# La distribuzione campionaria della media

# La distribuzione campionaria della media

- È una distribuzione teorica
- Diversa dalla distribuzione del campione
- Diversa dalla distribuzione della popolazione
- Serve nell'inferenza statistica per la stima puntuale e intervallare (per stimare valori singoli o intervalli di valori probabili)

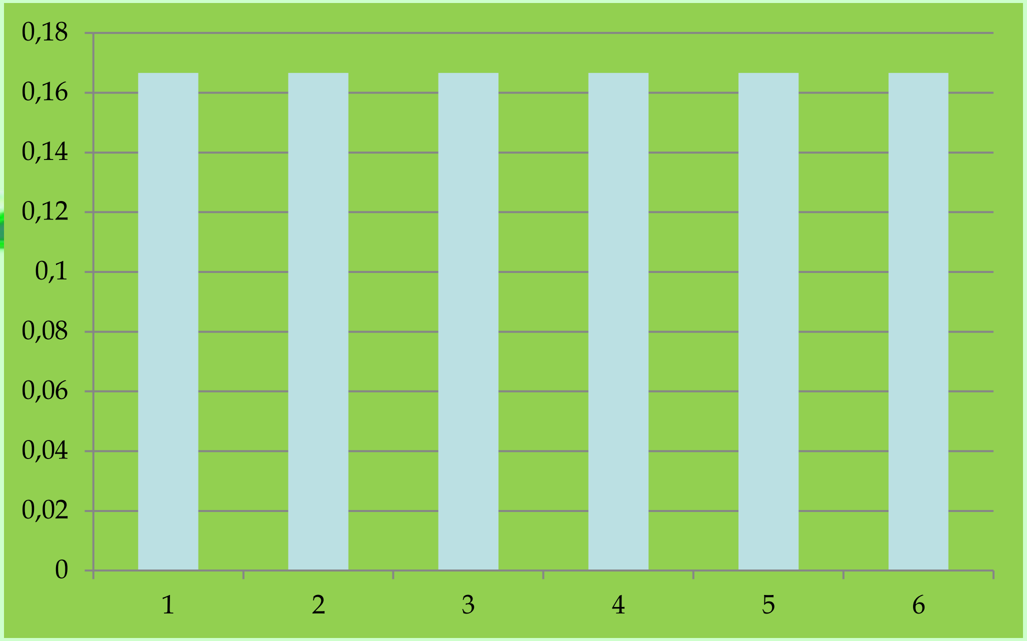
# Per spiegare la DCM..

- Useremo una distribuzione discreta
- Simuleremo tantissime estrazioni casuali

Useremo una distruzione conosciuta, anche se del tutto inutile ai fini pratici, come quella di un dado da gioco (sei facce, con punti da 1 a 6)

- Un dado dà un punteggio discreto compreso fra 1 e 6
- Due dadi, tirati assieme, danno un punteggio compreso fra 2  e 12.
- La media invece varia sempre da 1 a 6, anche per cinque o dieci dadi.

- Il punteggio di un dado segue una distribuzione teorica, uniforme, e discreta



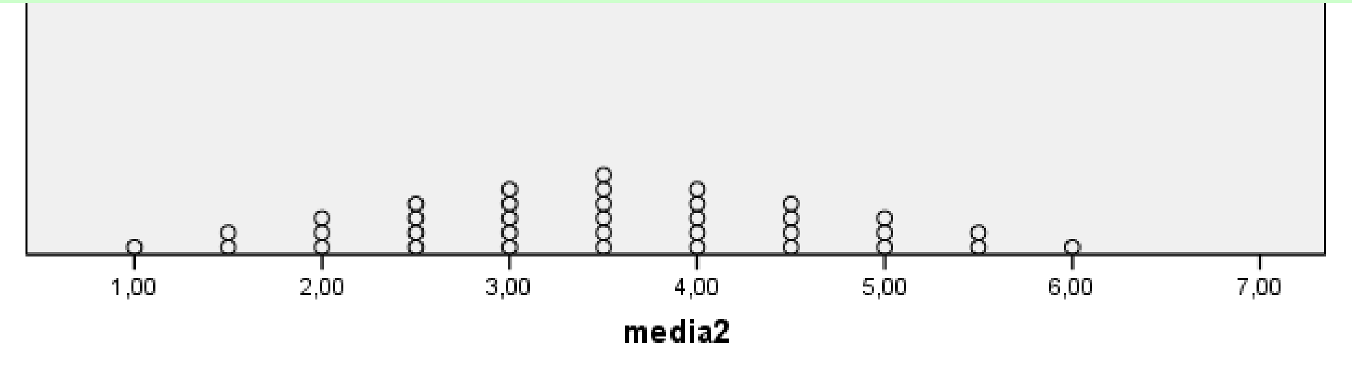

La media di 1, 2, 3, 4, 5, 6 è pari a 3,5

La varianza (media dei quadrati meno quadrato della media) è pari a

(1+4+9+16+25+36 )/6-3,5x3,5= 2,9166

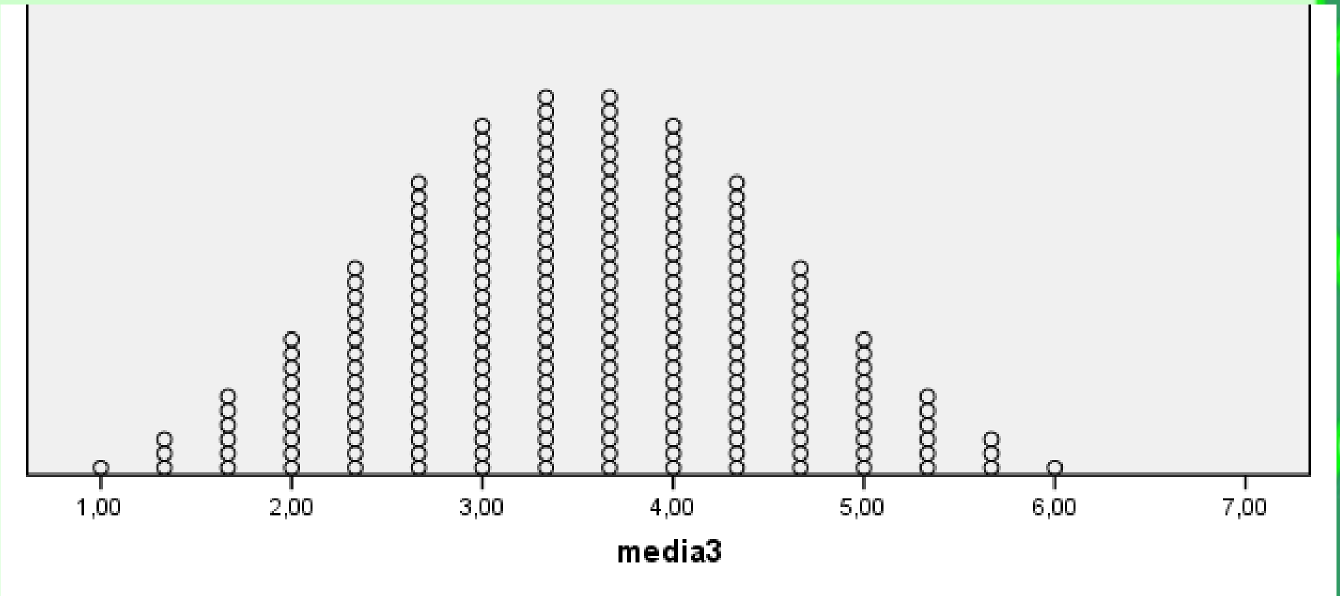
La deviazione standard è pari a 1,7078

# La distribuzione campionaria delle medie dei punteggi di due dadi




- Combinando il punteggio di due dadi, otteniamo 6 x 6 = 26 combinazioni di punteggi, (1,1; 1,2; 1,3…6,5; 6,6) le cui medie sono presentate in questo grafico

