



Last updated Marzo 6, 2024

Scale di Misurazione - Indici Statistici Lezione 2

G. Bacaro

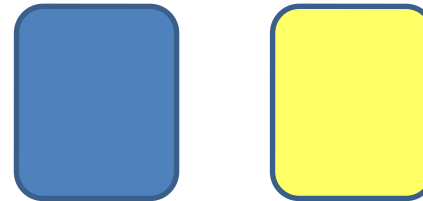
Statistica
CdL in Scienze e Tecnologie per l'Ambiente e la Natura
Il anno, Il semestre

Tipi di Variabili

1

Scale di Misurazione

1. Variabile categoriche (qualitative)



2. Variabile quantitativa continua

3. Variabile quantitativa discreta (conteggi o proporzioni)

Le misure possono essere raggruppate in 4 tipi di scale, che godono di proprietà formali differenti; di conseguenza, esse ammettono operazioni differenti.

- 1) **nominale o classificatoria;**
- 2) **ordinale o per ranghi;**
- 3) **ad intervalli;**
- 4) **di rapporti.**

Nominale o Classificatoria

La **scala nominale** o **classificatoria** è il livello più basso di misurazione. E' utilizzata quando i risultati possono essere classificati o raggruppati in categorie qualitative, dette anche nominali ed eventualmente identificate con simboli. I caratteri nominali, detti anche "sconnessi", costituiscono variabili le cui modalità o attributi non assumono alcun ordine precostituito.

Nella **scala nominale o qualitativa**, esiste una sola relazione, quella di identità: gli individui attribuiti a classi diverse sono tra loro differenti, mentre tutti quelli della stessa classe sono tra loro equivalenti, rispetto alla proprietà utilizzata nella classificazione.

Ordinale o per Ranghi

La scala ordinale o per ranghi rappresenta una misurazione che contiene una quantità di informazione immediatamente superiore a quella nominale; essa assume modalità logicamente sequenziali, non importa se in ordine crescente o decrescente.

Questa misura ha un limite fondamentale. In una scala ordinale, **non è possibile quantificare le differenze di intensità tra le osservazioni.**

Con la scala per ranghi, le differenti classi possono essere ordinate sulla base dell'intensità del fenomeno. (es.: Si supponga che il risultato di un reagente sia di colorare in verde una serie di provette, secondo la quantità di sostanza contenuta. E' possibile mettere in ordine le provette secondo l'intensità del colore, per avere una stima approssimata della quantità di sostanza contenuta. Se si confrontano tre o più provette con intensità di colore differente, è facile stabilirne l'ordine; rimane impossibile misurare ad occhio la quantità di colore di ognuna e la differenza esistente tra esse).

Scala per Intervalli

6

Scale di Misurazione

Intervalli

La scala ad intervalli **aggiunge la proprietà di misurare le distanze o differenze tra tutte le coppie di valori**. La scala di intervalli si fonda su una misura oggettiva e costante, anche se il punto di origine e l'unità di misura sono arbitrari

In una scala ad intervalli, ***solo le differenze tra i valori sono quantità continue e isomorfe alla struttura dell'aritmetica***. Solo per le differenze sono permesse tutte le operazioni: possono essere tra loro sommate, elevate a potenza oppure divise, determinando le quantità che stanno alla base della statistica parametrica.

Esempi classici di scale ad intervalli sono la temperatura (misurata in gradi Celsius o Fahrenheit, ma non Kelvin) ed il tempo (misurato secondo calendari differenti).

6



Scala di Rapporti

7

Scale di Misurazione

Scala di Rapporti

La scala di rapporti ha il vantaggio di avere **un'origine reale**. Sono tipiche scale di rapporti l'altezza, la distanza, la velocità, l'età, il peso, il reddito; più in generale, tutte quelle misure in cui **0 (zero) significa quantità nulla**

Non solo le differenze, ma gli stessi valori possono essere moltiplicati o divisi per quantità costanti. Alle variabili misurate con una scala di rapporti, il tipo di misurazione più sofisticato e completo, ***può essere applicato qualsiasi test statistico.***

Pure con una scala di rapporti **è possibile scendere nella scala di misurazione,** trasformandola in una scala di rango o addirittura qualitativa.

7



Proprietà delle Scale

- nella scala nominale, esistono solo relazioni di equivalenza;

La scala nominale gode delle seguenti *proprietà*:

- *riflessiva*: $A = A$
- *simmetrica*: se $A = B \Rightarrow B = A$
- *transitiva*: se $A = B$ e $B = C \Rightarrow A = C$.

- in quella ordinale, alla precedente si aggiungono relazioni di minore o maggiore di;

Considerati due qualsiasi elementi, A e B , appartenenti a classi diverse, la scala ordinale consente di stabilire se A precede B ($A < B$), o se B precede A ($B < A$); oppure, invertendo la direzione dell'ordinamento se A segue B ($A > B$) o B segue A ($B > A$). L'insieme delle categorie ordinate costituisce una *graduatoria* e il numero d'ordine assegnato all'unità statistica in seno alla graduatoria è chiamato *rango*.

Le scale ordinali, oltre che delle proprietà delle scale nominali, godono della *proprietà della tricotomia*, secondo cui:

$$\forall A, B \in E: \text{sempre } A = B \vee A < B \vee B < A$$

dove E è l'insieme sul quale è stata definita la *relazione d'ordine* $<$.

Proprietà delle Scale

- in quella ad intervalli alle due precedenti si aggiunge la relazione di rapporto tra ogni coppia d'intervalli;

Equivalenza

1. se $x = y$, non è vero che $x < y$ e non è vero che $y < x$;
2. se $x = y$ e $x < z \Rightarrow y < z$;
3. se $x = y$ se non è vero che $x < z$, non è vero neppure che $y < z$.

Insieme di proprietà per definire la **proprietà additiva**:

4. se $x + y = z \Rightarrow y + x = z$ (*proprietà commutativa*);
5. se $x = x'$ e $0 < y \Rightarrow x' < x + y$: 0 precede y , cioè y è più di zero, quindi, aggiungendo y ad una delle due quantità equivalenti, le due quantità non saranno più equivalenti;
6. se $x = x'$ e $y = y' \Rightarrow x + y = x' + y'$ (*assioma degli uguali*);
7. $x + (y + z) = (x + y) + z$ (*proprietà associativa*).

Proprietà delle Scale

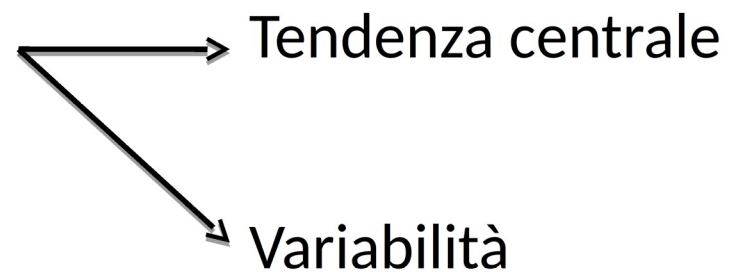
- nella scala di rapporti si ha anche la quarta relazione di rapporto conosciuto tra ogni coppia di valori.

Statistica descrittiva

Come descrivere una variabile in un insieme di osservazioni

1. Utilizzare rappresentazioni grafiche dei dati

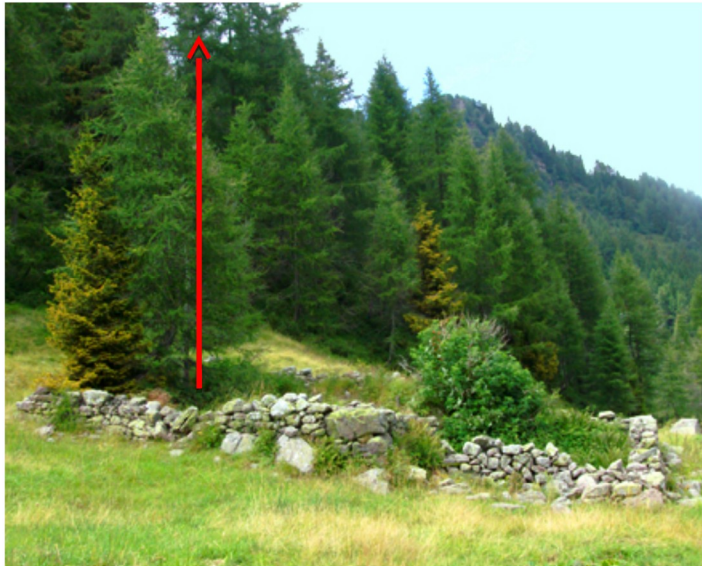
2. Sintetizzare i dati con degli indici



Rappresentazione grafica di una variabile quantitativa

Esempio: misura dell'altezza di 400 alberi

15.1, 17.2, 18, 21, 22, 25, 24, 15, 25.5, 16, 25, 24, 14, 26, 24.3,
27, 23, 22, 22.3, 25, 20, 18, 21.4, 24, 24, 28, 26, 22, 27.3, 21,
26, 18, 25, 20.4, 14, 25, 24, 19, 17,
15.4, 24, 27, 23.7, 19, 18, 15, 25.3, 18, 13.1, 24, 26, 22...



Rappresentazione grafica di una variabile quantitativa

Esempio: misura dell'altezza di 400 alberi

15.1, 17.2, 18, 21, 22, 25, 24, 15, 25.5, 16, 25, 24, 14, 26, 24.3,
27, 23, 22, 22.3, 25, 20, 18, 21.4, 24, 24, 28, 26, 22, 27.3, 21,
26, 18, 25, 20.4, 14, 25, 24, 19, 17,
15.4, 24, 27, 23.7, 19, 18, 15, 25.3, 18, 13.1, 24, 26, 22...

