

Corso di Statistica Psicometrica

Canale 1 - Matricole Dispari

CdS triennale in
Scienze e Tecniche
Psicologiche

Domenico Vistocco

Dipartimento di Scienze Politiche
Università degli Studi di Napoli Federico II



Attenzione...



Lei ha il diritto di rimanere in silenzio: qualsiasi cosa dirà / farà potrà essere usata contro di lei...

Nella
puntata
precedente...



- Tipologia di variabili
- Tipologia di tabelle
- Frequenza assolute, relative e percentuali
- Frequenze cumulate
- Densità di frequenza
- Istogramma e funzione di ripartizione empirica
- Indici assoluti ed indici relativi
- Classificazione degli indici univariati in base all'obiettivo di studio
- Quantile e rango quantile (più comunemente: percentile e rango percentile)
- Moda, mediana e media aritmetica (definizione e tecnicismi)
- Proprietà di internalità degli indici di posizione
- Proprietà particolari della media
- Gli scarti come misura di dispersione intorno al centro e come segnali per la misura della forma di una distribuzione
- **Criterio di ottimalità della mediana**
- **Modelli statistici**
- **I centri come "semplici" modelli**
- **Misura dell'adattamento di un modello ai dati (alla realtà?)**

Un'equazione per tutte le stagioni

$$\text{esito}_i = \text{modello} + \text{errore}_i$$

$$\text{risposta}_i = \text{modello} + \text{errore}_i$$

$$\text{risultato}_i = \text{modello} + \text{errore}_i$$

Un'equazione per tutte le stagioni

$$\text{esito}_i = \text{modello} + \text{errore}_i$$



dato osservato

Un'equazione per tutte le stagioni

$$\text{esito}_i = \text{modello} + \text{errore}_i$$



rappresentazione
semplificata della realtà



media, mediana, moda, ...

Un'equazione per tutte le stagioni

$$\text{esito}_i = \text{modello} + \text{errore}_i$$



scarto tra il
valore osservato e
il modello scelto

Un'equazione per tutte le stagioni

$$\text{esito}_i = \text{modello} + \text{errore}_i$$



L'approssimazione che
commettiamo nell'utilizzare
un modello per rappresentare
la realtà è negli errori

$$\sum_i \text{errore}_i = ?$$

Un'equazione per tutte le stagioni

$$\text{esito}_i = \text{modello} + \text{errore}_i$$



la media come modello?

$$\sum_i \text{errore}_i^2 = \sum_i (x_i - \bar{x})^2$$

minimo

misura dell'adattamento
del modello ai dati

Un'equazione per tutte le stagioni

$$\text{esito}_i = \text{modello} + \text{errore}_i$$

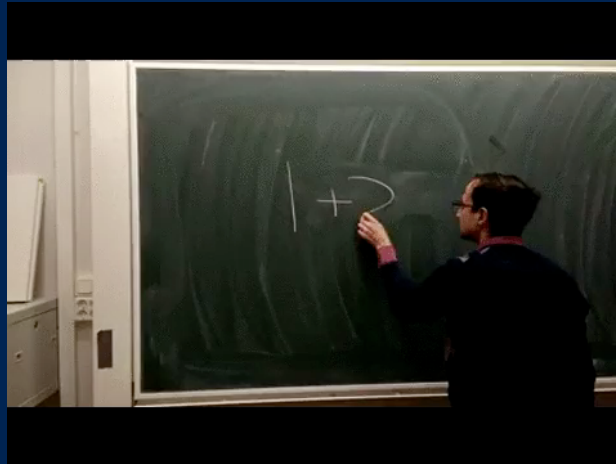


la mediana come modello?

$$\sum_i |\text{errore}_i| = \sum_i |x_i - Me|$$

minimo

misura dell'adattamento
del modello ai dati



Si riparte?

Tre diversi accezioni di **variabilità**

disuguaglianza

differenze medie

dispersione

scarti o
scostamenti medi

distanza tra due
posizioni opposte

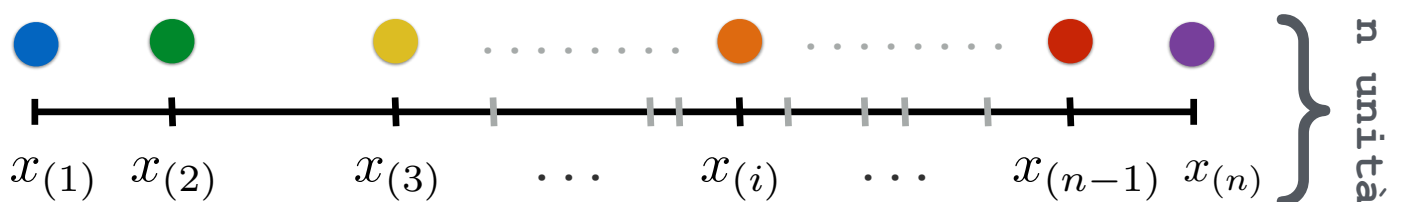
intervalli di
variazione (range)

Uguali o diversi?

