Intervalli di fiducia

Una parte al posto del tutto

- Per conoscere i fenomeni sociali o psicologici, non è necessario misurare o intervistare tutte le persone che hanno una certa caratteristica. Per sapere quanti sono gli studenti soddisfatti della loro scelta, non è necessario porre a tutti loro questa domanda. Un campione rappresentativo è molto più economico e ci dà gli stessi risultati, con un piccolo margine di errore.

Si prende un campione per stimare un parametro della popolazione

Esempio

- Quanti studenti iscritti al secondo anno hanno finito tutti gli esami del primo anno?
- Quanto sono soddisfatti i laureati degli studi che hanno compiuto?

Si calcola la media del campione
per stimare
la media della popolazione
(stima puntuale)

- Tuttavia, la variabilità statistica dei campioni farebbe sì che al prossimo campione rilevato, la media potrebbe essere leggermente diversa.

- Leggermente? Quanto leggermente? E se la variazione fosse enorme?

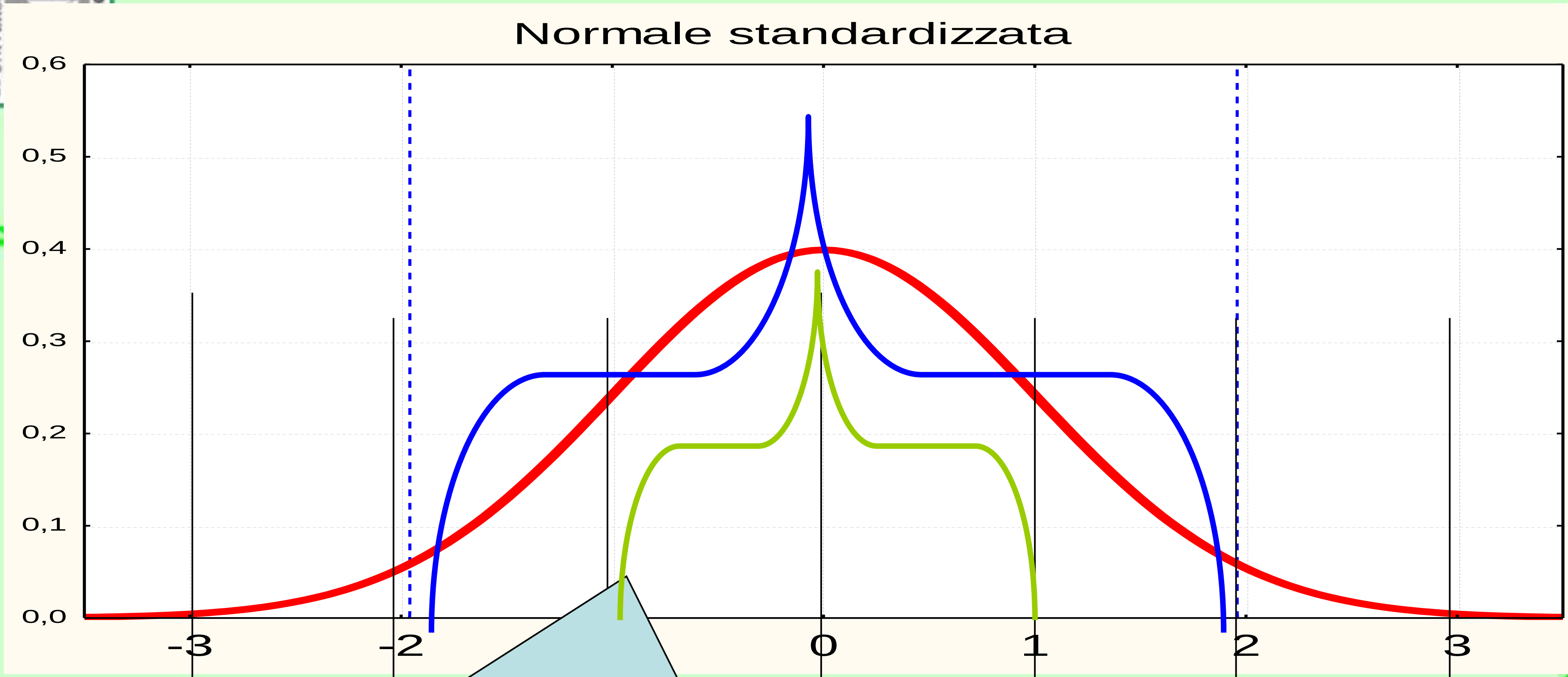
Possiamo essere più precisi ?

- Gli statistici possono riformulare il problema e dare una soluzione in questo modo:
- *Dopo aver calcolato la media di un campione, si può stabilire un intervallo entro cui ricade, con molta verosomiglianza, la media della popolazione*

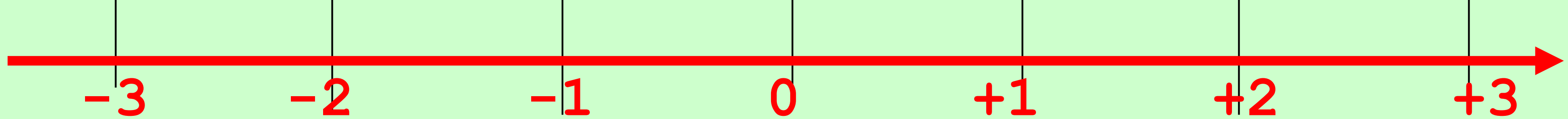
La verosimiglianza la chiameremo probabilità o sicurezza, e la quantificheremo, per esempio al 95%
(19 su 20 di probabilità)