Posso creare delle classi di altezze:

13-14: ●●●

14-15: ●●

15-16: ●●●●●●

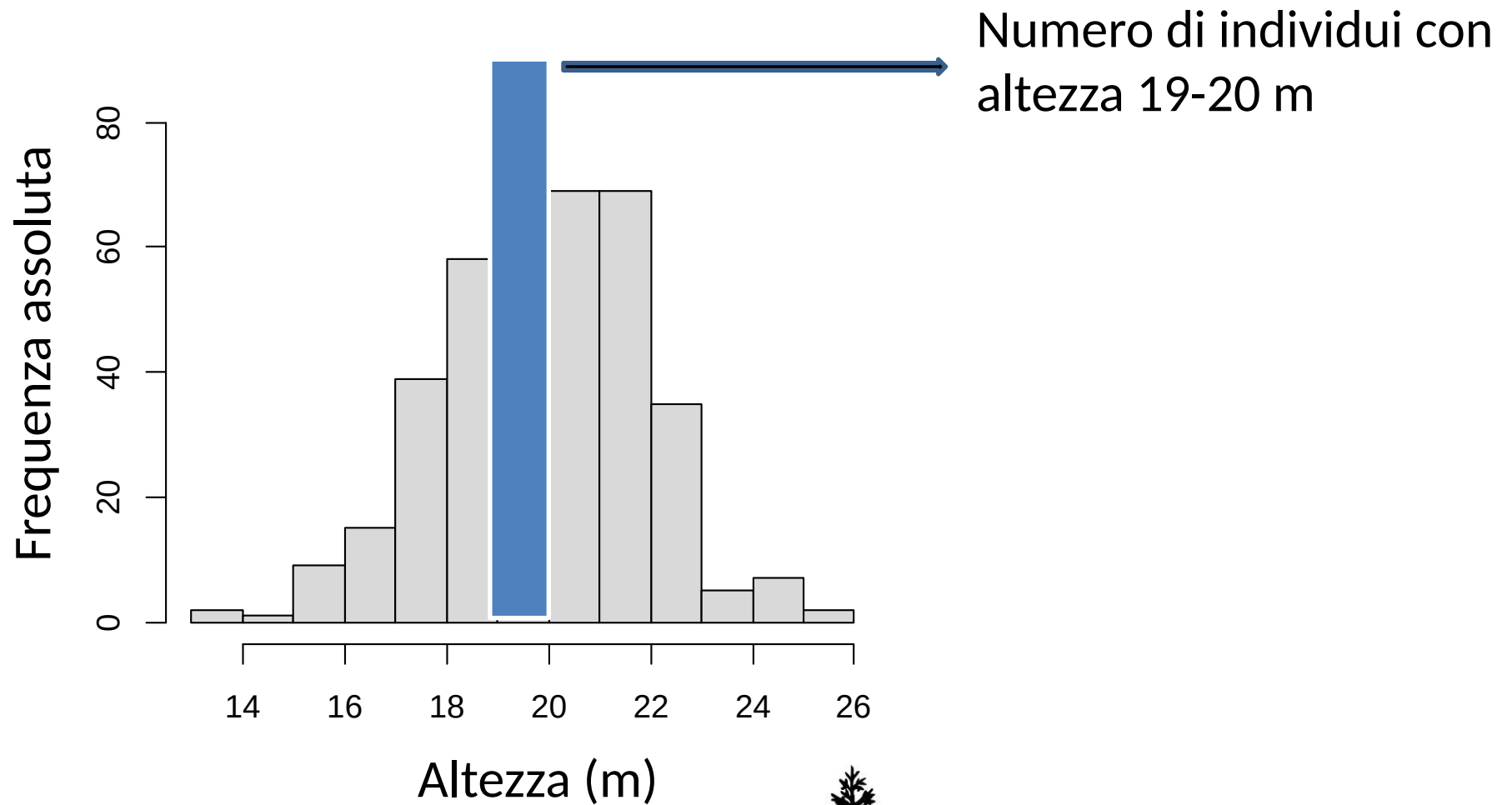
17-18: ●●●●●●●●●●●●●●●●

...

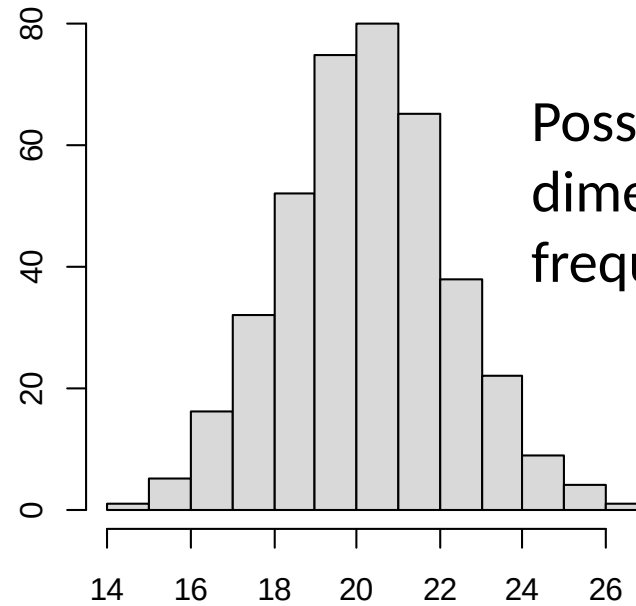
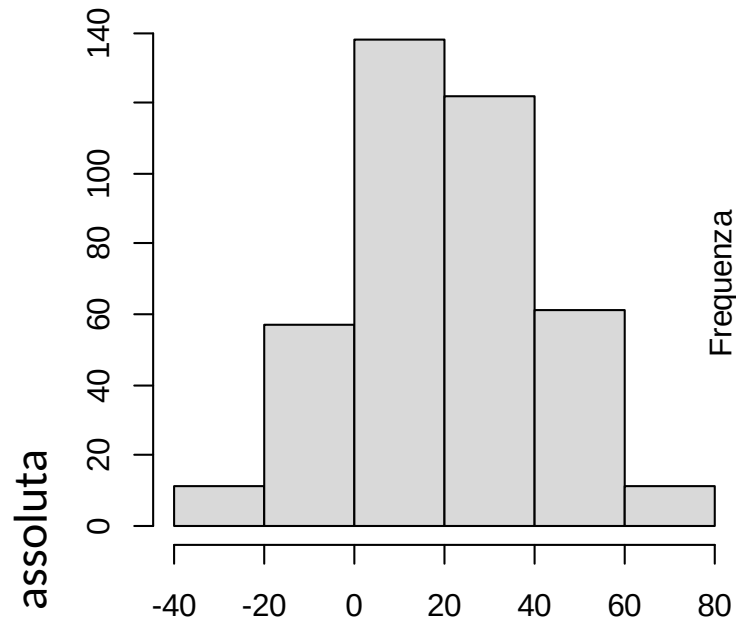


Frequenze
assolute

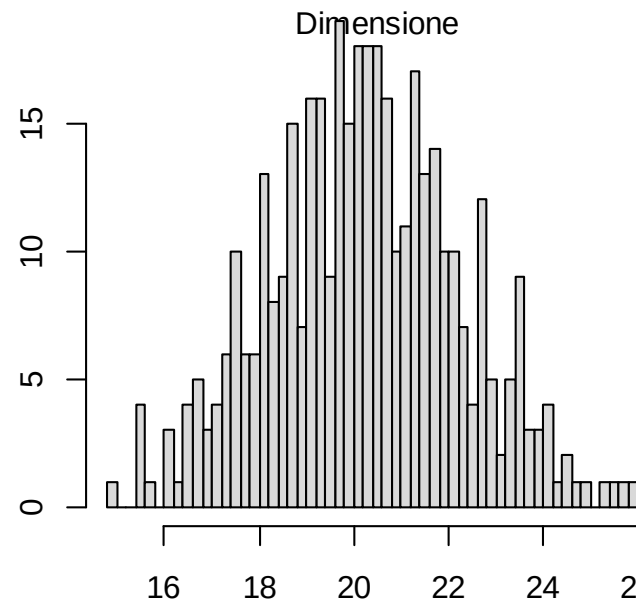
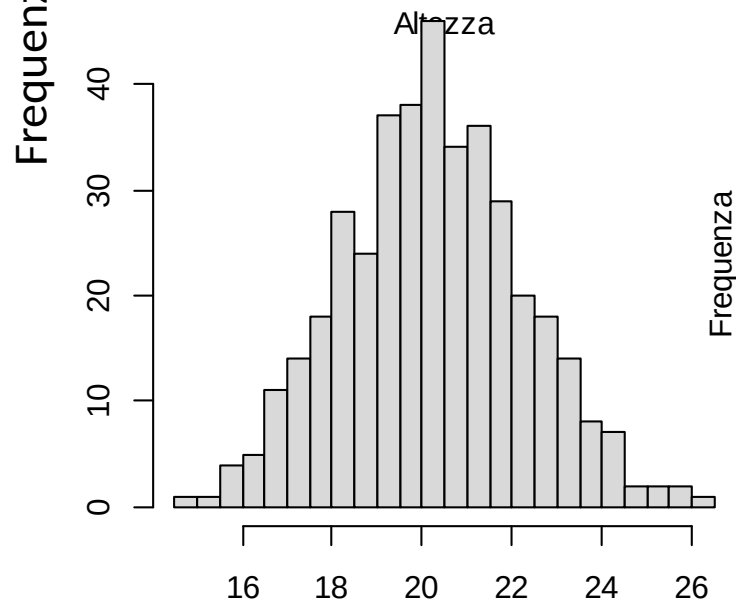
Istogramma delle frequenze assolute



Istogramma delle frequenze assolute

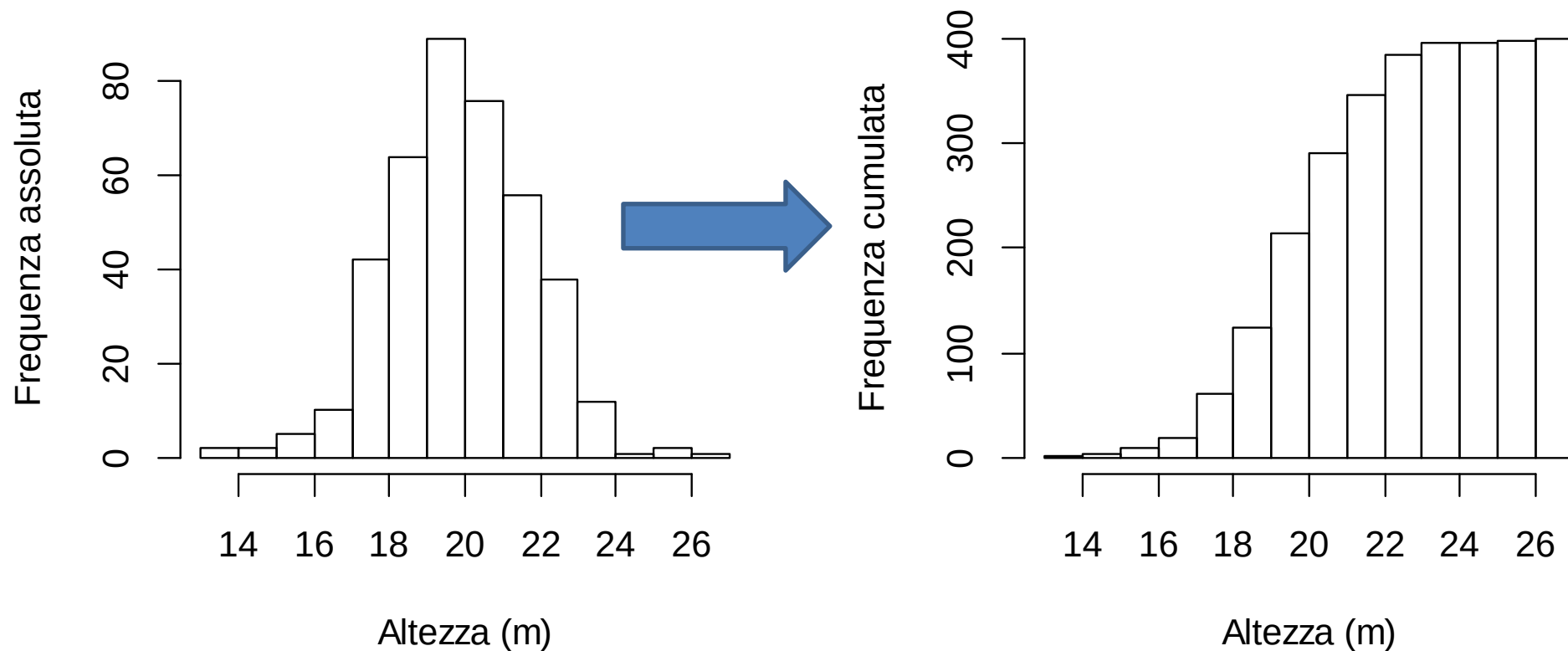


Posso variare la dimensione delle classi di frequenza



Altezza (m)

