRIBUZIONE DELLA FREQUENZA DEI NUMERI ESTRATTI



DISTRIBUZIONI CAMPIONARIE

Facciamo diventare questi 100 dati estratti casualmente la nostra popolazione ovvero quel gruppo di oggetti (la statistica) che vogliamo studiare e per farlo estrarremo da essa, non più dall'urna di prima, un certo numero di campio

CENTO numeri estratti casualmente tra 0 e 9									
9	1	0	7	5	6	9	5	8	8
1	0	5	7	6	5	0	2	1	2
1	8	8	8	5	2	4	8	3	1
6	5	5	7	4	1	7	3	3	3
2	8	1	8	5	8	4	0	1	9
2	1	6	9	4	4	7	6	1	7
1	9	7	9	7	2	7	7	0	8
1	6	3	8	0	5	7	4	8	6
7	0	2	8	8	7	2	5	4	1
8	6	8	3	5	8	2	7	2	4



Ora la nostra urna conterrà 100 palline
Con la distribuzione riportata qui a destra e ogni numero avrà una propria probabilità di scire.
Nel caso specifico il n. 4 ed il n. 6 hanno la stessa probabilità: p= 8/100 = 0.08

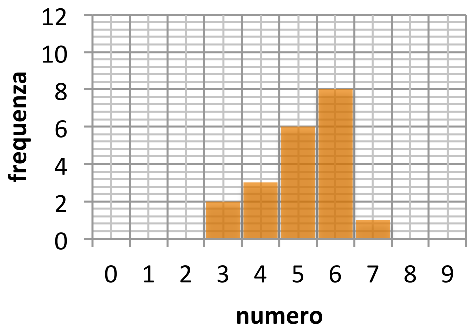
Estraiamo dall’urna, sempre reinserendo il numero uscito, 20 campioni ognuno composto da 4 numeri, per ogni gruppo calcoliamo la media ottenendo i risultati riportati in tabella

campione	1		2		3		4		5
	6		7		7		1		5
	4		8		9		8		2
	6		1		2		8		9
	1		8		7		4		5
media	4.25		6.00		6.25		5.25		5.25
campione	6		7		8		9		10
	5		4		7		2		8
	5		2		4		8		1
	7		7		0		7		2
	8		6		1		7		0
media	6.25		4.75		3.00		6.00		2.75
campione	11		12		13		14		15
	7		7		2		8		3
	8		3		5		0		7
	7		8		0		7		4
	2		7		8		7		8
media	6.00		6.25		3.75		5.50		5.50
campione	16		17		18		19		20
	4		5		4		4		7
	8		5		3		5		4
	7		8		1		8		6
	7		3		6		2		3
media	6.50		5.25		3.50		4.75		5.00

campione	media campionaria
1	2.75
2	3.00
3	3.50
4	3.75
5	4.25
6	4.75
7	4.75
8	5.00
9	5.25
10	5.25
11	5.25
12	5.50
13	5.50
14	6.00
15	6.00
16	6.00
17	6.25
18	6.25
19	6.25
20	6.50

distribuzione delle 20 medie campionarie

classe	numero	frequenza
0 - 0.5	0	0
0.5 - 1.5	1	0
1.5 - 2.5	2	0
2.5 - 3.5	3	2
3.5 - 4.5	4	3
4.5 - 5.5	5	6
5.5 - 6.5	6	8
6.5 - 7.5	7	1
7. 5- 8.5	8	0
8.5 - 9	9	0

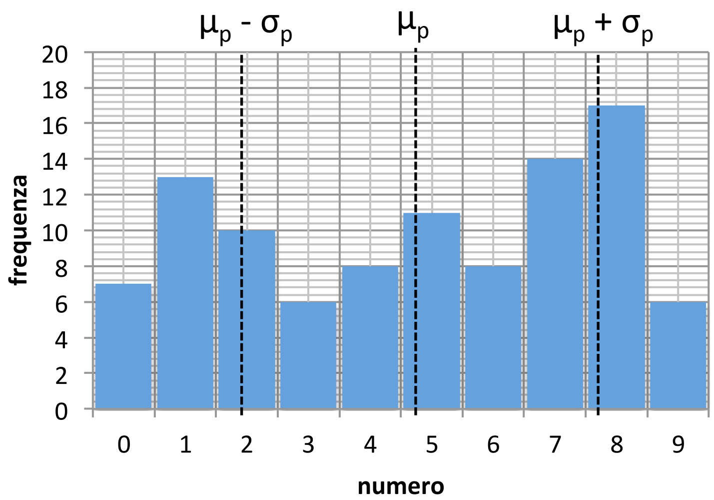


CONFRONTO

distribuzione della popolazione

numero	frequenza
0	7
1	13
2	10
3	6
4	8
5	11
6	8
7	14
8	17
9	6

DISTRIBUZIONE DELLA POPOLAZIONE

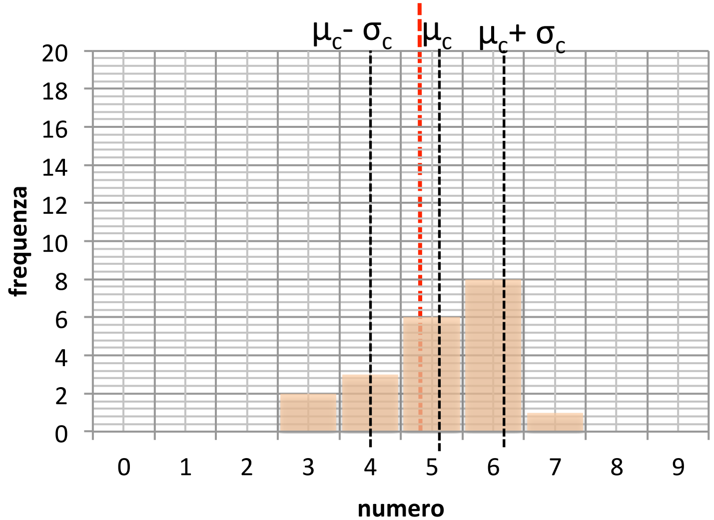


POPOLAZIONE	
Media (μ_p)	4.7
d.s. (σ_p)	2.9

distribuzione delle 20 medie campionarie

classe	numero	frequenza
0 - 0.5	0	0
0.5 - 1.5	1	0
1.5 - 2.5	2	0
2.5 - 3.5	3	2
3.5 - 4.5	4	3
4.5 - 5.5	5	6
5.5 - 6.5	6	8
6.5 - 7.5	7	1
7. 5- 8.5	8	0
8.5 - 9	9	0