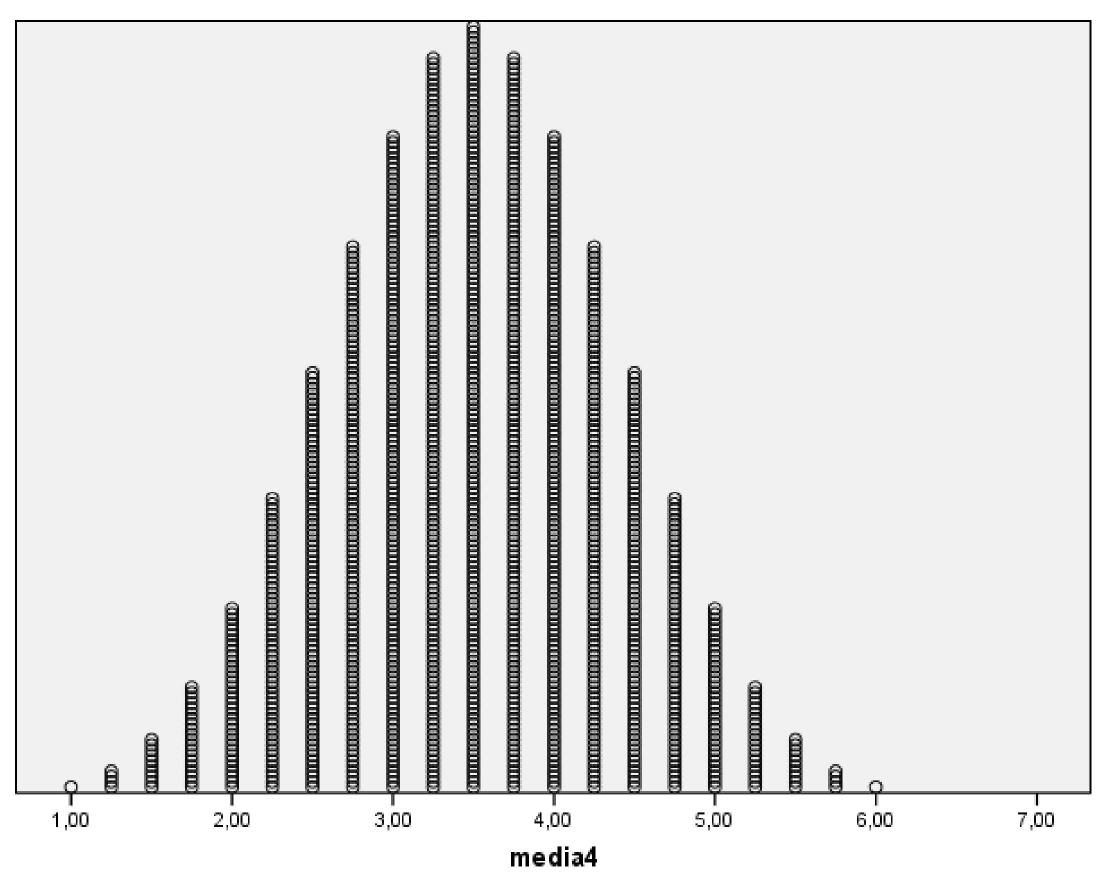
# Con tre dadi .




- Con tre dadi, ci sono 6x6x6 = 216 possibili disposizioni per un totale di 16 valori possibili di media.

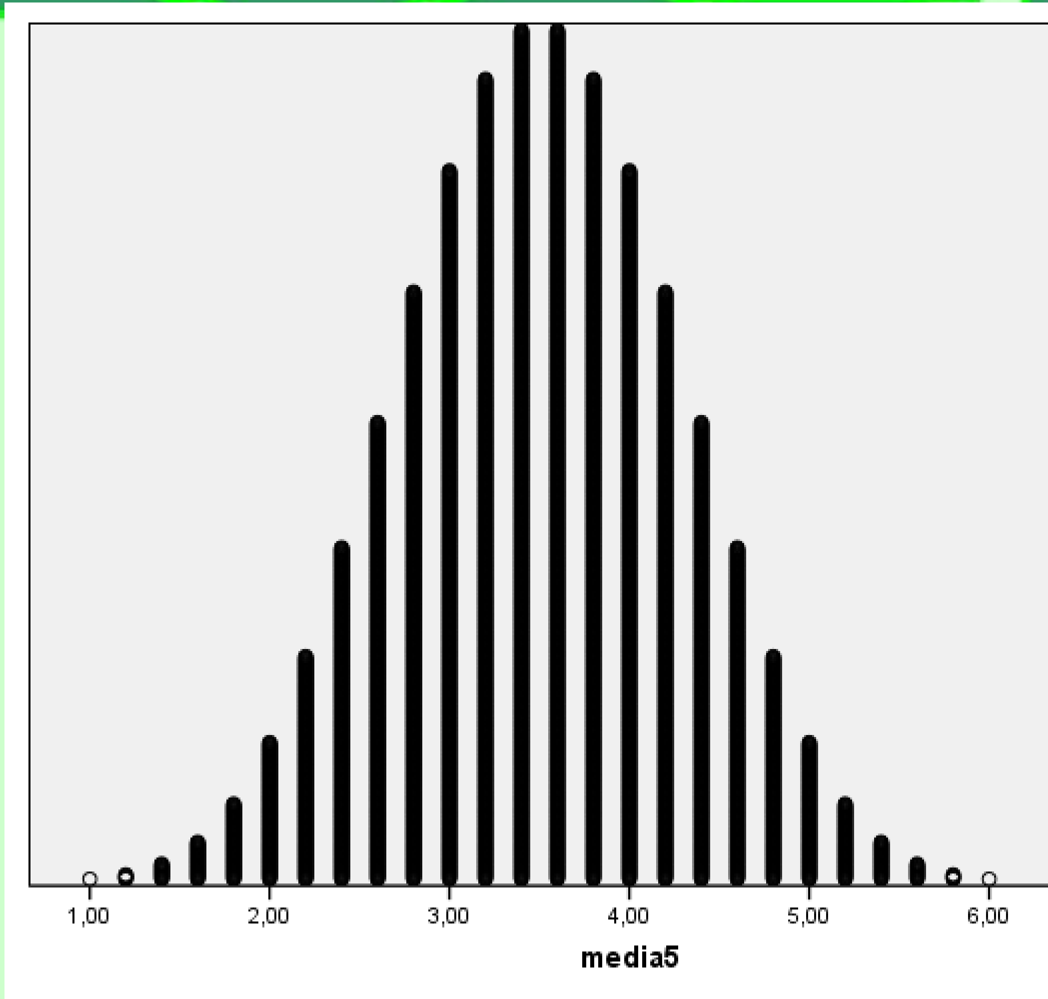
# Con quattro dadi.

- Con quattro dadi, ci sono 6x6x6x6 = 1296 possibili valori delle medie



# Con cinque dadi.

- Con cinque dadi, ci sono 6x6x6x6x6 = 7776 possibili valori delle medie, tutte rappresentate in questo grafico



# Conclusioni

L'andamento sembra ben chiaro: man mano che aumenta il numero di campioni (nel nostro caso dei dadi) , la distribuzione teorica delle medie si avvicina a quella normale

# Distribuzione empirica di molti più dadi

Facciamo un vero esperimento, lanciando un certo numero di dadi

- Lanciando 2, o 3 o 10 dadi un numero per esempio 100 volte, come si distribuiscono le medie?

# Immaginiamo un esperimento

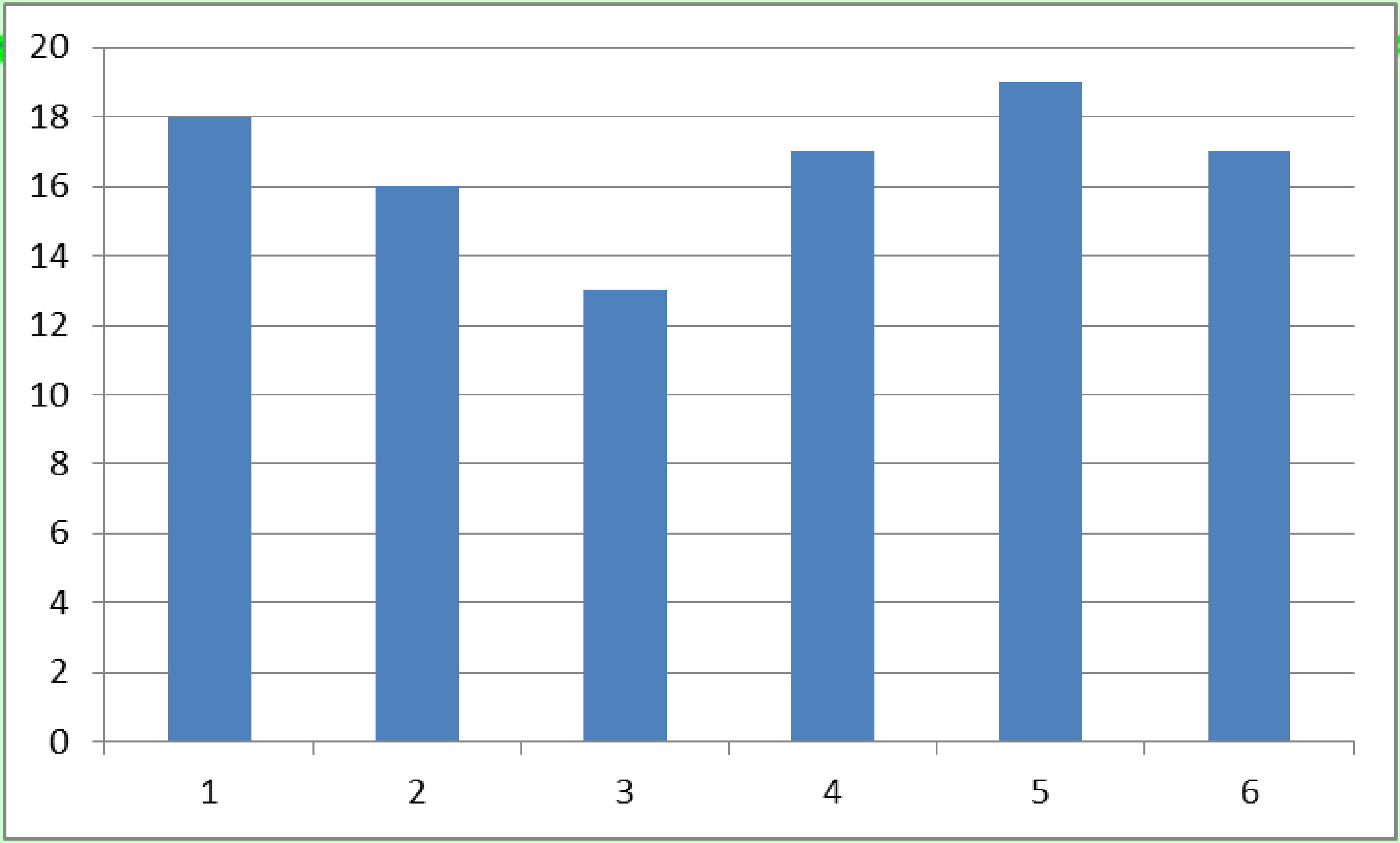
- Cento studenti lanciano un certo numero di dadi, calcolano la media dei punteggi e registrano il risultato.