Istogramma delle frequenze assolute



Se sommo le frequenze assolute ottengo una frequenza cumulata

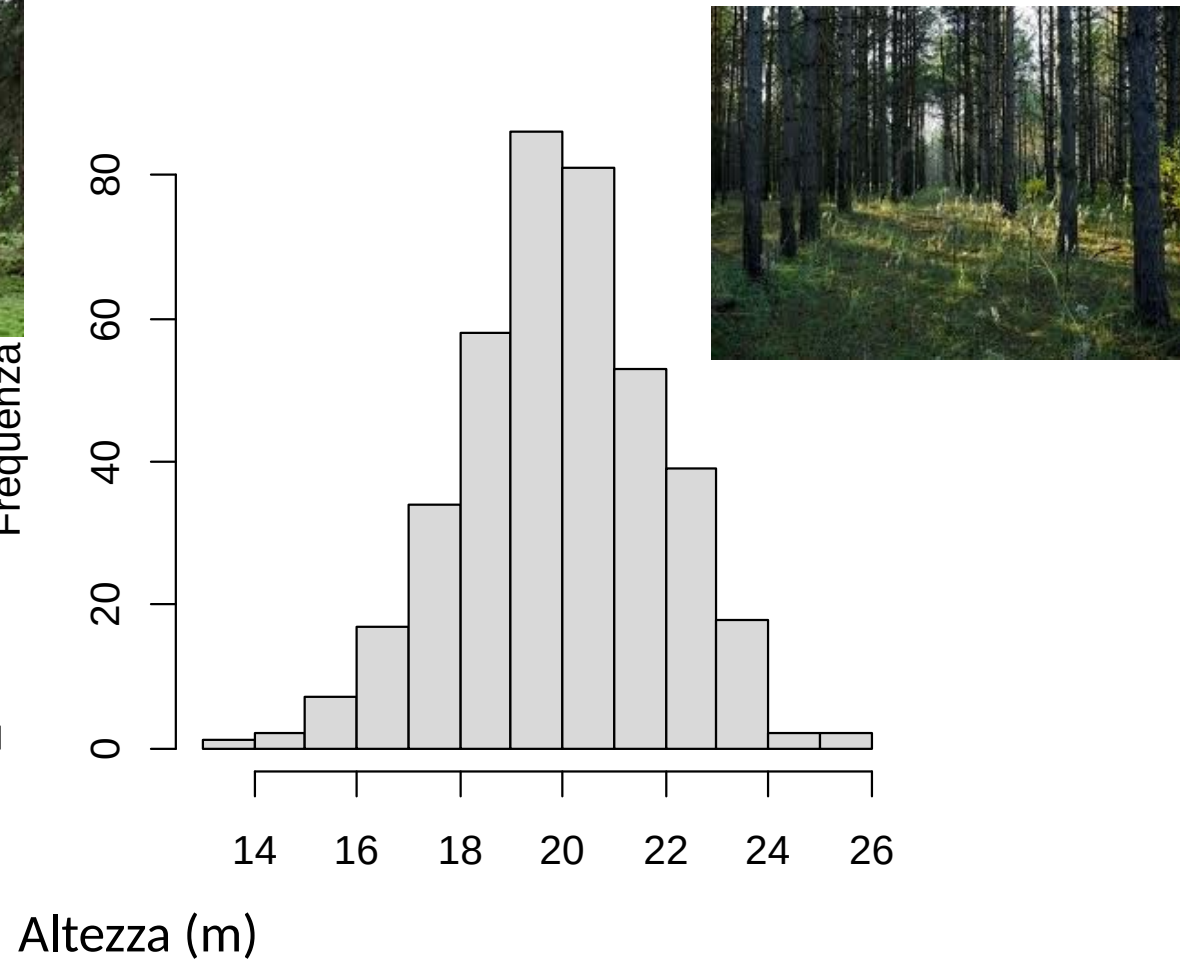
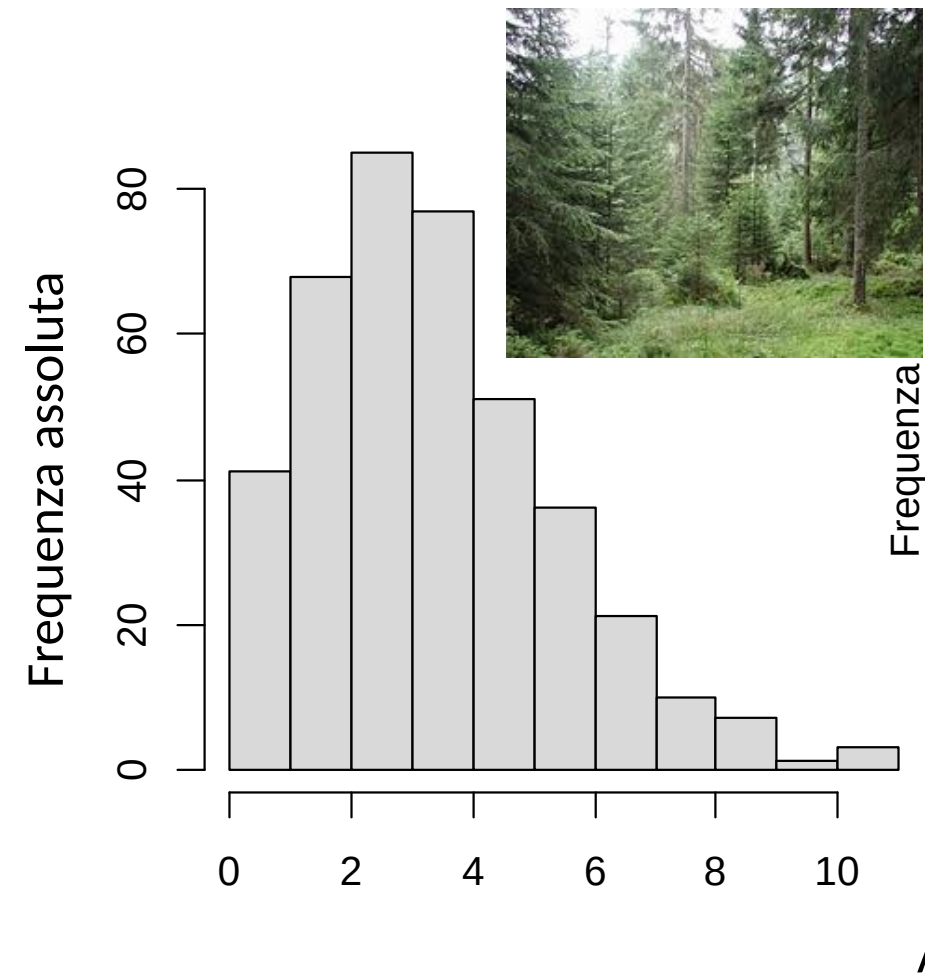
Rappresentazione grafica di una variabile quantitativa

Distribuzione di frequenza

La frequenza assoluta rappresenta il numero di osservazioni ottenute per un certo valore della variabile

La frequenza relativa (%) rappresenta il numero di osservazioni ottenute per un certo valore della variabile diviso per il numero totale di osservazioni

Istogramma delle frequenze



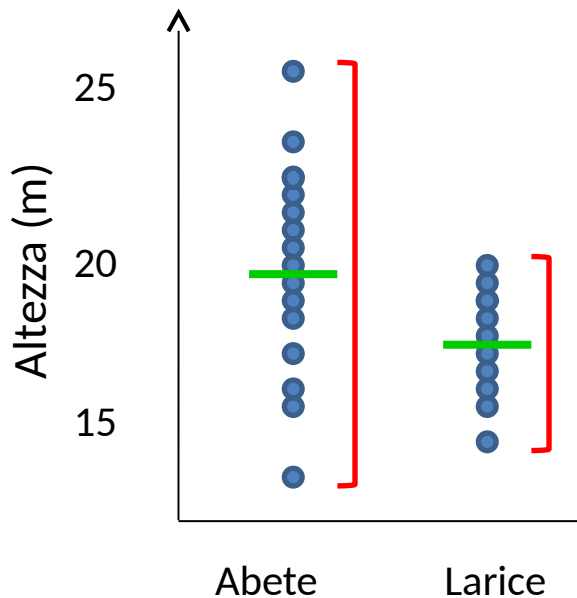
La forma dell'istogramma descrive la forma della distribuzione della variabile altezza albero

Indici di tendenza centrale e dispersione

Indici di tendenza centrale e dispersione

Indici di posizione: —
Descrivono il valore centrale di un gruppo di osservazioni

