Distribuzioni campionarie
Parte I

Inferenza

Statistica Descrittiva

Statistica Inferenziale:
affronta problemi di decisione in condizioni di incertezza basandosi sia su informazioni a priori sia sui dati campionari

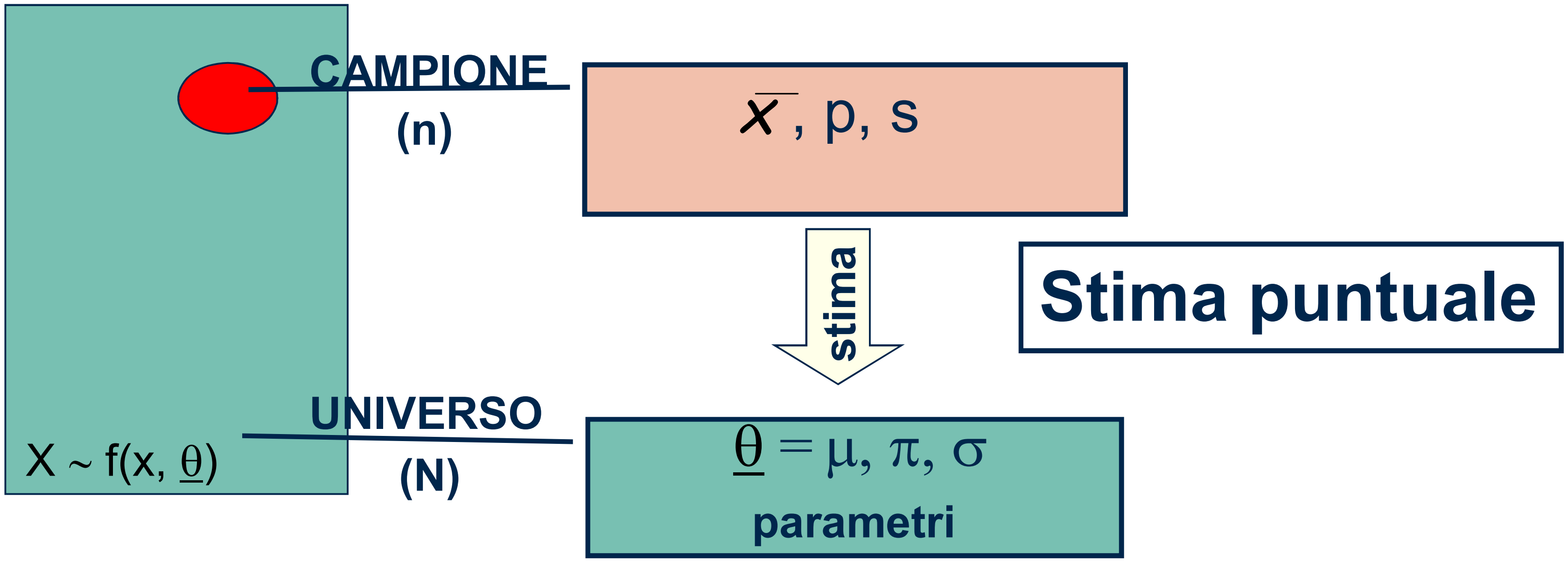
Inferenza

Inferenza

1) Stima puntuale
2) Stima intervallare
(intervalli di confidenza)

3) Test statistico
(verifica di ipotesi)

Processo di stima



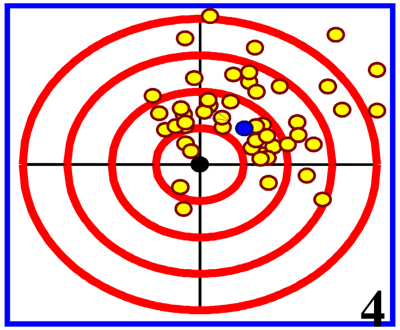
Processo di stima

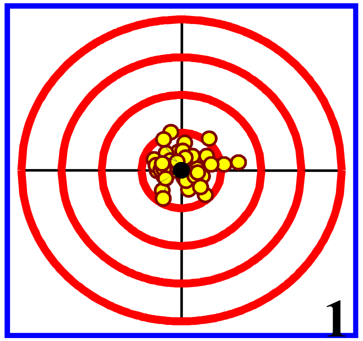
Obiettivo:

ottenere risultati accurati e precisi

Evitare **errori sistematici**

Minimizzare **errori casuali**





Rilevanza della fase di **pianificazione** di una ricerca

Processo di stima

Non è possibile valutare la bontà della stima ottenuta da un singolo campione.

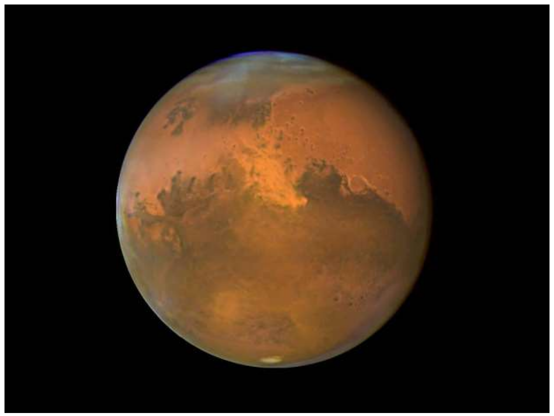
Si deve fare riferimento ad una situazione teorica in cui si considerano le stime ottenute da tutti i possibili campioni estraibili da una popolazione (universo).

Distribuzione della media campionaria

Trasferiamoci su Marte !

L'intera popolazione di marziani è piuttosto limitata

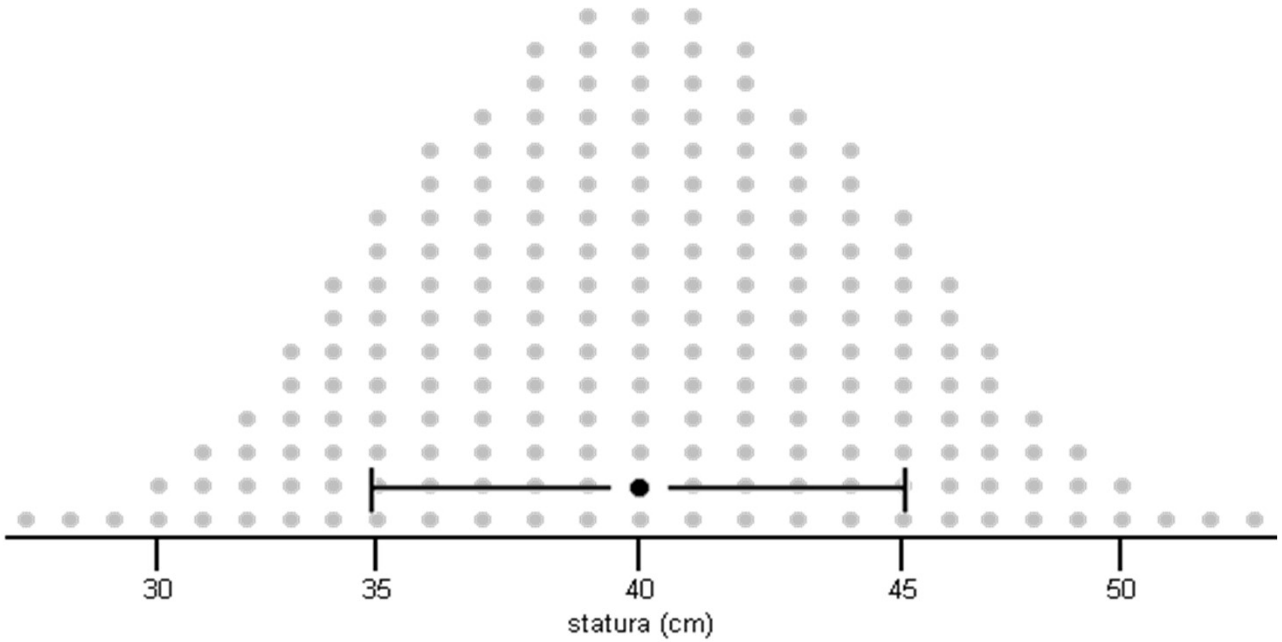
Immaginiamo sia pari a 200
Vogliamo studiarne la statura…