# Lanciando un solo dado…

…La media è uguale al punto del dado

E la rappresentazione grafica è la seguente

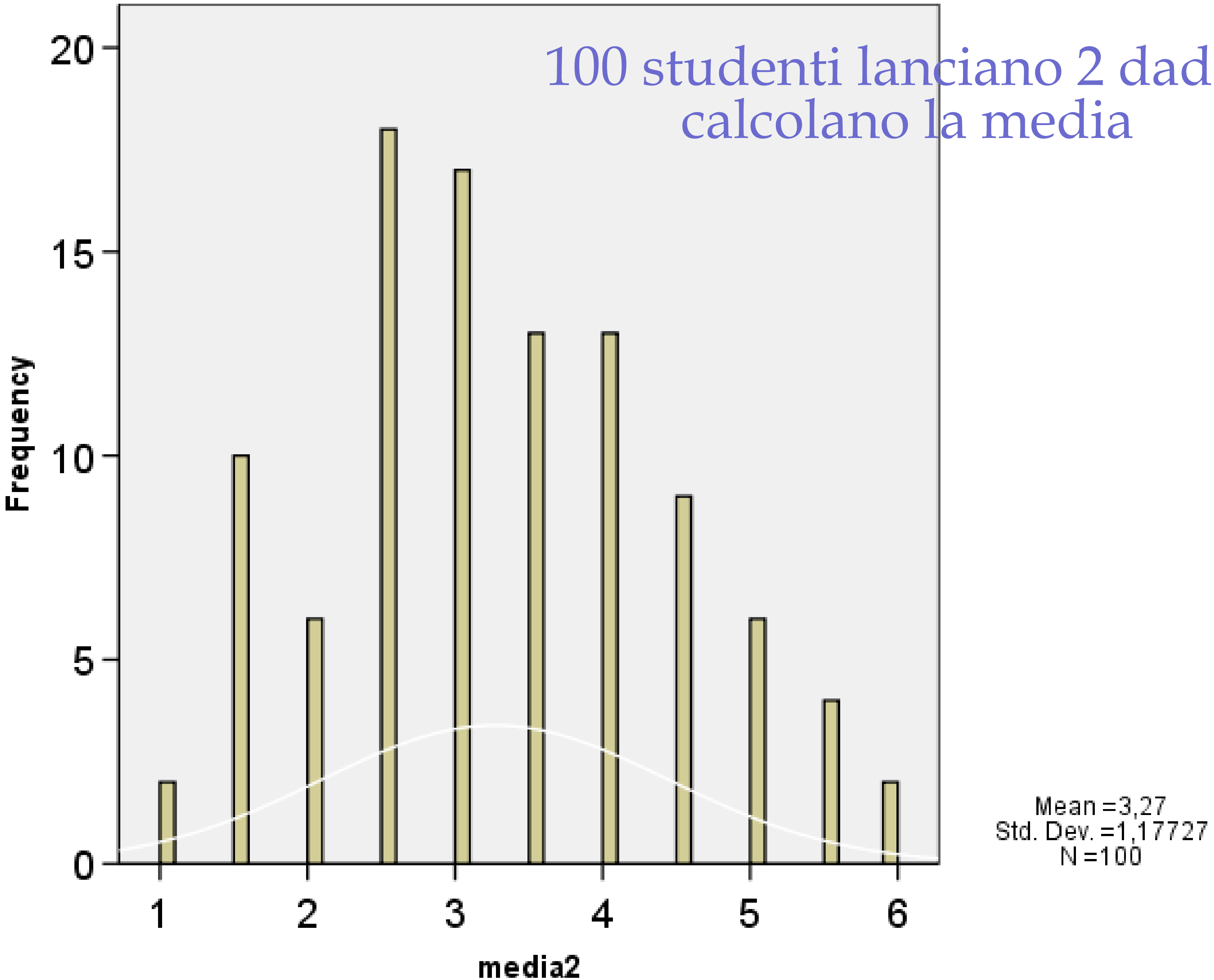
La distribuzione dei risultati non è perfettamente uniforme, ma varia per  fluttuazione casuale ( per la variabilità stocastica.



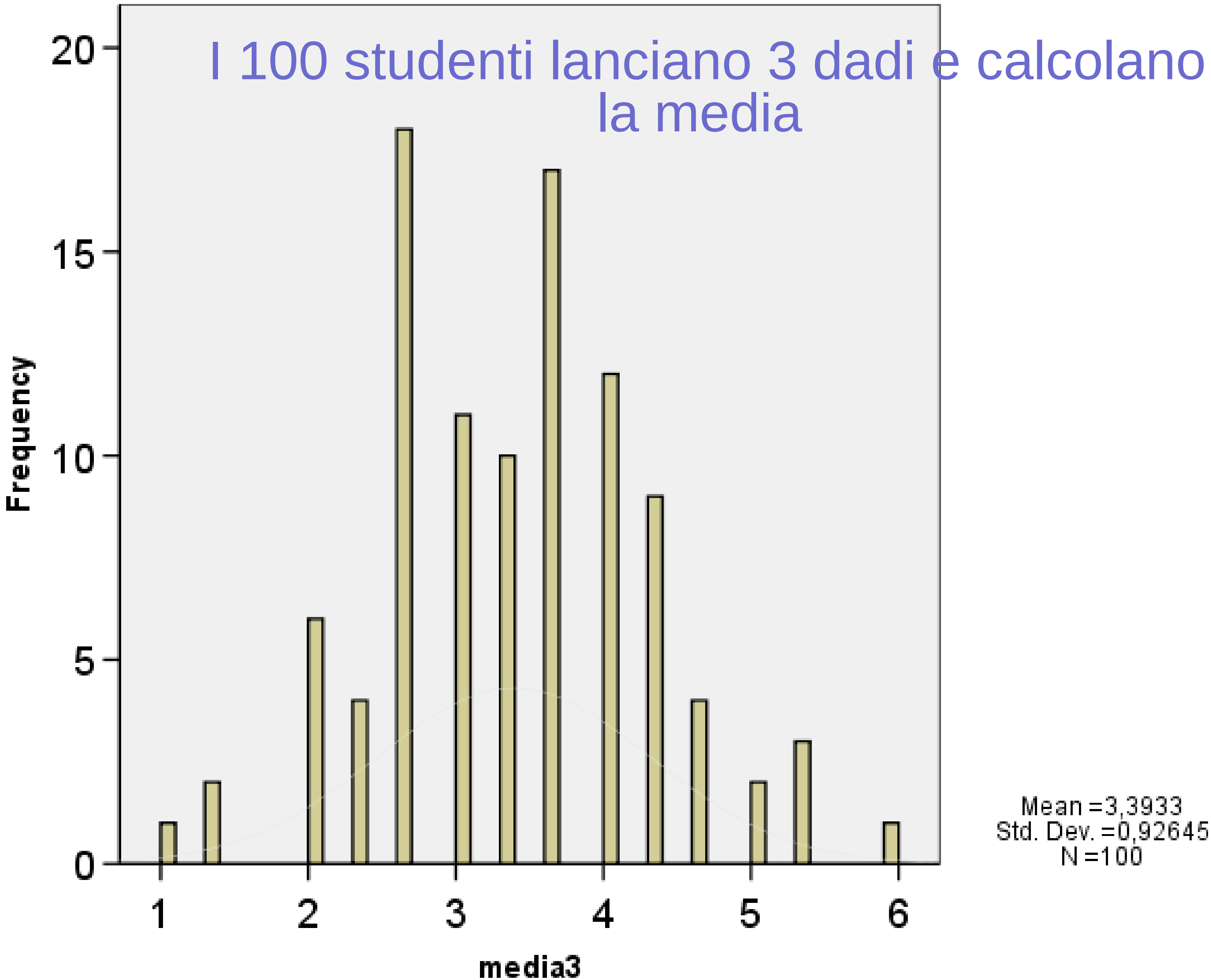
# Continuiamo l'esperimento

- I cento  studenti lanciano non uno, ma due,  poi tre, poi  cinque, dieci… secondo le indicazioni dello sperimentatore e calcolano la media dei valori ottenuti.
- Come si distribuiscono queste cento medie ?