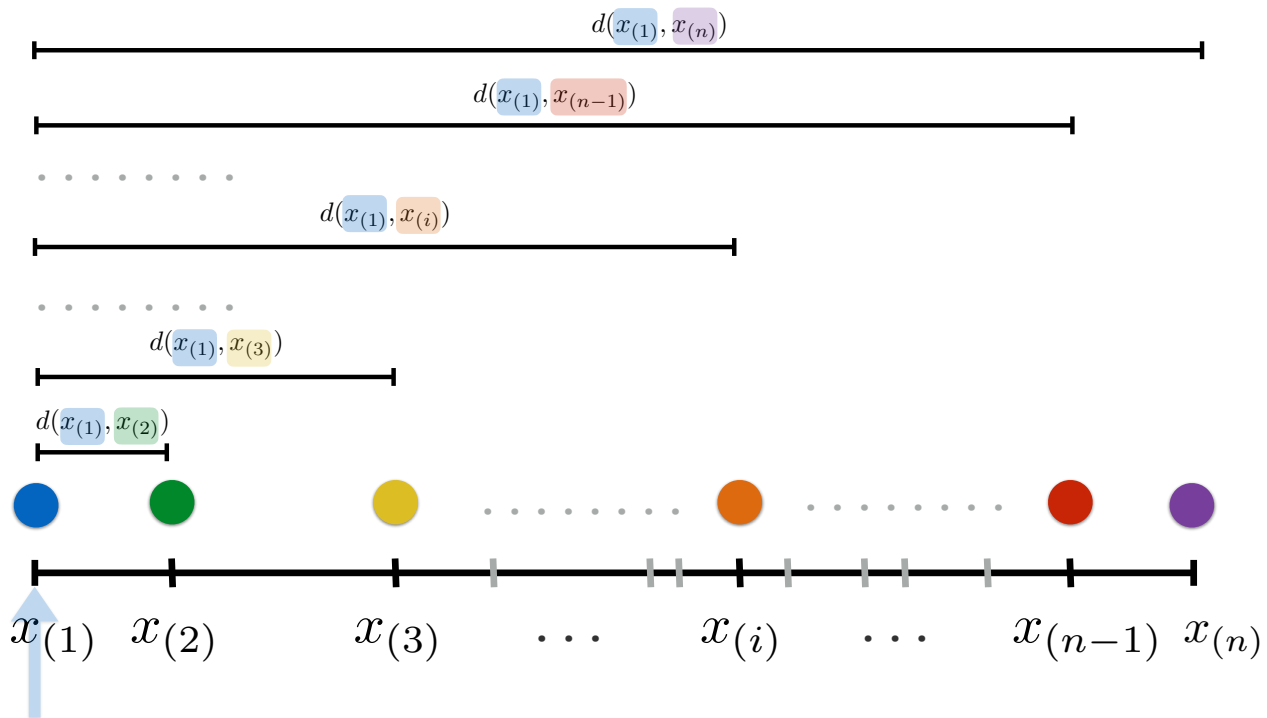


Variabilità come disuguaglianza

(mutua variabilità o variabilità reciproca)