DISTRIBUZIONE DELLE MEDIE CAMPIONARIE
20 campioni 4 dati/campione



MEDIE CAMPIONARIE	
Media (μ_c)	5.1
d.s. (σ_c)	1.1

Relazioni tra medie e d.s.		
μ_c / μ_p	5.1 /4.7	1.08
σ_p / σ_c	2.9/1.1	2.6

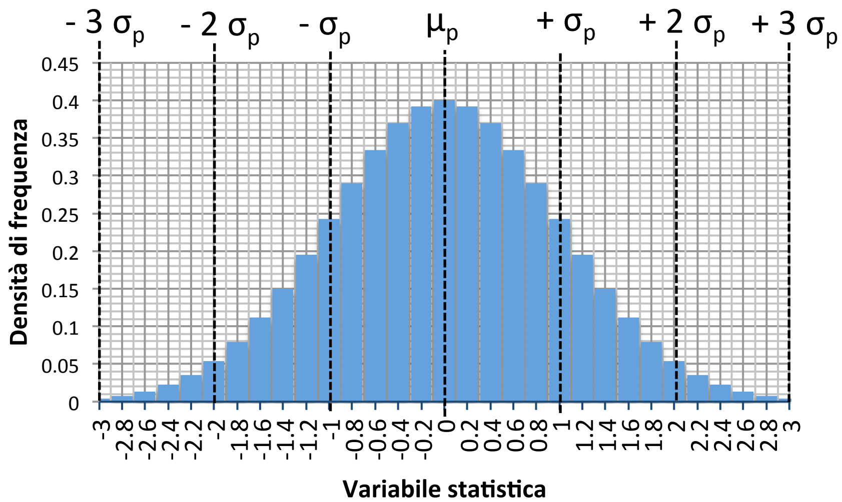
Importante notare che la media della popolazione μ_p cade dentro l'intervallo $\mu_c \pm \sigma_c$; questo sta ad indicare che la media campionaria è una buona stima della media della popolazione.
Meno accurata è la stima della d.s. della popolazione, possiamo solamente dire che è più del doppio della d.s. campoinaria

DISTRIBUZIONI CAMPIONARIE

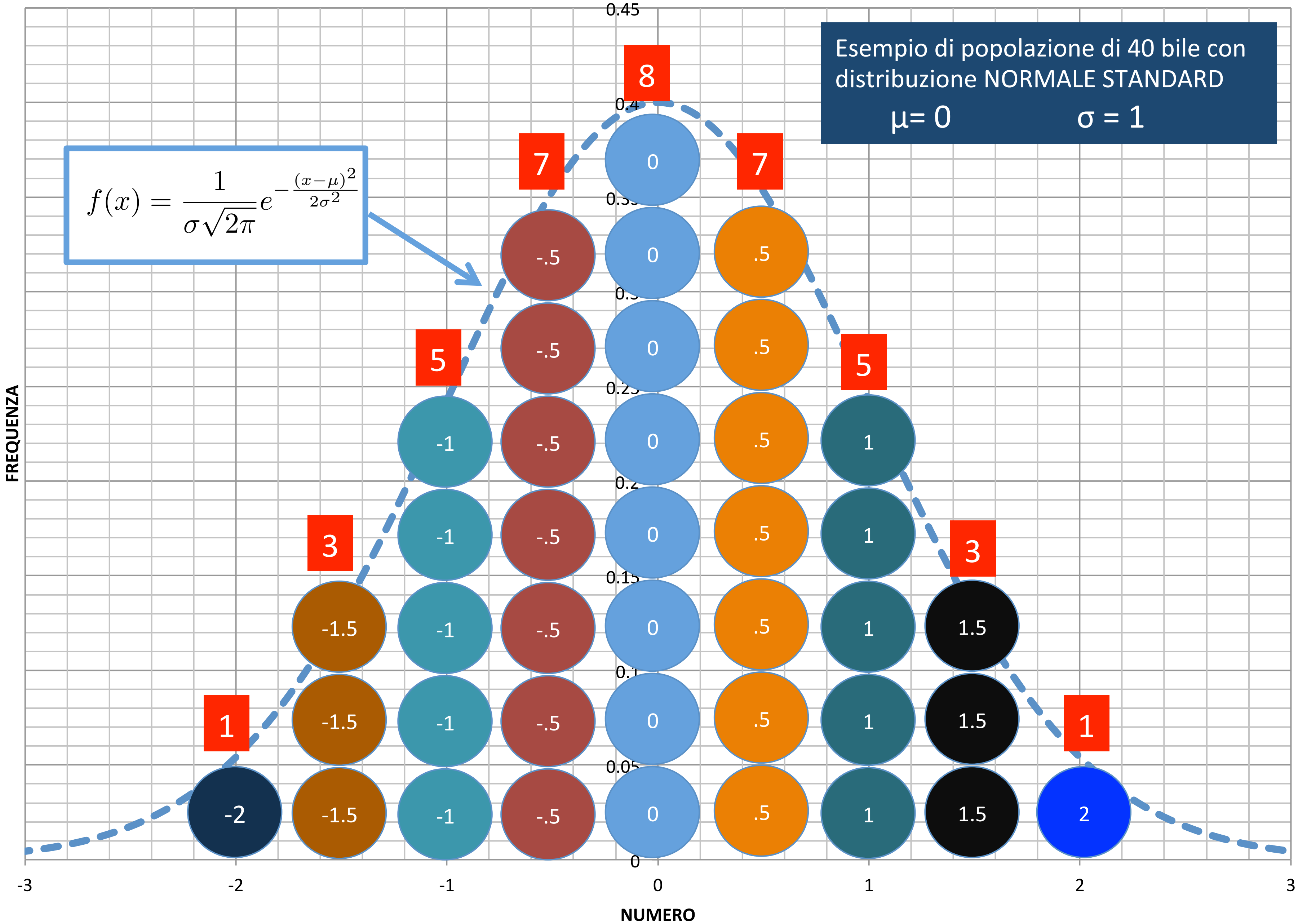
Abbiamo descritto il comportamento statistico dei campioni di dati raccolti da una popolazione originata da un variabile aleatoria con probabilità piatta $p_i = 0.1$