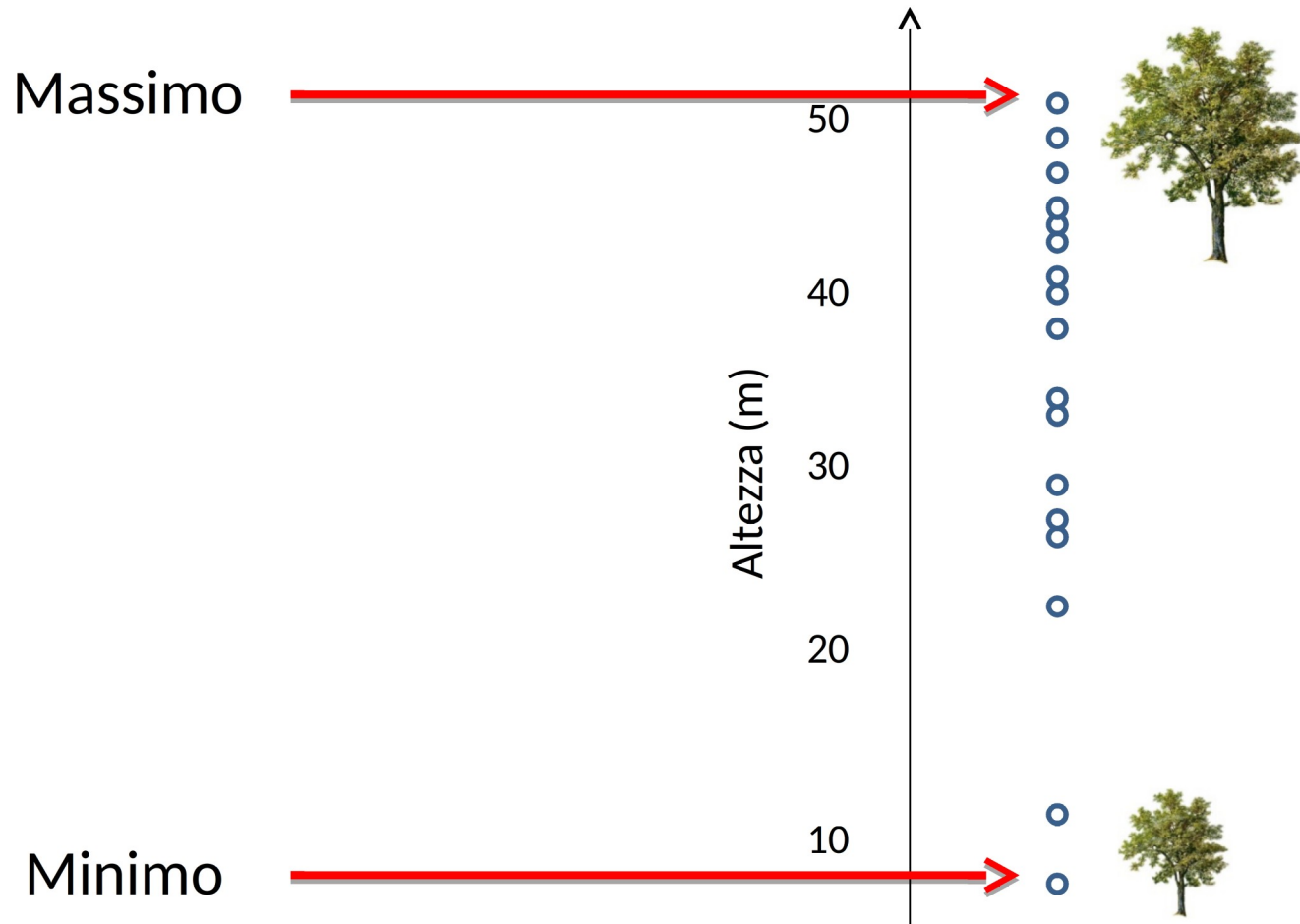
Indici di dispersione o variabilità:
quanto si discostano le singole osservazione?



Ogni popolazione presenta un certo grado di variabilità!!!

Indici di tendenza centrale: min e max

Intervallo dei valori



Indici di tendenza centrale: media

La media aritmetica

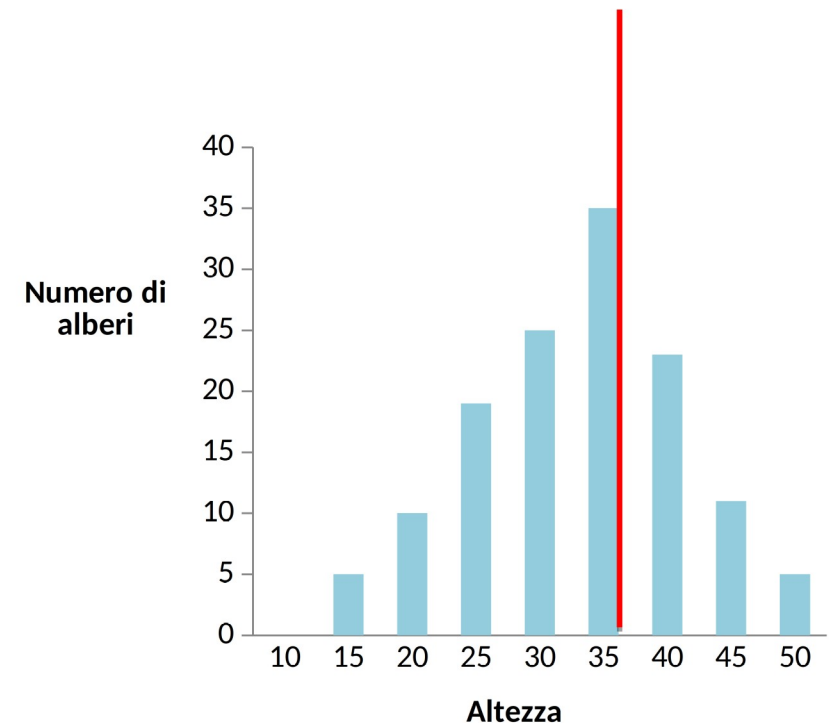
Intera popolazione

$$\mu = \frac{\sum x_i}{N}$$

Campione

$$\bar{x} = \frac{\sum x_i}{n}$$

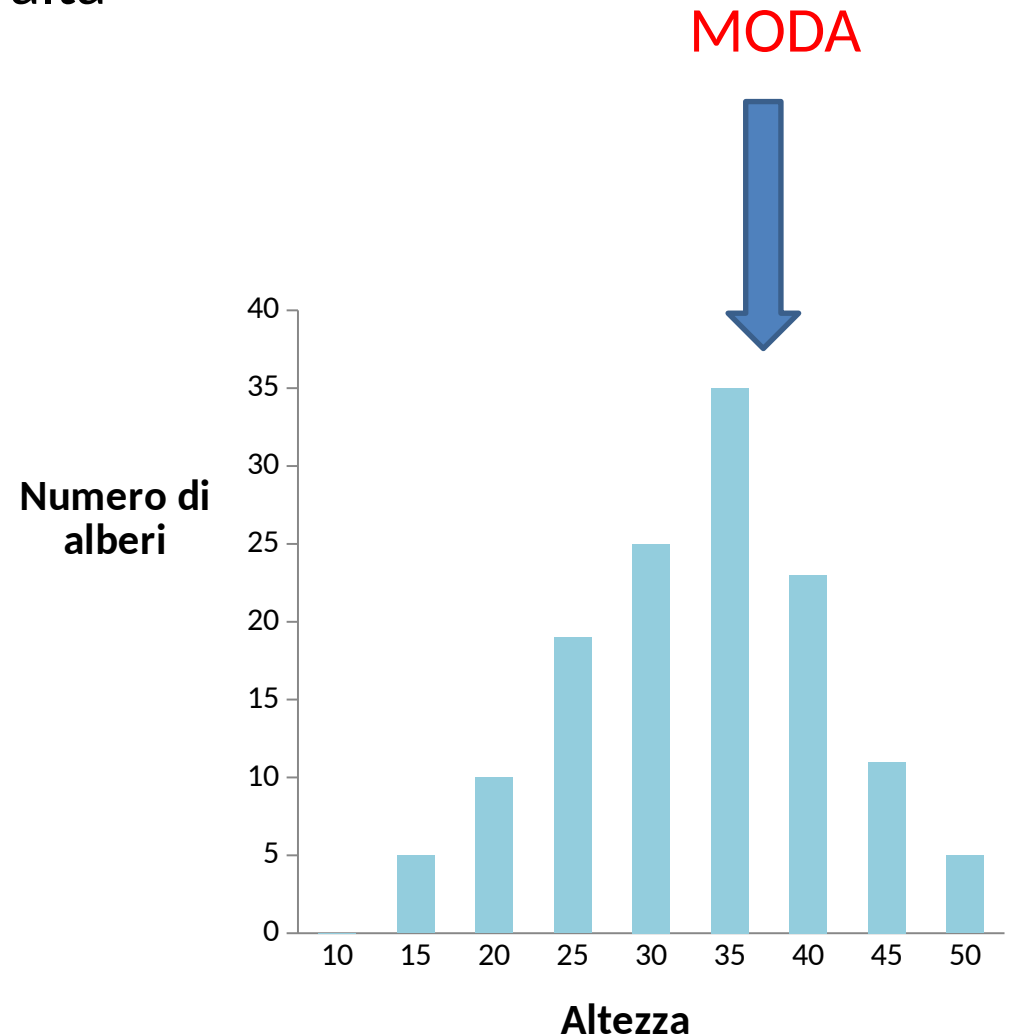
Uso della media: quando i valori si distribuiscono in modo più o meno simmetrico attorno ad un valore centrale



Indici di tendenza centrale: moda

La moda rappresenta il valore della variabile a cui è associata la frequenza più alta

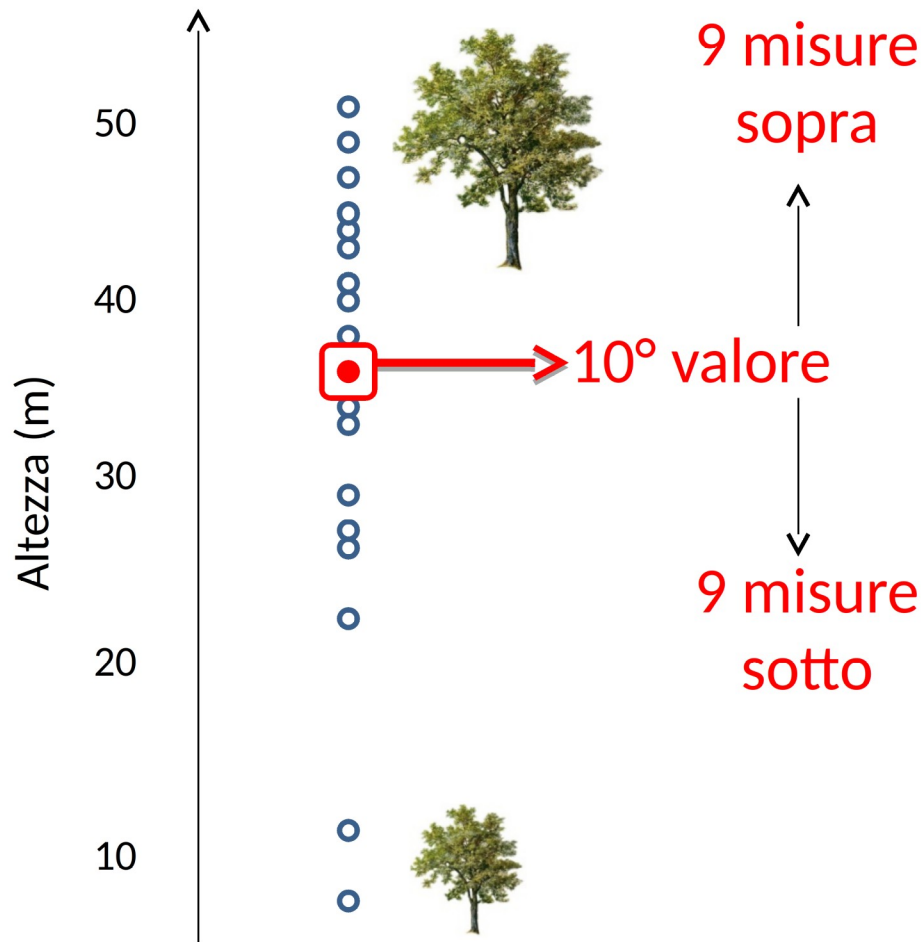
Nel caso in cui la classe non è discreta si prende il valore centrale della classe