La popolazione di marziani:

- Supponiamo che la distribuzione della statura (X) dell'intera popolazione di numerosità N=**200** sia nota e pari a :

$$\mathbf{X \sim N(\mu=40\ cm,\ \sigma^2=25\ cm^2)}$$

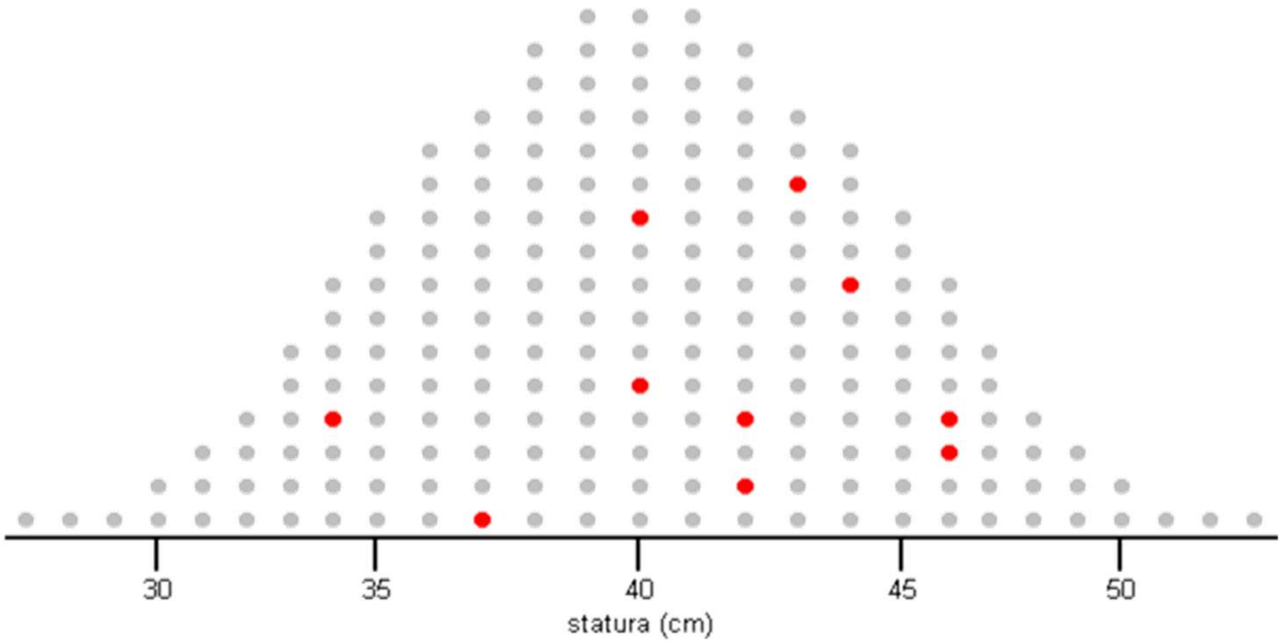


La popolazione di marziani: campionamento

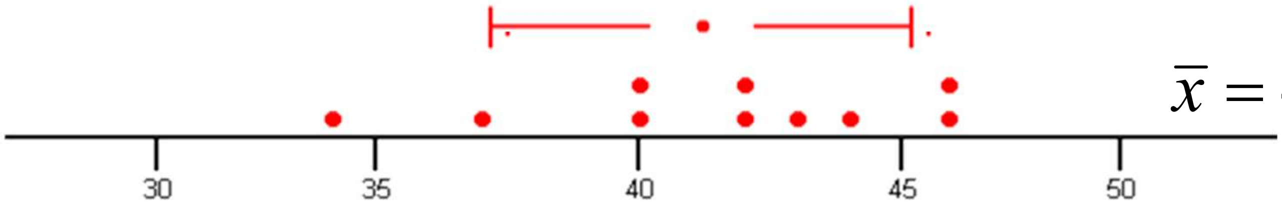
- Studiamo il processo di stima della statura media μ partendo dalle informazioni campionarie (come se non conoscessimo l'intera popolazione)
- Le nostre risorse ci permettono di osservare al massimo campioni di n=10 marziani
- Estraiamo un campione di n=10 in maniera casuale (e con reinserimento)
- Ricaviamo da esso una stima $\bar{x}$ del parametro μ

La popolazione di marziani: campionamento

primo campione: n=10



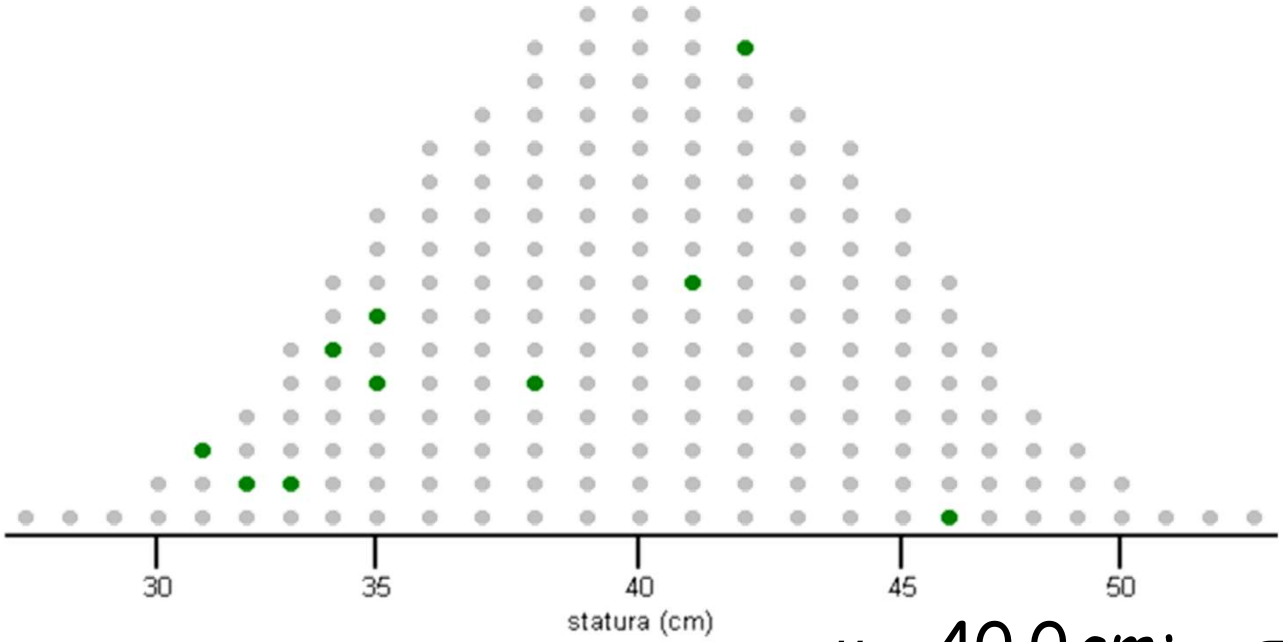
$$\mu = 40.0\,cm; \quad \sigma = 5\,cm$$