Ora ci chiediamo cosa succede alla media campionaria quando il parametro analizzato proviene da una distribuzione NORMALE ovvero GAUSSIANA
Per questo esempio utilizzeremo la funzione normale normalizzata ovvero la Gaussiana con:

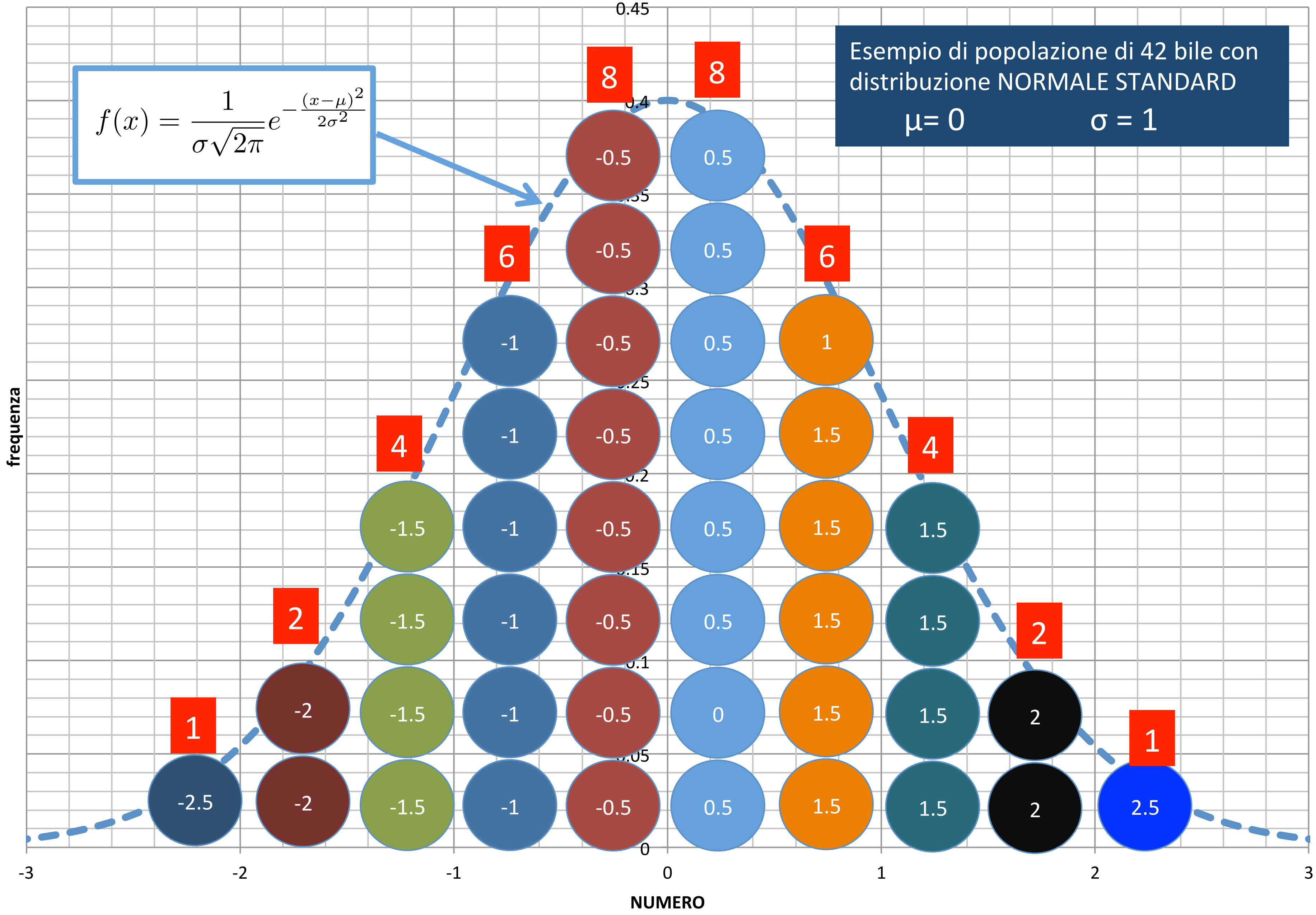
media = 0   e deviazione standard = 1



Come facciamo a produrre una distribuzione di probabilità con queste caratteristiche?