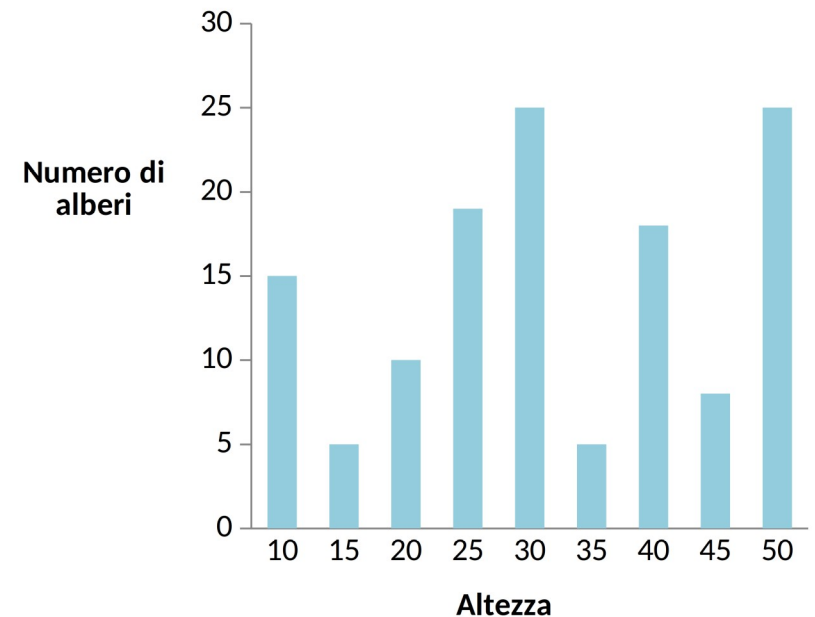
Indici di tendenza centrale: mediana

La mediana: la mediana costituisce il valore centrale di una serie di misure

Serie di 19 misure di altezza



Si usa quando la distribuzione dei valori non è simmetrica



Indici di tendenza centrale: mediana

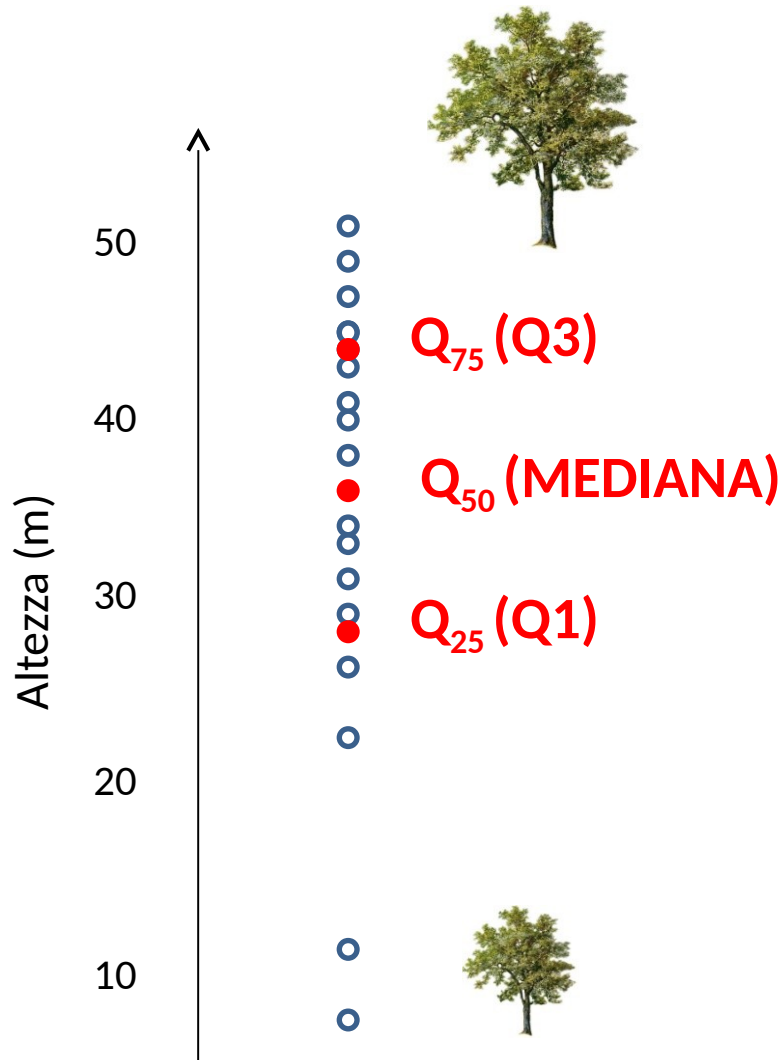
Se il numero di osservazione è pari la mediana è la media dei due valori centrali

Serie di 12 misure di altezza



Indici di tendenza centrale: quantili

Quantili: la mediana costituisce un caso specifico di quantile

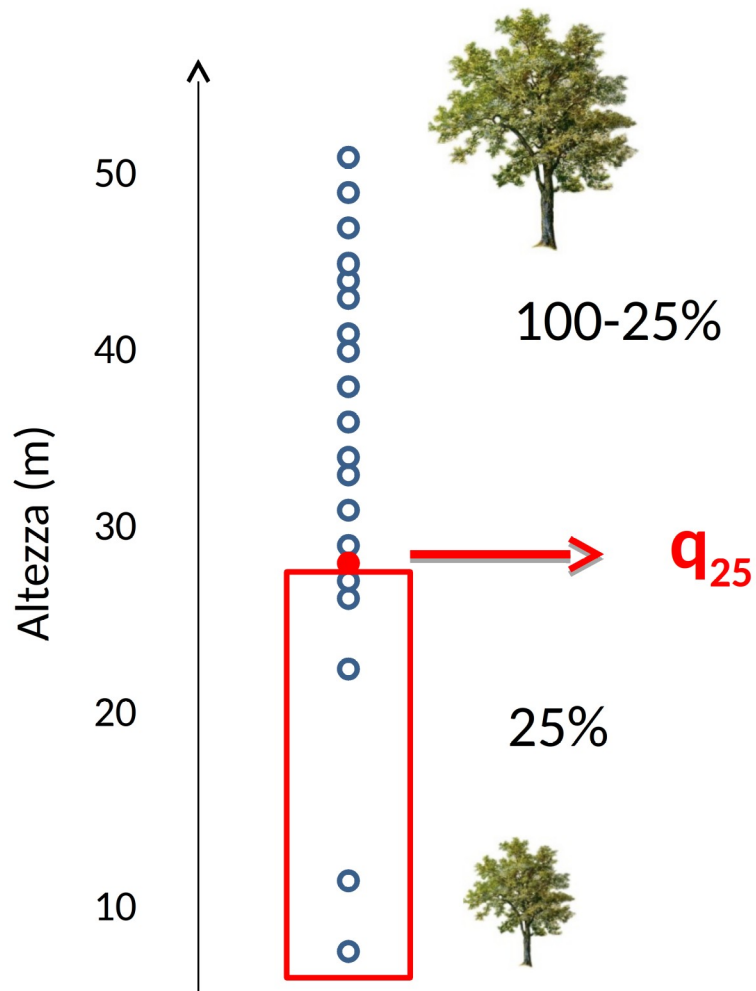


Il **quantile** è un valore q_α che divide le osservazioni in due parti, proporzionali ad α e $(1-\alpha)$ e caratterizzate da valori rispettivamente minori e maggiori di q_α

Indici di tendenza centrale: quantile 25%

q_{25}

È il valore che divide la serie in due parti: la prima contiene il 25% delle osservazioni mentre la seconda il restante 75%



Formule per i quartili (Q1 e Q3)

Formule per ottenere le posizioni di Q1 E Q3

$$Q1=(n+1)/4$$

$$Q3=3(n+1)/4$$

Le tre regole:

1. Se Q1 o Q3 sono un numero intero: **OK**
2. Se Q1 o Q3 sono ...,5: **media dei due valori**
3. Se Q1 o Q3 è ...,XX: **si approssima al più vicino**