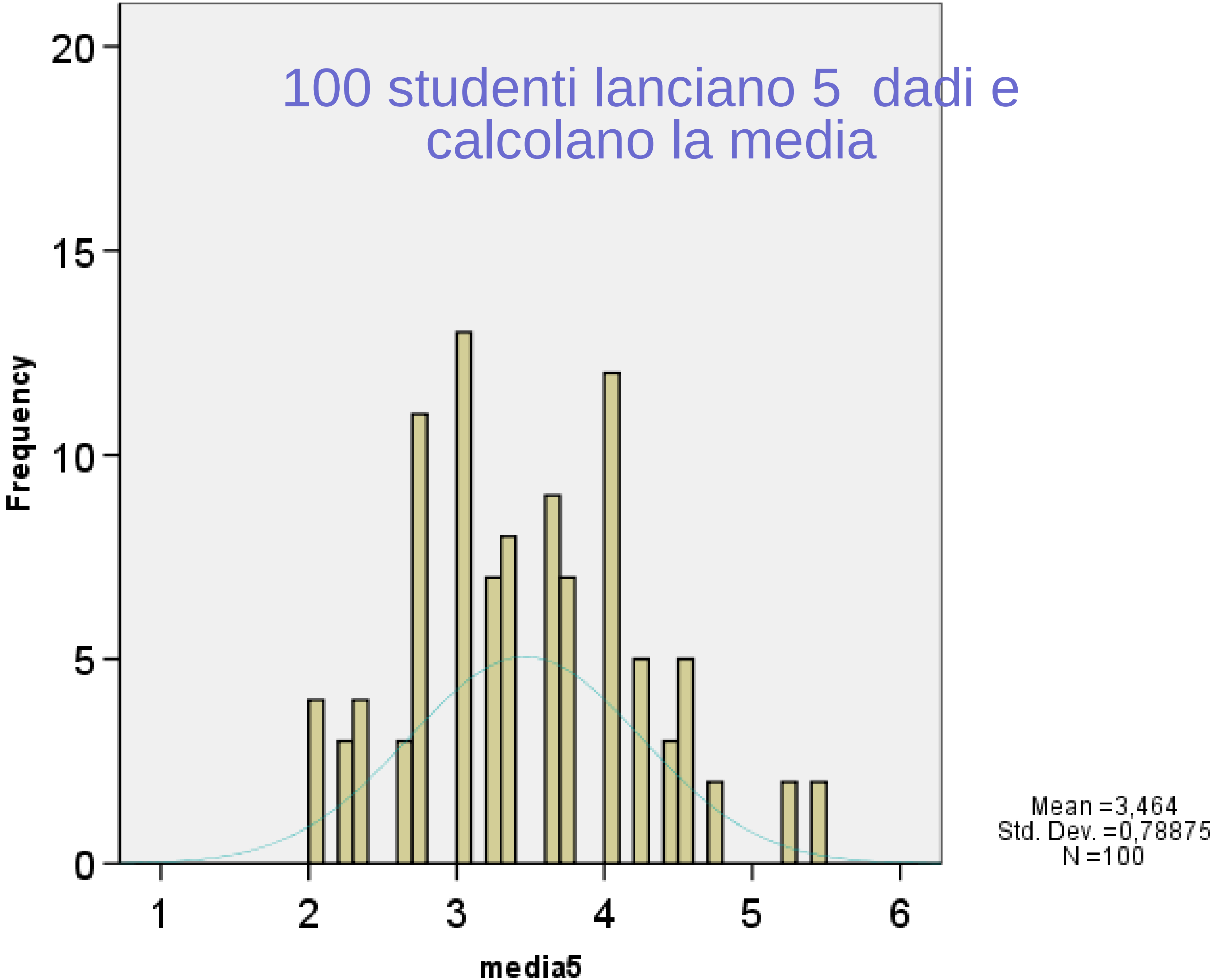
## 100 studenti lanciano 2 dadi e calcolano la media

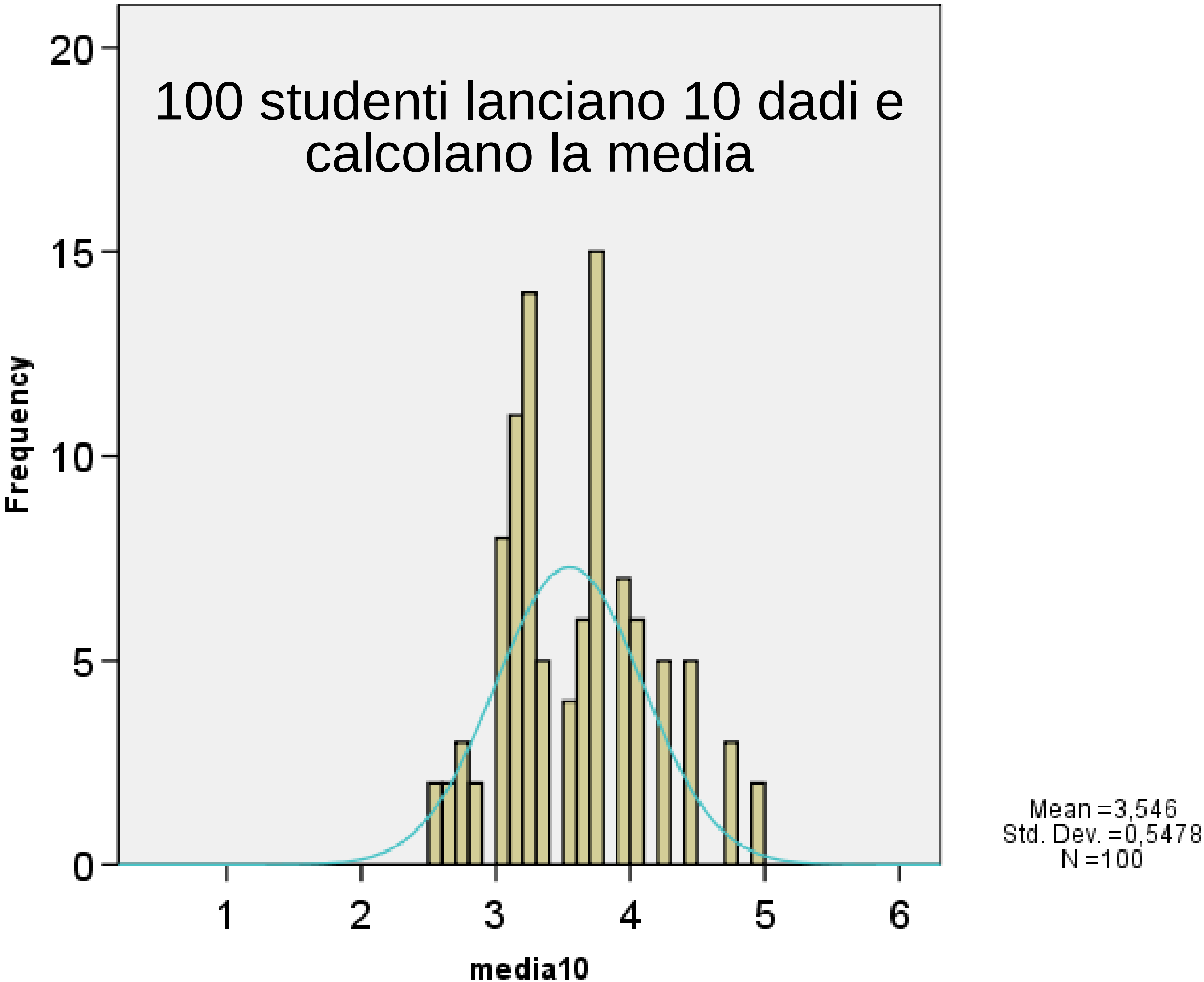


I 100 studenti lanciano 3 dadi e calcolano la media



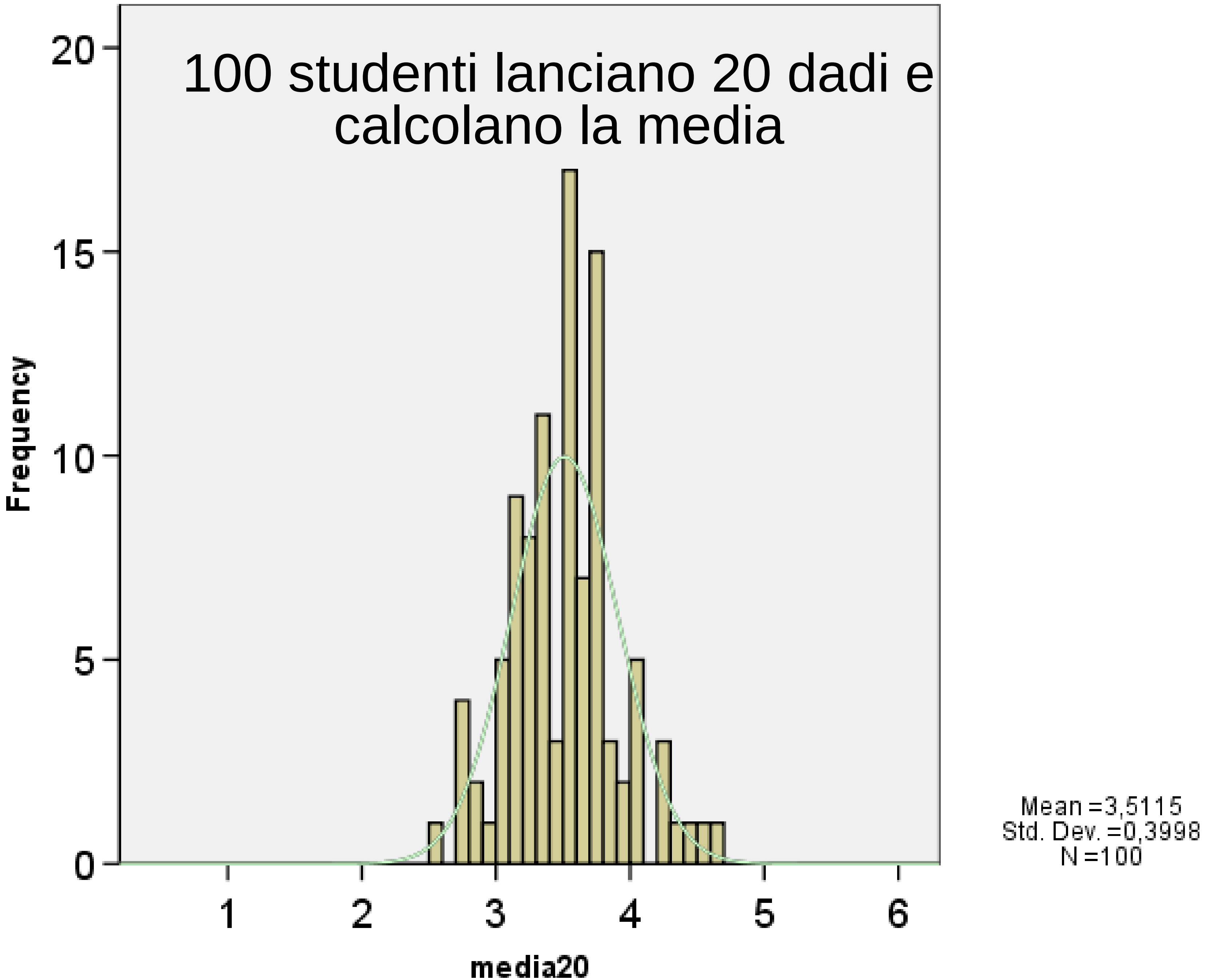
# 100 studenti lanciano 5 dadi e calcolano la media