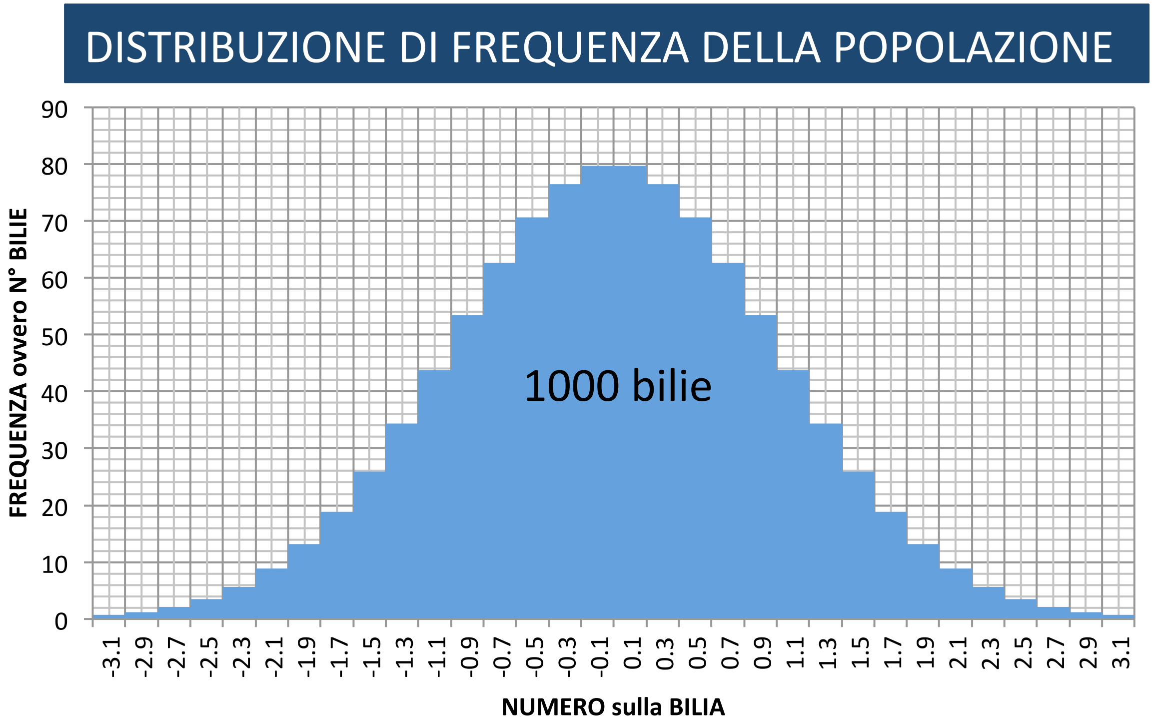


Esempio di popolazione di 40 bile con distribuzione NORMALE STANDARD
μ= 0 σ = 1
$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$
8
7
7
5
5
3
3
1
1
-2
-1.5
-1
-.5
0
.5
1
1.5
2
FREQUENZA
NUMERO
0.45
0.4
0.05
0
-3
-2
-1
0
1
2
3



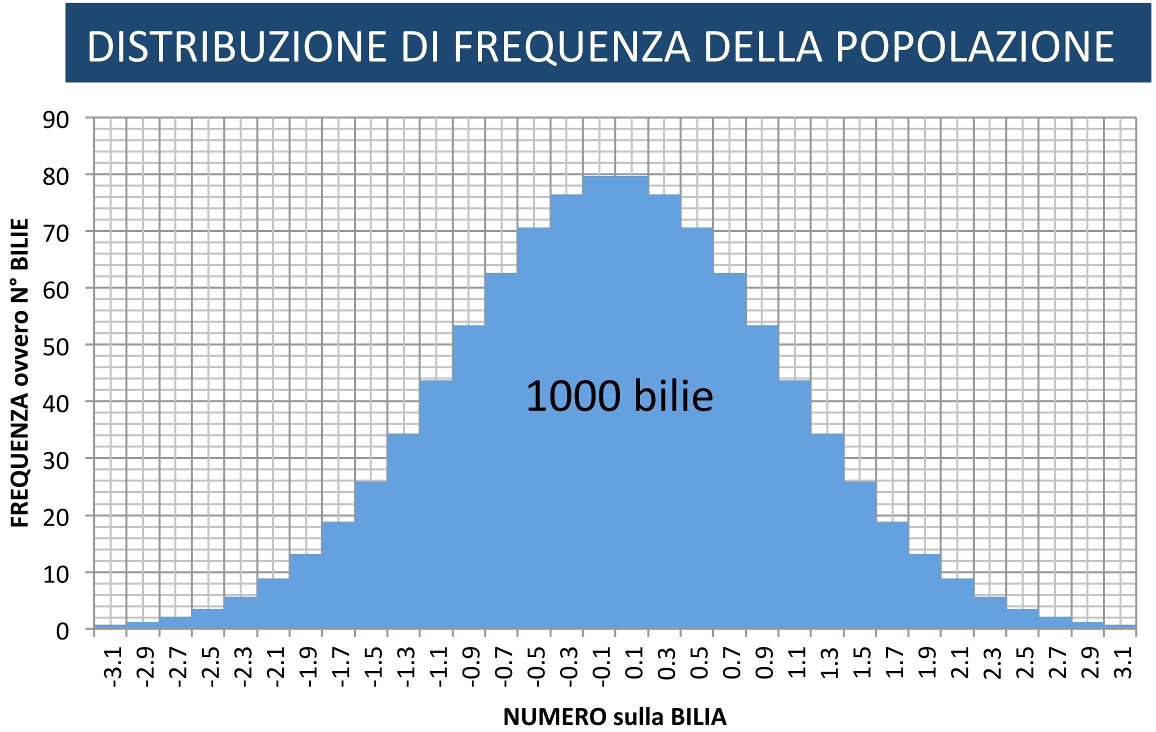
Esempio di popolazione di 42 bile con distribuzione NORMALE STANDARD
μ= 0
σ = 1
$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$
frequenza
NUMERO
0.45
0.4
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
-3
-2
-1
0
1
2
3
8
8
6
6
4
4
2
2
1
1
-0.5
0.5
-1
1
-1.5
1.5
-2
2
-2.5
2.5
0

La popolazione dalla quale estrarremo i campioni sarà costituita da 1000 bilie
su cui sono riportati i numeri tra – 3.1 a 3.1 con un differenza (ampiezza di classe) di 0.2