Esempio con Q1

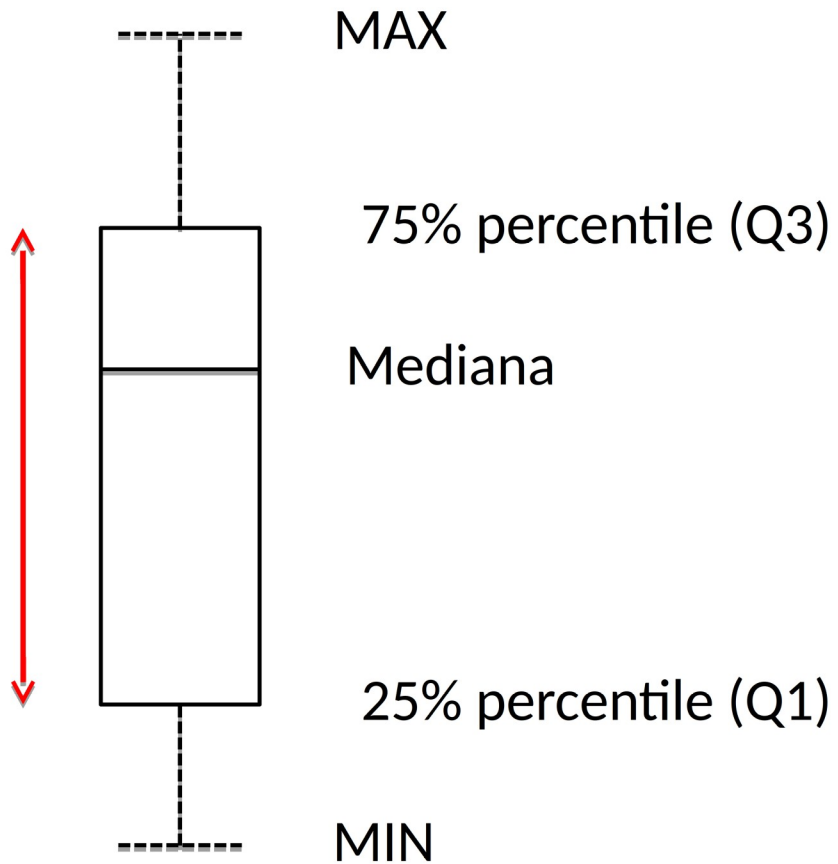
Q1=4 si prende il 4° valore

Q1=3.5 si media il 3° e 4° valore

Q1=3.45 si prende il 3° valore

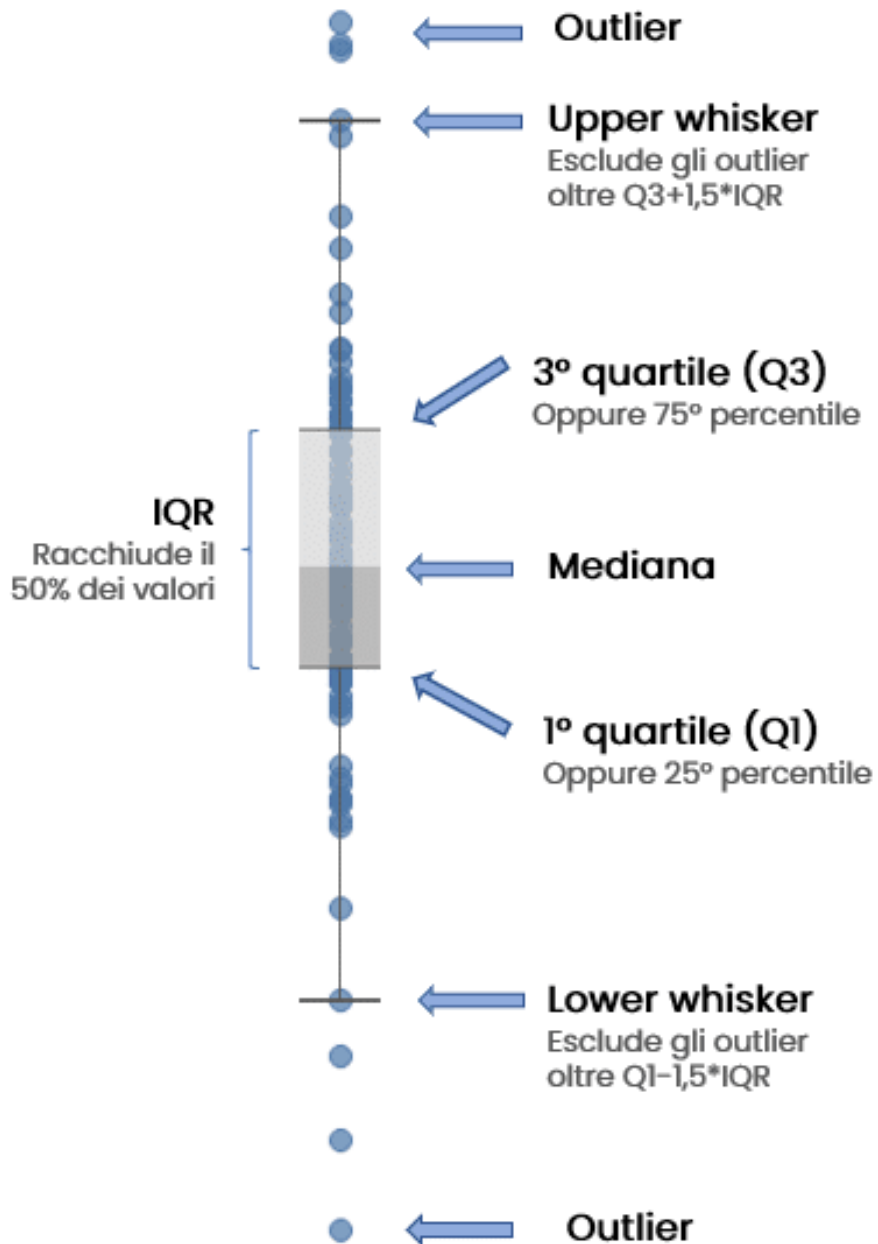
Q1=3.89 si prende il 4° valore

Box-plot

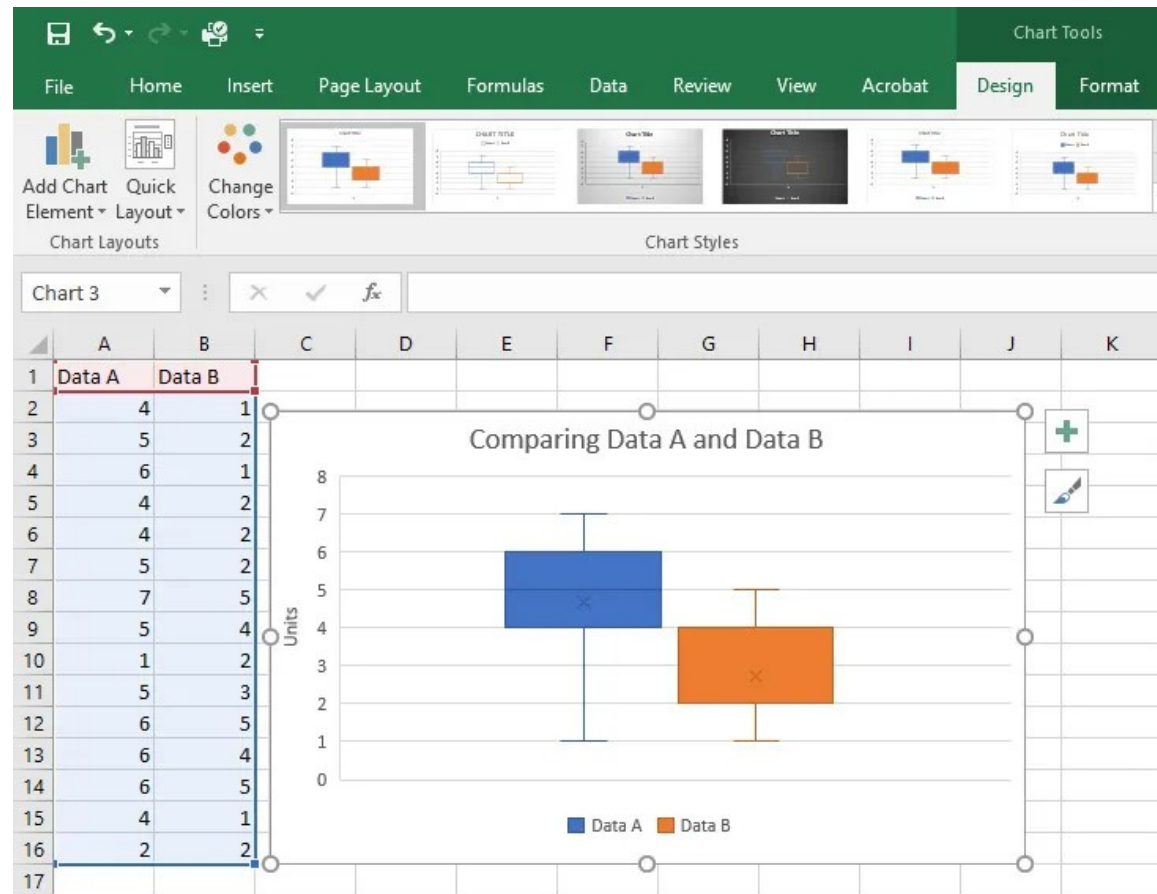


Range interquartile (IQR)

Box-plot e gli Outlier



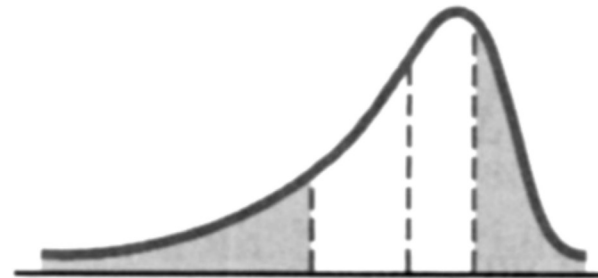
Utili anche per confrontare dati tra gruppi



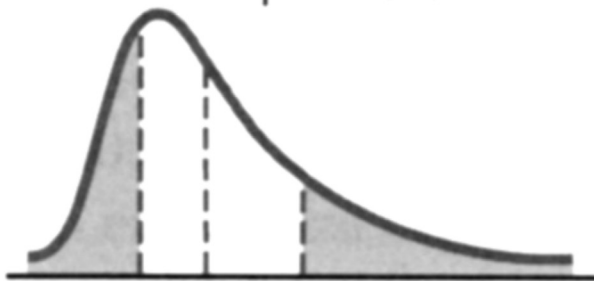
Box-plot



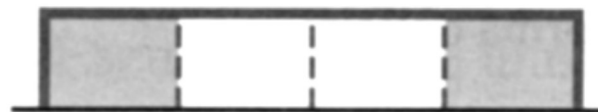
Riquadro A
Distribuzione simmetrica
campanulare