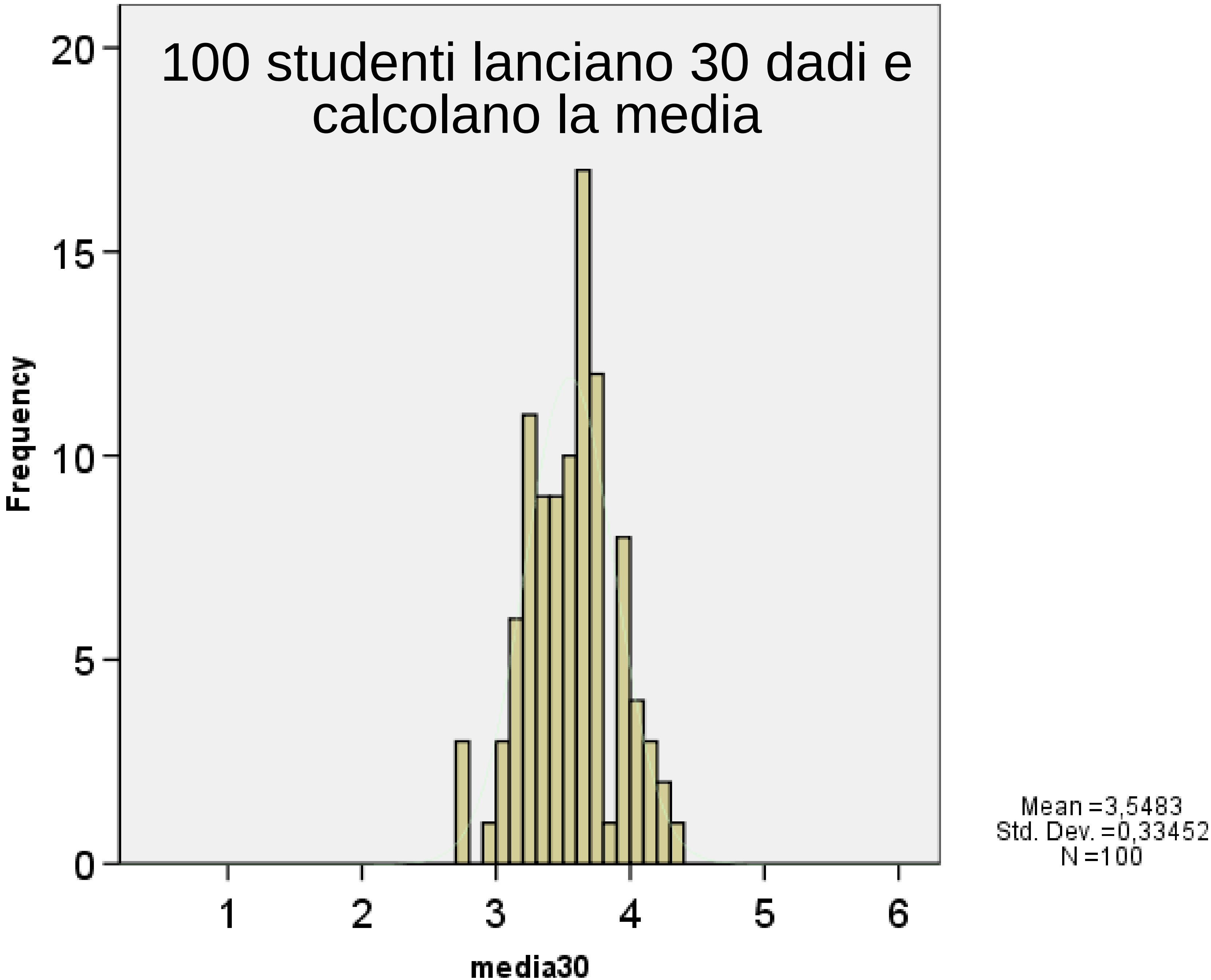


100 studenti lanciano 10 dadi e calcolano la media
20
15
10
5
0
Frequency
1
2
3
4
5
6
media10
Mean =3,546
Std. Dev. =0,5478
N =100

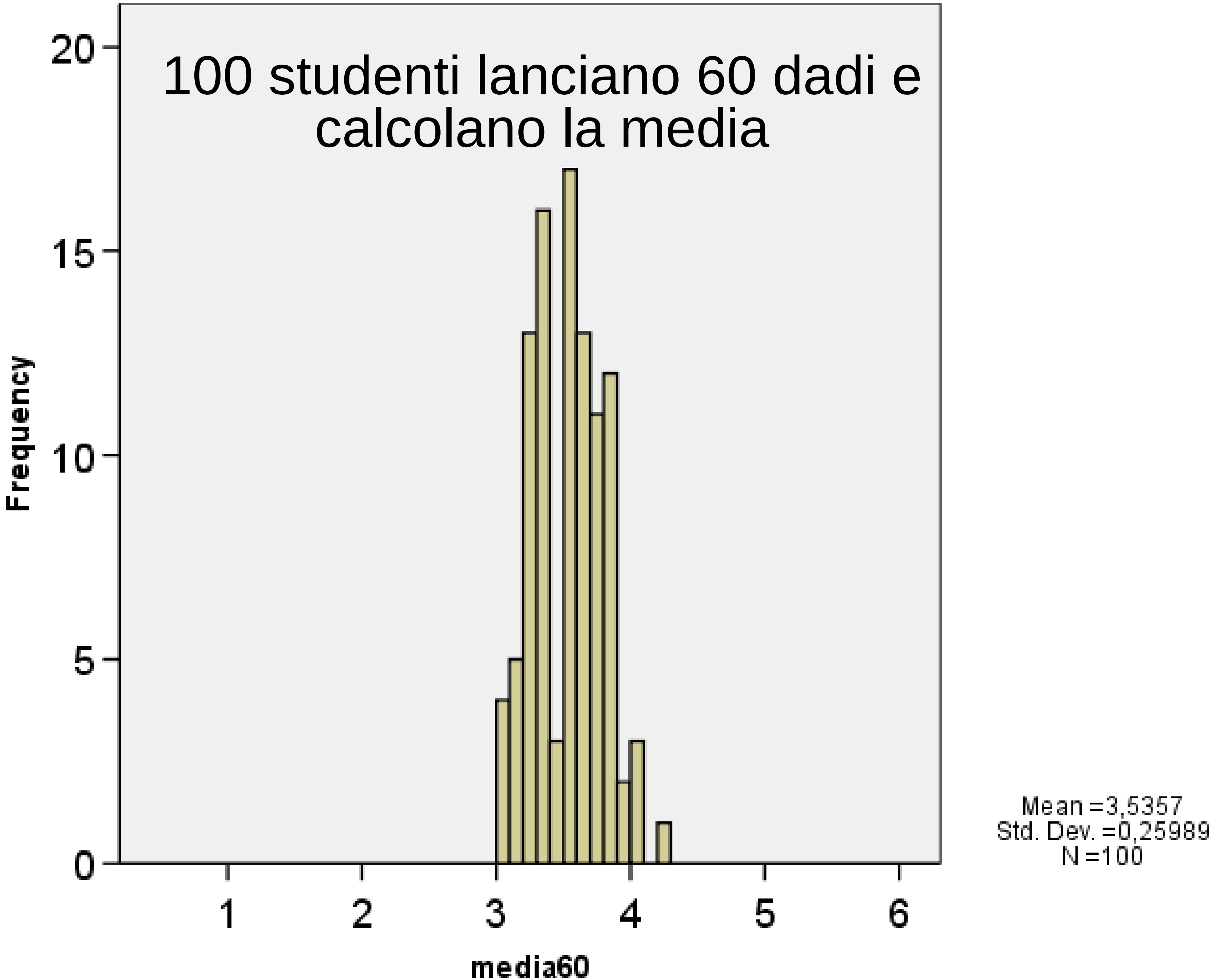
100 studenti lanciano 20 dadi e calcolano la media