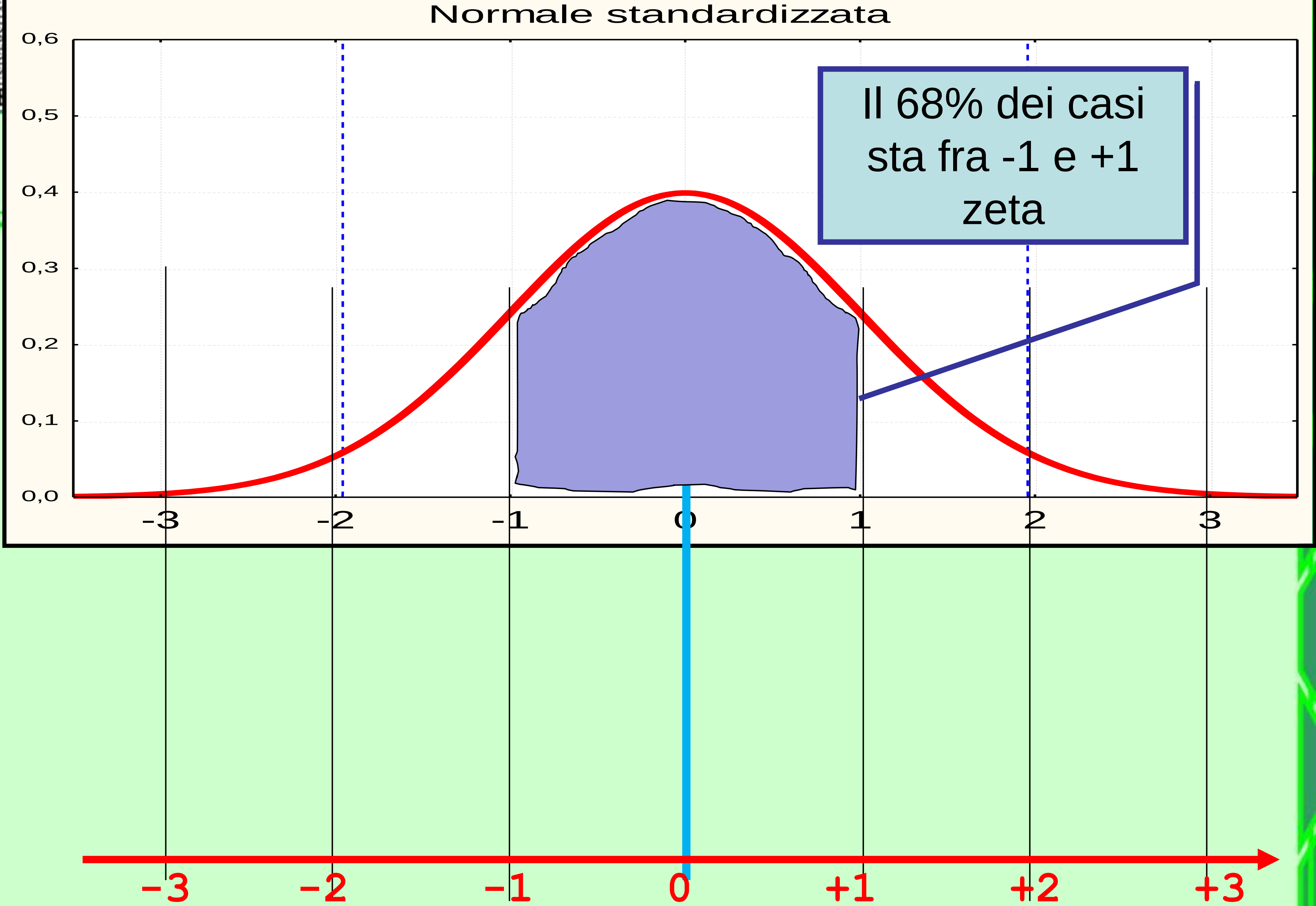
Per procedere, ricordiamo due concetti fondamentali