Riquadro B
Distribuzione asimmetrica
a sinistra



Riquadro C
Distribuzione asimmetrica
a destra

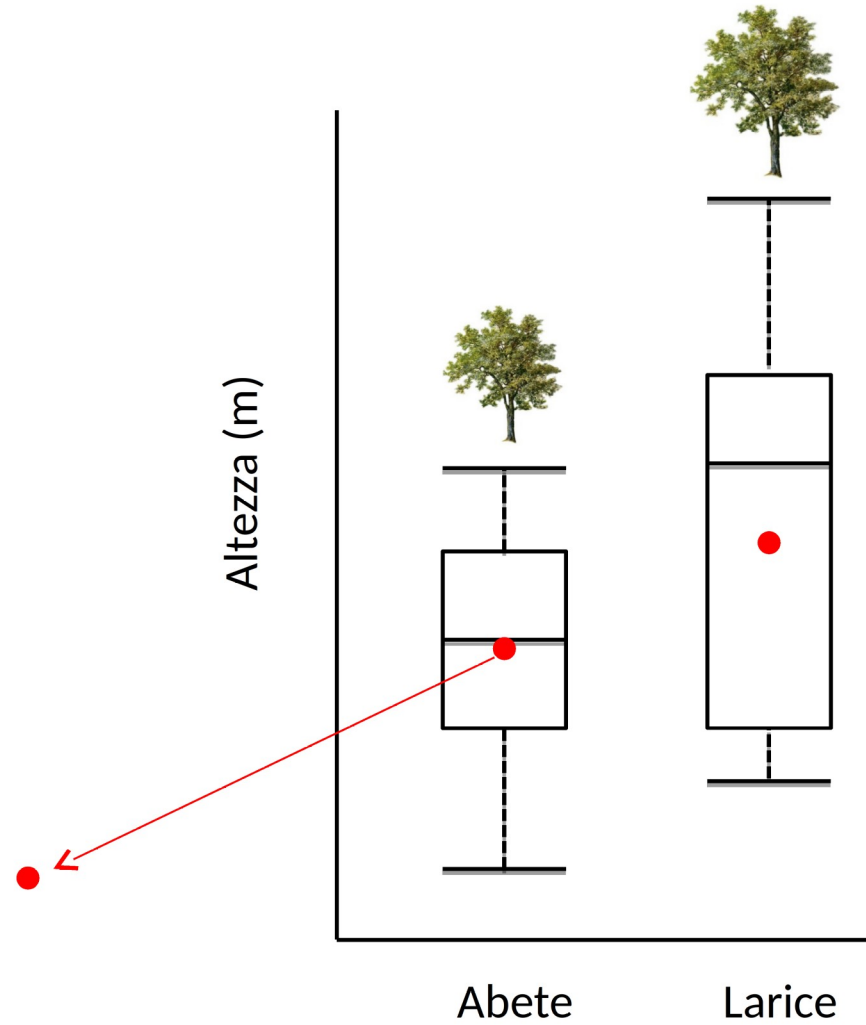


Riquadro D
Distribuzione uniforme
o rettangolare

I 5 valori di sintesi

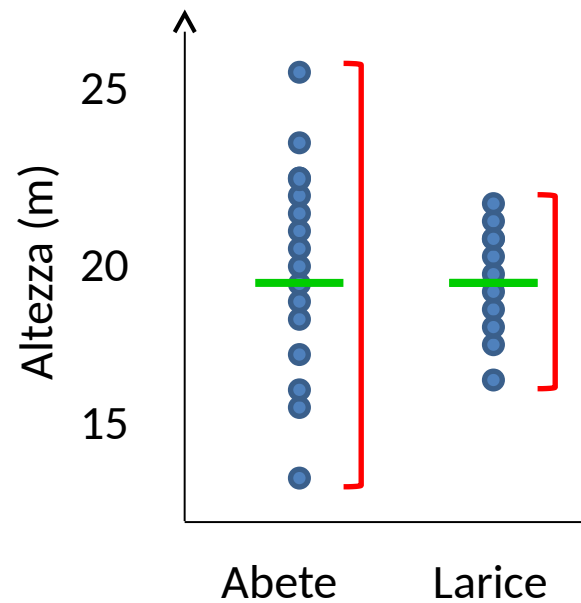
MAX
Q3
MEDIANA
Q1
MIN

+ MEDIA ARITMETICA

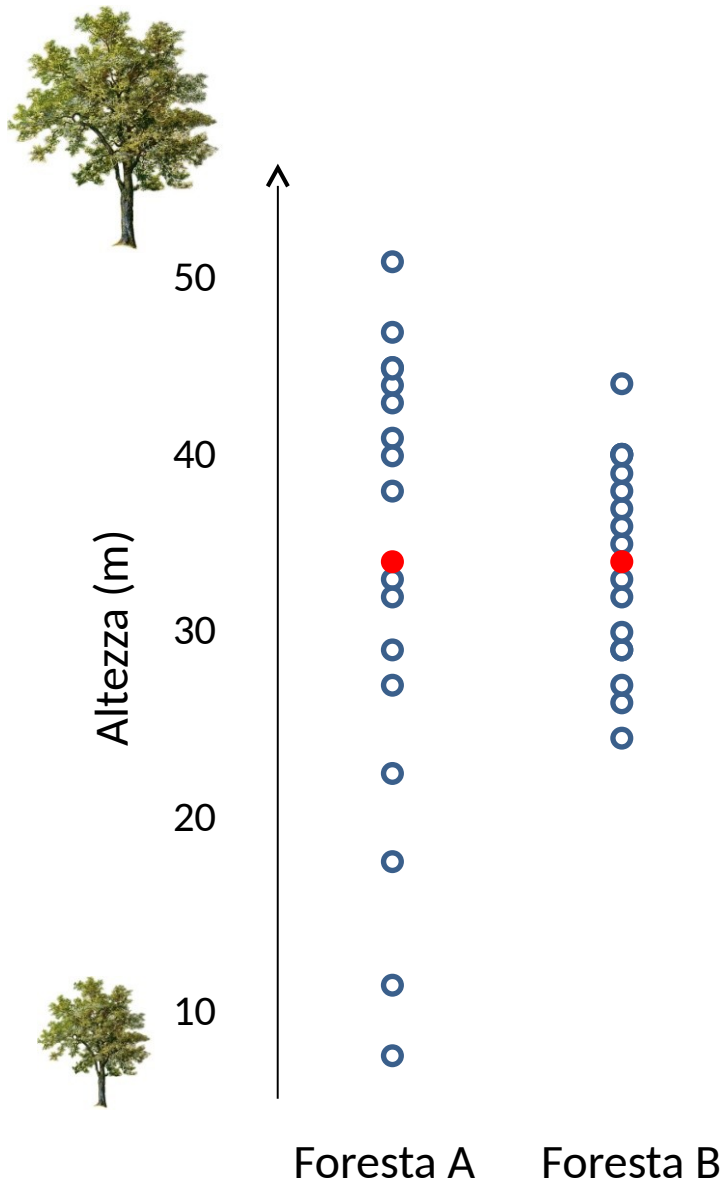


Indici di variabilità

Tutte le popolazioni hanno un grado più o meno alto di variabilità!



Indici di variabilità

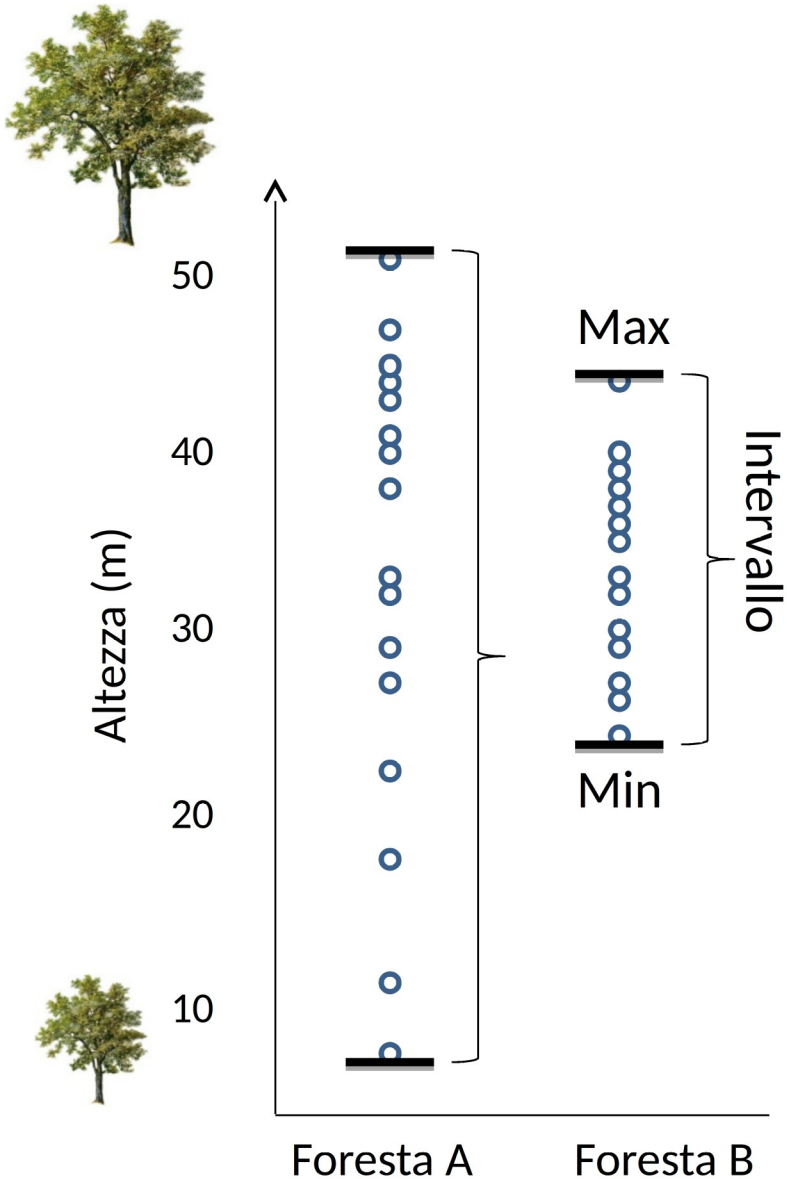


Gli indici di posizione non sono sufficienti per descrivere una variabile quantitativa

Gli indici di variabilità misurano il grado di dispersione dei dati attorno al valore centrale

Indici di variabilità: intervallo di variazione

L'intervallo di variazione è la più cruda misura di variabilità

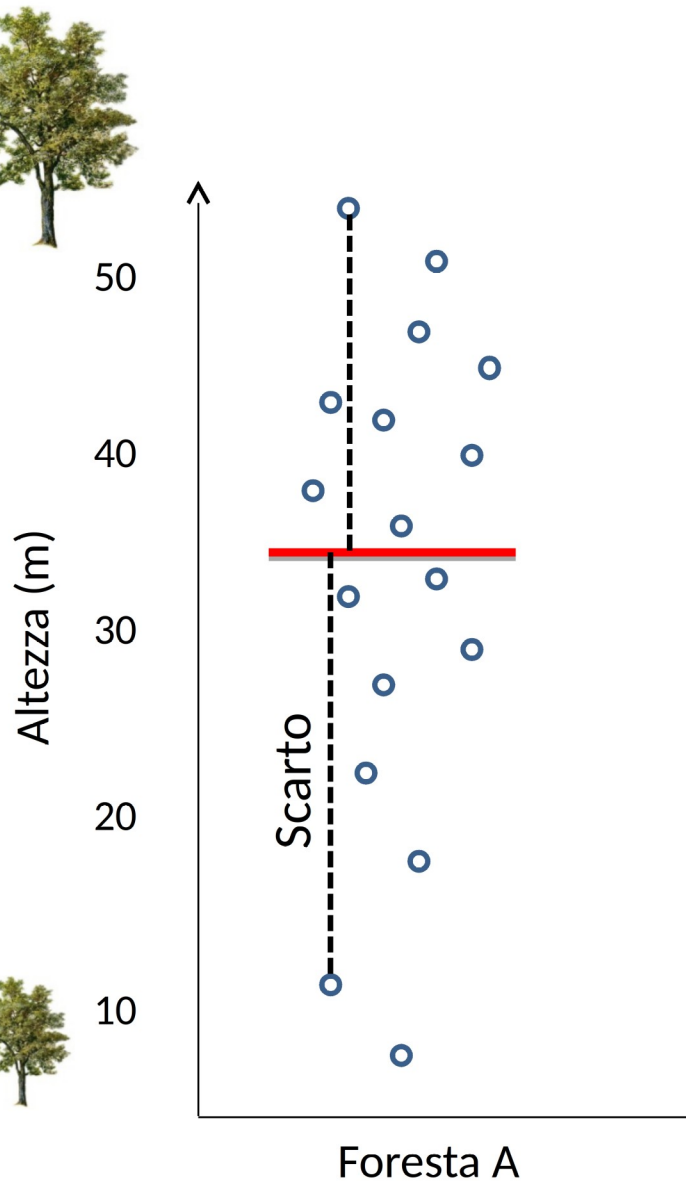


$$\text{Intervallo} = \text{Max} - \text{Min}$$

Maggiore l'intervallo maggiore la variabilità

Indici di variabilità: deviazione standard (I)