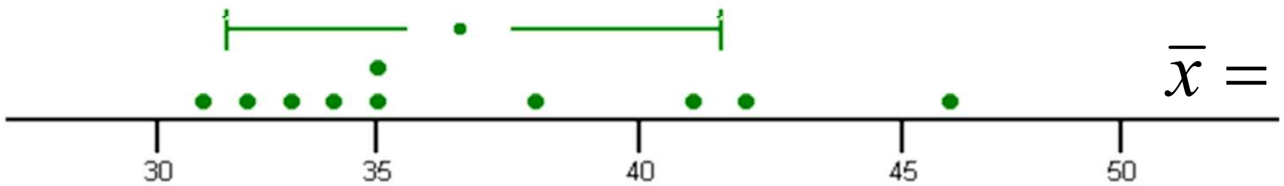
$$\bar{x} = 41.5\,cm$$

La popolazione di marziani: campionamento

secondo campione: n=10



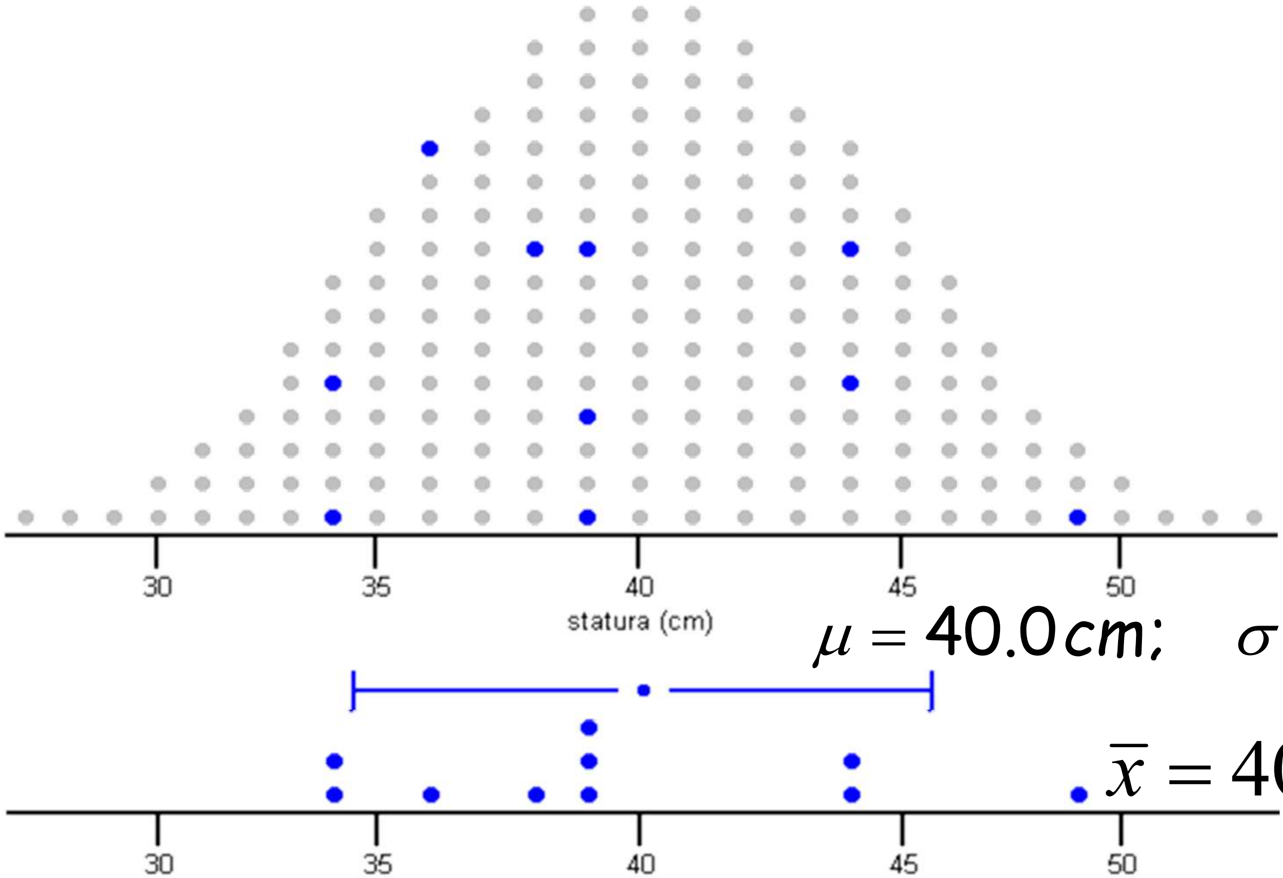
$\mu = 40.0\,cm; \quad \sigma = 5\,cm$



$\bar{x} = 36.7\,cm$

La popolazione di marziani: campionamento

terzo campione: n=10



$\mu = 40.0\,cm; \quad \sigma = 5\,cm$

$\bar{x} = 40.0\,cm$

La popolazione di marziani: 25 campioni da n=10

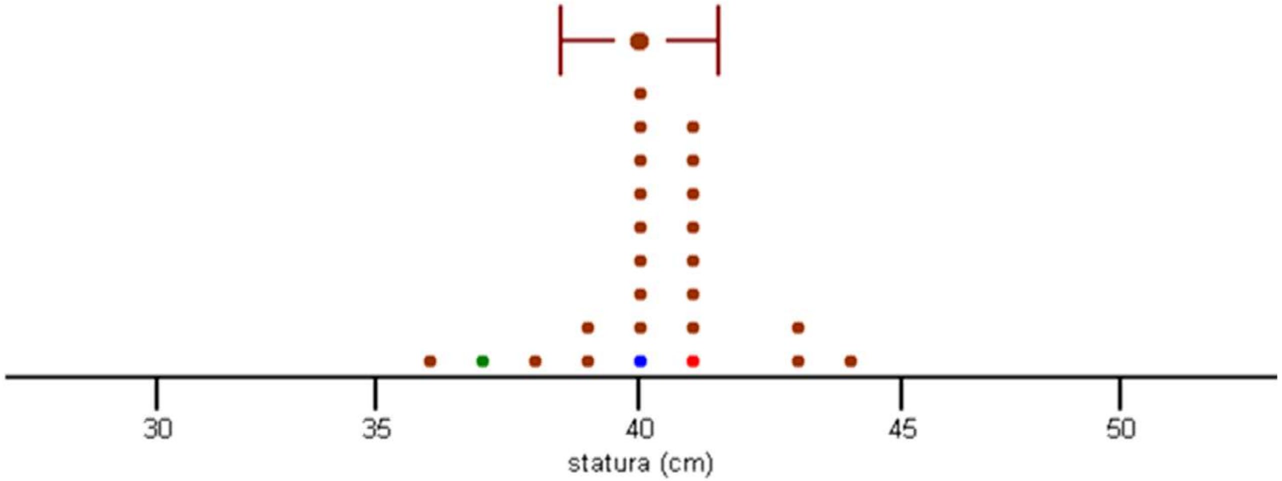
Estraiamo dalla popolazione altri 22 campioni da n=10 marziani, e poi tanti altri campioni ancora…