Ora la nostra urna conterrà 1000 palline con la distribuzione sopra riportata e ogni numero avrà una propria probabilità di essere estratto ad esempio la probabilità di estrarre la bilia con il numero - 0,5 ovvero $p_{0.5}$ = 70/1000 = 0.07 --> 7%
Poiché la distribuzione è simmetrica anche la bilia col numero 0.5 avrà la stessa probabilità di essere estratta e così via

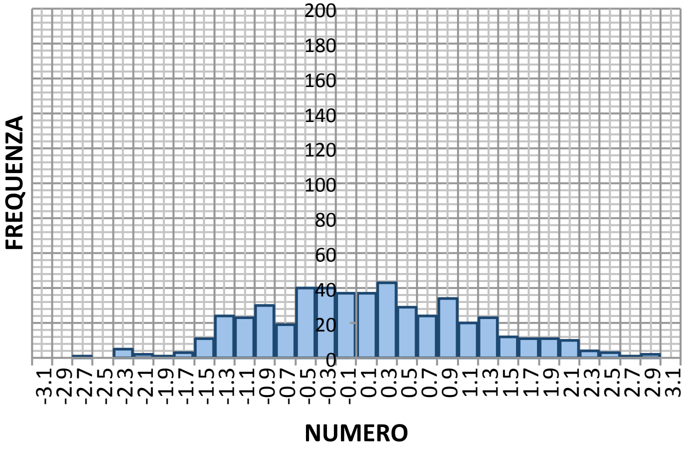
DISTRIBUZIONI CAMPIONARIE



Faremo 4 campionamenti :

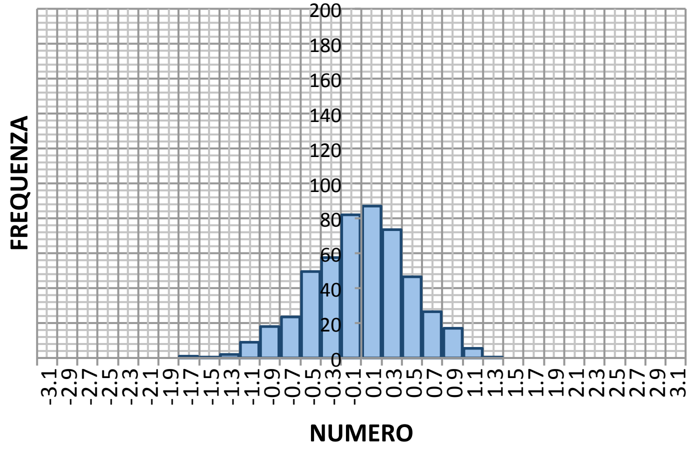
1) 500 estrazioni di una bilia	(campione a 1 elemento)
2) 500 estrazioni di gruppi da 4 bilie	(campione a 4 elementi)
3) 500 estrazioni di gruppi da 9 bilie	(campione a 9 elementi)
4) 500 estrazioni di gruppi da 16 bilie	(campione a 16 elementi)

500 osservazioni da una
Normale standard



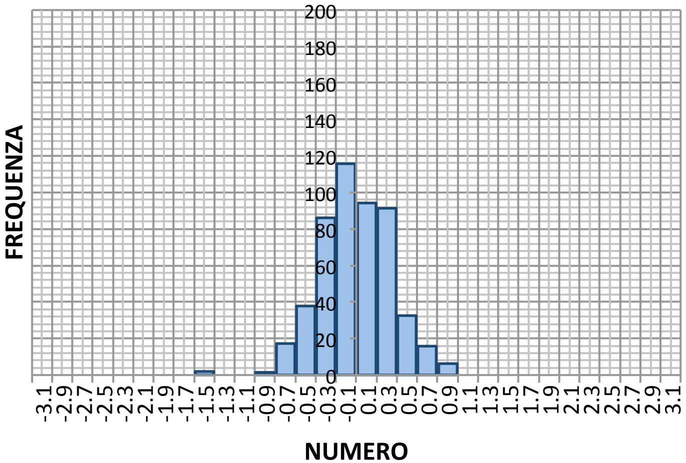
Media = -0.01 , dev. std. = 1.04 (≈ σ_p)

500 medie di 4 osservazioni
da una Normale standard



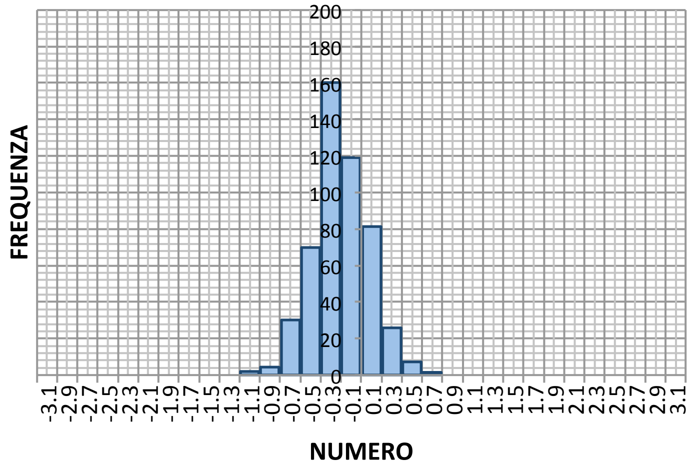
Media = 0.0 , dev. std. = 0.48 (≈ ½ σ_p)

500 medie di 9 osservazioni
da una Normale standard



Media = 0.0 , dev. std. = 0.35 (≈ 1/3 σ_p)

500 medie di 16 osservazioni
da una Normale standard



Media = 0.0 , dev. std. = 0.26 (≈ 1/4 σ_p)

All’aumentare del numero di osservazioni del campione la media campionaria Corrisponde alla media della popolazione mente la varianza invece diminuisce di un fattore pari all’inverso al numero delle osservazioni del campione