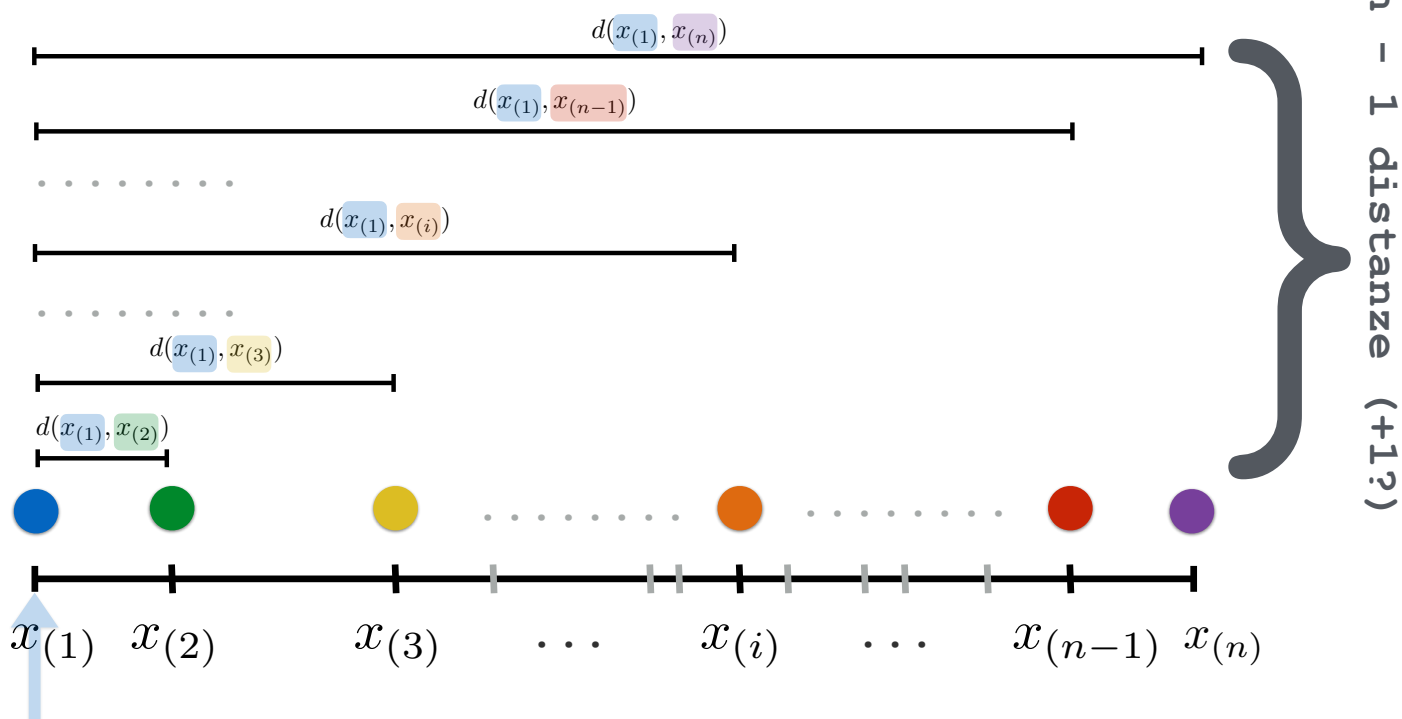


Variabilità come disuguaglianza

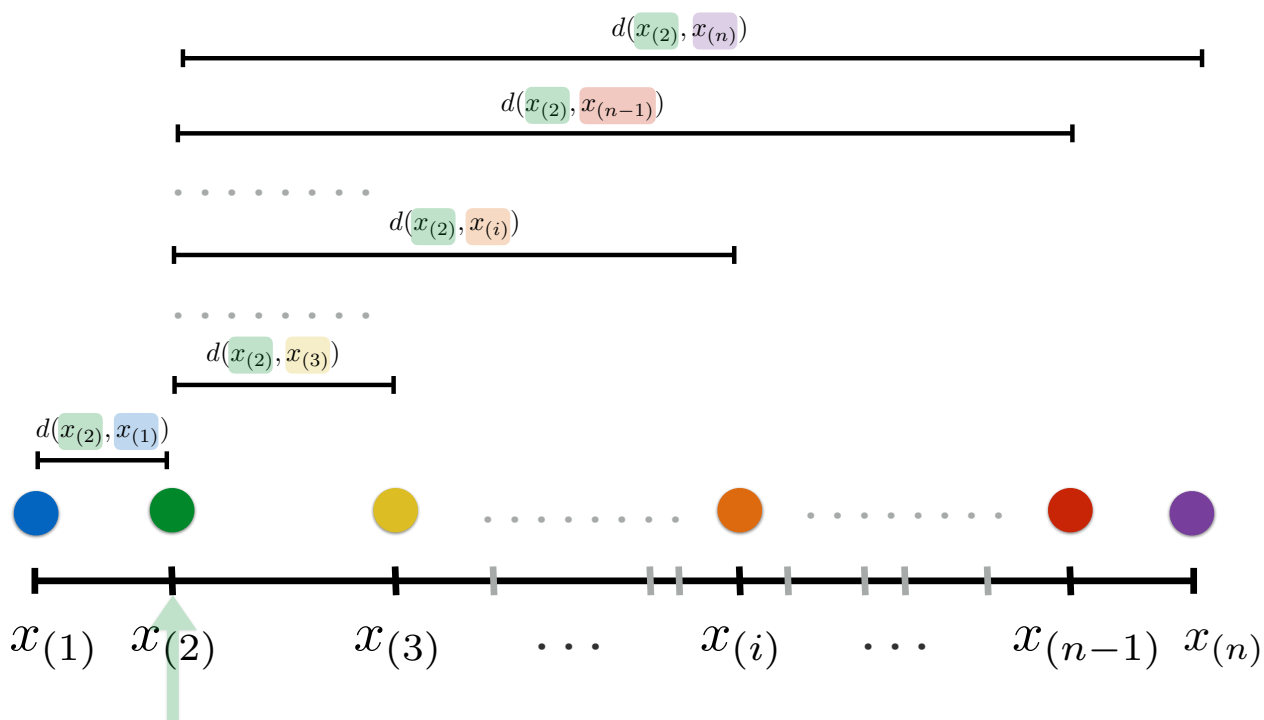


Quante
distanze
(o differenze) ?