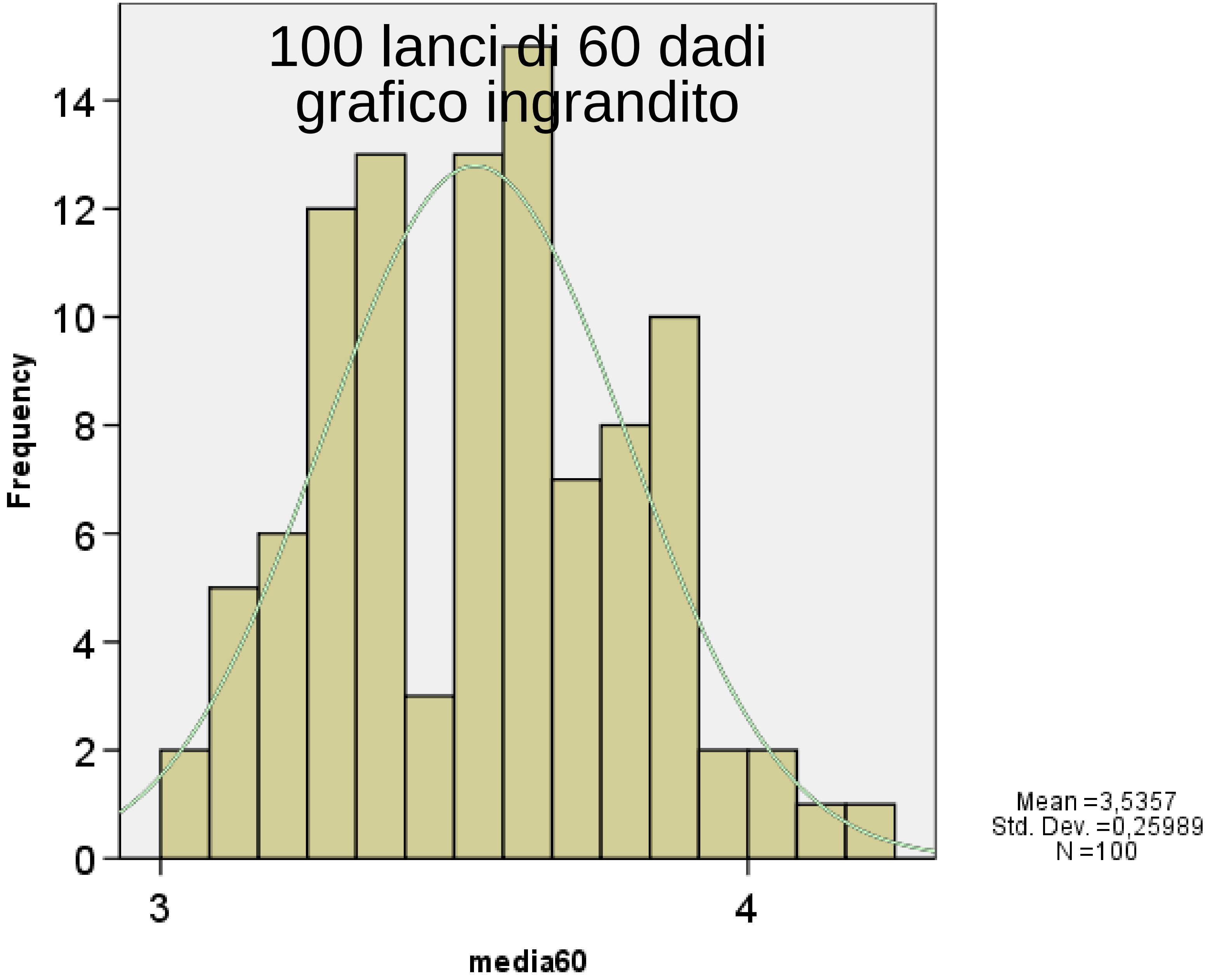
# 100 studenti lanciano 30 dadi e calcolano la media



# 100 studenti lanciano 60 dadi e calcolano la media



# 100 lanci di 60 dadi grafico ingrandito



# Conclusione 1

- All'aumentare della numerosità del campione
- (a) la media si avvicina alla media della popolazione
- (b) la dispersione diminuisce
- (c)Le medie seguono la distribuzione normale

$$M = \mu$$

# Calcolo della varianza

- La varianza e la deviazione standard diminuiscono all'aumentare della numerosità

$$\text{deviazione standard} = \frac{\sigma}{\sqrt{n}}$$

$$\text{varianza} = \frac{\sigma^2}{n}$$

# La legge dei grandi numeri

- All'aumentare di N aumenta la probabilità che la media campionata si avvicini alla media della popolazione

# Proprietà della distribuzione campionaria delle medie

- (1) La distribuzione campionaria delle medie segue approssimativamente la curva normale
- (2) La media delle medie campionarie tende a essere uguale al valore medio della popolazione
- (3) La deviazione standard è inversamente proporzionale alla numerosità del campione
- (4) la deviazione standard della distribuzione campionaria della media prende il nome di errore standard

# Teorema del limite centrale

Se si estraggono da una popolazione infiniti campioni di dimensione n, la media si distribuisce come una normale con

- Media=media della popolazione
- Deviazione standard (errore standard $\frac{\sigma}{\sqrt{n}}$) ,ottenuta dividendo la varianza della popolazione per la numerosità del campione e calcolando la radice quadrata .

# Forma della distribuzione della popolazione

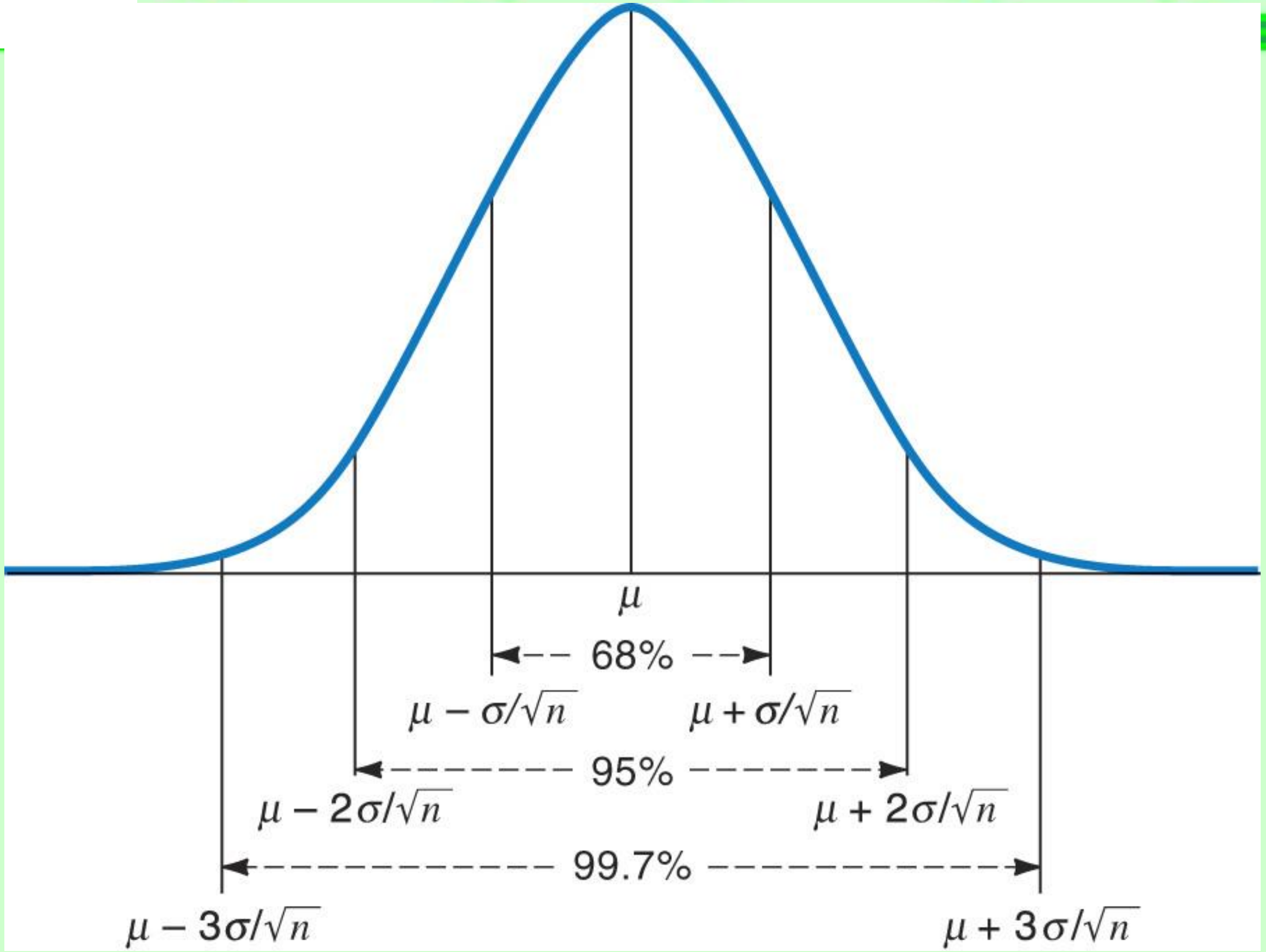
Se la distribuzione della popolazione è normale, il campione può essere di qualsiasi dimensione