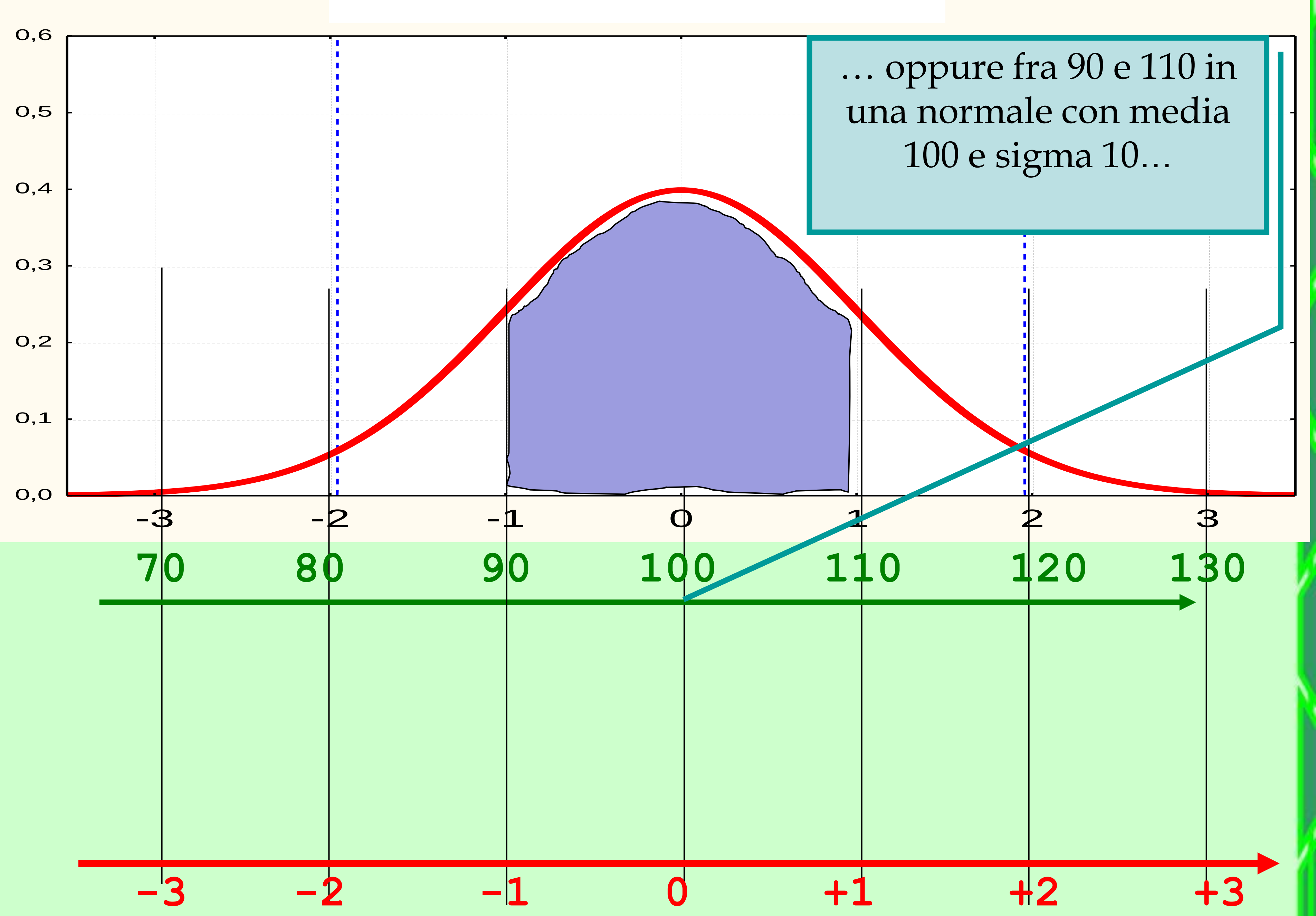
- 1 Le caratteristiche della curva normale
- 2 La distribuzione campionaria delle medie



Se la distribuzione è normale, si possono individuare le percentuali (per esempio 2/3, oppure il 95%), di casi limitati da due valori, qualunque sia la media e la dev stan