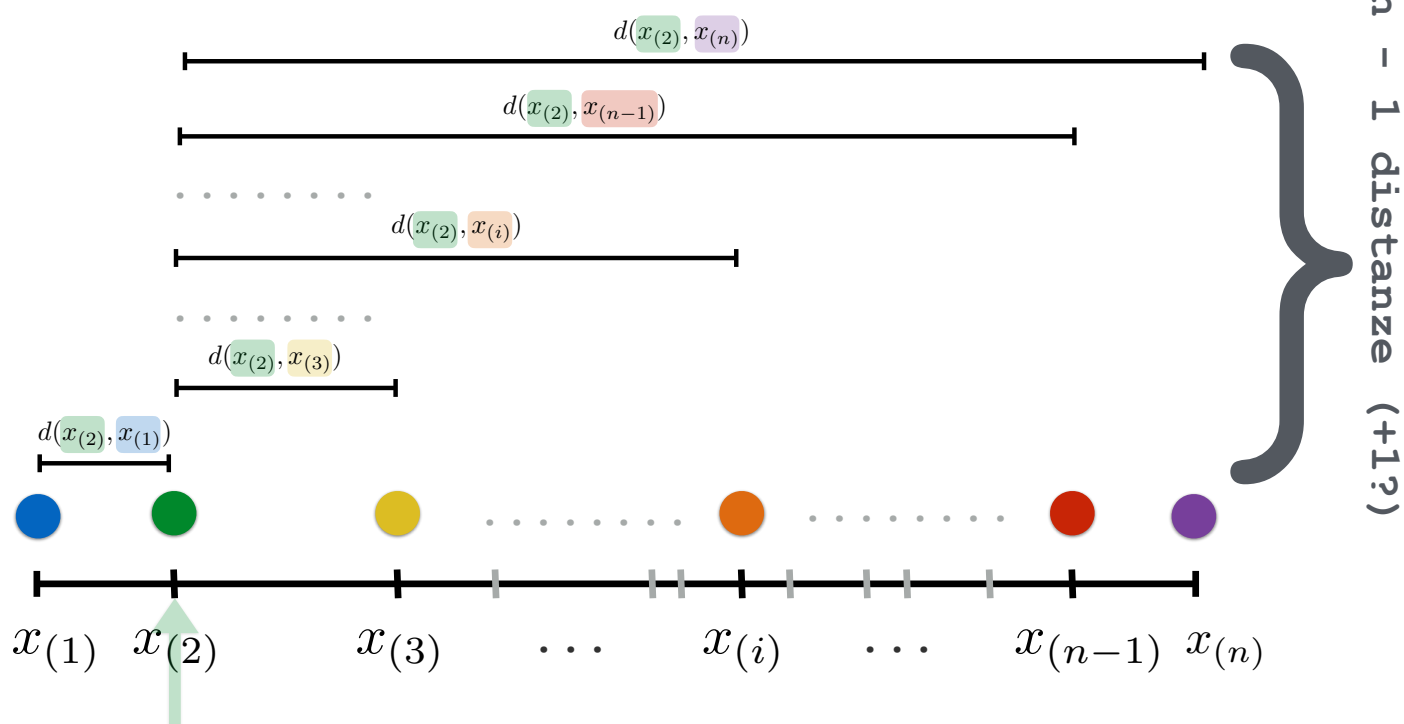
Variabilità come disuguaglianza



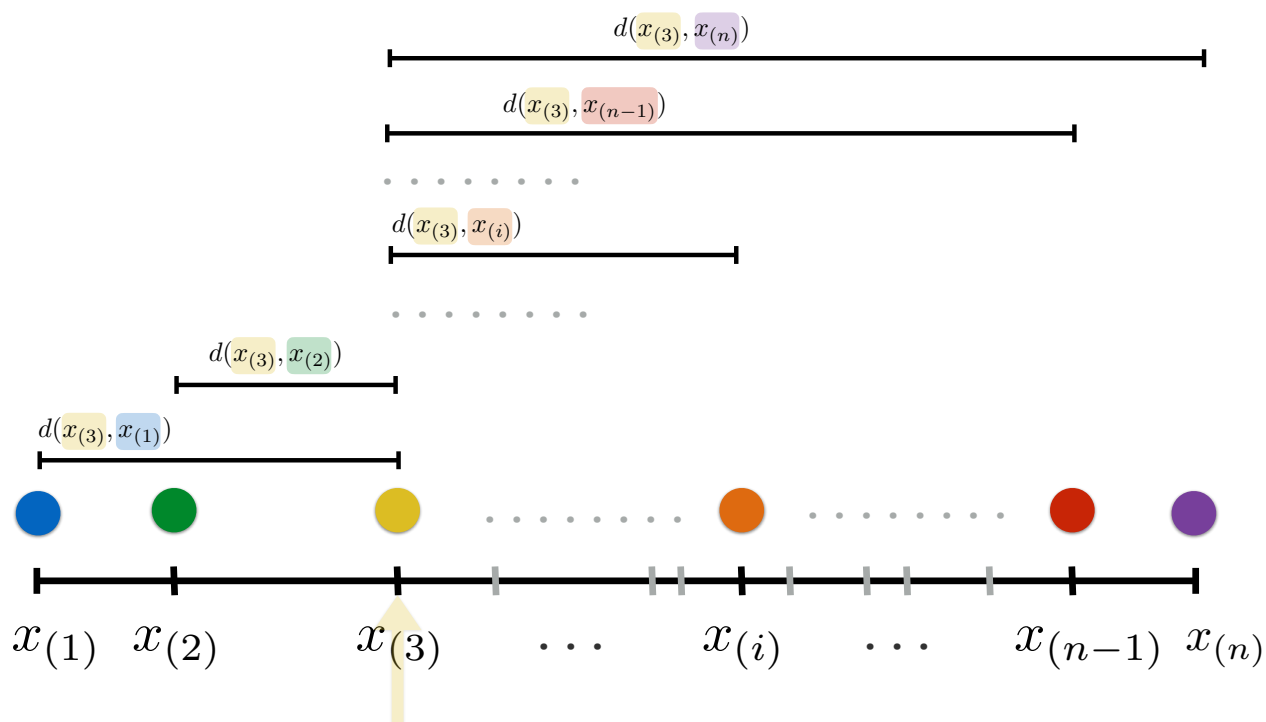
Variabilità come disuguaglianza



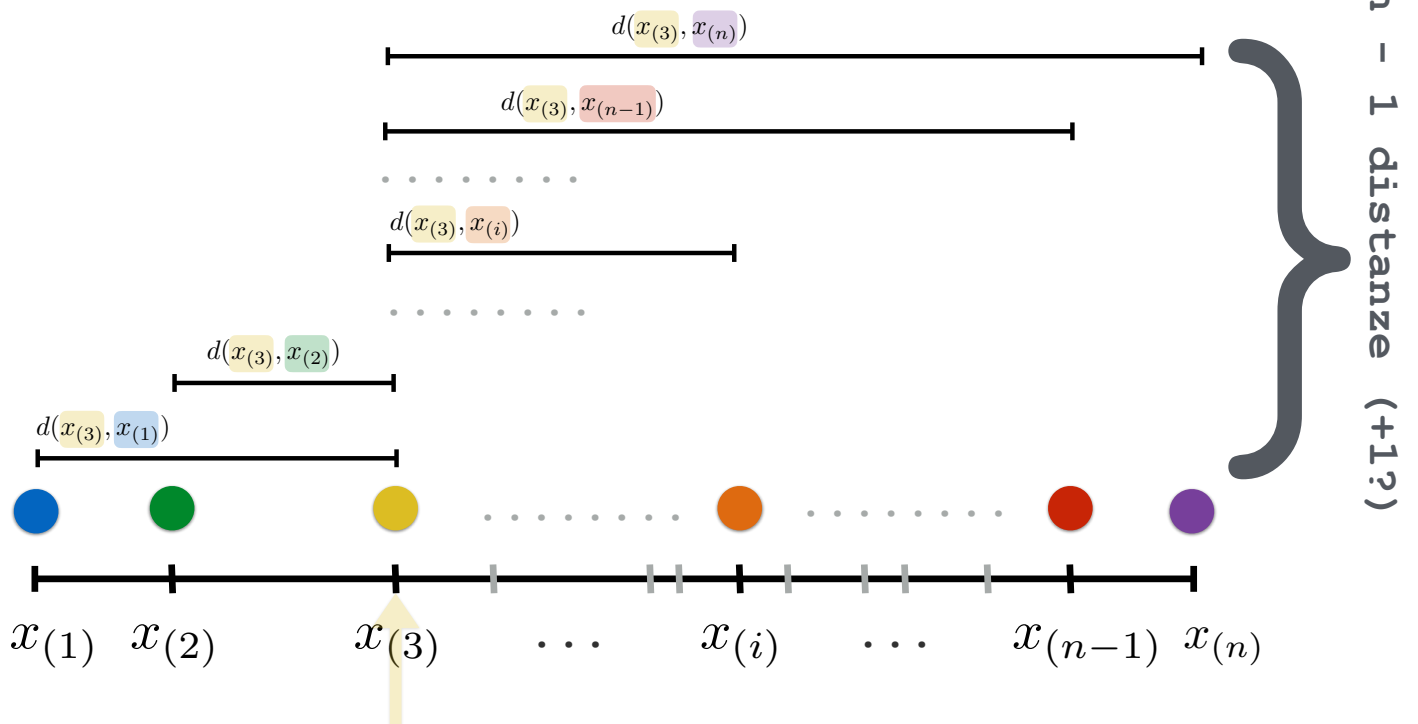
Variabilità come disuguaglianza



