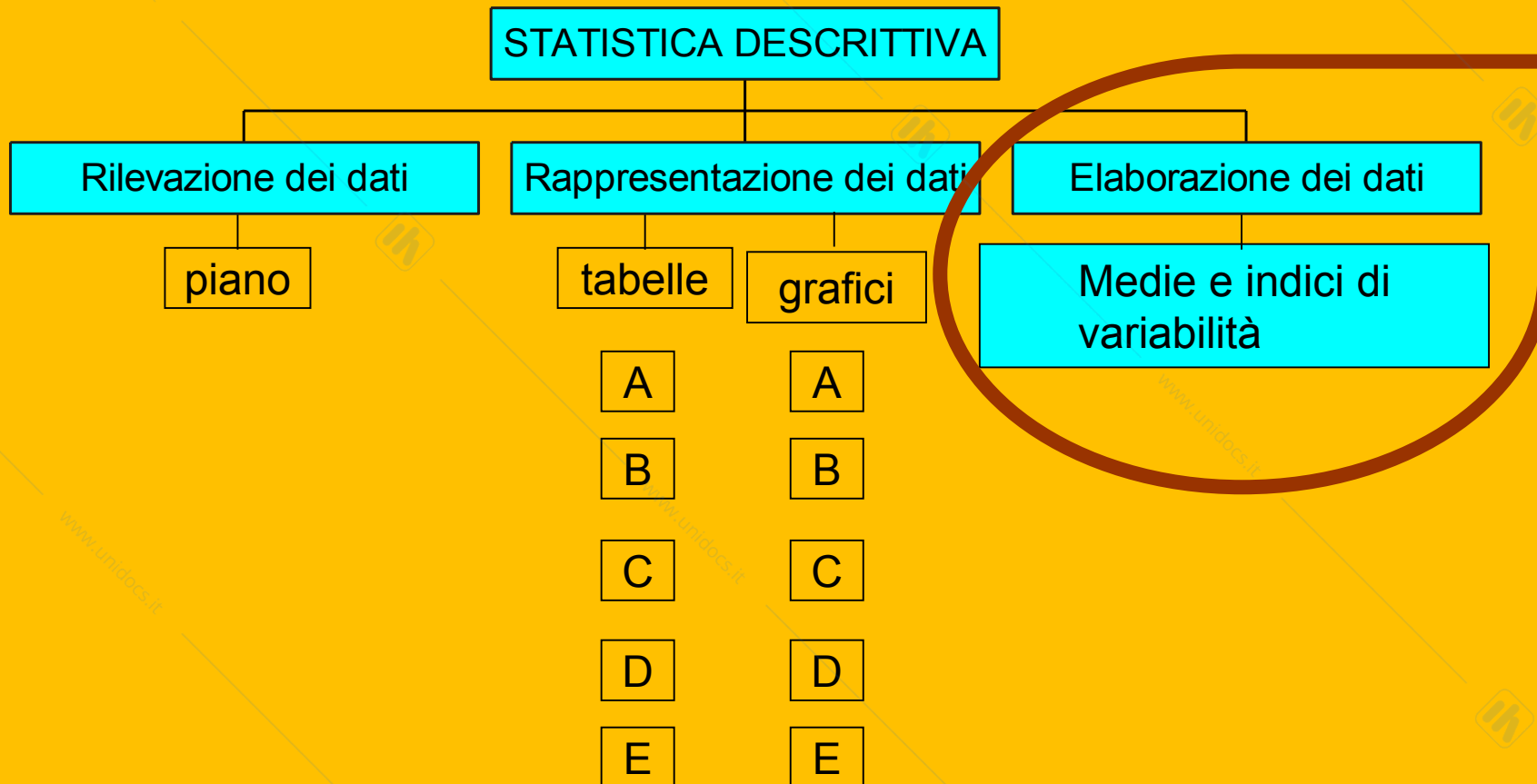


Statistica descrittiva



Indici Statistici

Per sintetizzare i dati ed evidenziare una certa caratteristica:

- **Indici di tendenza centrale**
- Indici di dispersione
- La forma

La Media

A questo punto bisogna dare dei criteri pratici per calcolare tale valor medio; i più importanti, quindi quelli più usati, sono i seguenti:

- a) si può calcolare il valor medio come **funzione matematica** dei dati rilevati e in tal caso si parla di **media analitica**;
- b) si possono ordinare i dati rilevati e ottenere la media in relazione alla **posizione** che occupa fra essi e in tal caso si parla di **media di posizione**.

