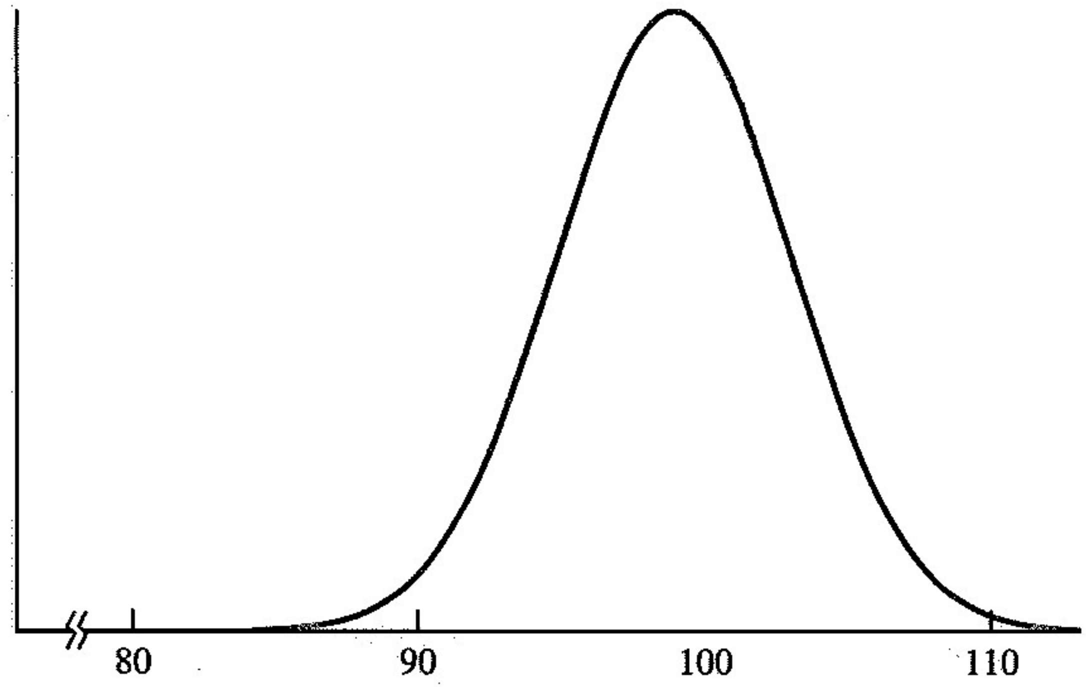
IL TEOREMA DEL LIMITE CENTRALE

Teorema del limite centrale

Consideriamo la distribuzione del QI nella popolazione degli studenti di scuola superiore:

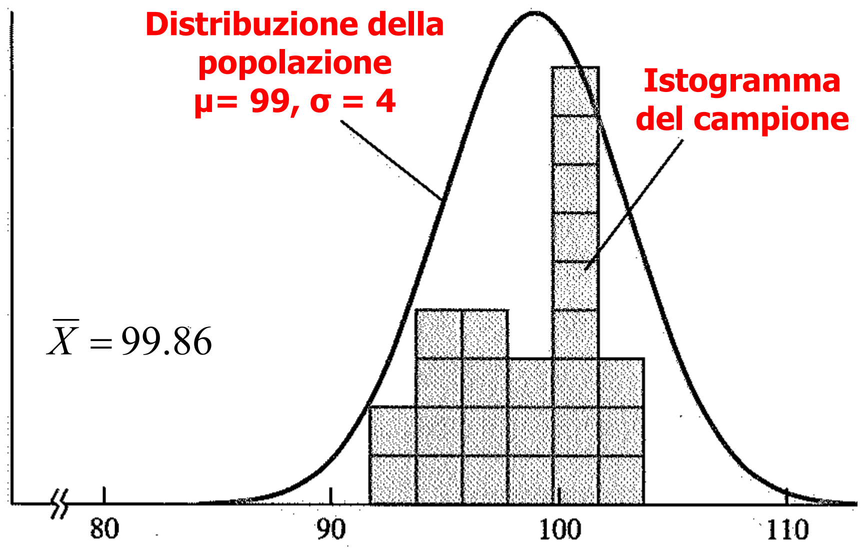


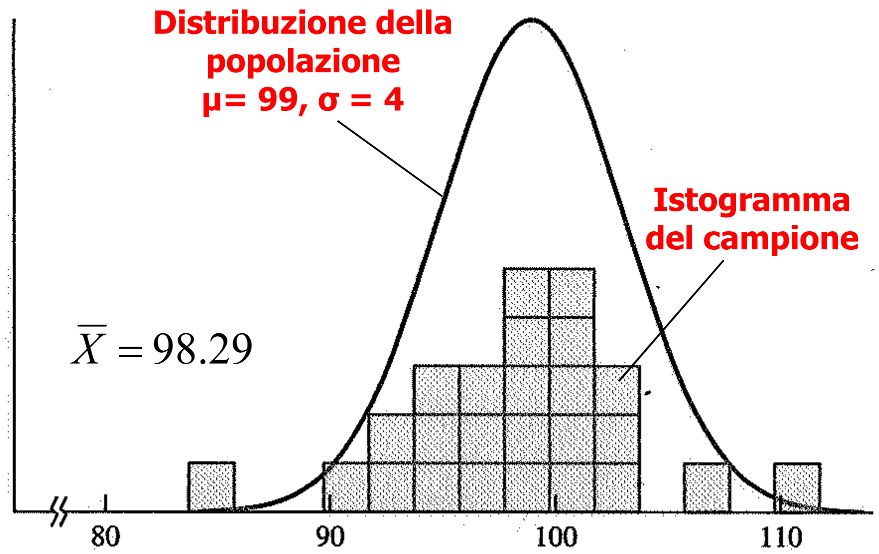
Immaginiamo di conoscere le caratteristiche della popolazione, che ha media $\mu = 99$ e dev.st $\sigma = 4$

Assumiamo inoltre che la distribuzione sia normale

Consideriamo le seguenti figure:

Immaginiamo ora di estrarre una serie di campioni dalla popolazione di studenti, e di misurare il QI…

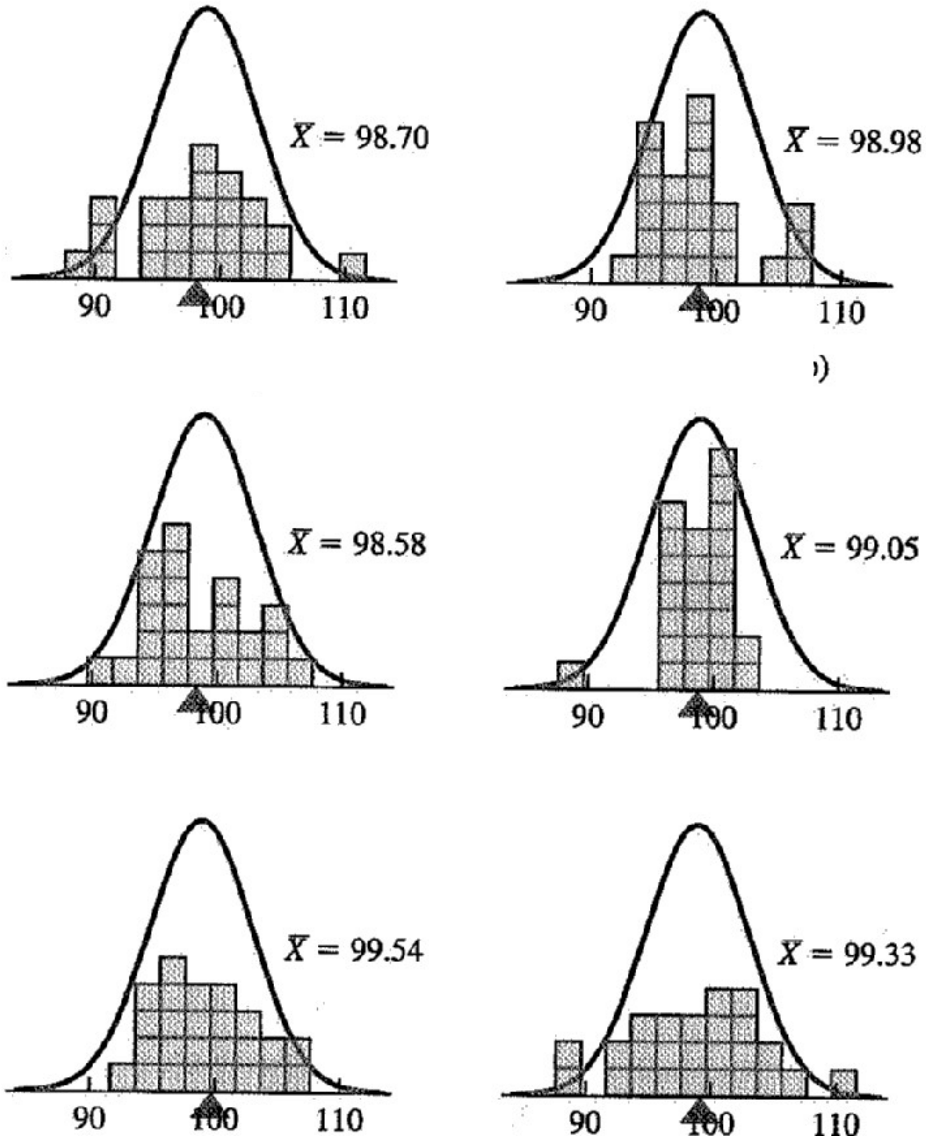




Le figure riportano due campioni di ampiezza 25, estratti casualmente dalla stessa popolazione