Se la popolazione non è normale (asimmetrica o discreta) il campione dovrà essere di almeno 30 osservazioni

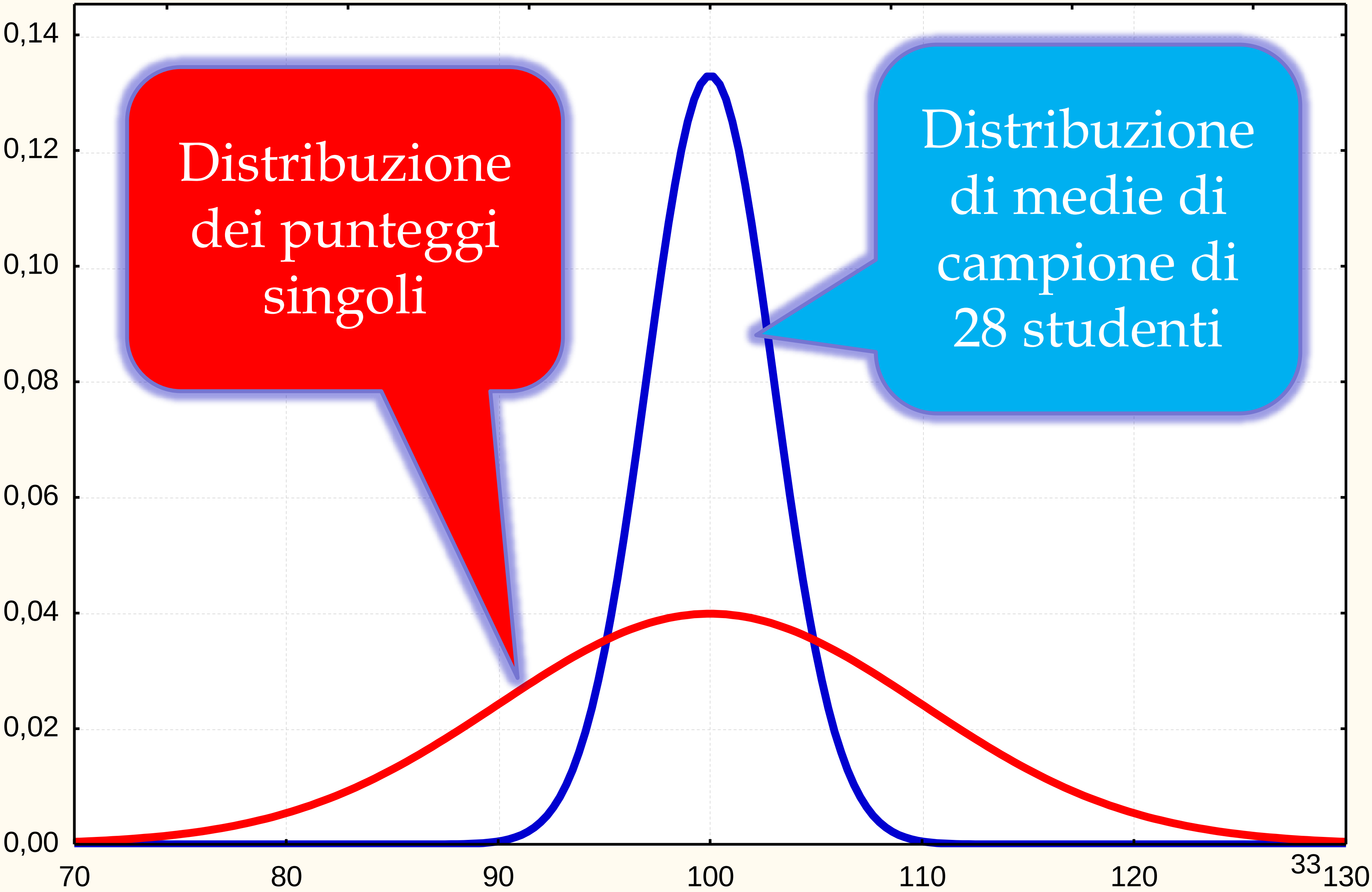
# Esempio

- In una classe il QI dei 28 studenti è pari a 108. Il test usato per misurare l'intelligenza ha una media di 100 e una d.s. pari a 16.
- Qual è la probabilità di avere un risultato simile per caso? Si tratta di una classe normale o molto abile?

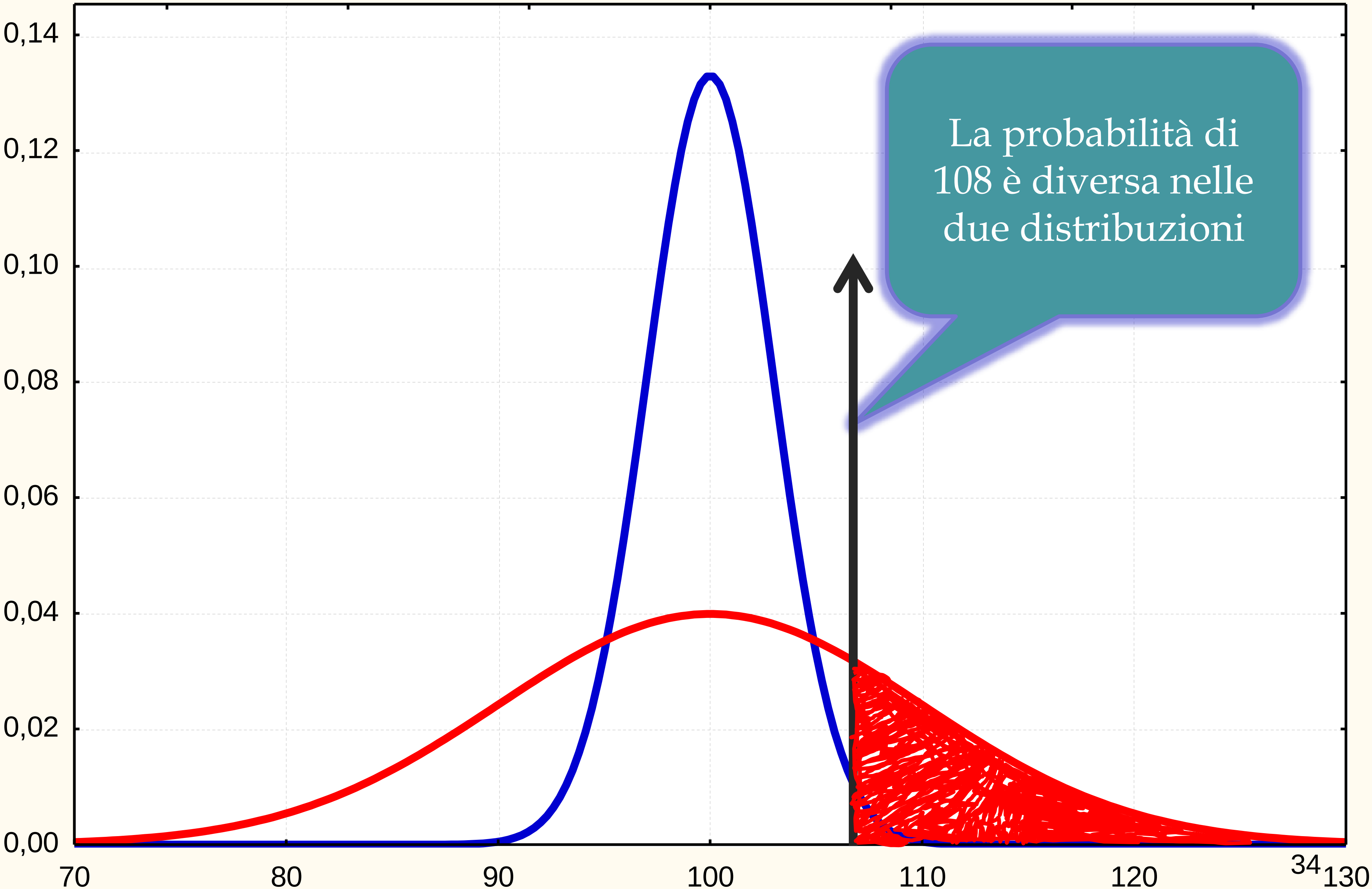
# Grafico della distribuzione della media campionaria da popolazione normale