Attenzione!

Non sempre il calcolo della media aritmetica rappresenta in modo significativo l'insieme dei valori a cui si riferisce. Per esempio, assegnati i valori:

$$a) 5, 6, 7, 6, 5, 6, 7, 6 \quad M = \frac{5+6+7+6+5+5+7+6}{8} = \frac{48}{8} = 6$$

$$b) 2, 3, 3, 2, 9, 9, 10, 10 \quad M = \frac{2+3+3+2+9+9+10+10}{8} = \frac{48}{8} = 6$$

$$c) 1, 2, 1, 3, 1, 1, 2, 13 \quad M = \frac{1+2+1+3+1+1+2+13}{8} = \frac{24}{8} = 3$$

È opportuno allora definire altri valori medi che non siano frutto di calcolo matematico, ma che siano individuati in base alla loro posizione nella sequenza dei valori osservati.

Tali medie si dicono **medie di posizione** le più utilizzate sono:

- La moda
- La mediana

La MEDIANA

Mediana: è il valore divisorio in quanto bipartisce la successione dei dati in due gruppi ugualmente numerosi; è il valore che taglia in due parti uguali la distribuzione dei dati ordinati, cioè il termine preceduto e seguito dallo stesso numero di dati.

La sua utilizzazione è **indispensabile nel caso di scale ordinali o di ranghi.**

Come la media è la misura di tendenza centrale nella statistica parametrica, la mediana è la misura di posizione o tendenza centrale utilizzata in quasi tutti i test non parametrici.

Mediana Me di n valori ordinati in modo crescente

- *sen è dispari* il termine che occupa la posizione centrale $\frac{n+1}{2}$
- *sen è pari* abbiamo due valori mediani $\frac{n}{2}$ e $\frac{n}{2} + 1$
e si usa la semisomma di $\frac{n}{2}$ e $\left(\frac{n}{2} + 1\right)$

Esempio: dati i valori ordinati:

1, 2, 2, 3, 4, 5, 6 Me = 3

i valori sono 7 la mediana è il termine che occupa il 4° posto $\blacktriangle (7+1)/2=4$

Altri indici “robusti”

Altri indici “robusti”

- ✓ I quartili
- ✓ I decili
- ✓ I percentili
- ✓ I quantili
- ✓ La moda
- ✓ Le medie troncate
- ✓ La semisomma interquartile

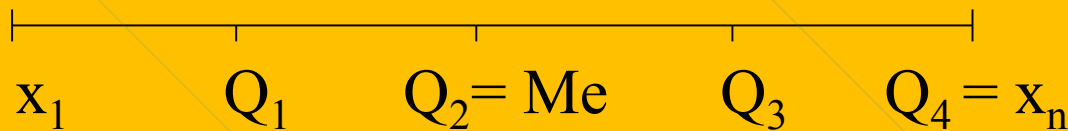
Caratteri, informazione e indici

Variabili quantitative	➔	Media, Mediana, Moda
Mutabili ordinabili	➔	Mediana, Moda
Mutabili sconnesse	➔	Moda

I Quartili

Il concetto di mediana si può facilmente generalizzare ottenendo altri valori divisori fra i quali i più usati sono i **quartili**. Tali indici di posizione si fondano sempre sul concetto di divisione della distribuzione.

I **Quartili** dividono la serie ordinata in quattro parti contenendo ciascuna lo stesso numero di dati.



La MODA

Voto (modalità)	Allievi (frequenza)
4	3
5	5
6	8
7	5
8	3

Moda di un fenomeno è la modalità con frequenza più elevata.

$$M_o = 6$$

MODA

non è influenzata dalla presenza di nessun valore estremo;
tuttavia viene utilizzata solamente a scopi descrittivi, perché **è meno stabile e meno oggettiva delle altre misure di tendenza centrale**

Misure di posizione: moda

E' l'osservazione che si verifica con maggior frequenza; quindi la moda puo' essere assente (specie se le osservazioni sono poche) o puo' essere plurima (es. curve bimodali con 2 picchi).