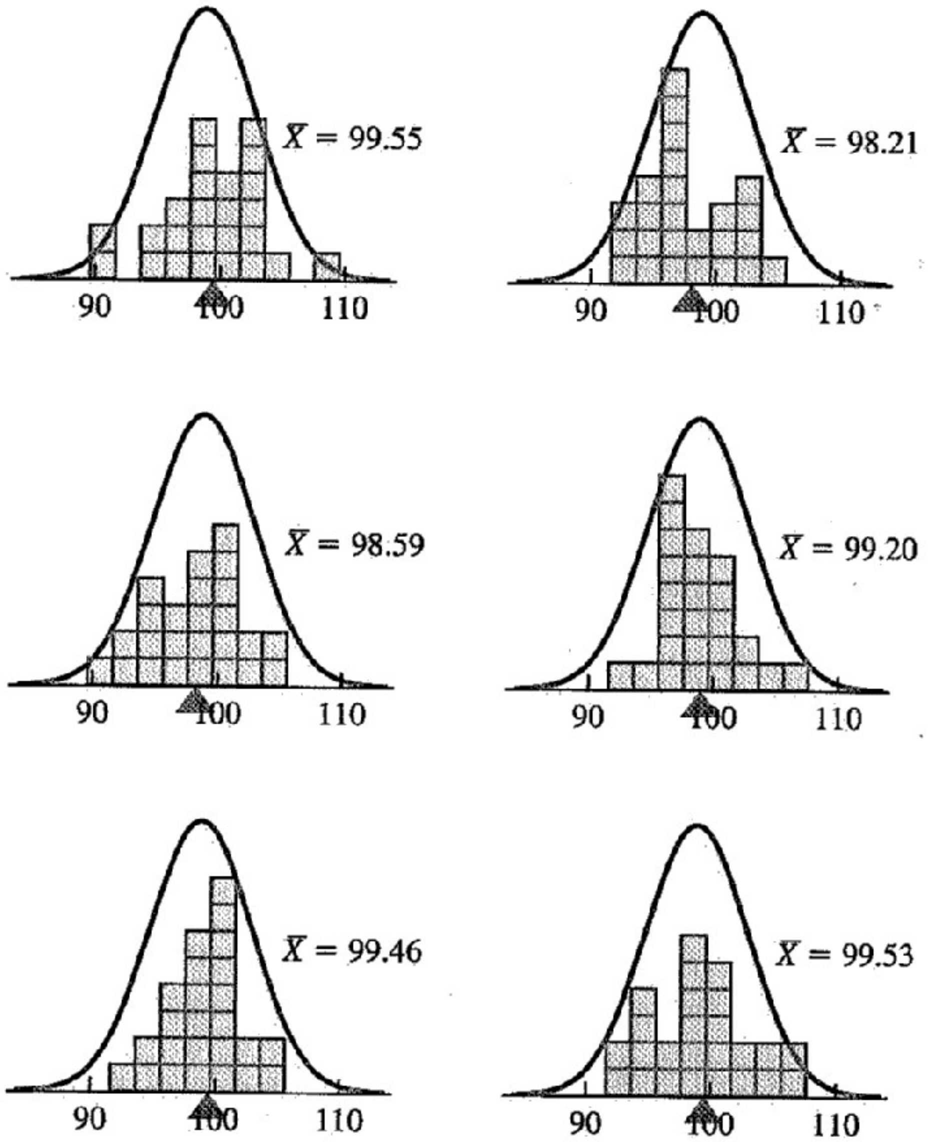
Consideriamo le seguenti figure:

Nelle figure vengono rappresentati altri sei campioni (n =25) estratti casualmente dalla stessa popolazione ...



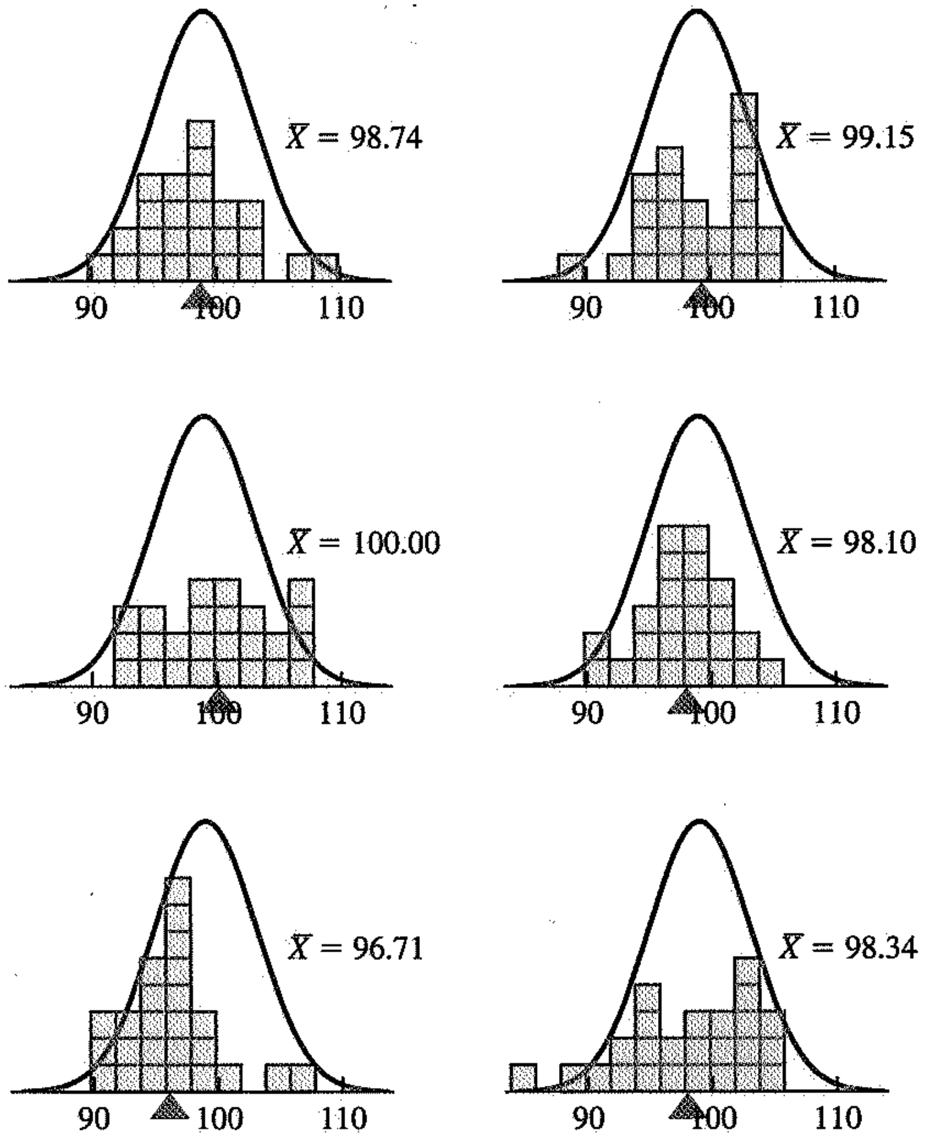
Consideriamo le seguenti figure:

Altri sei campioni casuali di ampiezza 25 …



Consideriamo le seguenti figure:

… ancora sei campioni casuali di ampiezza 25



I campioni sono identici alla popolazione da cui sono stati estratti?

Teorema del limite centrale

Abbiamo estratto in tutto 20 campioni di $n = 25$

Su ogni campione abbiamo calcolato la media:
le medie variano tra 96.71 e 100.00

In particolare, le medie sono pari a:

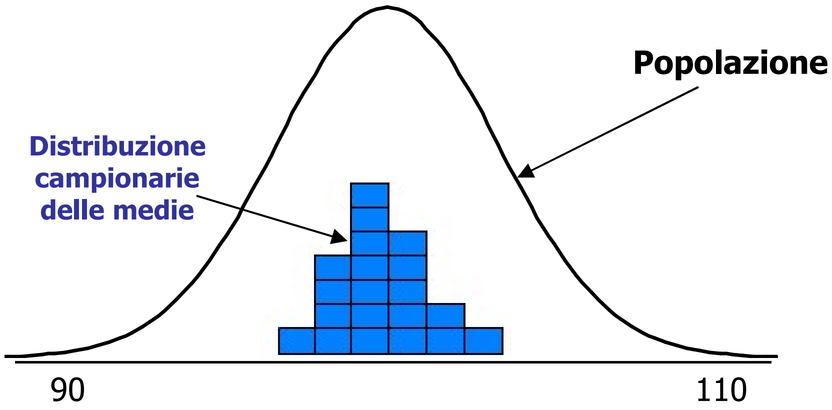
99.86	98.29	98.70	98.98	98.58	99.05	99.54
99.33	99.55	98.21	98.59	99.20	99.46	99.53
98.74	99.15	100.00	98.10	96.71	98.34	

Questa distribuzione viene definita distribuzione campionaria delle medie (rappresenta un concetto fondamentale nell'inferenza statistica)

Teorema del limite centrale

La distribuzione campionaria delle medie

In questo caso abbiamo una distribuzione campionaria delle medie di ampiezza 25 (n = 25)

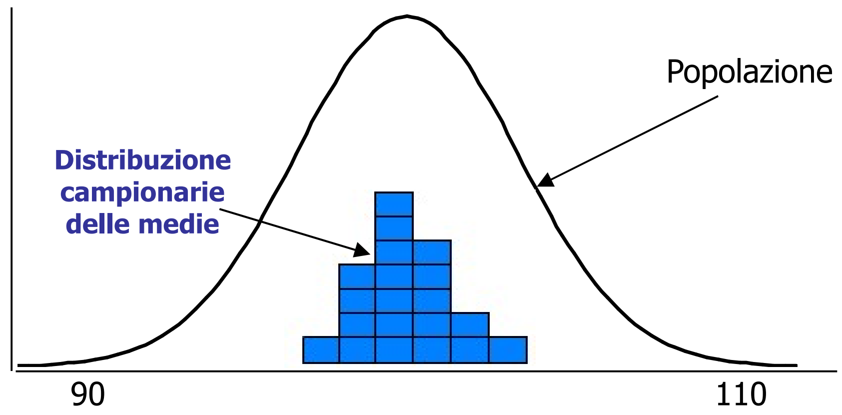


Le unità statistiche in questa distribuzione non sono i soggetti ma le medie (calcolate su campioni di ampiezza *n* estratti da una stessa popolazione)