Ogni campione ha una sua **media campionaria** ed una sua **deviazione standard**

Ogni media campionaria è una stima la media della popolazione μ

Ogni deviazione standard campionaria è una stima della deviazione standard della popolazione σ

Distribuzione delle medie di 25 campioni

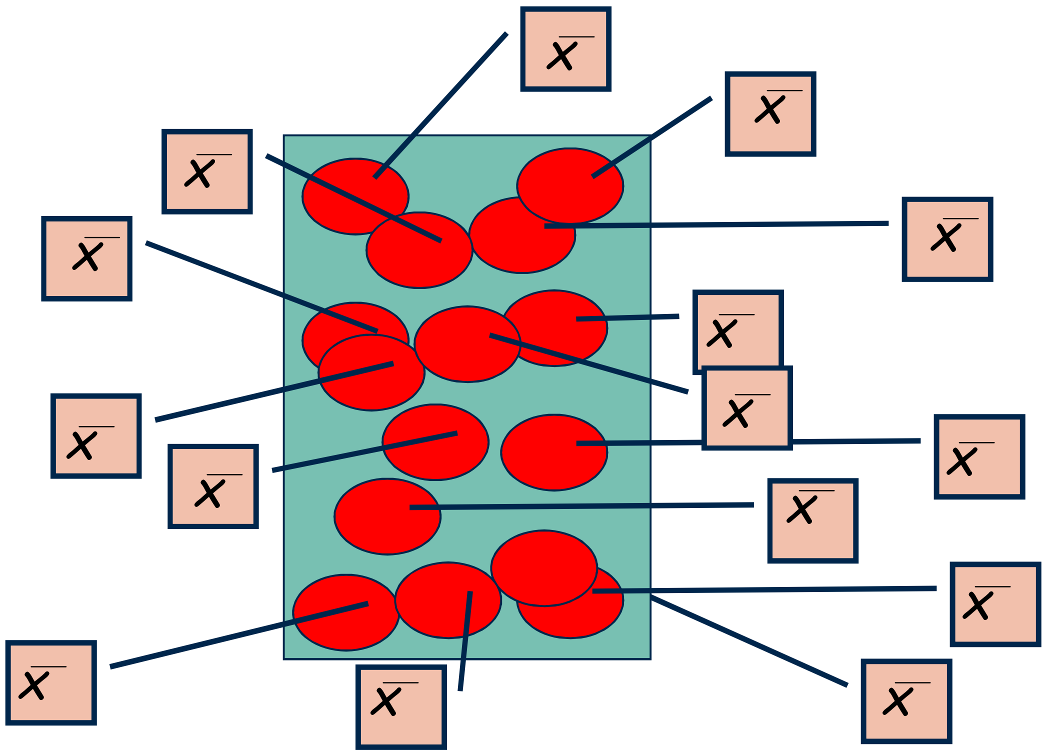


La media delle medie dei 25 campioni da n=10 marziani risulta $\overline{\overline{x}} = 40cm$

La deviazione standard di queste 25 medie è pari a $s_{\overline{x}} = 1,6cm$

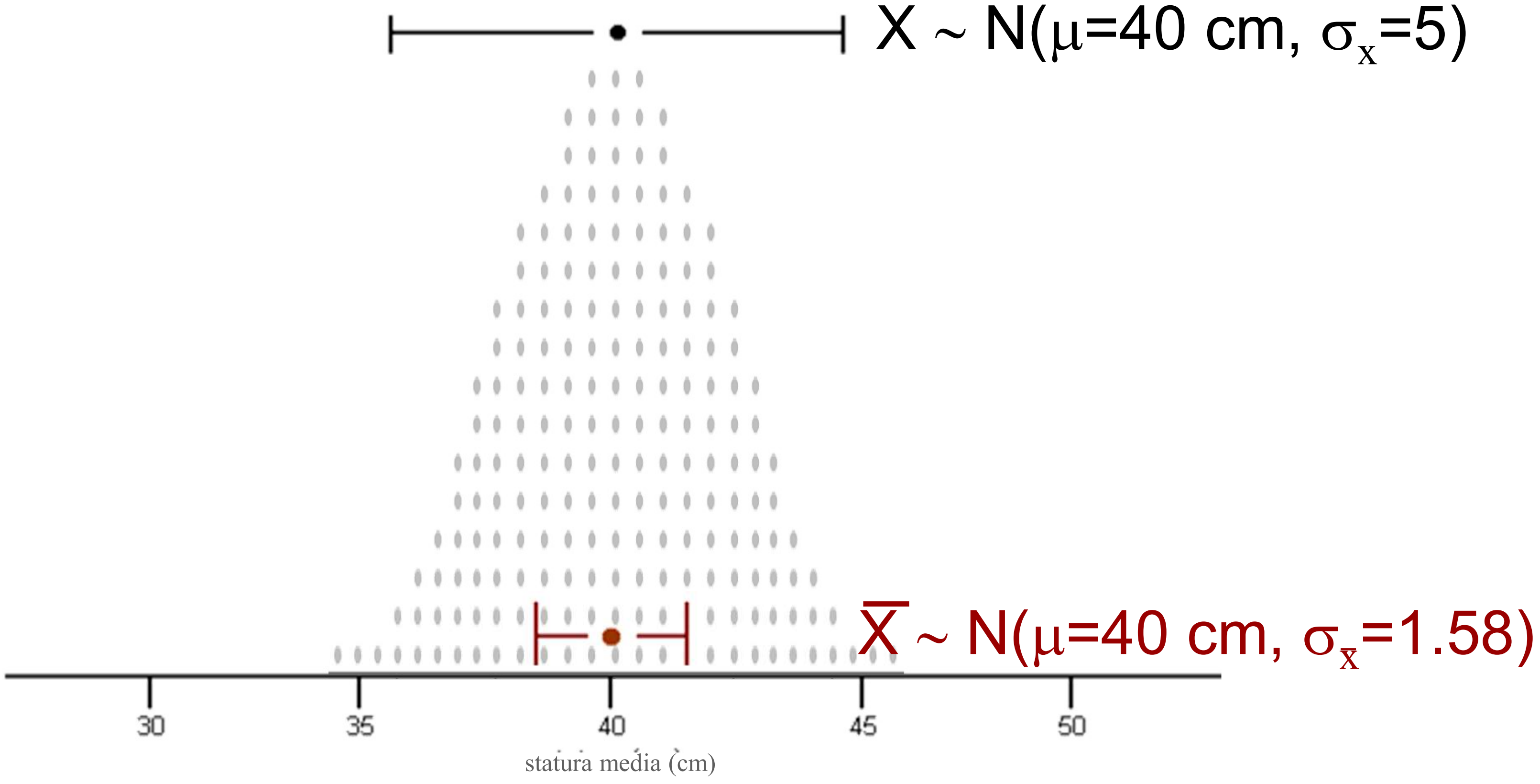
La popolazione di marziani: campionamento

Ripetiamo l’estrazione di un campione di numerosità 10 molte altre volte e calcoliamo la media campionaria delle altezze



Distribuzione di campionamento

Se si stimasse la media su tutti i possibili campioni di numerosità n=10 estraibili dalla popolazione, la distribuzione delle medie campionarie sarebbe...