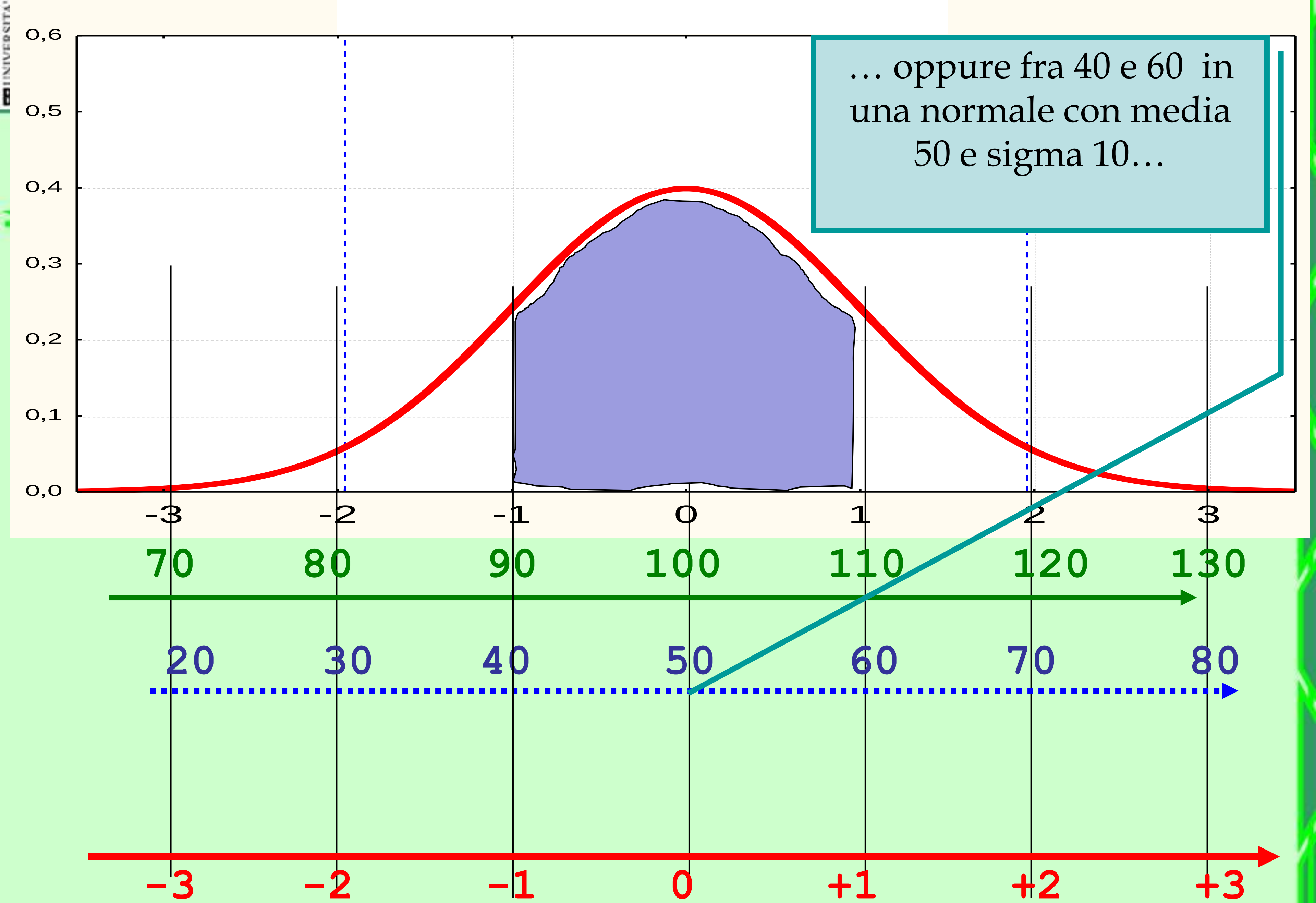


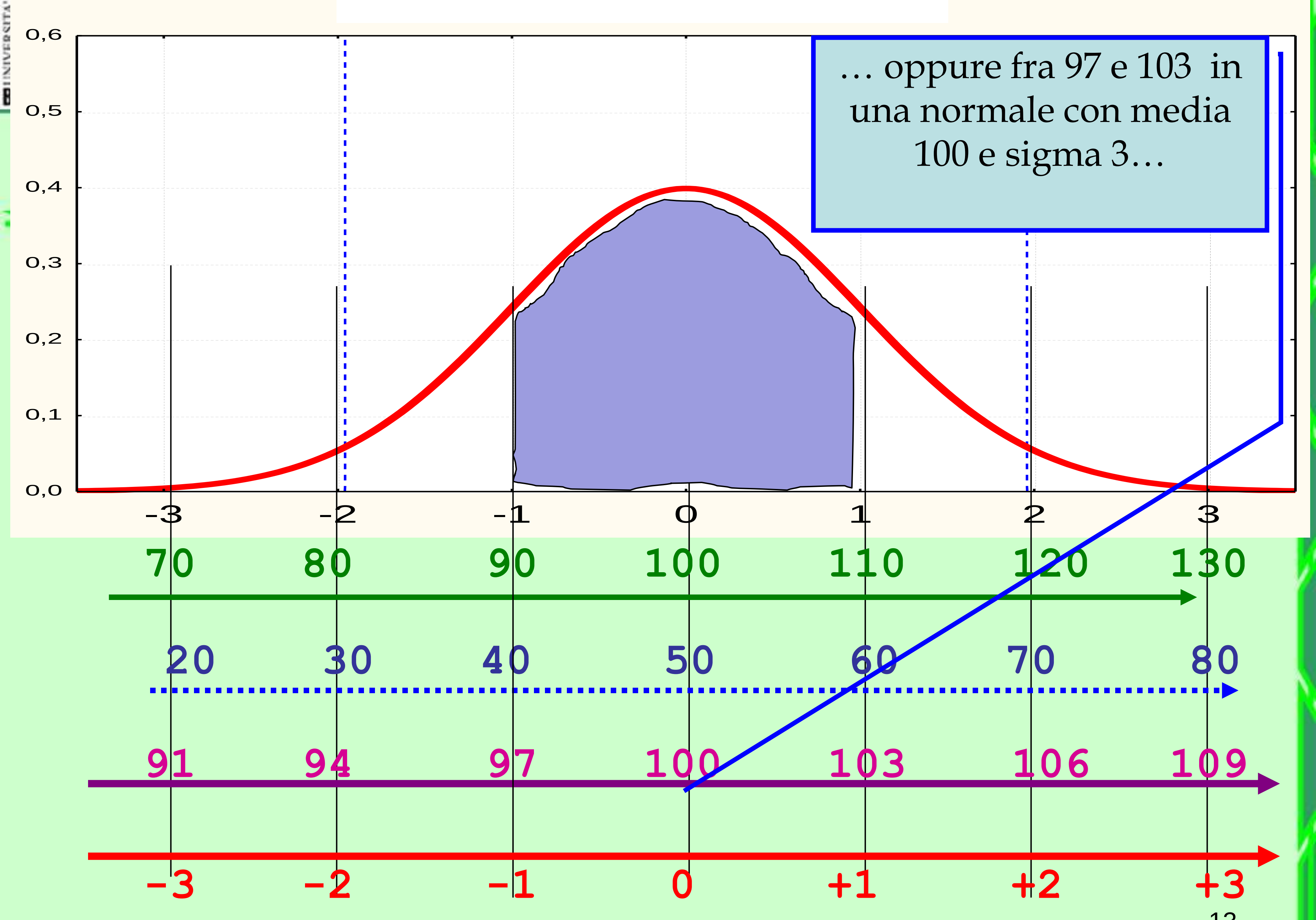

Normale standardizzata
0,6
0,5
0,4
0,3
0,2
0,1
0,0
-3
-2
-1
0
1
2
3
Il 68% dei casi sta fra -1 e +1 zeta
-3
-2
-1
0
+1
+2
+3