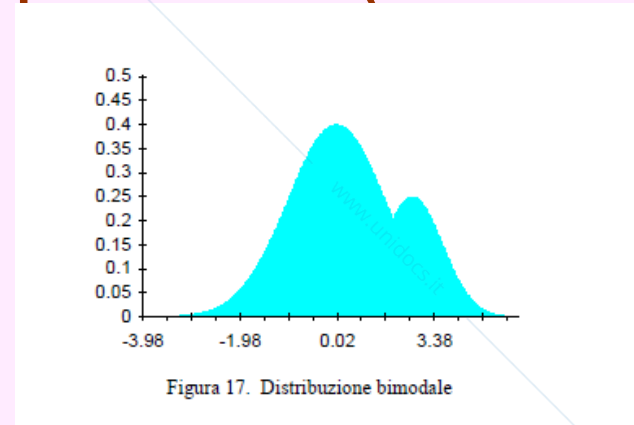


Figura 17. Distribuzione bimodale

Nelle distribuzioni "normali" (cioe' unimodali e simmetriche) media, mediana e moda coincidono.

MODA

Quando la distribuzione dei dati evidenzia due o più mode, il ricercatore deve quindi sospettare che i dati non siano omogenei, ma formati da altrettanti gruppi con differenti tendenze centrali.

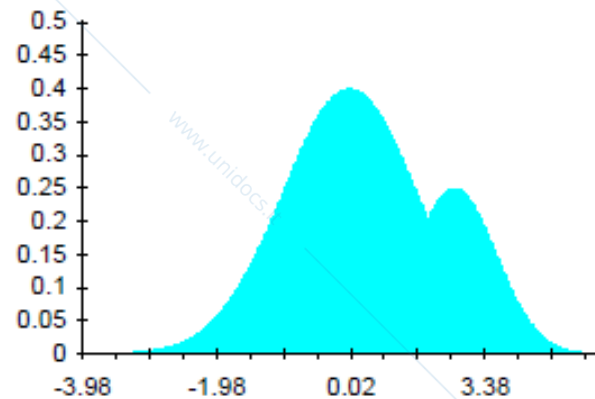
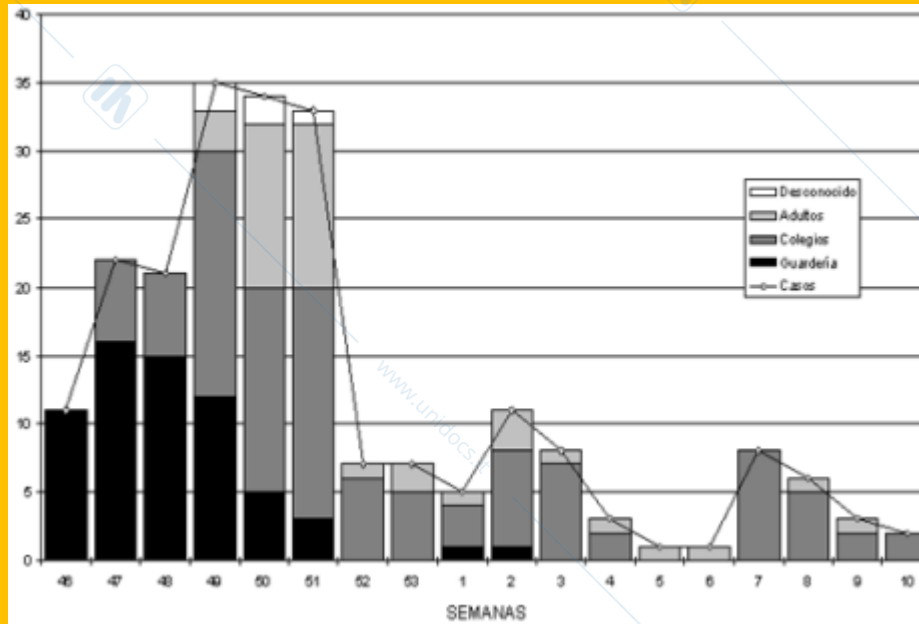


Figura 17. Distribuzione bimodale

Curva plurimodale



Indici Statistici

Per sintetizzare i dati ed evidenziare una certa caratteristica:

- **Indici di tendenza centrale**
- **Indici di dispersione**
- La forma

MISURE DI DISPERSIONE O DI VARIABILITA'

La dispersione o variabilità è la seconda importante caratteristica di una distribuzione di dati.