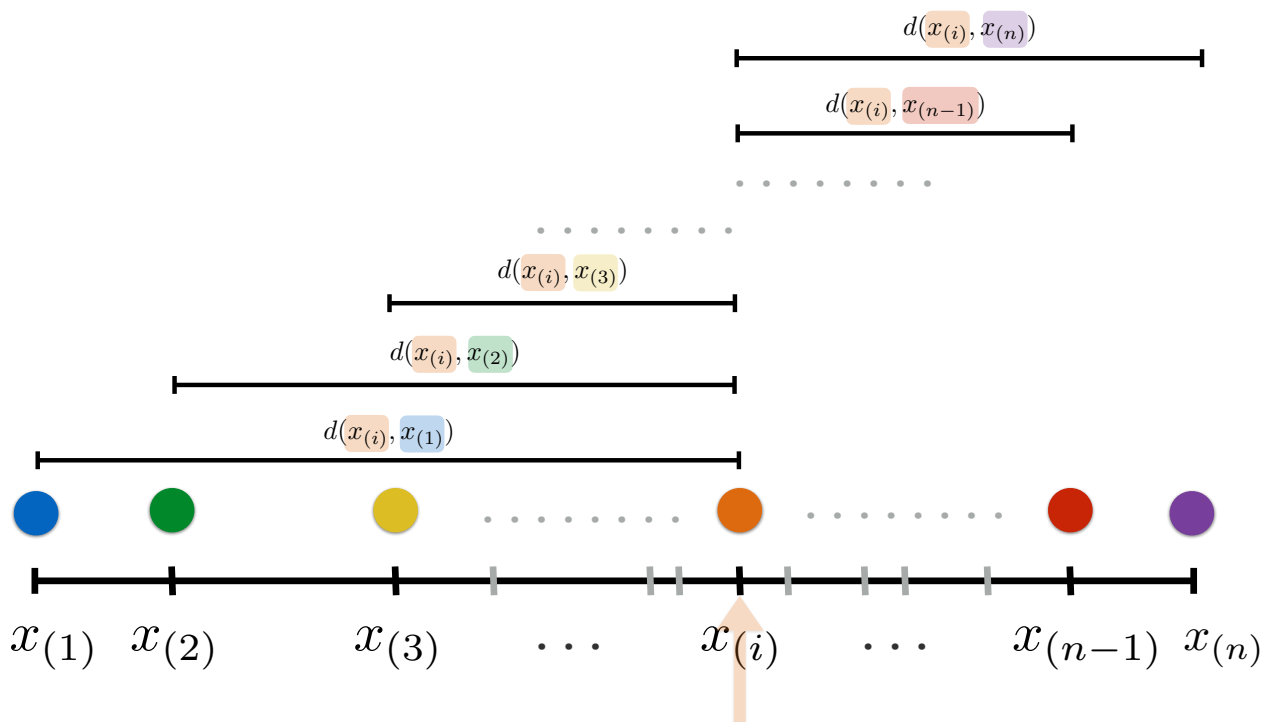
Variabilità come disuguaglianza



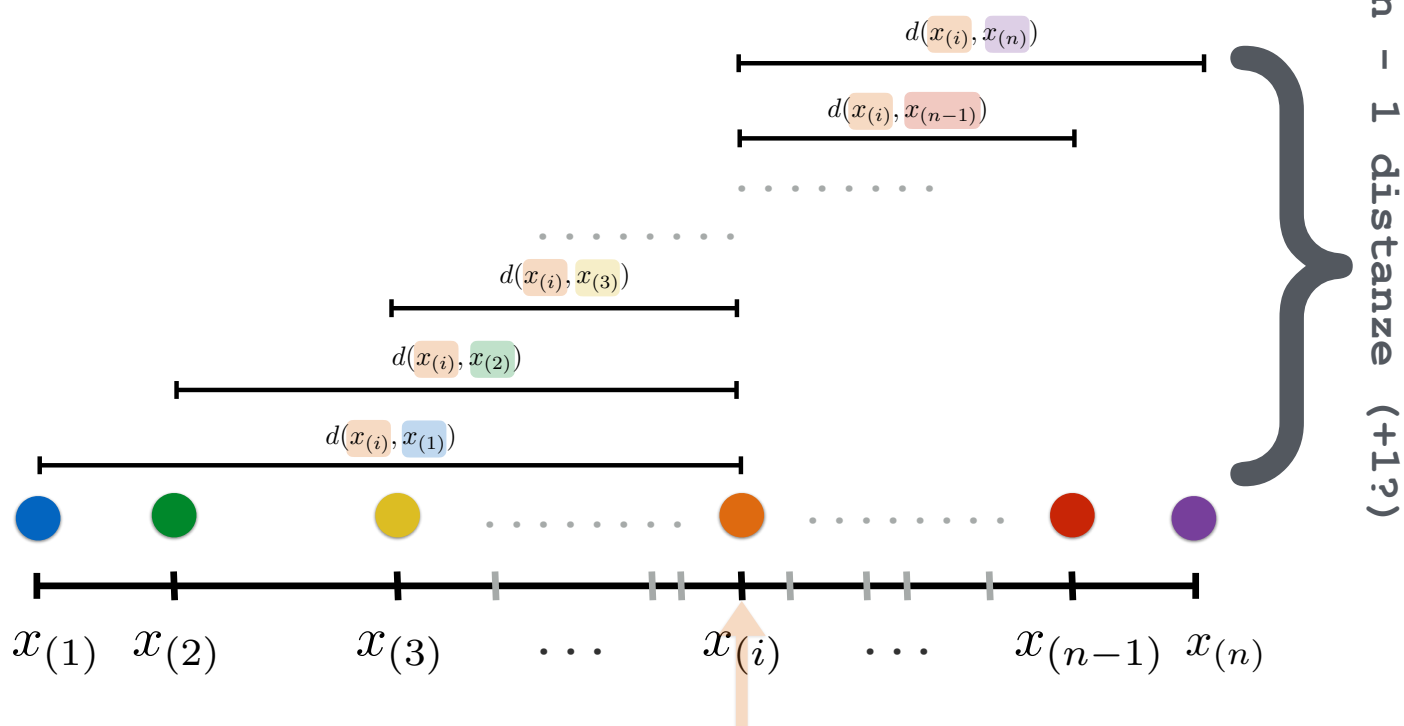
Variabilità come disuguaglianza



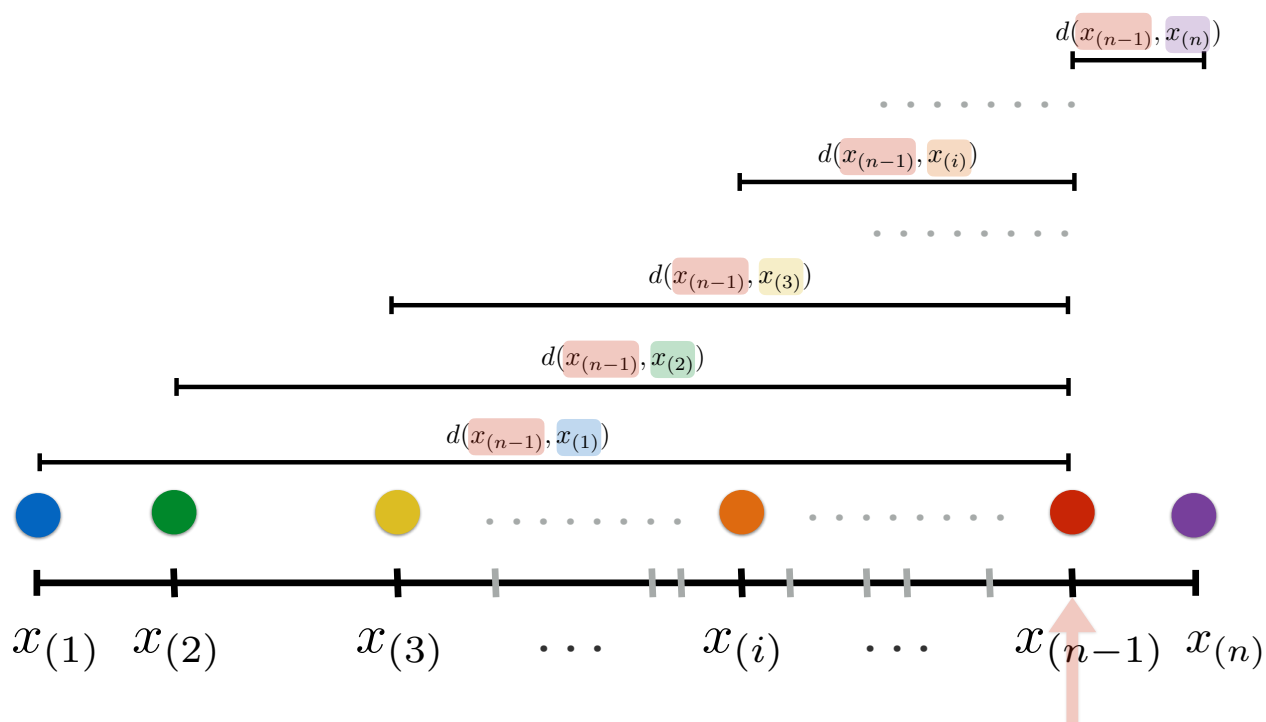
Variabilità come disuguaglianza