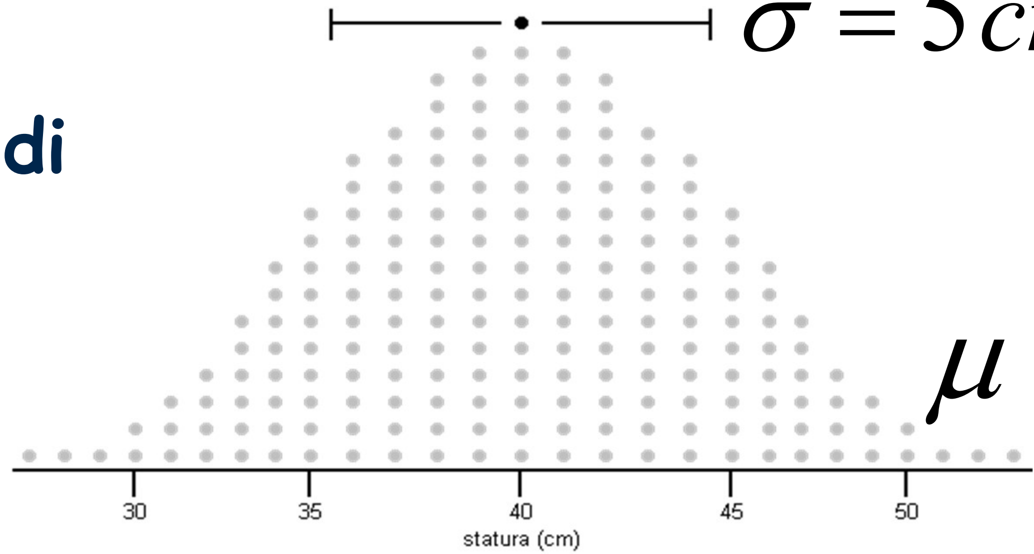
Distribuzione della media campionaria

La media delle medie coincide con la media di popolazione $\mu_{\bar{x}} = \mu = 40cm$

La deviazione standard delle medie ($\sigma_{\bar{x}}$) è **molto inferiore** quella di popolazione σ

$$\sigma_{\bar{x}} = \sigma / \sqrt{n} = 5\,cm / \sqrt{10} = 1{,}6\,cm$$

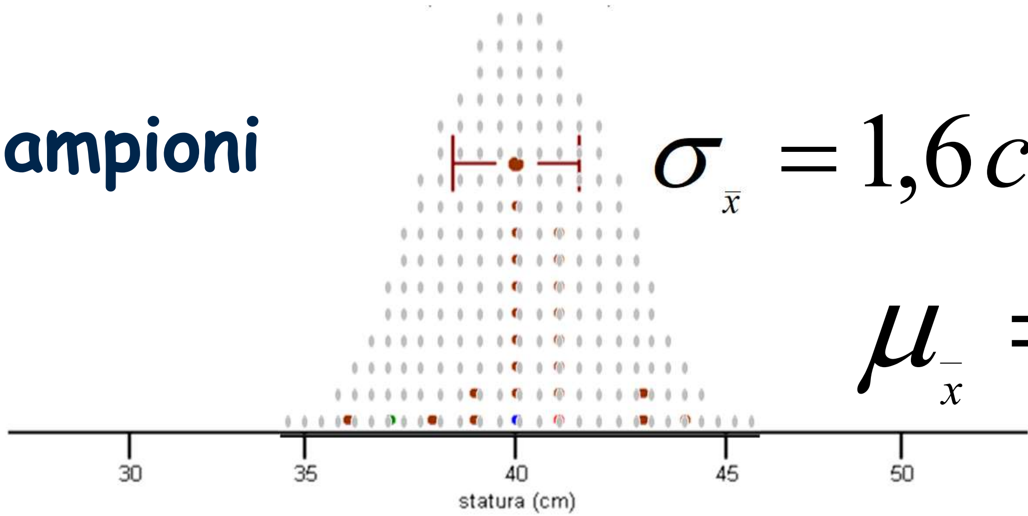
Popolazione di marziani



$$\sigma = 5\,cm$$

$$\mu = 40cm$$

Medie dei campioni (n=10)