Essa definisce la forma più o meno raccolta della distribuzione intorno al valore centrale e fornisce indicazioni sul tipo di test da applicare

MISURE DI DISPERSIONE O DI VARIABILITA'

Servono a esprimere il grado di variabilità o dispersione attorno al valore medio.

Naturalmente le fonti di variabilità possono essere tante e si suole raggrupparle in tre grandi gruppi (v. oltre):

- ***variazione biologica***
- ***temporale***
- ***errori di misura.***

SORGENTI DI VARIAZIONE

In medicina e biologia le misure cliniche quantitative sono esposte a tre tipi fondamentali di variazione:

- 1) biologica:** tutti i fattori che rendono diversi i vari individui: età, sesso, razza, dieta, ecc; la variazione dovuta alla variabilità biologica è detta anche "variabilità tra soggetti" (between subjects)
- 2) temporale :** tutti i fattori che determinano variazioni nello stesso individuo da un momento all'altro;
- 3) errore di misura:** tutti i fattori che tendono a produrre differenze in misure diverse dello stesso fenomeno (strumenti di misura, errori tecnici, diversità dell'osservatore, ecc); tale variabilità è detta anche "variabilità entro soggetti" (within subjects).

Indici di Variabilità

I valori medi sono indici importanti per la descrizione sintetica di un fenomeno statistico

Hanno però il limite di non darci alcuna informazione sulla distribuzione dei dati

Indici statistici di variabilità

- Campo di variazione o range R
- Varianza
- Scarto quadratico medio
 - Scarto dalla media
 - Coefficiente di variazione

Permettono di valutare le disuguaglianze dei dati rilevati in relazione al loro scostamento o dispersione da una media.