Come per qualsiasi distribuzione, anche per la distribuzione campionaria delle medie è possibile definire tendenza centrale (media), dispersione (deviazione standard) e forma della distribuzione

La distribuzione campionaria delle medie

Popolazione e distribuzione campionaria delle medie: confrontiamo le due distribuzioni

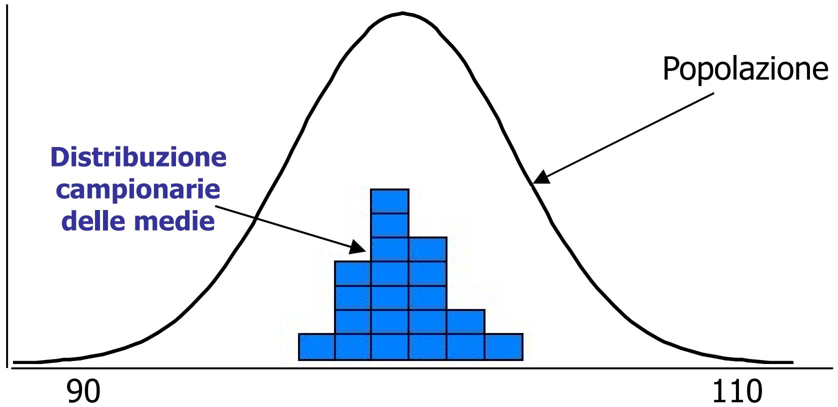


Tendenza centrale: la media della distribuzione campionaria delle medie (98.90) è molto simile alla media della popolazione (μ= 99.00)

Teorema del limite centrale

La distribuzione campionaria delle medie

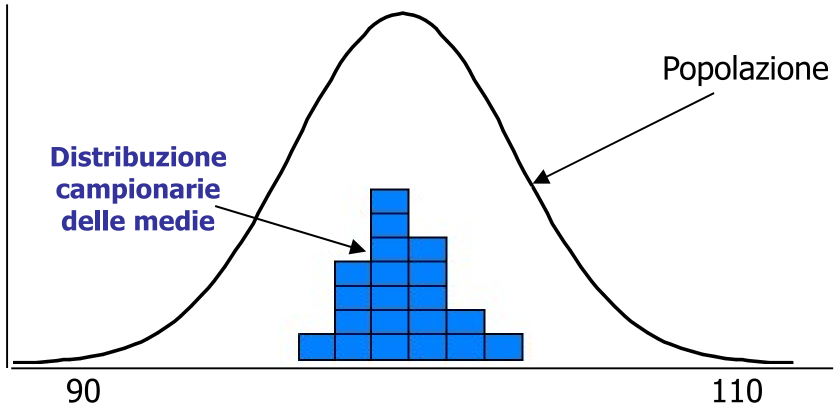
Popolazione e distribuzione campionaria delle medie: confrontiamo le due distribuzioni



Dispersione: poiché le medie dei 20 campioni estratti dalla popolazione sono tutte molto simili tra loro, la distribuzione campionaria delle medie ha una deviazione standard ridotta ($s = 0.76$), inferiore a quella della popolazione ($\sigma = 4.00$)

La distribuzione campionaria delle medie

Popolazione e distribuzione campionaria delle medie: confrontiamo le due distribuzioni



Forma: la forma della distribuzione campionaria delle medie approssima la forma della popolazione. Entrambe le distribuzioni hanno forma normale

Alcune distinzioni terminologiche

Abbiamo parlato di medie, riferendole a tre diverse distribuzioni

1. **La media del campione** $\rightarrow \overline{X}$

2. **La media della popolazione** $\rightarrow \mu$

3. **La media della distribuzione campionaria delle medie** $\rightarrow \mu_{\overline{X}}$

Alcune distinzioni terminologiche

Anche la deviazione standard è stata riferita a tre diverse distribuzioni

1. **La deviazione standard del campione** $\rightarrow s$

2. **La deviazione standard della popolazione** $\rightarrow \sigma$

3. **La deviazione standard della distribuzione campionaria delle medie** $\rightarrow \sigma_{\overline{X}}$

   Quest'ultima viene definita anche "errore standard della media"

Teorema del limite centrale

La distribuzione campionaria delle medie

Come detto, la distribuzione che abbiamo descritto si basa solo su 20 campioni

In teoria, la distribuzione campionaria delle medie si riferisce alle medie di tutti i possibili campioni di ampiezza n (nel nostro esempio 25) che è possibile estrarre

In questo caso dovremmo:

1→estrarre dalla popolazione tutti i possibili campioni di ampiezza 25 (un numero infinito di combinazioni)

2→calcolare la media di ciascun campione

3→calcolare la media e la deviazione standard della distribuzione che include tutte queste medie [distribuzione campionaria della media]

La distribuzione campionaria delle medie

Ovviamente non è possibile calcolare in questo modo la media e la deviazione standard della distribuzione campionarie delle medie

Non è possibile, infatti, estrarre un numero infinito di campioni!

Per calcolare la distribuzione campionaria delle medie possiamo sfruttare le proprietà del teorema del limite centrale

Teorema del limite centrale

Il teorema del limite centrale è uno dei teoremi fondamentali in statistica

Se conosciamo media e deviazione standard della popolazione, il teorema del limite centrale ci consente di conoscere le caratteristiche della distribuzione campionaria delle media senza dover estrarre dalla popolazione tutti i possibili campioni di ampiezza *n*

Il teorema del limite centrale specifica tre importanti caratteristiche della distribuzione campionaria delle medie:

(1) Media

(2) Dispersione

(3) Forma

1. Media della distribuzione campionaria delle medie

La media della distribuzione campionaria delle medie coincide con la media della popolazione dalla quale i campioni sono stati estratti:

$$\mu = \mu_{\overline{X}}$$

Media della popolazione

Media della distribuzione campionaria delle medie

Teorema del limite centrale

2. Variabilità della distribuzione campionaria delle medie

La deviazione standard della distribuzione campionaria delle medie (errore standard della media) è pari a:

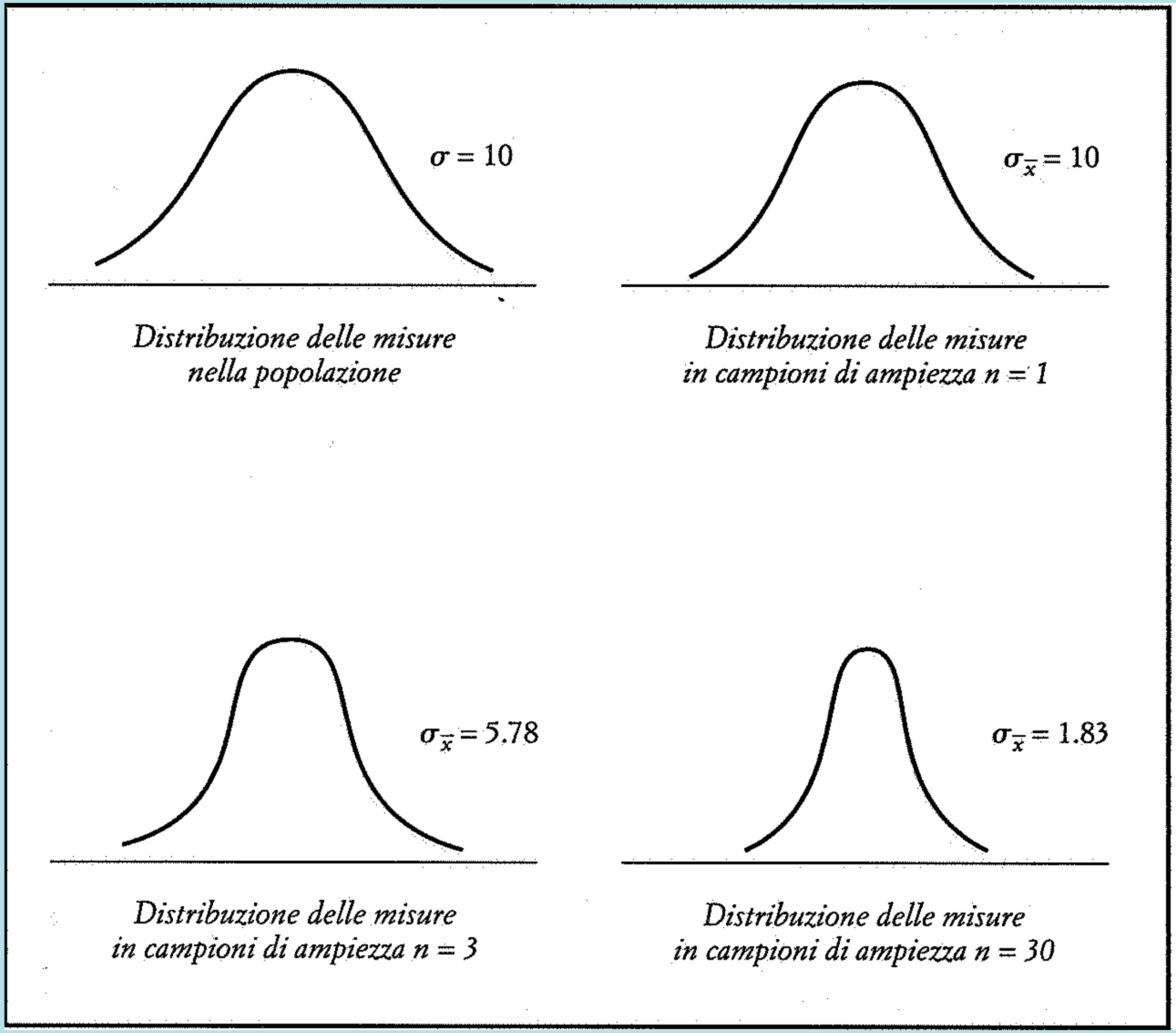
$$\sigma_{\overline{X}} = \frac{\sigma}{\sqrt{n}}$$