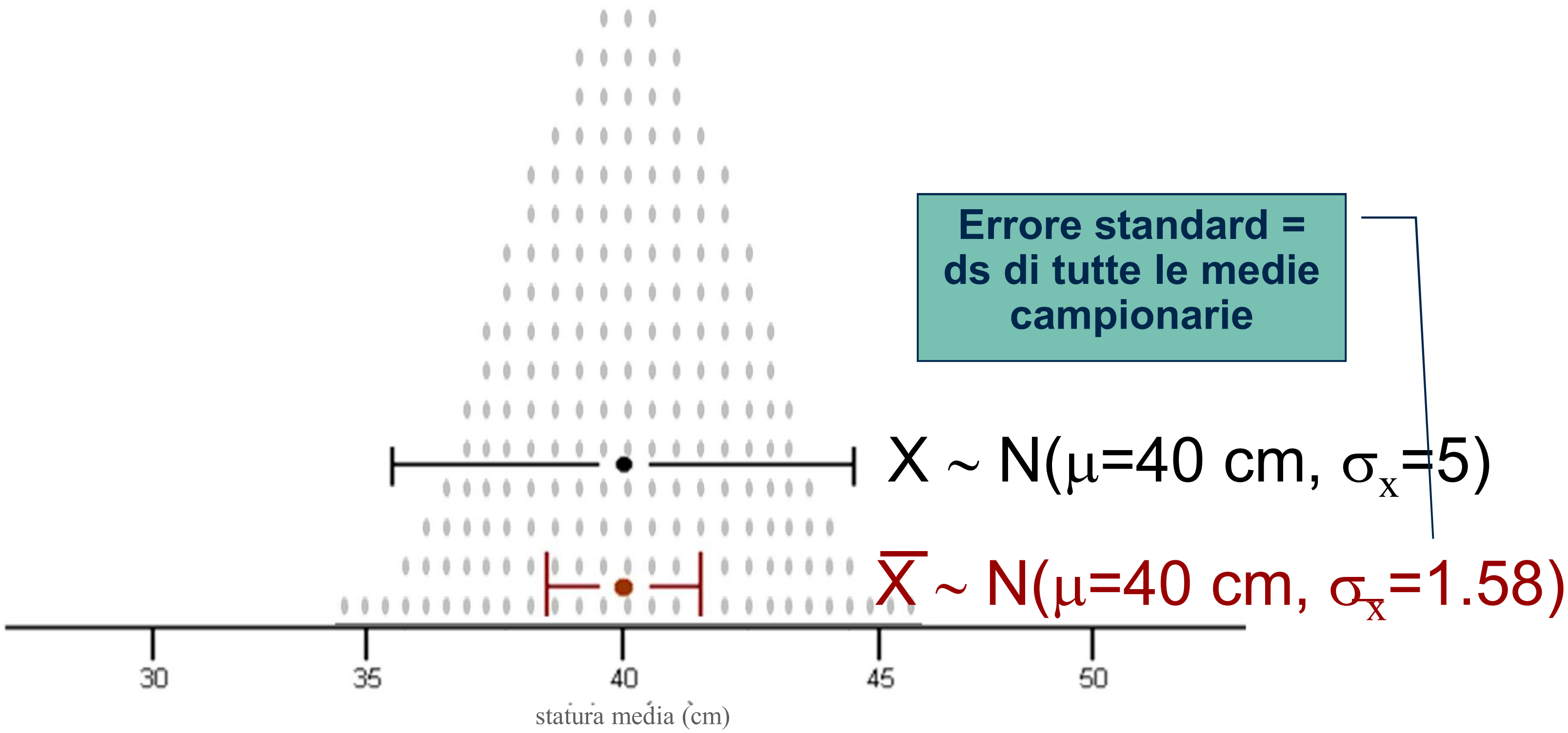


$$\sigma_{\bar{x}} = 1{,}6\,cm$$

$$\mu_{\bar{x}} = 40cm$$

Distribuzione di campionamento

- Lo stimatore della media campionaria ($\overline{X}$) segue una distribuzione:
  - Normale
  - centrata sulla media vera (μ=40)
  - con una deviazione standard pari a $\sigma/\sqrt{n}$ (1.58)



La distribuzione di campionamento

Nello specifico:
le medie calcolate sui diversi campioni sono le realizzazioni della v.c. media campionaria $\overline{X}$ (stimatore di μ).

In generale:
ogni statistica campionaria è una v.c. caratterizzata da una specifica distribuzione di probabilità: distribuzione campionaria dello stimatore

Processo di stima

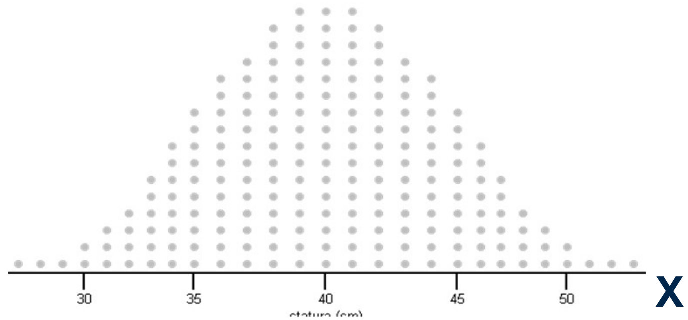
Non è possibile valutare la bontà della stima ottenuta da un singolo campione.

Si deve fare riferimento ad una situazione teorica in cui si considerano le stime ottenute da tutti i possibili campioni.

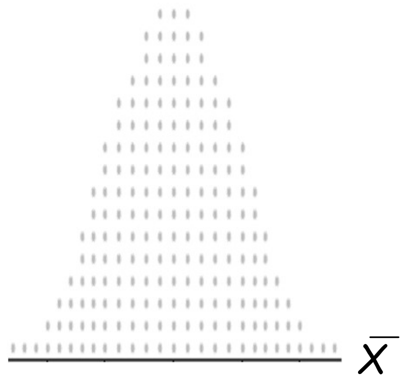
L'inferenza si basa dunque sulla conoscenza delle caratteristiche teoriche della distribuzione di campionamento di uno stimatore.

Deviazione standard vs Errore standard

La **deviazione standard** è un indice di **variabilità del fenomeno**. Fornisce informazioni su come si distribuiscono i dati intorno alla media



L'errore standard è un indice di **variabilità degli stimatori**. Fornisce informazioni su come si distribuiscono le stime (ad esempio della media) intorno al valore vero.



Errore standard

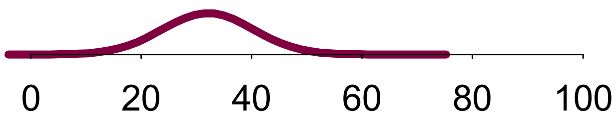
L'errore standard quantifica il grado di incertezza dello stimatore, ovvero la sua precisione nello stimare il parametro.

In generale, il suo valore dipende:

- dalle caratteristiche della variabile misurata sulla popolazione, in particolare dal grado di variabilità (σ);
- dalla dimensione del campione (n);
- dalla strategia di campionamento.

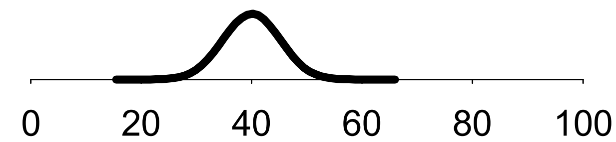
Errore standard – al variare di σ^2

a parità di N(=200) ed n(=10)



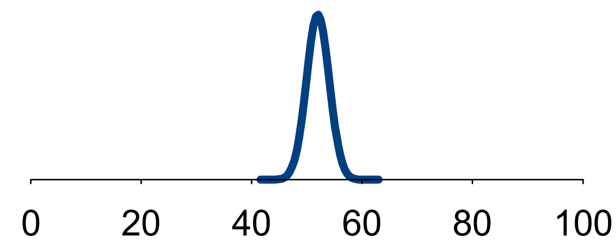
Altezza dei venusiani (2-50 cm)

$X_V \sim N(\mu=32\text{ cm}, \sigma=9\text{ cm})$



Altezza dei marziani (25-55 cm)

$X_M \sim N(\mu=40\text{ cm}, \sigma=5\text{ cm})$

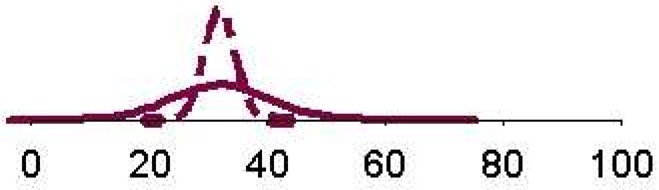


Altezza dei saturniani (42-60 cm)

$X_S \sim N(\mu=52\text{ cm}, \sigma=3\text{ cm})$

Errore standard – al variare di σ^2

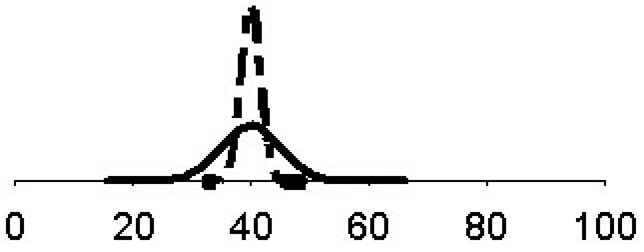
a parità di dimensione del campione n = 10 (N=200)



Altezza dei venusiani (2-50 cm)

$X_V \sim N(\mu=32\text{ cm}, \sigma=9\text{ cm})$

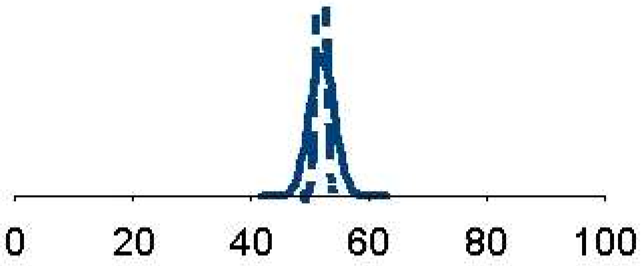
$\overline{X}_V \sim N(\mu=32\text{ cm}, \sigma_{\overline{x}}=9/\sqrt{10}\text{ cm})$