MEDIA $\mu_c = \mu_p$

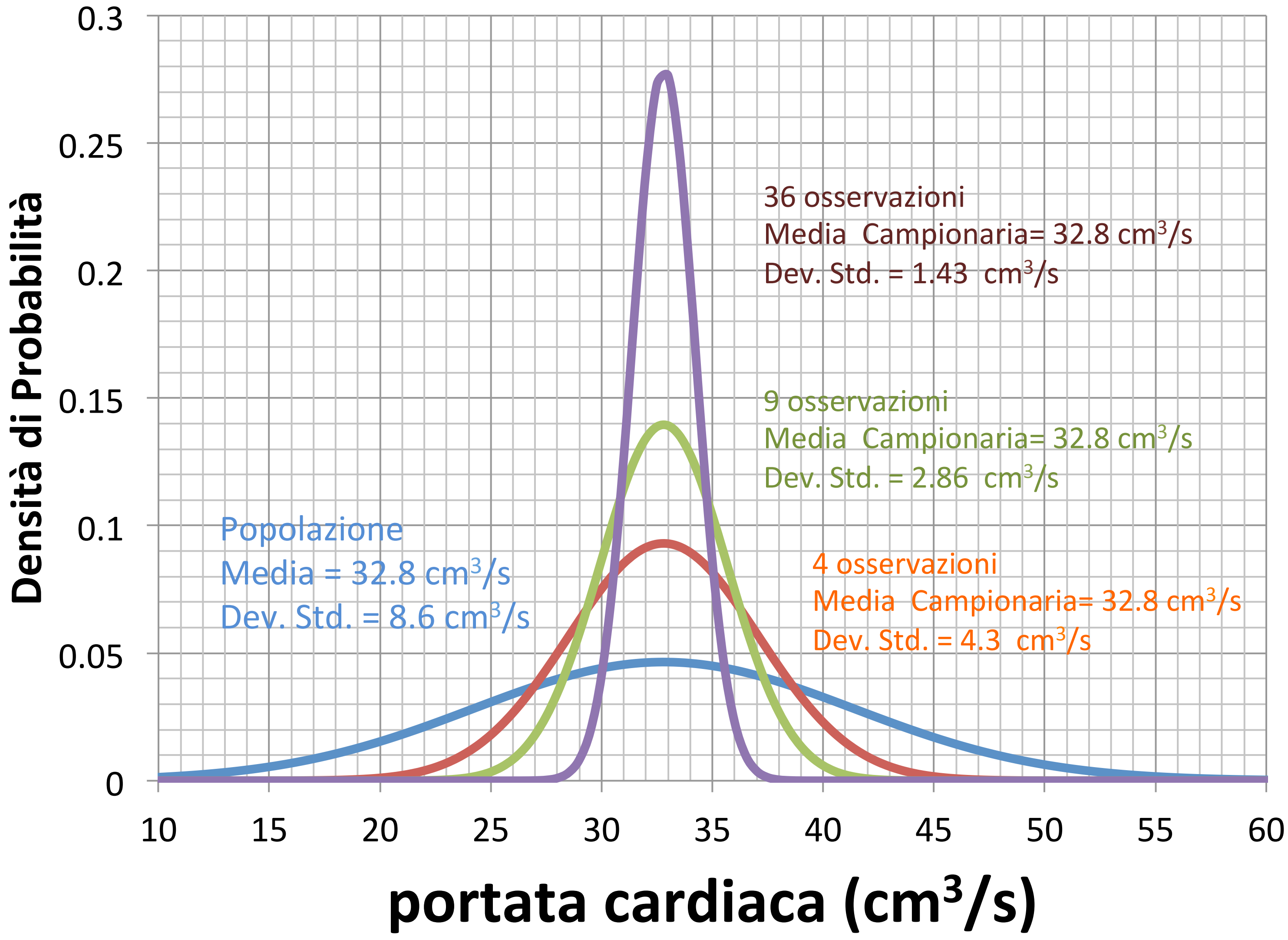
VARIANZA $\sigma_c^2 = \sigma_p^2 / N_c$