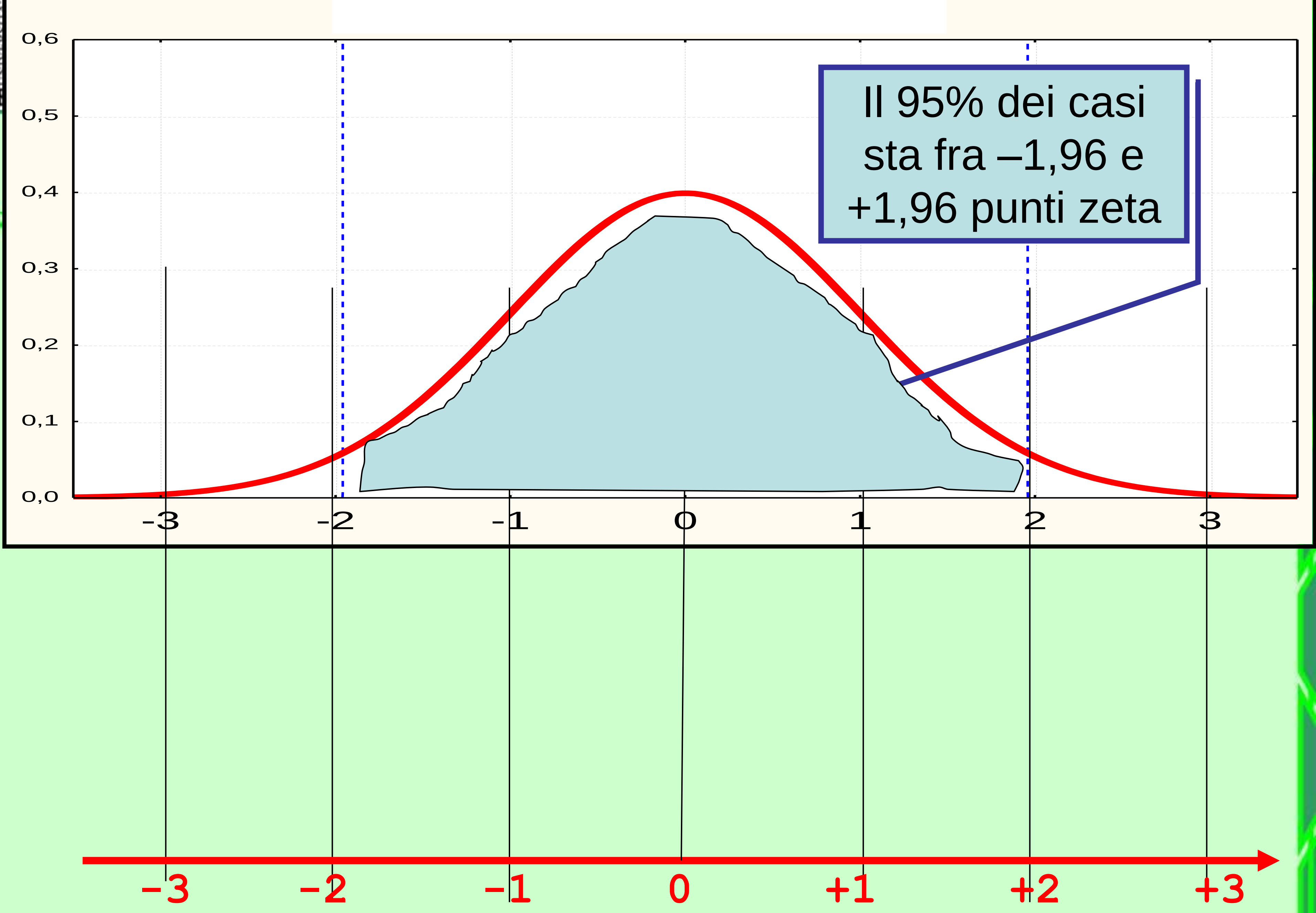


… oppure fra 90 e 110 in una normale con media 100 e sigma 10…
0,6
0,5
0,4
0,3
0,2
0,1
0,0
-3
-2
-1
0
1
2
3
70
80
90
100
110
120
130
−3
−2
−1
0
+1
+2
+3