Rispetto alla popolazione, dunque, la deviazione standard della distribuzione campionaria delle medie è ridotta di un fattore pari a $\sqrt{n}$

Se consideriamo la varianza:

$$\sigma^2_{\overline{X}} = \frac{\sigma^2}{n}$$

Distribuzioni campionarie delle medie di diverse ampiezza



SE n = 1, L'ERRORE STANDARD DELLA MEDIA È PARI ALLA DEVIAZIONE STANDARD DELLA POPOLAZIONE

ALL'AUMENTARE DI n, L'ERRORE STANDARD DELLA MEDIA SI RIDUCE

2. Variabilità della distribuzione campionaria delle medie

All'aumentare di *n* la variabilità della distribuzione campionaria delle medie diminuisce fino a tendere a zero (legge dei grandi numeri)

Man mano che l'ampiezza dei campioni aumenta, infatti, la media di ciascuno di essi diventa una stima sempre più «precisa» della media della popolazione, coincidendo con essa quando *n* = N

Quando *n* = N tutte le medie (calcolate su campioni di ampiezza uguale alla popolazione) sono identiche e la variabilità della distribuzione campionaria delle medie è pari a zero

3. Forma della distribuzione campionaria delle medie

Se la popolazione di riferimento ha forma normale, anche la distribuzione campionaria delle medie si distribuisce normalmente

E quando la popolazione non ha forma normale?

All'aumentare di n, la forma della distribuzione campionaria delle medie approssima la forma normale

Se il campione è sufficientemente grande (in genere $n > 30$), la distribuzione campionaria delle medie si distribuisce normalmente (anche quando la popolazione non ha forma normale!)

3. Forma della distribuzione campionaria delle medie

Per concludere …

La distribuzione campionaria delle medie ha forma normale quando:

- **la popolazione ha forma normale (indipendentemente da n)**
- **n > 30 (indipendentemente dalla forma della popolazione)**

La distribuzione campionaria delle medie NON ha forma normale quando:

- **la popolazione non ha forma normale e n < 30**

Teorema del limite centrale

Vediamo una serie di esempi tratti da:

http://onlinestatbook.com/stat_sim/sampling_dist/index.html

Esempio 1

LA POPOLAZIONE HA FORMA NORMALE

ESTRAIAMO DALLA POPOLAZIONE UN NUMERO SEMPRE MAGGIORE DI CAMPIONI DI AMPIEZZA 25

Esempio 1

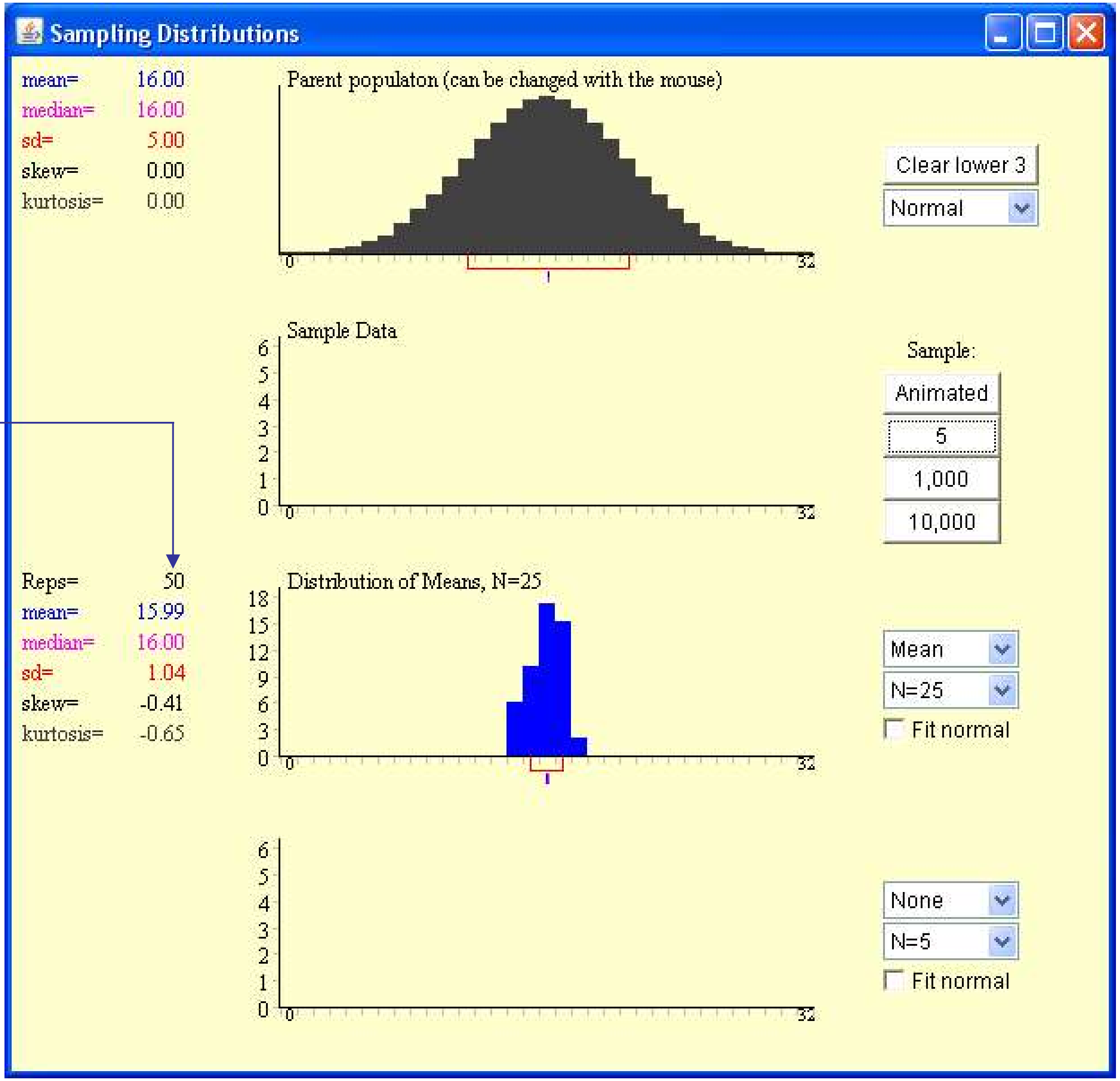
La distribuzione della popolazione ha forma normale

n = 25

Numero di campioni estratti = 50

Distribuzione campionaria delle medie:

Asimmetria: – .041
Curtosi: – 0.65



Teorema del limite centrale

Esempio 1

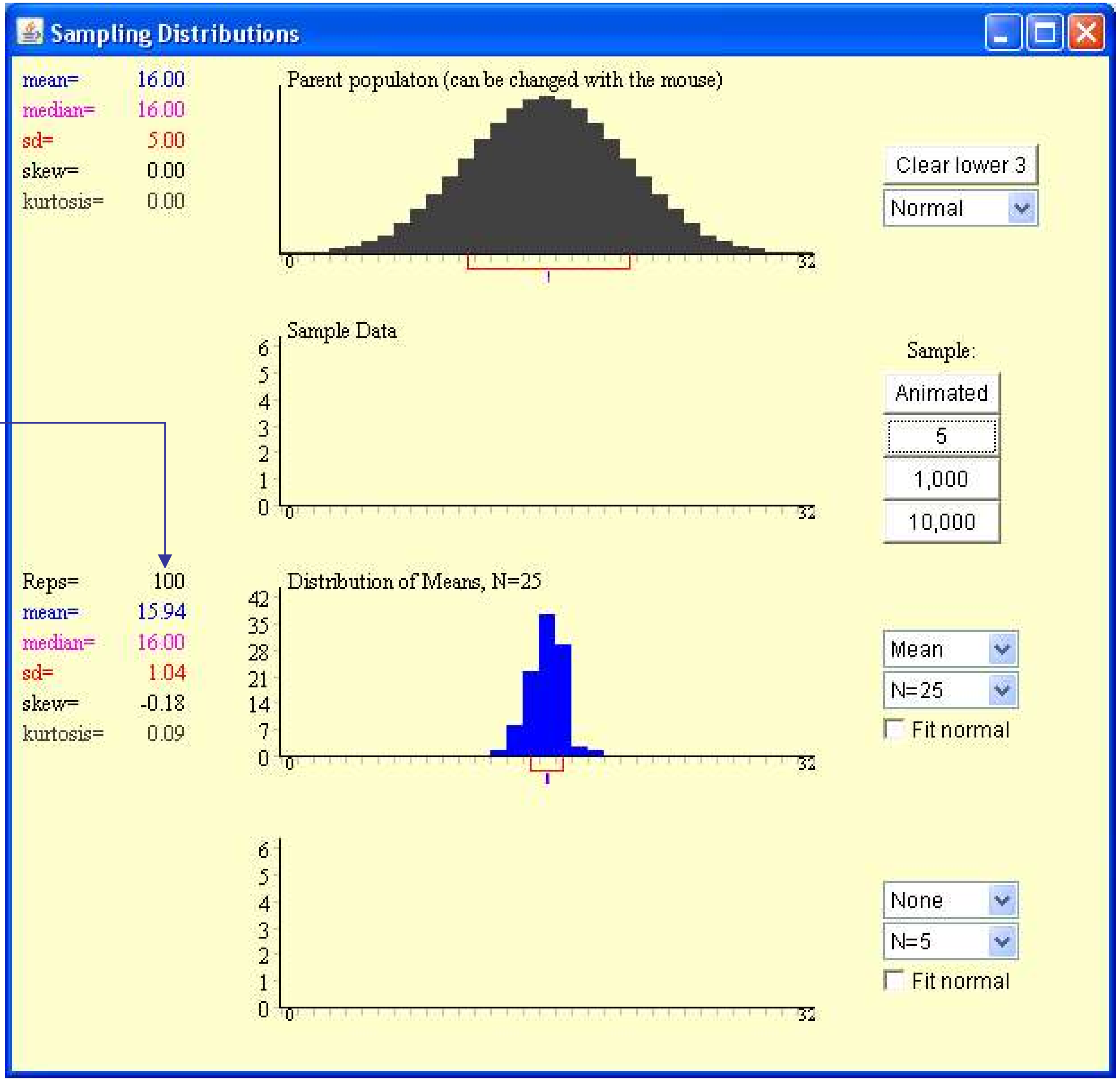
La distribuzione della popolazione ha forma normale