Altezza dei marziani (25-55 cm)

$X_M \sim N(\mu=40\text{ cm}, \sigma=5\text{ cm})$

$\overline{X}_M \sim N(\mu=40\text{ cm}, \sigma_{\overline{x}}=5/\sqrt{10}\text{ cm})$



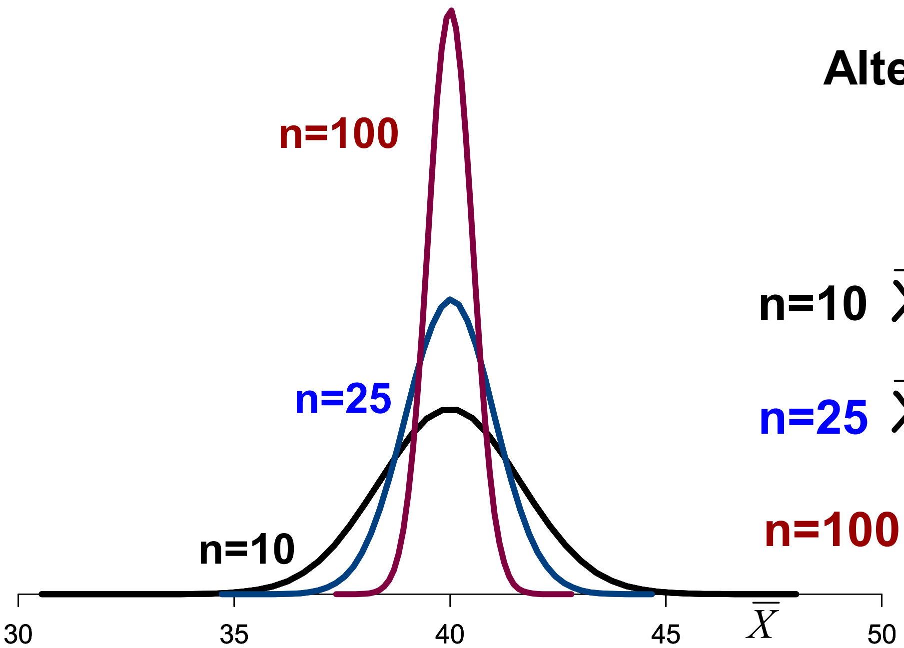
Altezza dei saturnini (42-60 cm)

$X_S \sim N(\mu=52\text{ cm}, \sigma=3\text{ cm})$

$\overline{X}_S \sim N(\mu=52\text{ cm}, \sigma_{\overline{x}}=3/\sqrt{10}\text{ cm})$

Errore standard – al variare di n

a parità di σ^2



Altezza dei marziani

$X_M \sim N(\mu=40\text{ cm}, \sigma=5\text{ cm})$

n=10 $\bar{X}_M \sim N(\mu=40\text{ cm}, \sigma_{\bar{x}}=5/\sqrt{10}\text{ cm})$

n=25 $\bar{X}_M \sim N(\mu=40\text{ cm}, \sigma_{\bar{x}}=5/\sqrt{25}\text{ cm})$

n=100 $\bar{X}_M \sim N(\mu=40\text{ cm}, \sigma_{\bar{x}}=5/\sqrt{100}\text{ cm})$

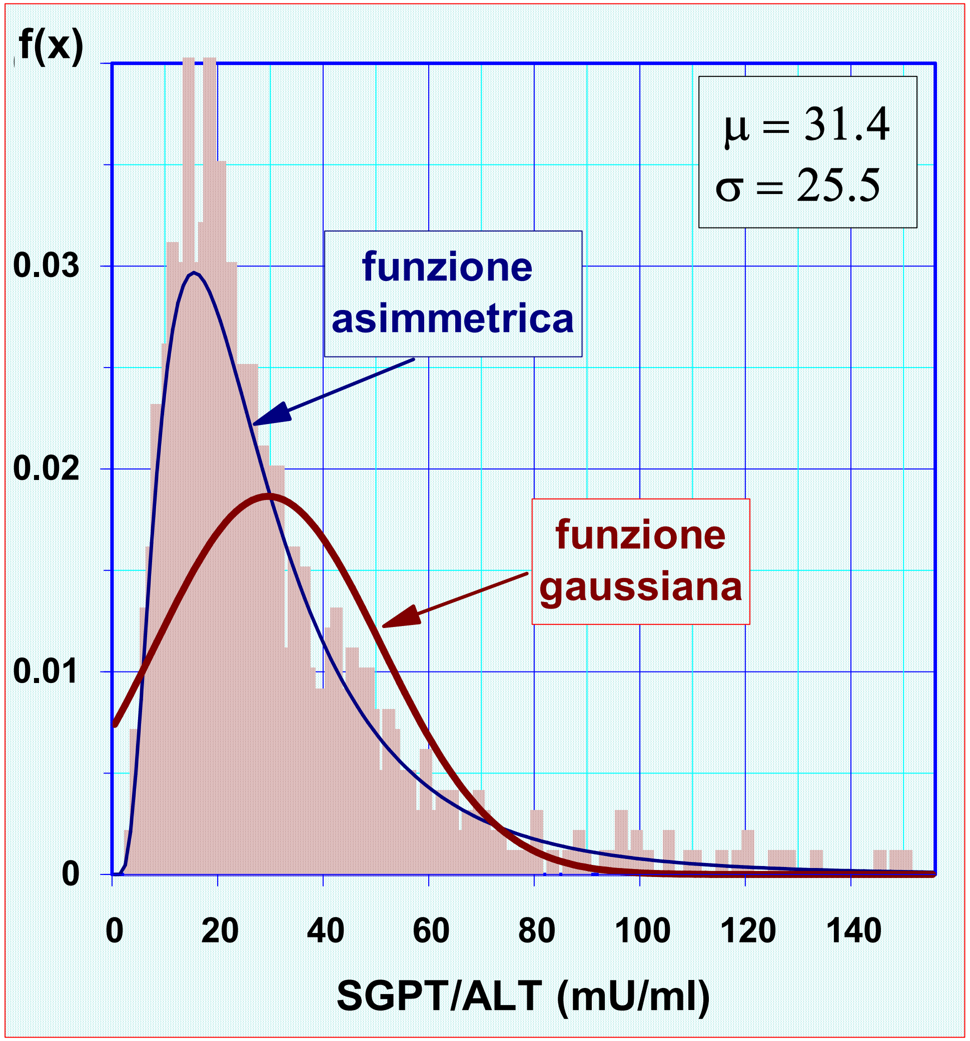
Distribuzioni campionarie
Parte II

Ma che succede sul pianeta Terra?

- E nel mondo reale ?
- Che succede se la variabile casuale che studiamo, come spesso accade, non ha una distribuzione gaussiana ?