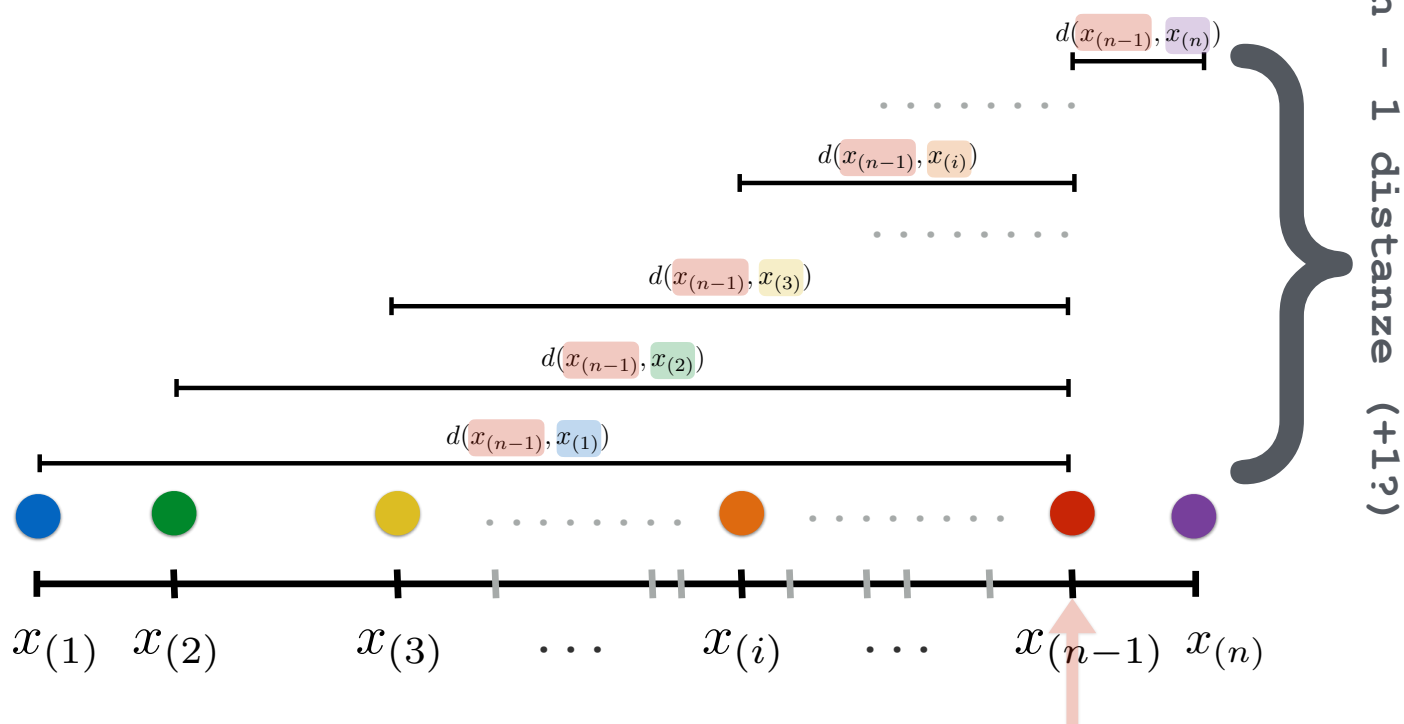
Variabilità come disuguaglianza



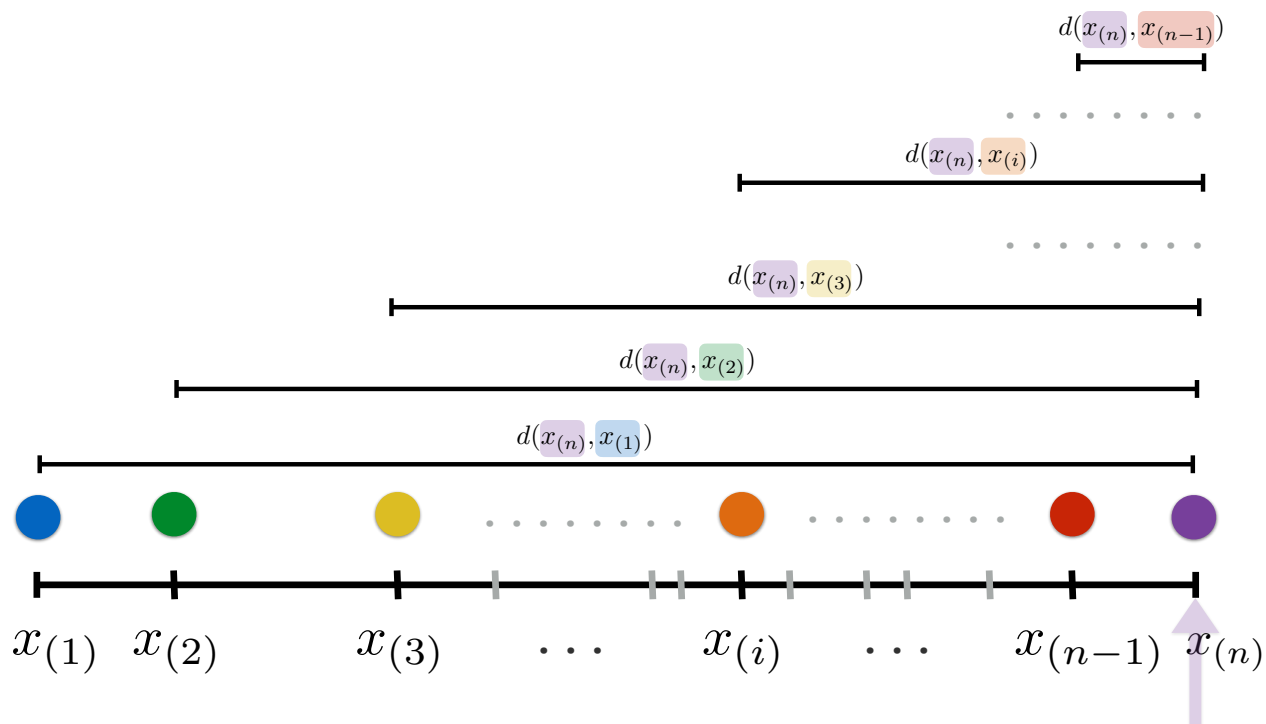
Variabilità come disuguaglianza



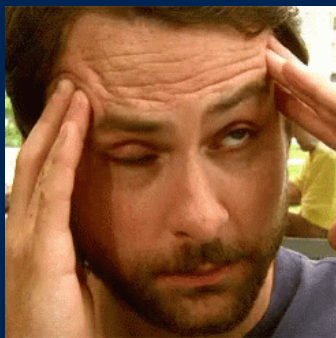
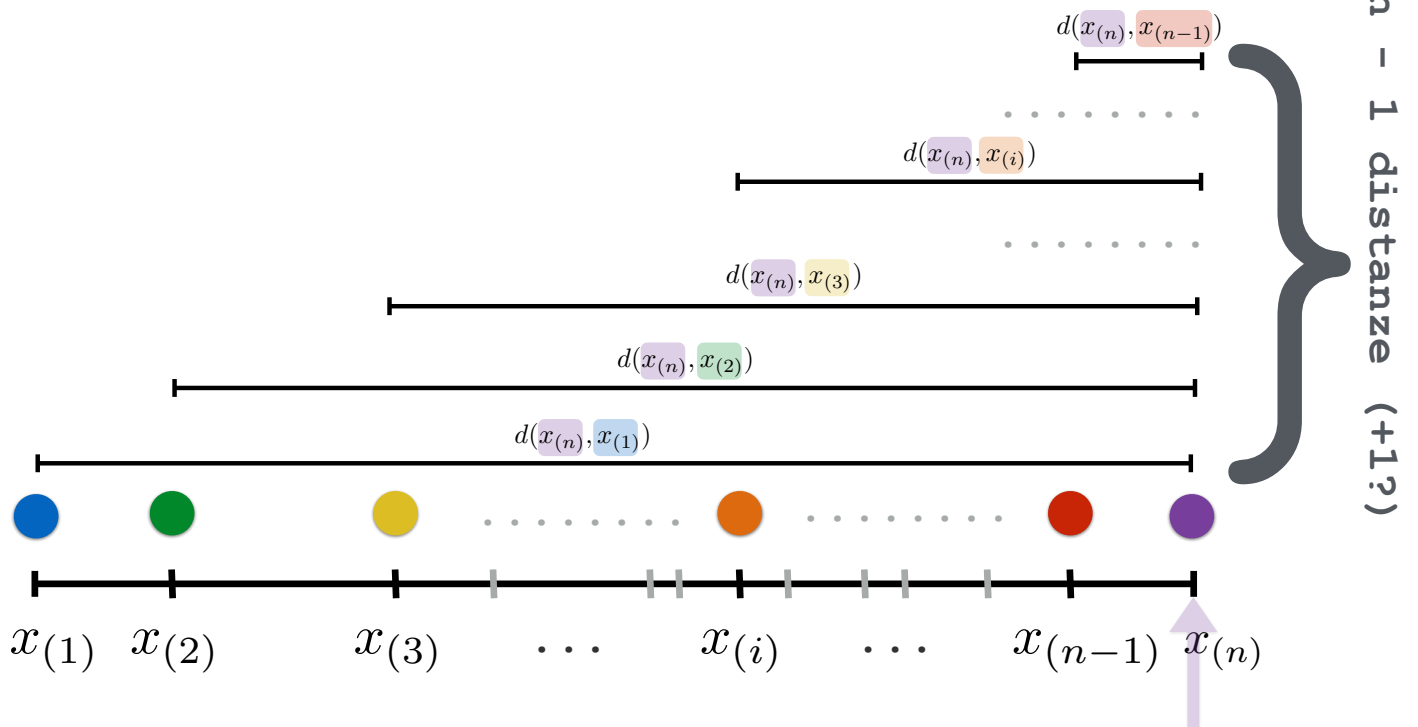