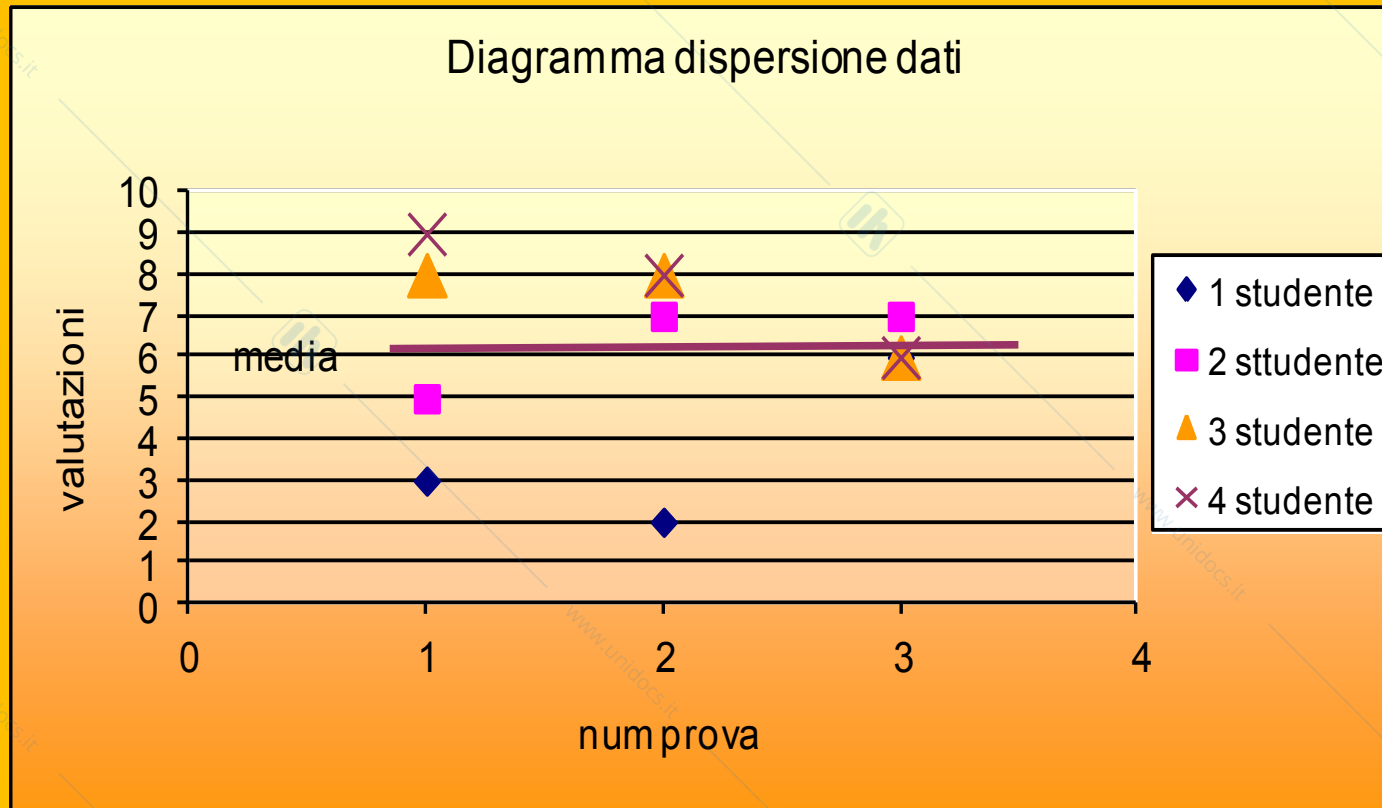
Esempio

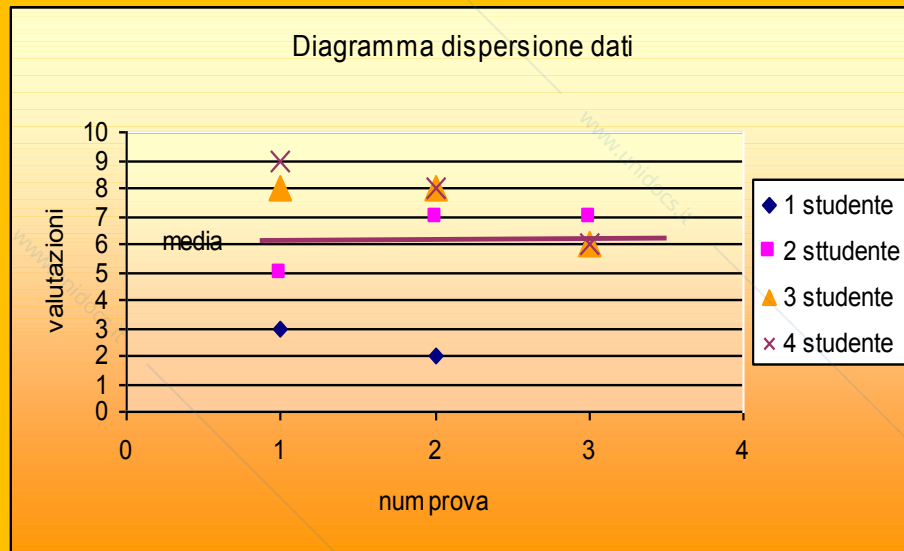
In tre differenti prove di matematica 4 studenti hanno riportato le seguenti valutazioni

	1a Prova	2a Prova	3a Prova
1° studente	3	5	6
2° studente	5	7	7
3° studente	8	6	6
4° studente	9	7	6
media	6,25	6,25	6,25

In tutte e tre le prove la media è **6,25**
ma i dati sono chiaramente distribuiti in modo
diverso

Diagramma di distribuzione delle tre prove





RANGE

$$1^{\circ} \text{ prova} = (9-3) = 6$$

$$2^{\circ} \text{ prova} = (7-5) = 2$$

$$3^{\circ} \text{ prova} = (7-6) = 1$$

- nel caso della 1^a prova e 2^a prova sarà opportuno fare un recupero per alcuni studenti
- nel caso della 3^a prova l'insegnante può ritenere che gli obiettivi siano stati raggiunti dalla classe, anche se ad un livello solo sufficiente

Indici statistici di variabilità

- Campo di variazione o range R
- Varianza
- Scarto quadratico medio
 - Scarto dalla media
 - Coefficiente di variazione

Permettono di valutare le disuguaglianze dei dati rilevati in relazione al loro scostamento o dispersione da una media.