n = 25

Numero di campioni estratti = 100

Distribuzione campionaria delle medie:

Asimmetria: – 0.18
Curtosi: 0.09



Teorema del limite centrale

Esempio 1

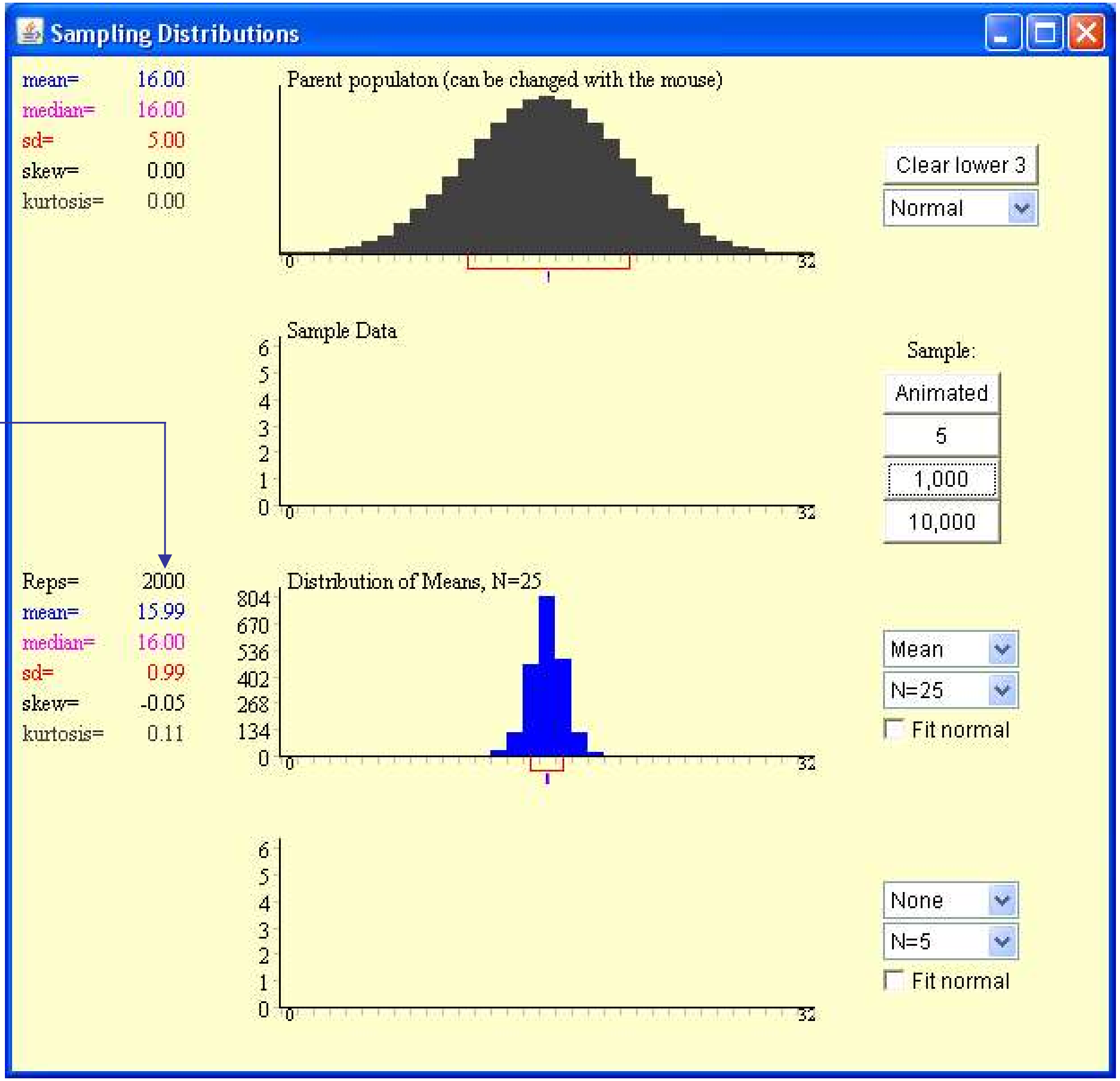
La distribuzione della popolazione ha forma normale

n = 25

Numero di campioni estratti = 2000

Distribuzione campionaria delle medie:

Asimmetria: – 0.05
Curtosi: 0.11



Teorema del limite centrale

Esempio 1

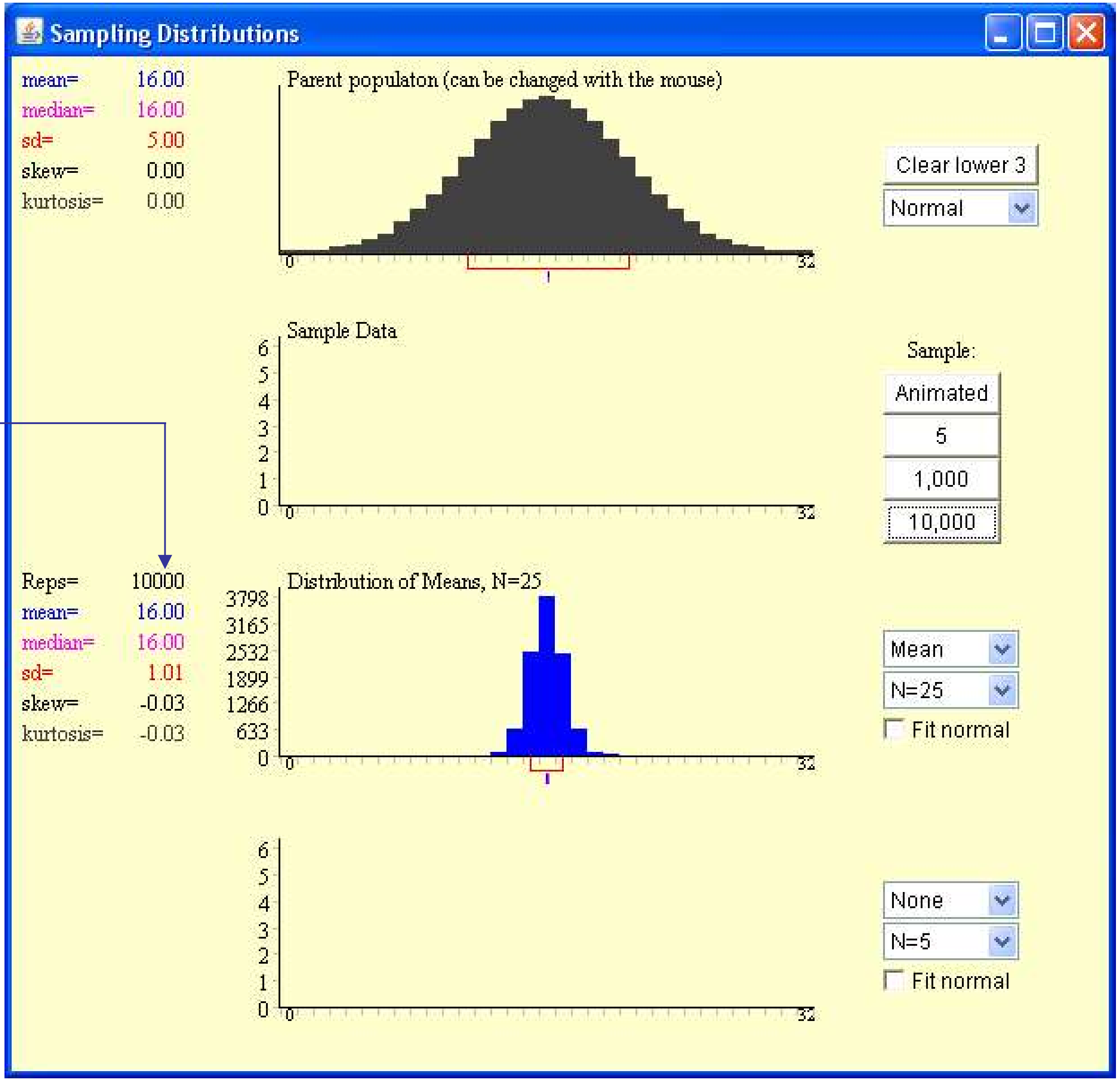
La distribuzione della popolazione ha forma normale

n = 25

Numero di campioni estratti = 10000

Distribuzione campionaria delle medie:

Asimmetria: – 0.03
Curtosi: – 0.03



Teorema del limite centrale

<u>All'aumentare del numero di campioni estratti, la forma della distribuzione campionaria delle medie approssima la forma normale</u>

Esempio 2

LA POPOLAZIONE HA FORMA NORMALE

ESTRAIAMO DALLA POPOLAZIONE UN NUMERO SEMPRE MAGGIORE DI CAMPIONI DI AMPIEZZA 5

Teorema del limite centrale

Esempio 2

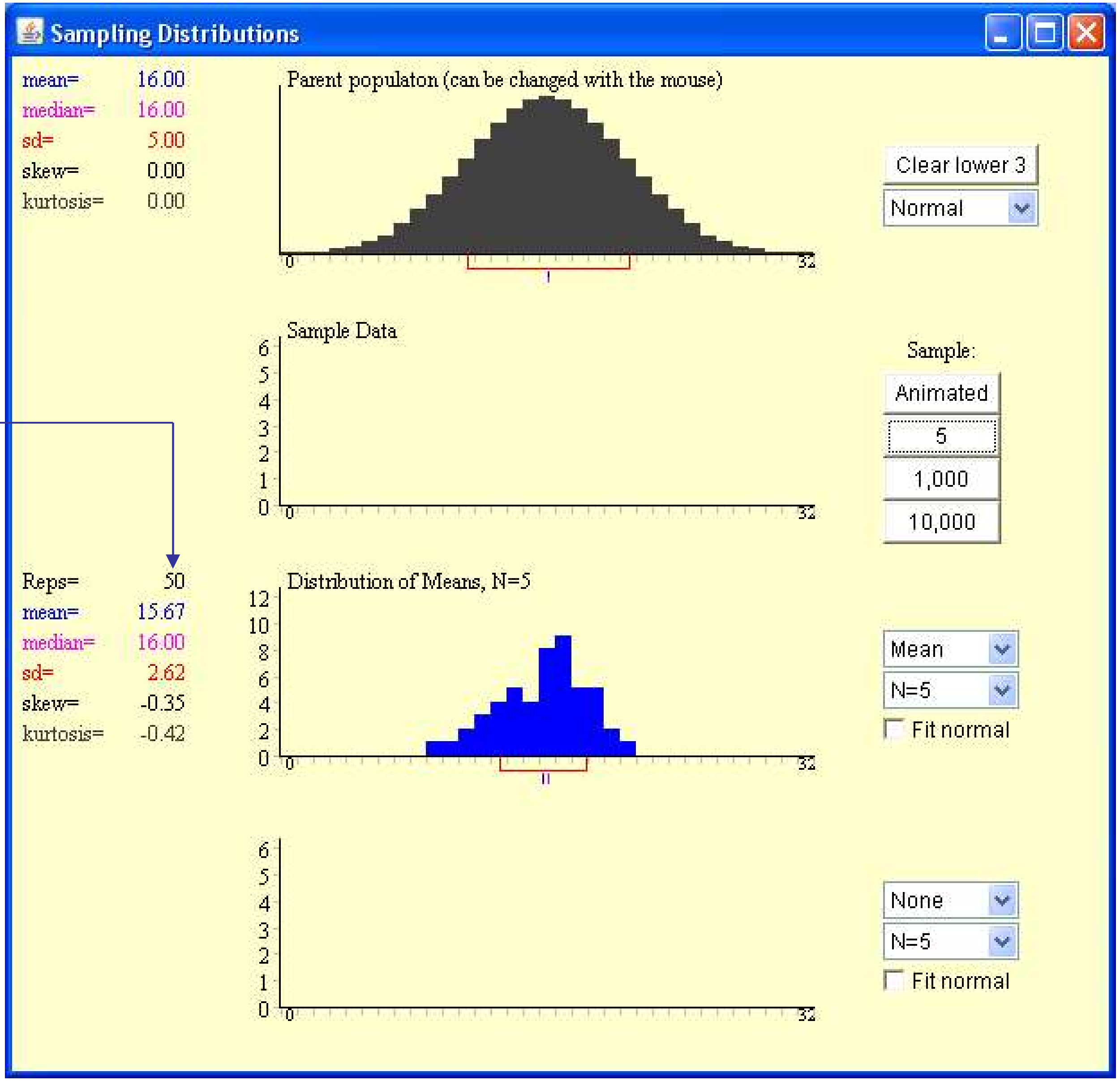
La distribuzione della popolazione ha forma normale

n = 5

Numero di campioni estratti = 50

Distribuzione campionaria delle medie:

Asimmetria: – 0.35
Curtosi: – 0.42



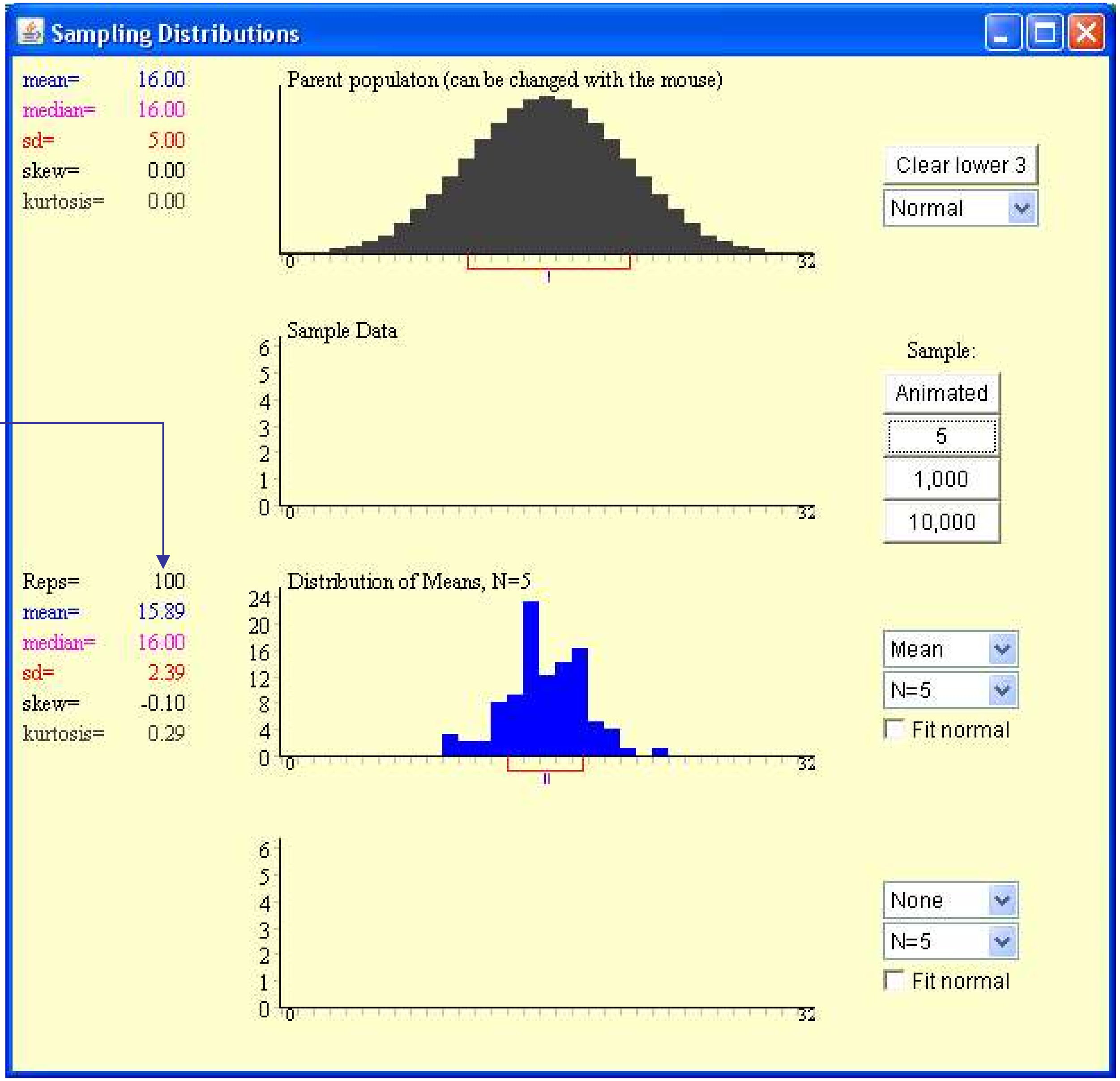
Teorema del limite centrale

Esempio 2

La distribuzione della popolazione ha forma normale

n = 5

Numero di campioni estratti = 100



Distribuzione campionaria delle medie:

Asimmetria: – 0.10
Curtosi: + 0.29

Teorema del limite centrale

Esempio 2

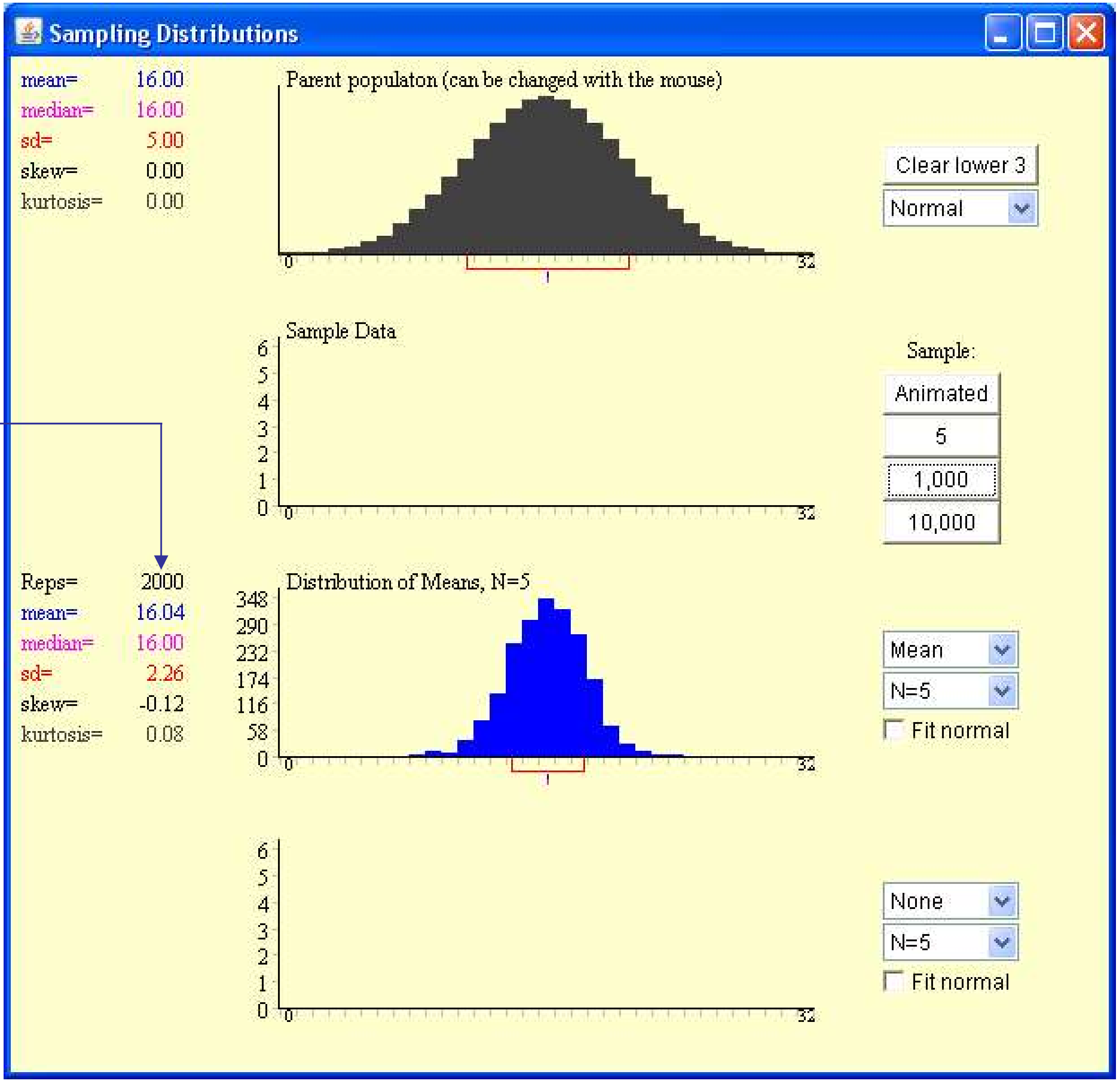
La distribuzione della popolazione ha forma normale

n = 5

Numero di campioni estratti = 2000

Distribuzione campionaria delle medie:

Asimmetria: – 0.12
Curtosi: + 0.08



Teorema del limite centrale

Esempio 2

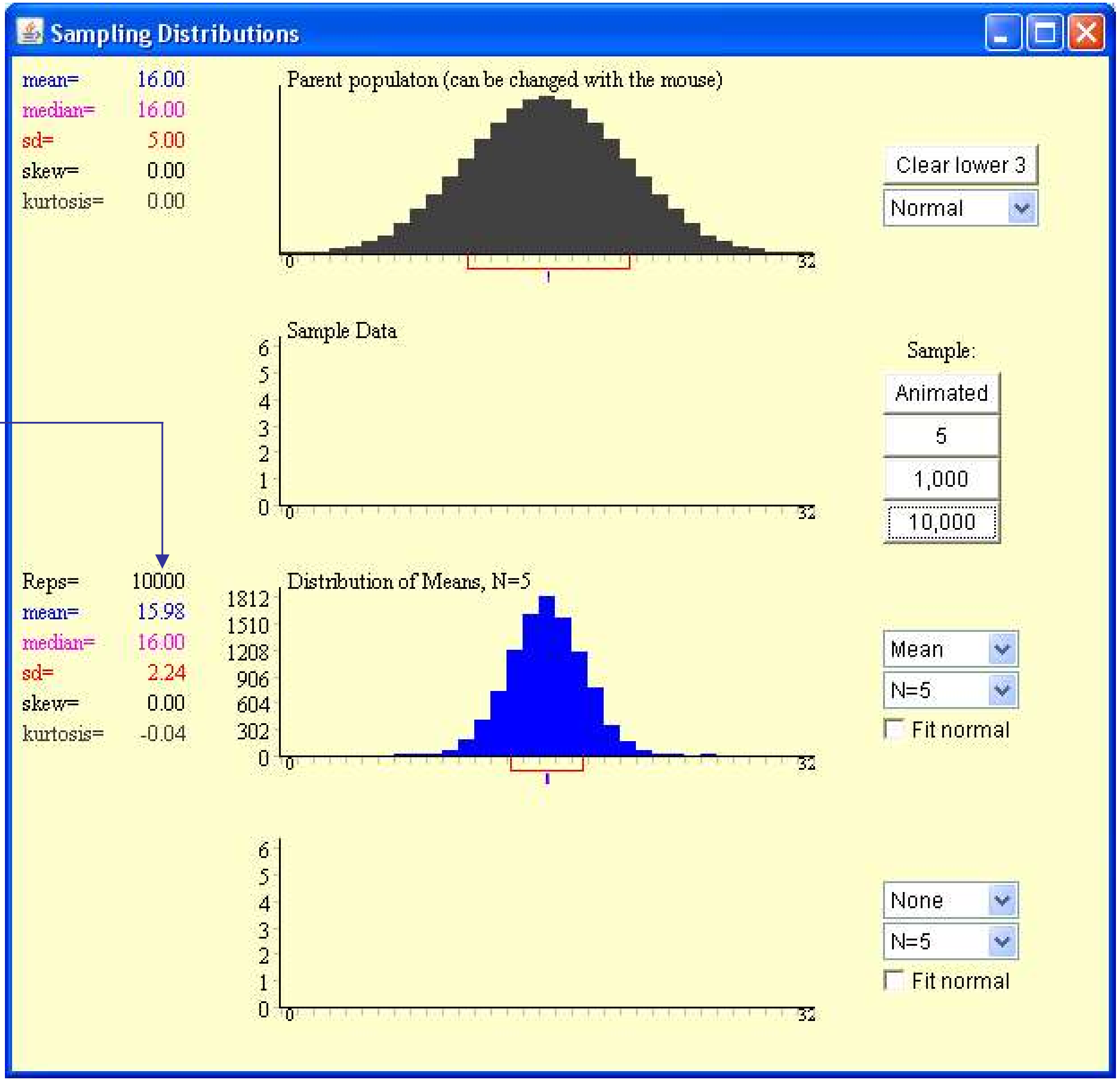
La distribuzione della popolazione ha forma normale

n = 5

Numero di campioni estratti = 10000

Distribuzione campionaria delle medie:

Asimmetria: 0.00
Curtosi: – 0.04



Teorema del limite centrale

<u>Anche con campioni più piccoli (*n* = 5), la distribuzione campionaria delle medie finisce con l'approssimare la curva normale (se la popolazione ha forma normale)</u>

Esempio 3

LA POPOLAZIONE HA FORMA NON NORMALE (ASIMMETRICA)

ESTRAIAMO UN NUMERO MOLTO ELEVATO DI CAMPIONI DI AMPIEZZA 25 DALLA POPOLAZIONE

Teorema del limite centrale

Esempio 3

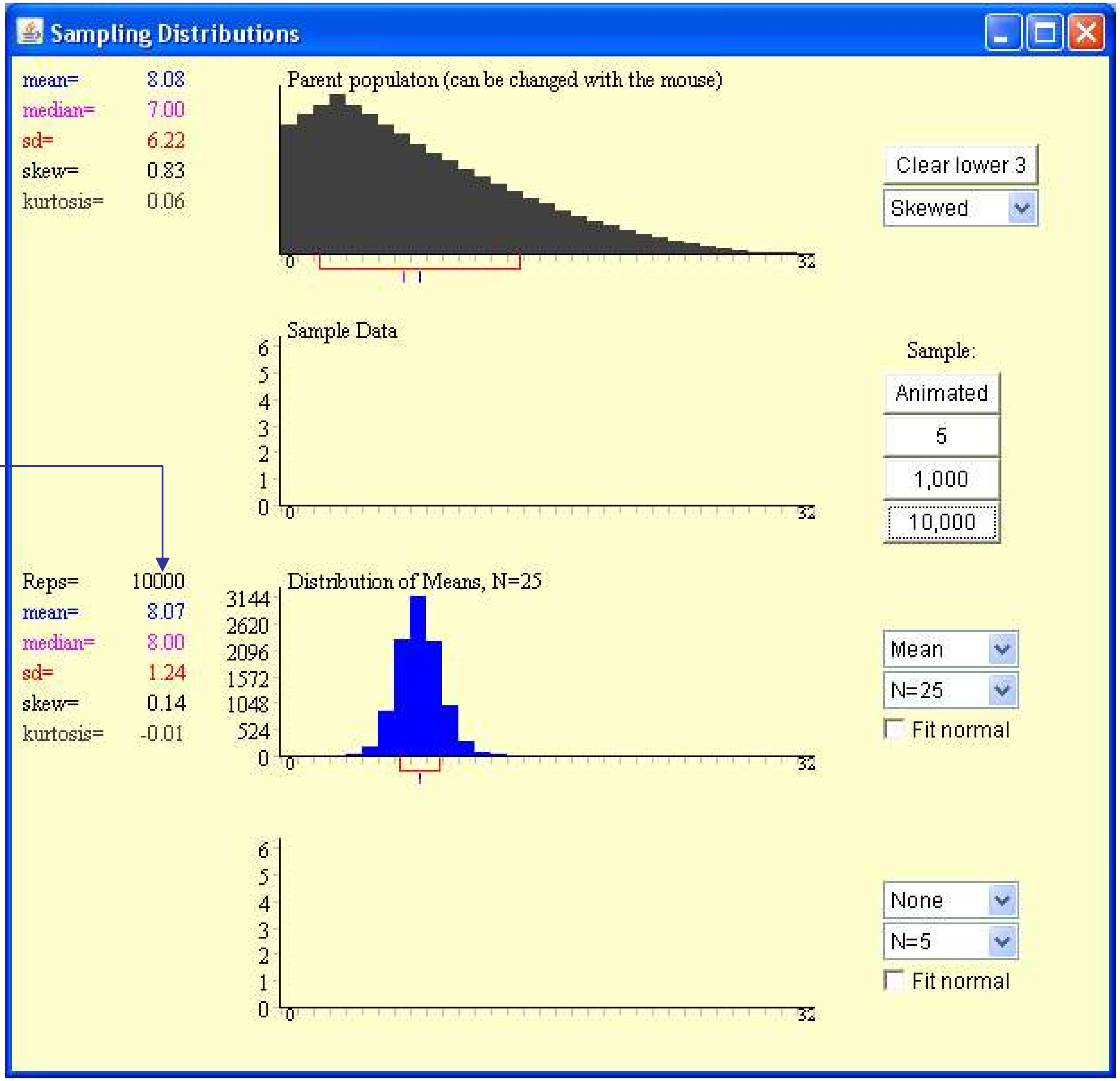
La distribuzione della popolazione ha forma non normale (asimmetrica)