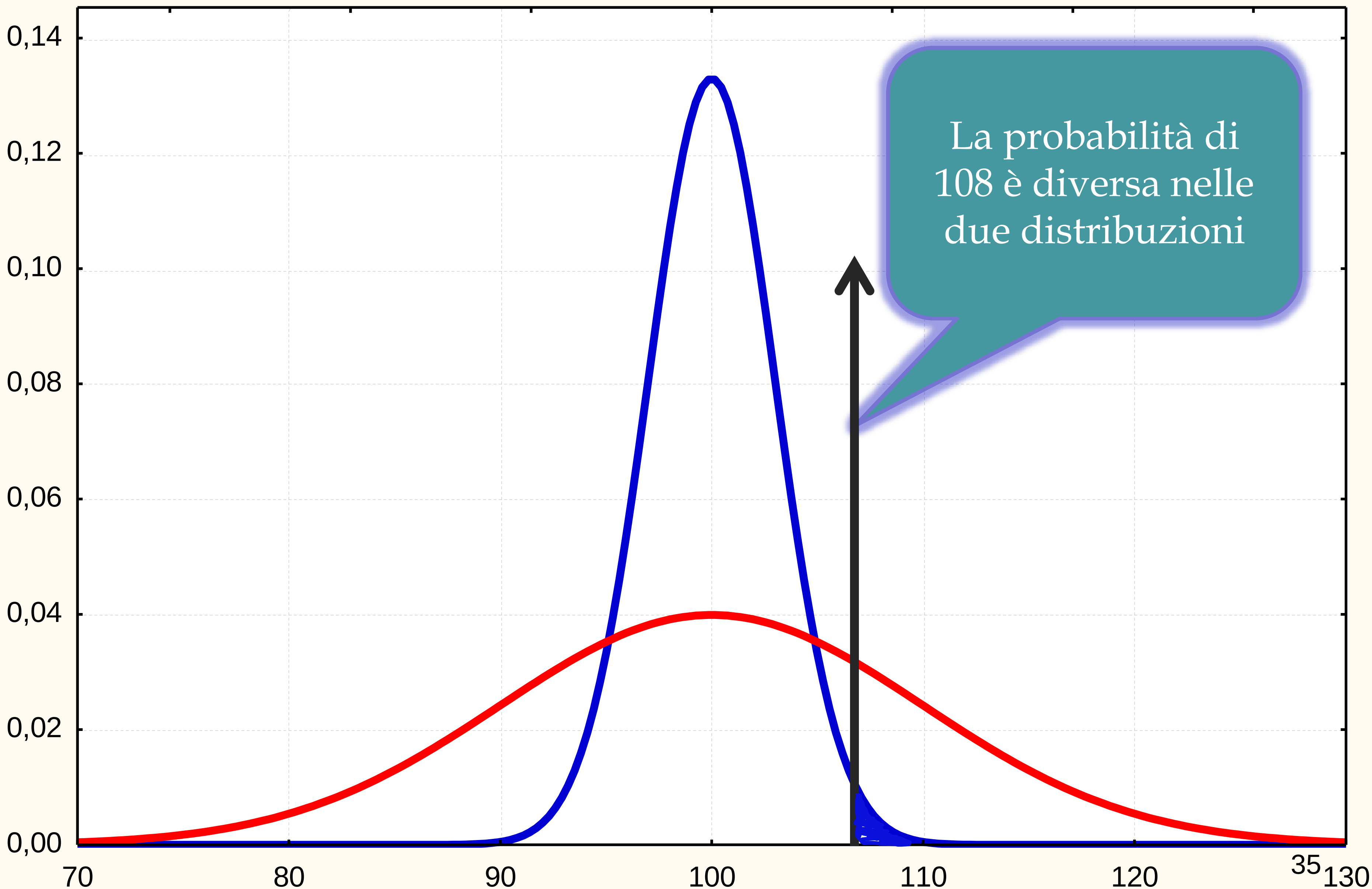
Blu =normal(x;100;3)
Rossa =normal(x;100;10)



Blu =normal(x;100;3)
Rossa =normal(x;100;10)



Blu =normal(x;100;3)
Rossa =normal(x;100;10)



# Trasformiamo il valore 108 in punto zeta…

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

- Media= 100
- Ds= 16/√28 = 3,023

$$z = \frac{108 - 100}{16 / \sqrt{28}}$$

$$z = \frac{8}{16 / \sqrt{28}} = 2,64$$

# Trasformiamo il valore 108 in punto zeta…

| | dev stan | Formula | Punto zeta | Perc sotto | perc sopra |
|---|---|---|---|---|---|
| popolazione | ds = 16 | Z=(108-100)/16 | 0,5 | 0,6915 | 0,3085 |
| Distribuzione campionaria | ds= 3.023 | Z=(108-100)/(16/radq(28)) | 2,64 | 0,9959 | 0,0041 |

# Il valore osservato 108 è raro o comune?

Non si tratta di una osservazione tratta da una popolazione con media 100 e deviazione standard 16, bensì di una media di un campione.

La media è 100, ma la deviazione standard è pari a 3, e quindi il 108 è molto sopra la media.

Dalle tavole della normale si legge che il punto zeta uguale o superiore a 2,64 si ottiene solo raramente, solo 4 volte su 1000 ($p=0,004$)

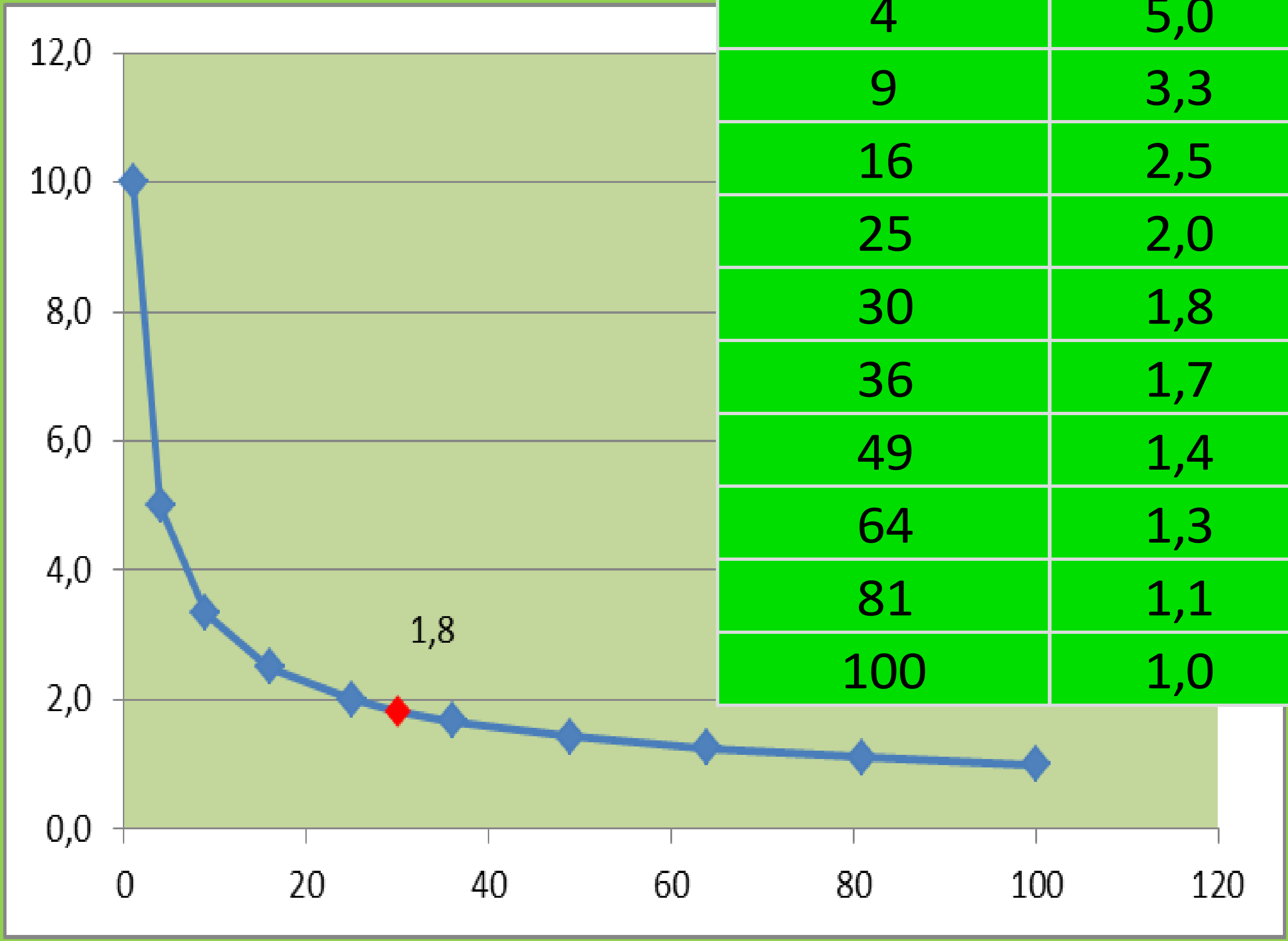
Si conclude che la classe è più dotata della norma

# Inferenza con l’uso della distribuzione campionaria delle medie

- (1) Calcolare il valore medio e l’errore standard delle medie campionarie
- (2) Disegnare il grafico
- (3) Calcolare il valore standardizzato
- (4) Cercare la probabilità sulla tavola della normale e trarre le conclusioni

# Variazione dell'e.s. all'aumentare della numerosità del campione

La variazione dell'E.S (in ordinata) al variare della numerosità di N (in orrizontale) per una popolazione con la deviazione standard pari a 10. Il guadagno di precisione diminuisce notevolmente dopo un N di 30




| Numerosità del campione | DS = 10 |
|---|---|
| 1 | 10,0 |
| 4 | 5,0 |
| 9 | 3,3 |
| 16 | 2,5 |
| 25 | 2,0 |
| 30 | 1,8 |
| 36 | 1,7 |
| 49 | 1,4 |
| 64 | 1,3 |
| 81 | 1,1 |
| 100 | 1,0 |