DEVIAZIONE STANDARD $\sigma_c = \sigma_p / \sqrt{N_c}$

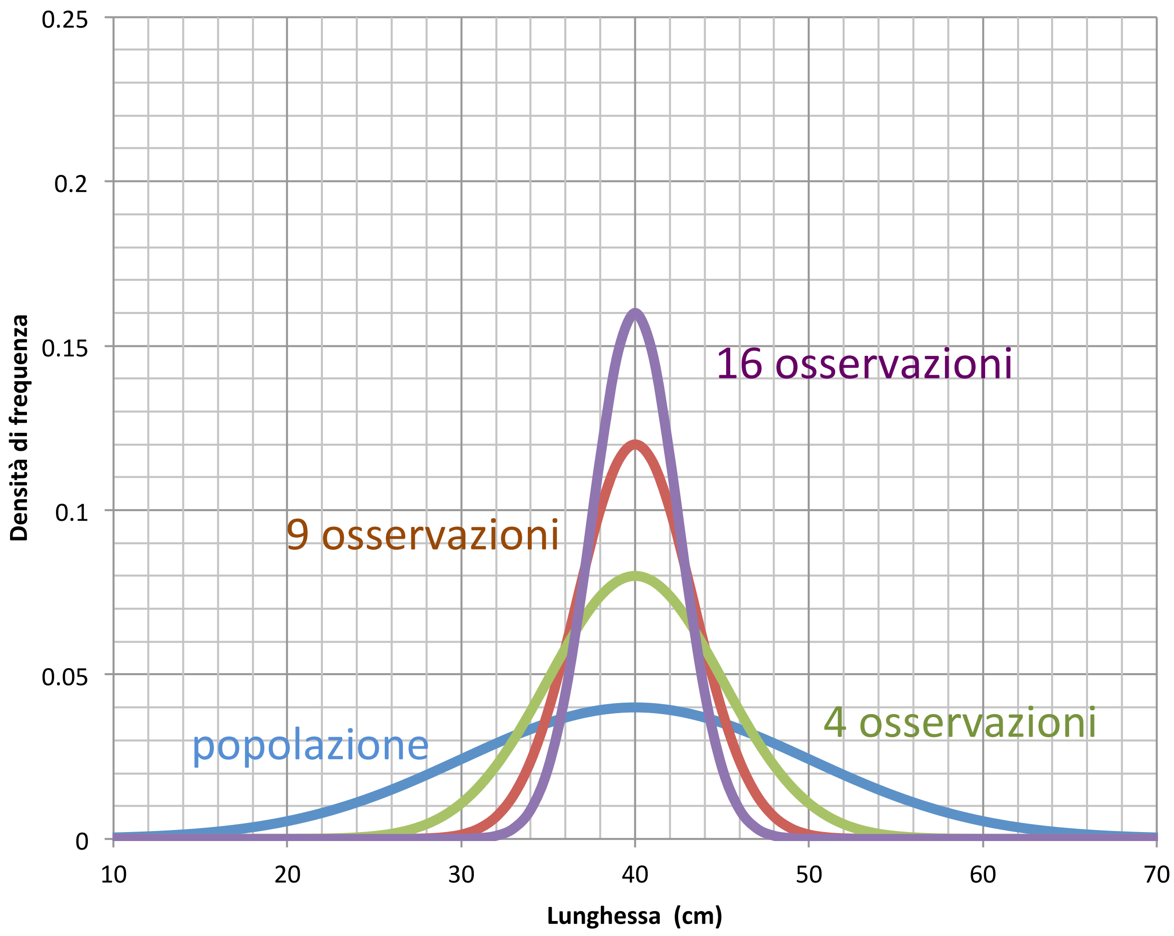
n.osservazioni	media	dev.std (σ_c)	$\sigma_p = \sigma_c \sqrt{N_c}$
1	-0.01	1.04	1.04 x 1 = 1.04
4	0.0	0.48	0.48 x 2 = 0.96
9	0.0	0.35	0.35 x 3 = 1.05
19	0.0	0.26	0.24 x 4 = 0.98
popolazione	**0**	**1**	

Abbiamo effettuato una buona stima della media e la dev. std. della popolazione

Andamento delle distribuzioni delle **medie campionarie** all'aumentare del numero delle osservazioni nel campionamento



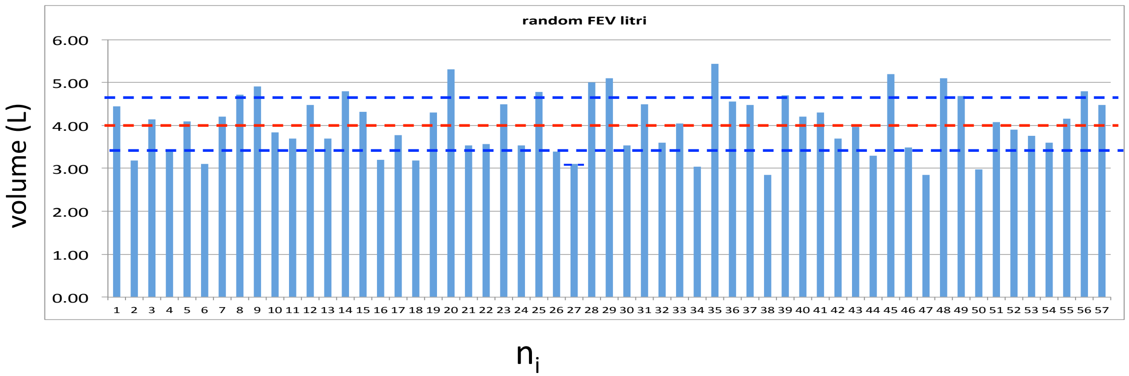
Andamento delle distribuzioni delle **medie campionarie** all'aumentare del numero delle osservazioni nel campionamento



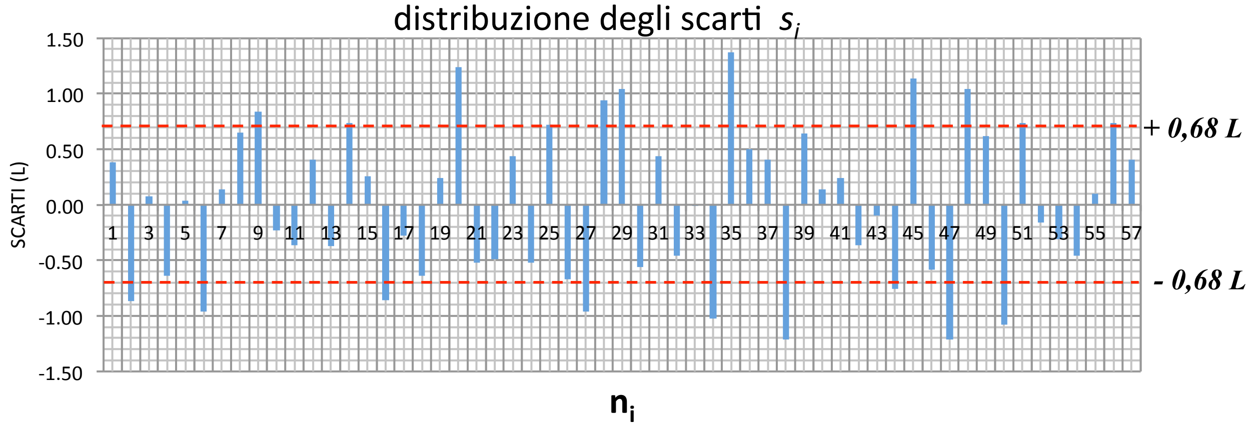
MEDIA la Varianza e la Deviazione Standard