… oppure fra 40 e 60 in una normale con media 50 e sigma 10…



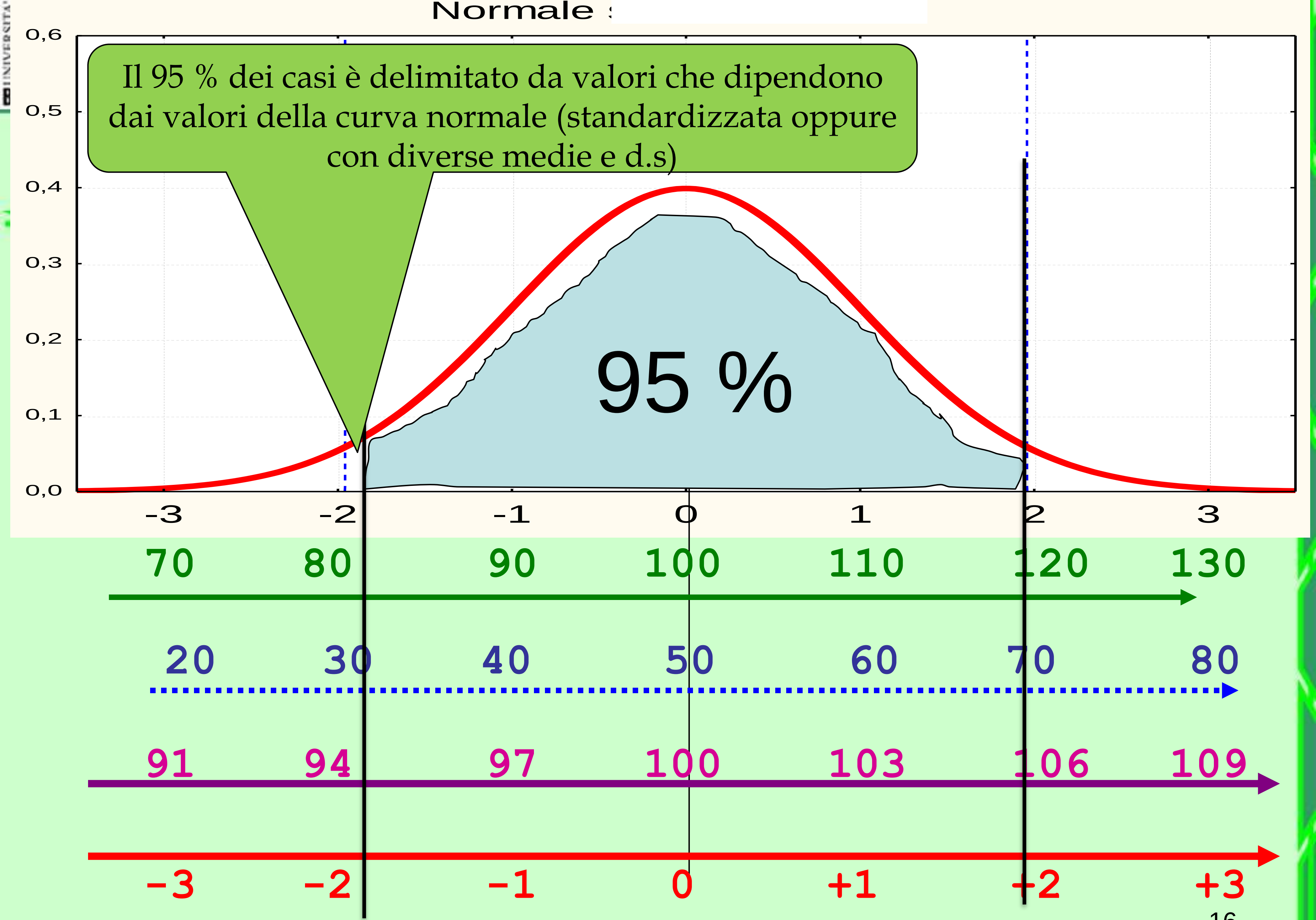

… oppure fra 97 e 103 in una normale con media 100 e sigma 3…
0,6
0,5
0,4
0,3
0,2
0,1
0,0
-3 -2 -1 0 1 2 3
70 80 90 100 110 120 130
20 30 40 50 60 70 80
91 94 97 100 103 106 109
−3 −2 −1 0 +1 +2 +3


0,6
0,5
0,4
0,3
0,2
0,1
0,0
-3
-2
-1
0
1
2
3
Il 95% dei casi sta fra –1,96 e +1,96 punti zeta
–3
–2
–1
0
+1
+2
+3

- Dobbiamo però convertire questi due punti zeta nei valori della nostra normale con la formula inversa della trasformazione dei punti zeta:
- Zeta= (X – media)/ dev.stan
- y = z x dev.stan + media

Perciò

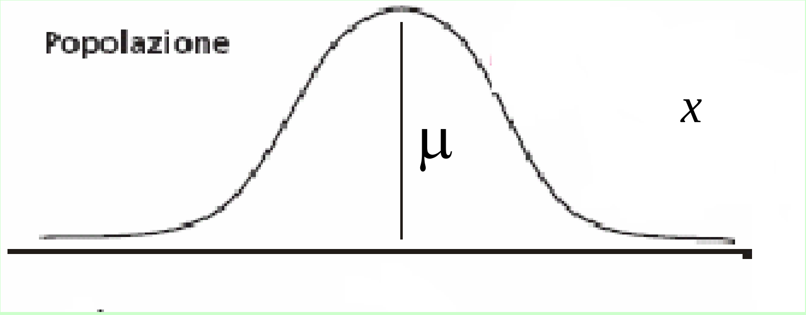
- Lim inf= -1,96 x dev stan + media
- Lim sup = +1,96 x dev.stan + media


Normale
Il 95 % dei casi è delimitato da valori che dipendono dai valori della curva normale (standardizzata oppure con diverse medie e d.s)
95 %
0,6
0,5
0,4
0,3
0,2
0,1
0,0
-3 -2 -1 0 1 2 3
70 80 90 100 110 120 130
20 30 40 50 60 70 80
91 94 97 100 103 106 109
−3 −2 −1 0 +1 +2 +3

- La deviazione standard della distribuzione campionaria delle medie si calcola con la formula