Scarto medio dalla media aritmetica

Un modo per calcolare la variabilità dei dati (tenendo conto di tutti i dati) consiste nel calcolare la distanza di tutti i dati dalla media e fare la media aritmetica di tali distanze

$$\text{Scarto medio} = S_m = \frac{|x_1 - \bar{x}| + |x_2 - \bar{x}| + \dots + |x_n - \bar{x}|}{n}$$

Scarto medio = Distanza media dei dati dalla media

Esempio

Consideriamo le valutazioni di una prova

	1a Prova
1° studente	3
2° studente	5
3° studente	8
4° studente	9
media	6,25

$$\begin{aligned} |\Delta x_1| &= |3 - 6,25| = 3,25; & |\Delta x_2| &= |5 - 6,25| = 1,25; \\ |\Delta x_3| &= |8 - 6,25| = 1,75; & |\Delta x_4| &= |9 - 6,25| = 2,75; \end{aligned}$$

$$S_m = \frac{3,25 + 1,25 + 1,75 + 2,75}{4} = 2,25$$

Calcoliamo lo Scarto medio per tutte le tre prove

	1a Prova	2a Prova	3a Prova
1° studente	3	2	6
2° studente	5	7	7
3° studente	8	8	6
4° studente	9	8	6
media	6,25	6,25	6,25
scarto medio	2,25	2,13	0,38

Scarto 1^a prova = 2,25 \Rightarrow dati più dispersi,
risultati più eterogenei

Scarto 3^a prova = 0,38 \Rightarrow dati più concentrati,
risultati più omogenei

Scarto 2^a pr. \neq Scarto 1^a pr.

“Le Distribuzioni Differiscono”