Il teorema del limite centrale: Distribuzione dei livelli ematici di ALT

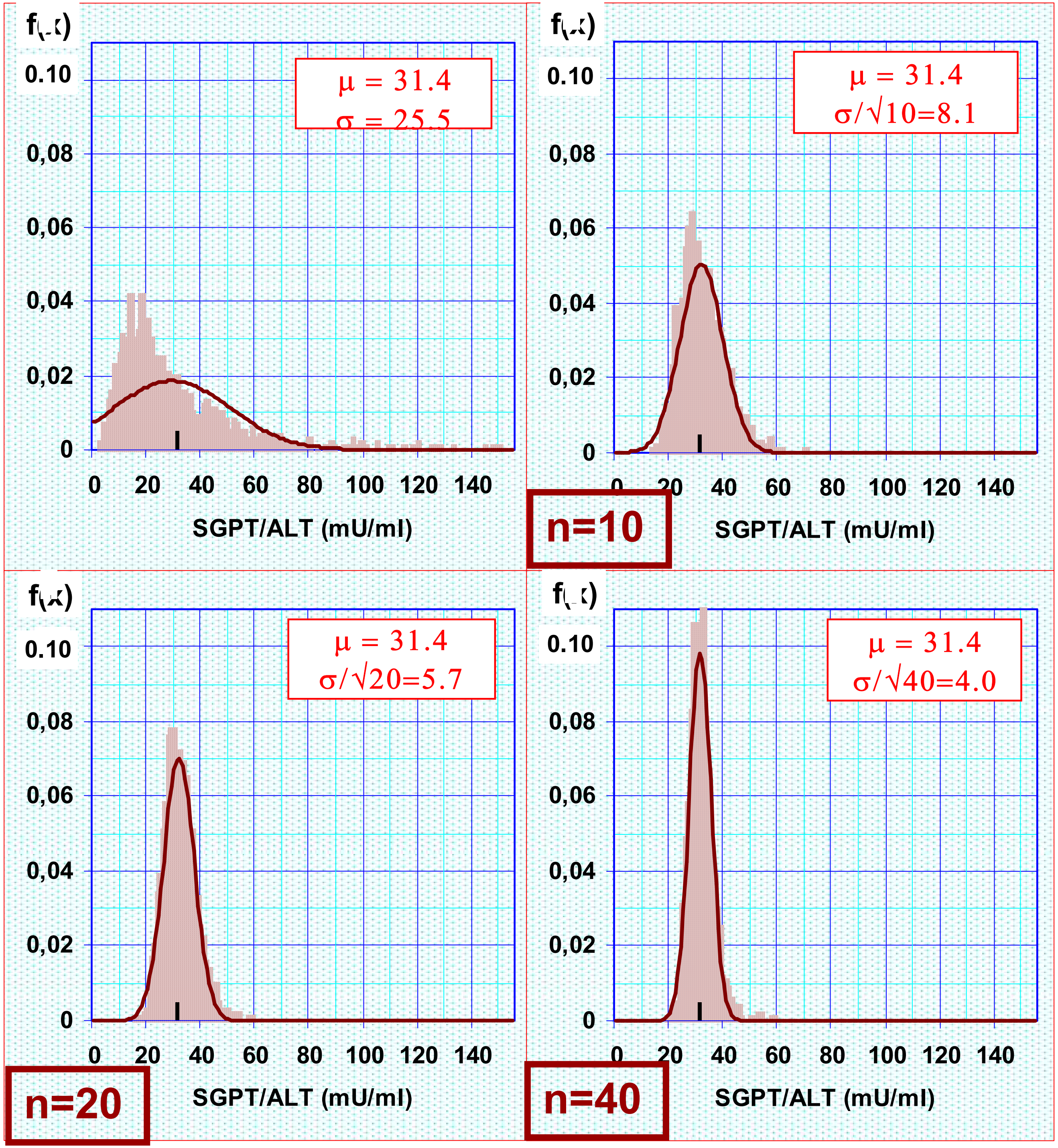
Nella popolazione dei maschi adulti, la distribuzione di alanina amino-transferasi (ALT) è fortemente asimmetrica per la presenza di individui con danni epatici causati da alcol, farmaci, infezioni virali…



Questa distribuzione è basata sullo studio del livello di ALT in 1000 soggetti maschi adulti

Al crescere di n, la distribuzione della media campionaria:

1. *riduce la sua dispersione;*
2. *tende ad una Normale*



Teorema centrale del limite

- Ciò che si è visto con questo esempio di campionamento, relativo ai livelli ematici di ALT, è formalizzato da un teorema detto **« teorema centrale del limite »**
- Dato un campione di dimensione n, tratto da una **variabile casuale qualunque** X con media μ e deviazione standard σ
- **la variabile casuale media campionaria** $\bar{x}$, al crescere di n, ha un **distribuzione gaussiana** con media μ e deviazione standard $\sigma/\sqrt{N}$ (errore standard)

Esercizio

Si supponga che il peso misurato su una certa popolazione di soggetti ha media 65 e deviazione standard 20. Si calcoli:

1) Il peso medio atteso e il suo errore standard in un campione di 25 individui

$$E(\overline{X})=65 \qquad es(\overline{X})=20/\sqrt{25}=10/5=4$$

1) Il peso medio atteso e il suo errore standard in un campione di 100 individui

$$E(\overline{X})=65 \qquad es(\overline{X})=20/\sqrt{100}=20/10=2$$

1) Se si estraesse un campione di 25 individui e il loro peso medio fosse di 73 kg, quale sarebbe la probabilità di osservare un valore superiore a quello effettivamente osservato?

$$Pr(\overline{X}>73)=Pr(Z>2)=0.023$$

La popolazione di marziani: la proporzione di blu

- Sino a questo momento…il colore verde dei marziani lo abbiamo dato per scontato…
- Immaginiamo che su Marte ci siano anche una minoranza di marziani di colore blu
- Essi sono 24 su 200, π=24/200=0,12
- π è la probabilità che incontrando per caso un marziano questo sia blu….

La popolazione di marziani: la proporzione di blu

- La proporzione π è la media della caratteristica *'essere blu'* definita come

  x_n = 1 per i marziani blu

  x_n = 0 per i marziani verdi

- La deviazione standard di *'essere blu'* è

$$\sigma = \sqrt{\pi \cdot (1 - \pi)} = \sqrt{0{,}12 \cdot (1 - 0{,}12)} = 0{,}32$$

σ e' minima quando π=0, π=1 (non c'è variabilità, σ=0)

σ è massima quando π=0.5 (σ=0.5)

La popolazione di marziani: campionamento

- Ripetiamo il processo di stima della proporzione di marziani blu (π) già descritto su tutti i possibili campioni di numerosità n=10 estraibili dalla popolazione degli N=200 marziani
- Stimiamo π con la percentuale p di marziani blu nel campione (o come la media delle x).

Ad esempio:

primo campione: p=1/10=0.10

secondo campione: p=7/10=0.70

terzo campione: p=2/10=0.20

......

La popolazione di marziani: campionamento

- Qual è la distribuzione di campionamento della proporzione campionaria?
- Per valori di n sufficientemente elevati, la distribuzione è:
  - Normale
  - con media pari alla vera proporzione della popolazione π
  - e deviazione standard pari a

$$\sigma / \sqrt{n} = \sqrt{\pi \cdot (1-\pi) / n} \rightarrow$$ ERRORE STANDARD

Frequenze relative campionarie

All'aumentare della dimensione (n) del campione, i valori della frequenza relativa (p) dell'evento mostrano tendenza a crescere ed accentrarsi attorno al parametro π, approssimando la distribuzione gaussiana.