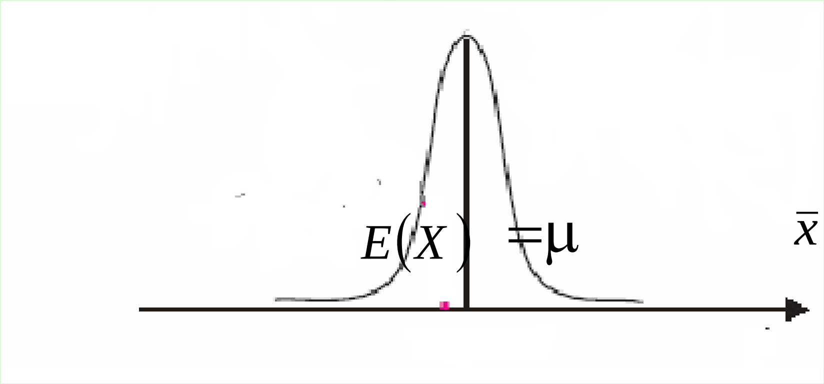
$$\frac{\sigma}{\sqrt{N}}$$

Grafico della distribuzione della media campionaria da popolazione normale



Popolazione

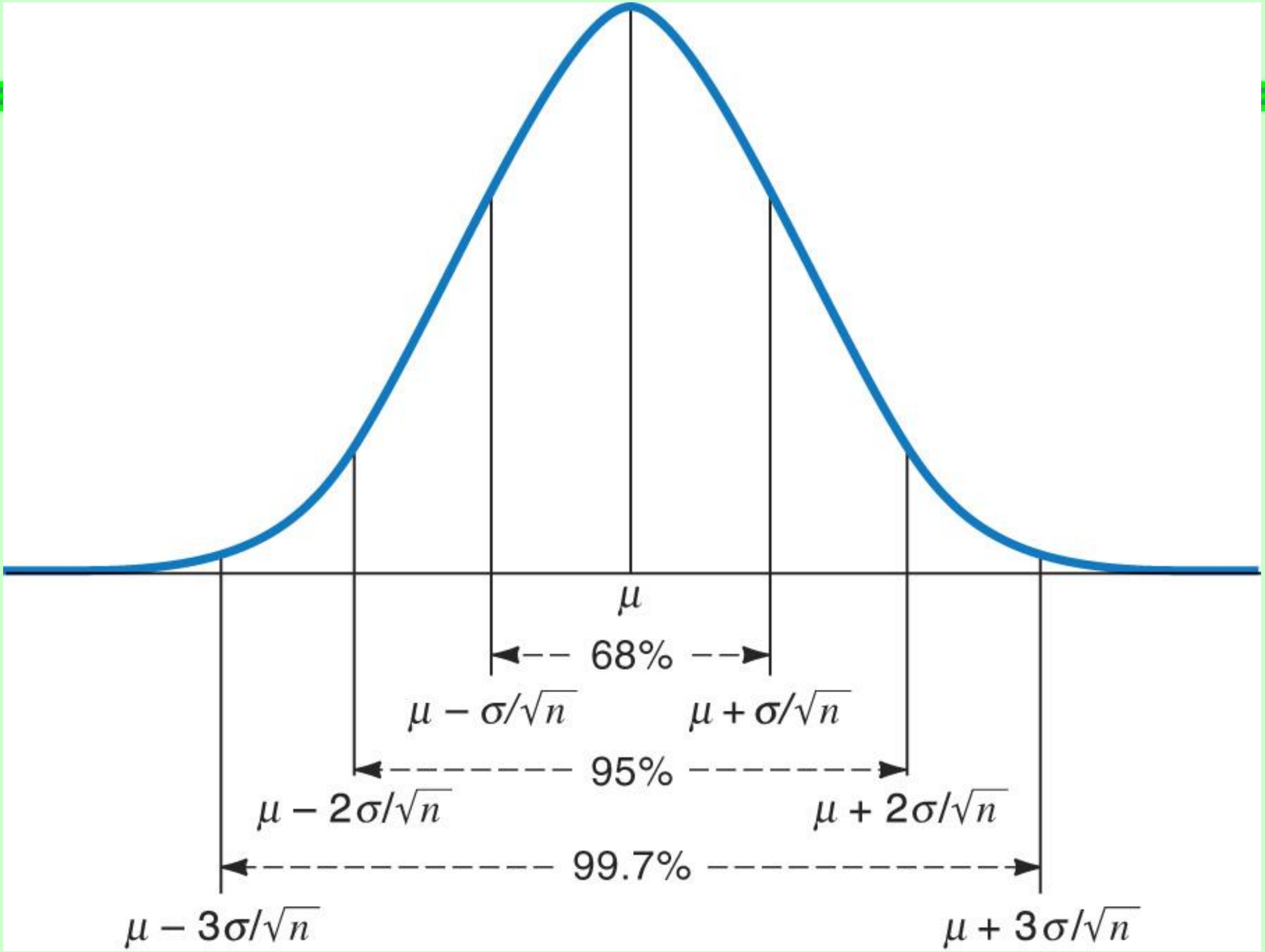
X~N(μ, σ²)



Stimatore

$\bar{X}$ ~N(μ, σ²/n)

Ricordando le proprietà della curva normale



Principio di calcolo

- Sappiamo che la media della popolazione è vicina (è simile) alla media del campione.
- Quanto vicina? *"E' molto probabile che sia molto vicina, è poco probabile che sia distante"*.
- Possiamo stabilire un intervallo di fiducia entro cui ricade il parametro della popolazione, perché sappiamo che
- 1) la distribuzione campionaria delle medie è normale
- 2) conosciamo la media e la deviazione standard della distribuzione campionaria delle medie

  Quindi…
- 3) stabiliamo un intervallo al 95% (o 68% o 90%) entro cui ricade la media della popolazione

Inoltre…

- Non abbiamo motivo di pensare che ci siano più probabilità che la media della popolazione sia maggiore della media del campione, o al contrario, che sia minore.
- Per questo facciamo ricorso ad un intervallo simmetrico attorno alla media.