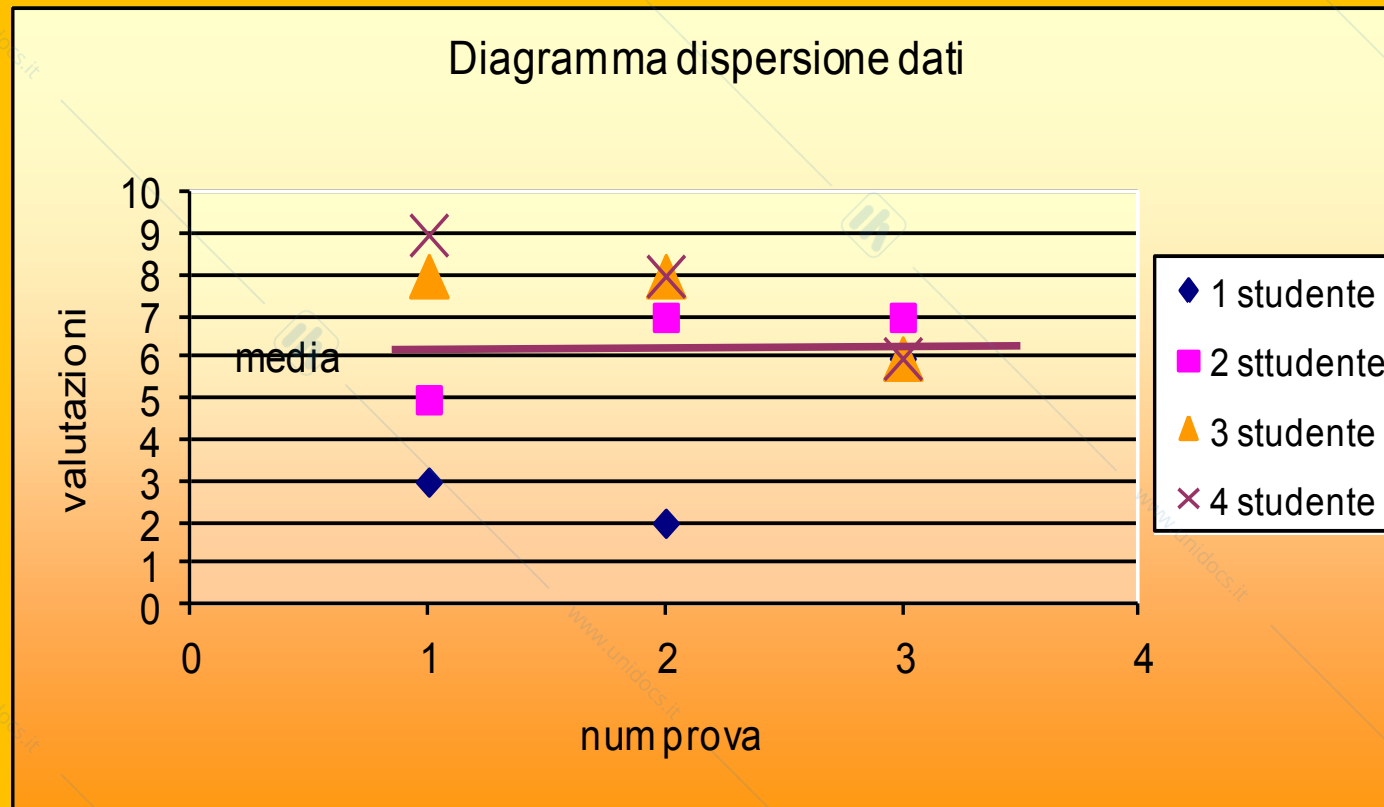
Esempio

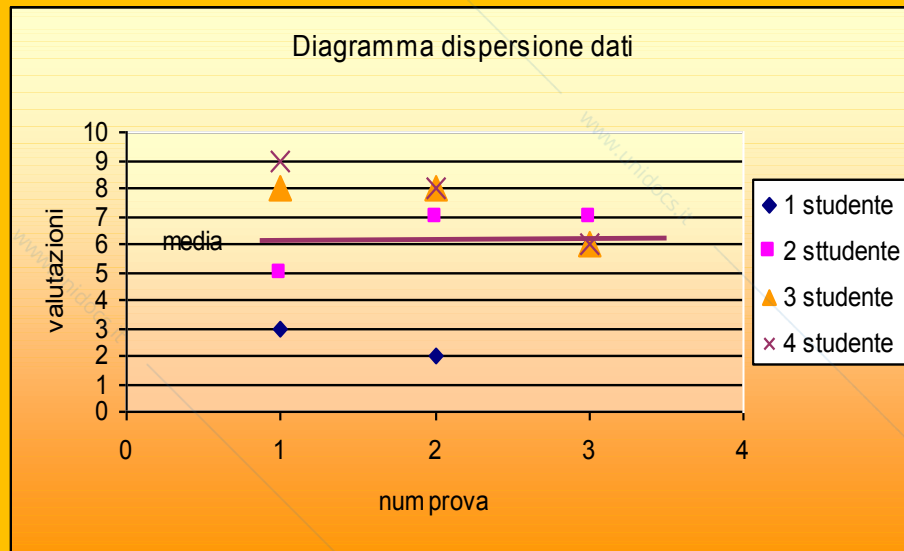
In tre differenti prove di matematica 4 studenti hanno riportato le seguenti valutazioni

	1a Prova	2a Prova	3a Prova
1° studente	3	5	6
2° studente	5	7	7
3° studente	8	6	6
4° studente	9	7	6
media	6,25	6,25	6,25

In tutte e tre le prove la media è **6,25**
ma i dati sono chiaramente distribuiti in modo
diverso

Diagramma di distribuzione delle tre prove





RANGE

$$1^{\circ} \text{ prova} = (9-3) = 6$$

$$2^{\circ} \text{ prova} = (7-5) = 2$$

$$3^{\circ} \text{ prova} = (7-6) = 1$$

- nel caso della 1^a prova e 2^a prova sarà opportuno fare un recupero per alcuni studenti
- nel caso della 3^a prova l'insegnante può ritenere che gli obiettivi siano stati raggiunti dalla classe, anche se ad un livello solo sufficiente