ERRORE STANDARD

VOLUME RESPIRATORIO FORZATO (litri)						
n	1	2	3	4	5	6
1	4.44	3.70	3.54	4.50	4.30	4.08
2	3.19	4.47	3.57	3.60	3.70	3.90
3	4.14	3.69	4.50	4.05	3.96	3.75
4	3.42	4.80	3.54	3.04	3.30	3.60
5	4.10	4.32	4.78	5.43	5.20	4.16
6	3.10	3.20	3.39	4.56	3.48	4.80
7	4.20	3.78	3.10	4.47	2.85	4.47
8	4.71	3.19	5.00	2.85	5.10	
9	4.90	4.30	5.10	4.70	4.68	
10	3.83	5.30	3.54	4.20	2.98	



$N=57$



MEDIA $\bar{x} = \frac{1}{N}\sum_{i=1}^{N} x_i$

MEDIA = ***4.06 L***

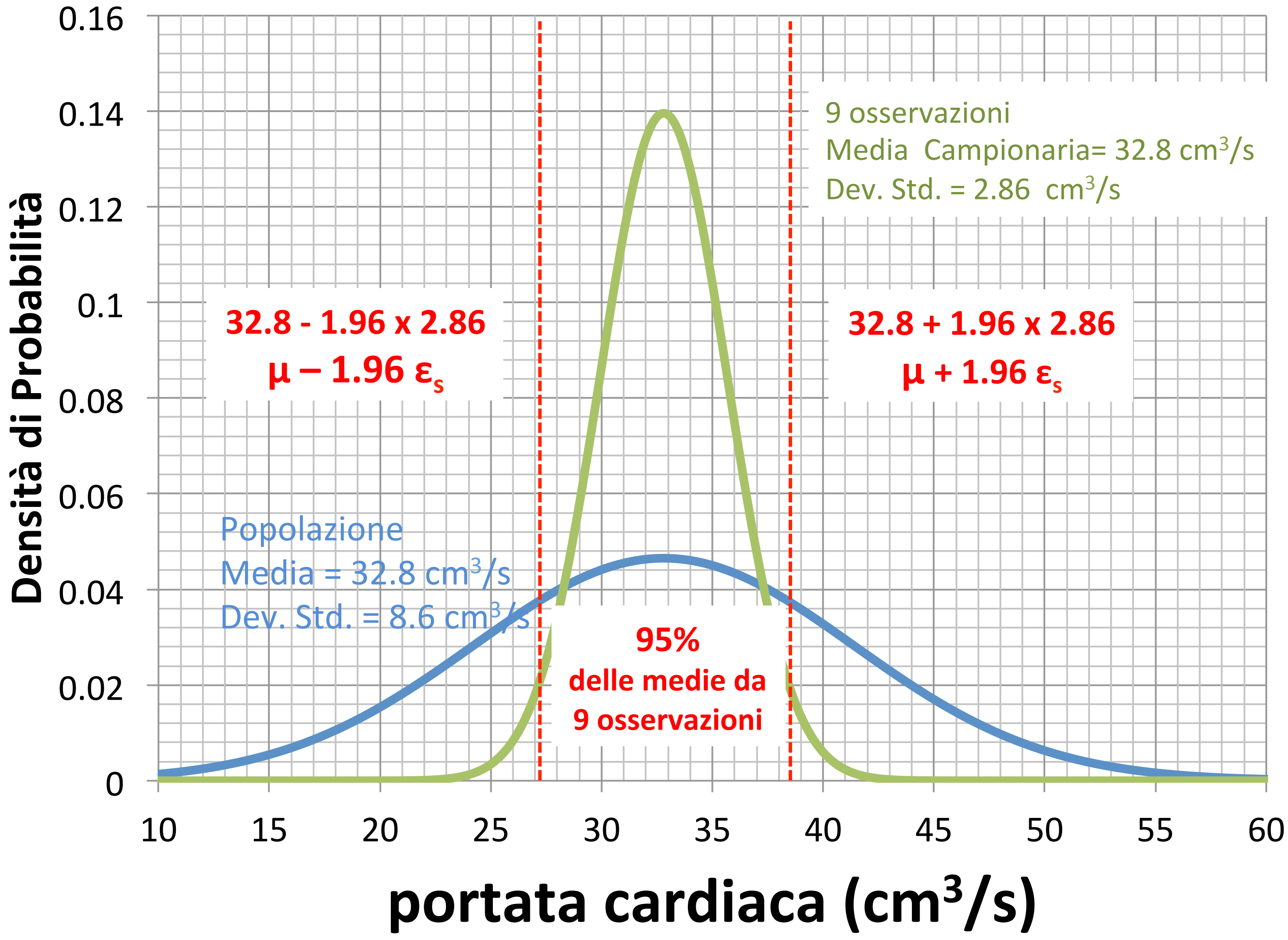
Deviazione Standard $\sigma_X = \sqrt{\frac{\sum_{i=1}^{N}(x_i-\bar{x})^2}{(N-1)}}$ $= \pm\, \mathbf{\mathit{0,68\ L}}$

Errore Standard $\varepsilon_s = \sigma_X / \sqrt{N} = 0{,}68/\sqrt{57} = 0{,}68 / 7{,}55 = \pm\, \mathbf{\mathit{0{,}09\ L}}$

Il 95% dei campioni con n_c= 57 avrà la media compresa tra ***4.06 L ± 1,96 x 0,09 L***

[$\mu \pm 1.96\ \varepsilon_s$] intervallo di confidenza al 95%

Andamento delle distribuzioni delle **medie campionarie** all'aumentare del numero delle osservazioni nel campionamento



FORMULARIO

Media della popolazione μ_p

Media dei campioni μ_c

Varianza della popolazione σ_p^2

Varianza campionaria σ_c^2

Deviazione standard della popolazione σ_p

Deviazione standard campionaria σ_c

Numero di osservazioni del campione N_c

MEDIA $\mu_c = \mu_p$

VARIANZA $\sigma_c^2 = \sigma_p^2 / N_c$

DEVIAZIONE STANDARD $\sigma_c = \sigma_p / \sqrt{N_c}$

ERRORE STANDARD $\varepsilon_S = \sigma_p / \sqrt{N_c}$