La deviazione standard si basa sulla misura degli scarti



— Media

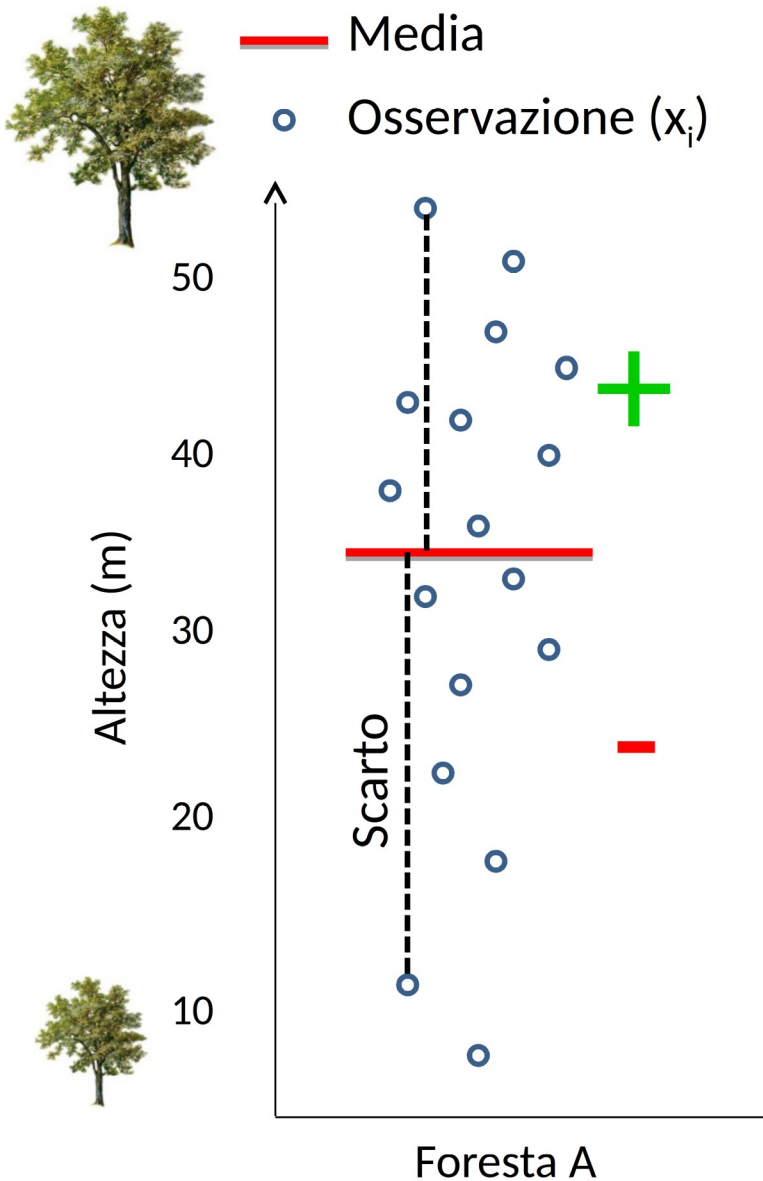
○ Osservazione (x_i)

Scarto o residuo = $x_i - \text{media}$

Maggiori gli scarti maggiore la variabilità

Indici di variabilità: deviazione standard (II)

Posso sommare gli scarti per misurare la variabilità?



$$\text{Somma dei quadrati} = \sum (x_i - \text{media})^2$$

Indici di variabilità: deviazione standard (III)

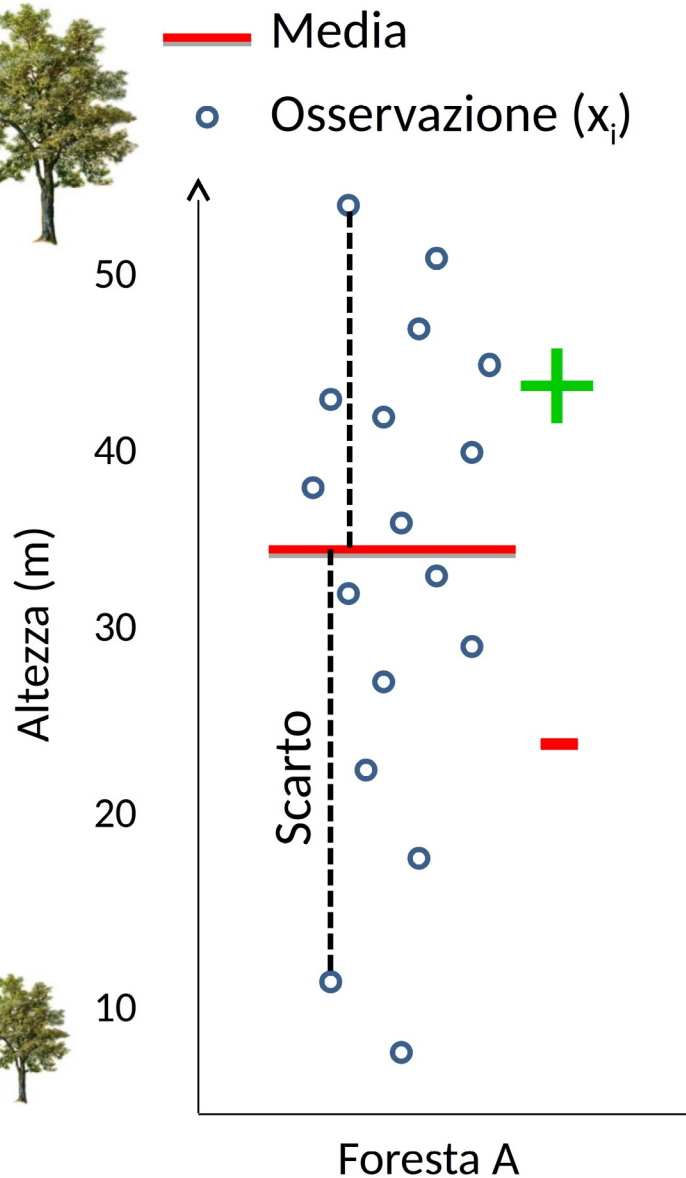
VARIANZA

CAMPIONE

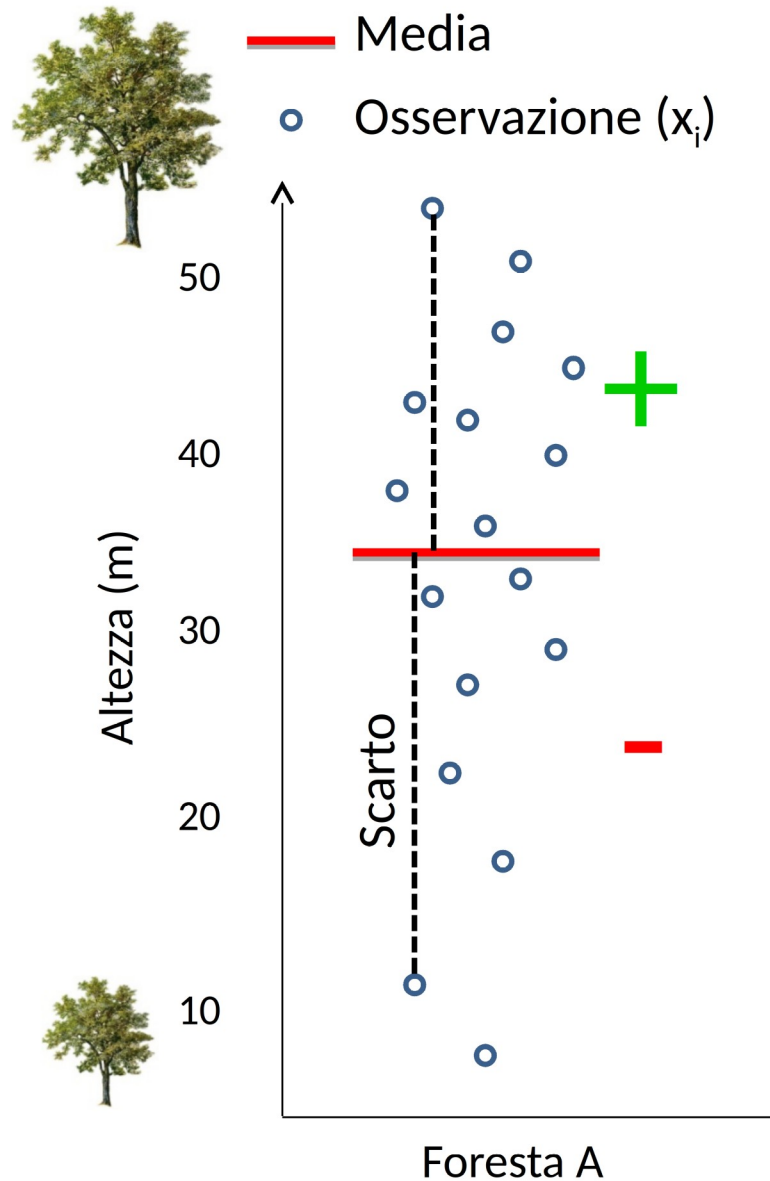
$$s^2 = \frac{\sum (x_i - \text{media})^2}{n - 1}$$

POPOLAZIONE

$$\sigma^2 = \frac{\sum (x_i - \text{media})^2}{N}$$



Indici di variabilità: deviazione standard (IV)



CAMPIONE

$$s = \sqrt{\frac{\sum (x_i - \text{media})^2}{n - 1}}$$

POPOLAZIONE

$$\sigma = \sqrt{\frac{\sum (x_i - \text{media})^2}{N}}$$

Indici di variabilità: Coefficiente di variazione

La deviazione standard misura la variabilità di una popolazione o campione

La deviazione standard dipende dal valore assoluto delle media

Il coefficiente di variazione ci permette di valutare la variabilità fra popolazioni con medie molto diverse

$$CV = \frac{s \times 100}{\bar{x}}$$

Indici di variabilità: Coefficiente di variazione

È più variabile l'altezza o l'età degli studenti?

Altezza media=1.70 (s=0.20) 11.7%

Età media=22 anni (s=1) 4.5%

$$CV = \frac{s \times 100}{\bar{x}}$$