n = 25

Numero di campioni = 10000

Distribuzione campionaria delle medie:

Asimmetria: 0.14
Curtosi: – 0.01



Esempio 4

LA POPOLAZIONE HA FORMA NON NORMALE (BIMODALE)

ESTRAIAMO UN NUMERO MOLTO ELEVATO DI CAMPIONI DI AMPIEZZA 25 DALLA POPOLAZIONE

Teorema del limite centrale

Esempio 4

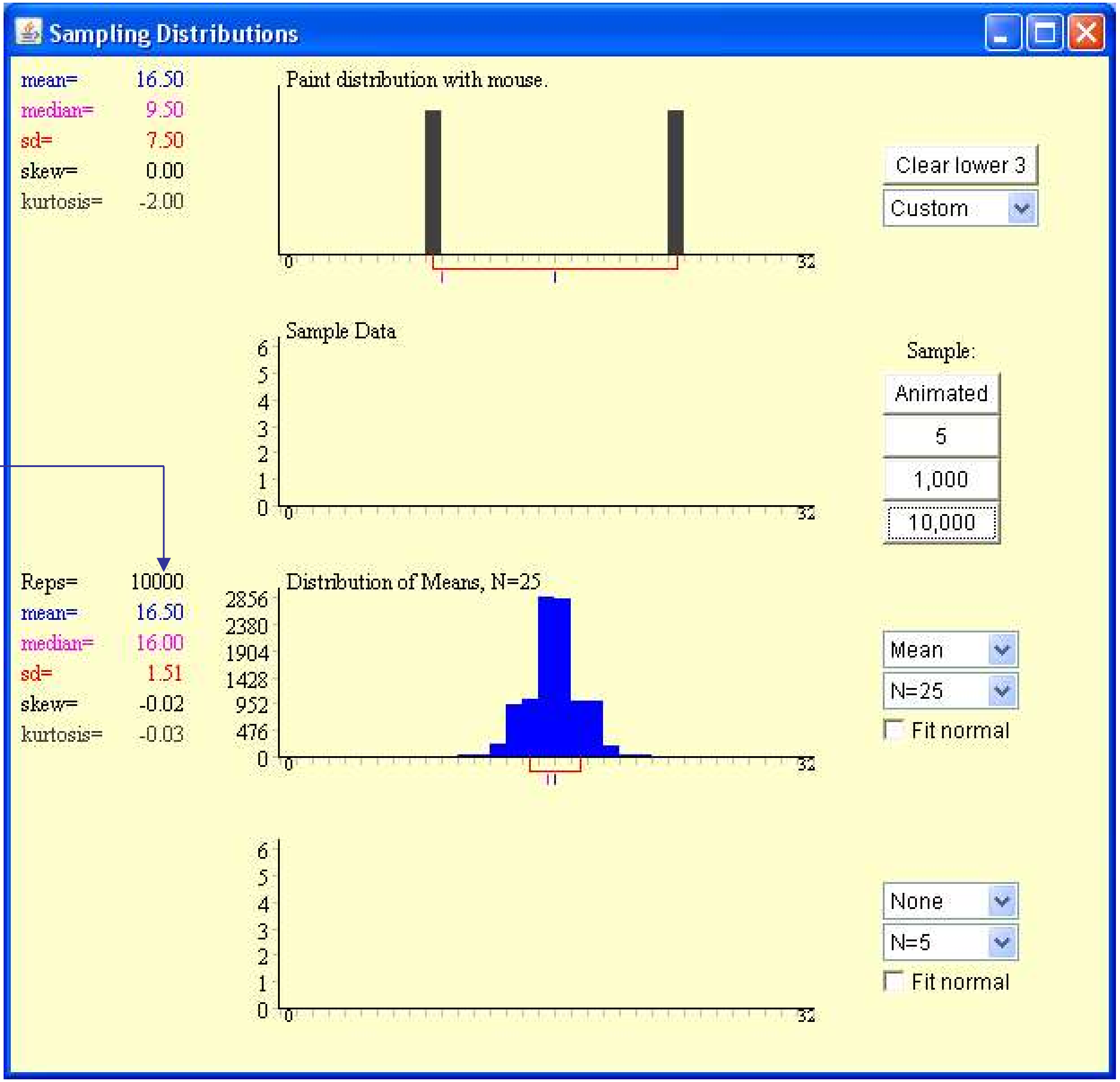
La distribuzione della popolazione ha forma non normale (bimodale)

n = 25

Numero di campioni = 10000

Distribuzione campionaria delle medie:

Asimmetria: – 0.02
Curtosi: – 0.03



Teorema del limite centrale

Se n è sufficientemente elevato, la forma della distribuzione campionaria delle medie approssima la curva normale anche se la popolazione NON si distribuisce normalmente

Esercizio 1

La variabile "apertura mentale" si distribuisce normalmente nella popolazione, con media $\mu=14.6$ e deviazione standard $\sigma=2.5$

Un ricercatore misura l'apertura mentale in un campione di 35 soggetti, estratto casualmente dalla popolazione. La media del campione è pari a 13.9

Descrivi le caratteristiche della distribuzione campionaria delle medie da cui proviene il campione:

Media: ________

Deviazione Standard: ____________________

Varianza: ___________________

Forma: _______________

Esercizio 2

Cosa cambia nella distribuzione campionaria delle medie se il campione è composto da 20 soggetti (invece di 35)?

Media: ______________________

Deviazione Standard:

____

Varianza: ____

Forma: __________________

Esercizio 3

Tra gli studenti universitari iscritti a «Psicologia e processi sociali» (popolazione) la coscienziosità, rilevata tramite il Big Five Questionnaire (BFQ) ha $\mu = 16.2$ e $\sigma = 4.1$

La distribuzione della variabile coscienziosità nella popolazione ha un'asimmetria pari a -1.82 e una curtosi pari a 0.00

Un ricercatore utilizza il BFQ per misurare la coscienziosità ad un campione di 15 studenti iscritti a «Psicologia e processi sociali», ottenendo i seguenti risultati:

Media $\bar{X} = 15.4$; deviazione standard $s = 4.9$

Teorema del limite centrale

1. In base a quanto riportato nella slide precedente, cosa si può dire a proposito della forma della variabile "coscienziosità" nella popolazione?

a) E normale b) Non è normale c) Non si può dire

2. Cosa si può dire a proposito della forma della "distribuzione campionaria delle medie"?

a) È sicuramente normale b) Probabilmente non è normale

3. A quanto è uguale n?

a) Un numero molto elevato o infinito b) 15

c) Non si può dire d) $\sqrt{15}$

4. A quanto è uguale N?

a) infinito b) 15

c) $\sqrt{15}$ d) Non si può dire

5. A quanto è uguale l'errore standard della media?

a) 4.10 b) 0.27 c) 1.06 d) 4.90

Teorema del limite centrale

6. A quanto è uguale la media della distribuzione campionaria delle medie?

a) 15.4 b) 16.2 c) 4.1 d) Non si può dire

7. Quanti campioni bisogna estrarre per calcolare la distribuzione campionaria delle medie di ampiezza 15?

a) 15 b) Un numero molto elevato o infinito
c) Almeno 30 d) Dipende da n

8. Secondo il teorema del limite centrale, si verifica una riduzione dell'errore standard della media se (segna le risposte corrette, anche più di una):

□ aumenta n
□ aumenta N
□ diminuisce n
□ diminuisce σ
□ aumenta il numero di campioni estratti
□ la forma della distribuzione è normale

Teorema del limite centrale

9. A quanto è uguale la varianza della popolazione?

a) 16.81
b) 4.1
c) 1.12
d) Non si può dire

10. Indica le condizioni in cui la forma della distribuzione campionaria delle medie è normale (segna le risposte corrette, anche più di una):

□ n = 150
□ n = 43
□ n = 10
□ n = 20