Esempio di calcolo

- Si rileva l'altezza di un gruppo di 25 studenti:
- Media = 178
- Dev stand 14,5

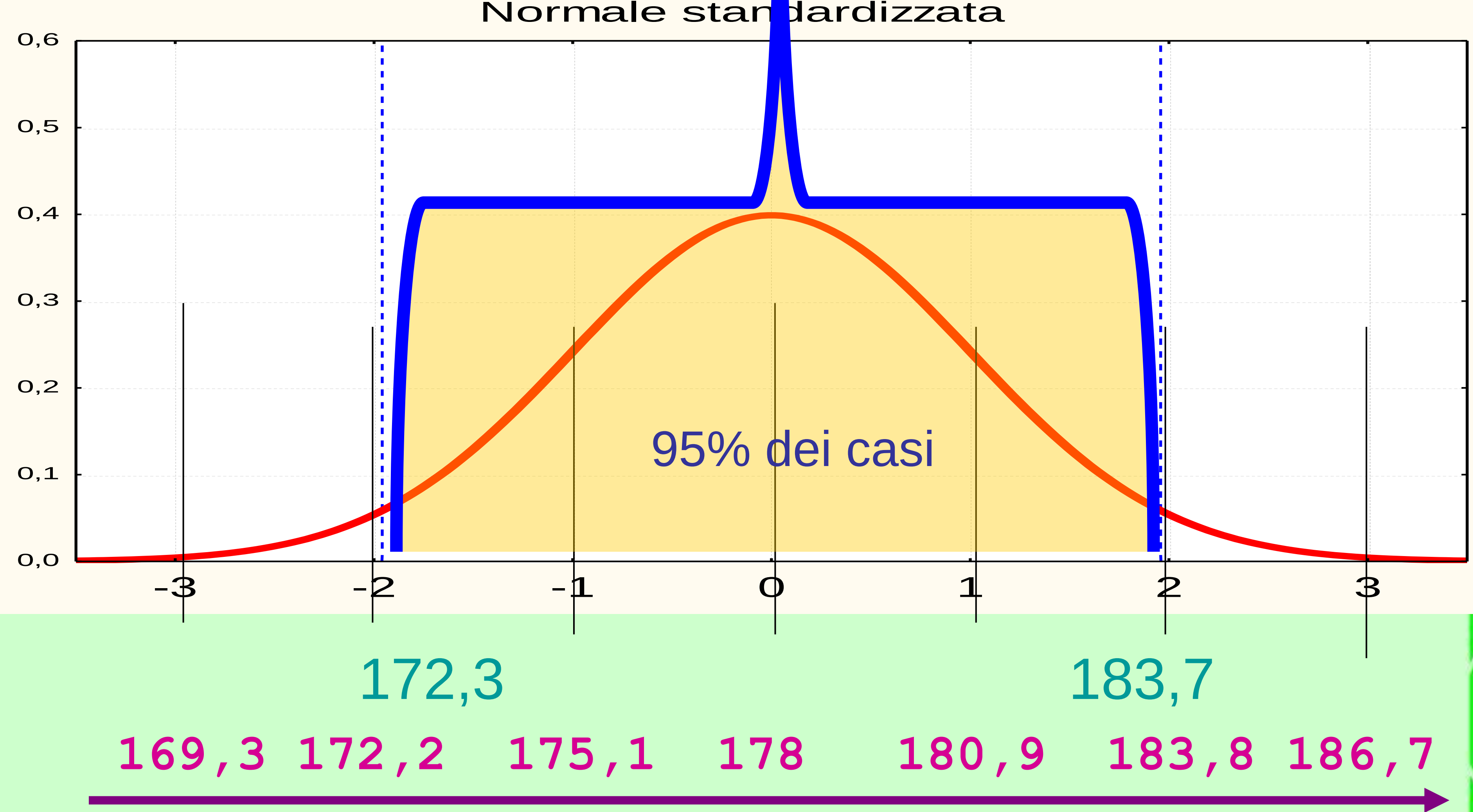
Calcoliamo l'errore standard della distribuzione campionaria delle medie

$$\frac{\sigma}{\sqrt{N}} = \frac{14,5}{\sqrt{25}} = 2,9$$

- Applicando la formula otteniamo:
- Lim inf= -1,96 x dev stan + media
- Lim sup = +1,96 x dev.stan + media
- Limite inferiore = 172,316
- Limite superiore = 183,684

Perciò possiamo affermare:

- La media della popolazione ha il 95% di probabilità di situarsi fra 172,3 e 183,7



$$95\%\,IC = Media\,campionaria \pm \left(1{,}96 \times \frac{\sigma}{\sqrt{N}}\right)$$

Per stimare la media della popolazione

- Possiamo affermare che c'è una probabilità di 0,95 (oppure una percentuale di riuscita) che la media della popolazione degli studenti sia situata fra 172,3 e 183,7.

Altri intervalli di fiducia

- Si prendono anche i due terzi di fiducia, perché 2/3 di probabilità corrispondono a una deviazione standard.
- la notazione diventa breve e comoda da comunicare

per esempio: media = 35 ± 4,5

- ± 4,5 corrisponde a ± 4,5 x 1 d.s. e tale notazione può essere semplificata.