Variabilità come disuguaglianza



Variabilità come disuguaglianza

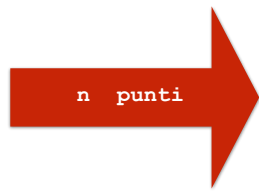


Variabilità come disuguaglianza



Quante
distanze
(o differenze)?

Variabilità come disuguaglianza



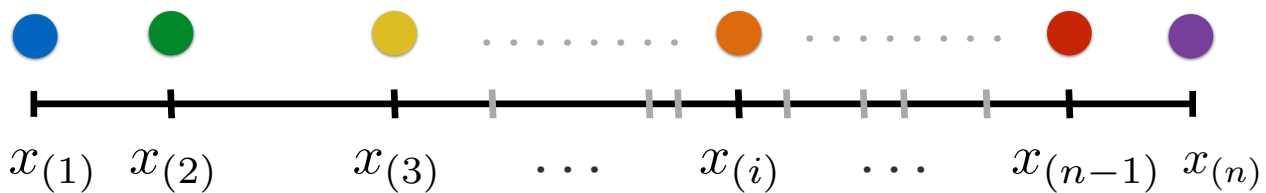
$n \times (n - 1)$ distanze

senza
ripetizione

$n \times n$

considerando anche
quelle "superflue"

con
ripetizione



Quante
distanze
(o differenze) ?

Variabilità come disuguaglianza

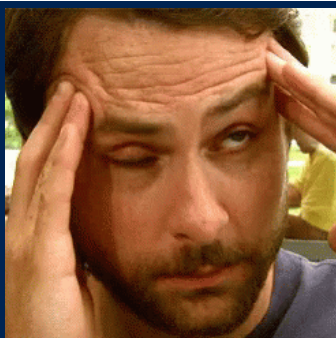
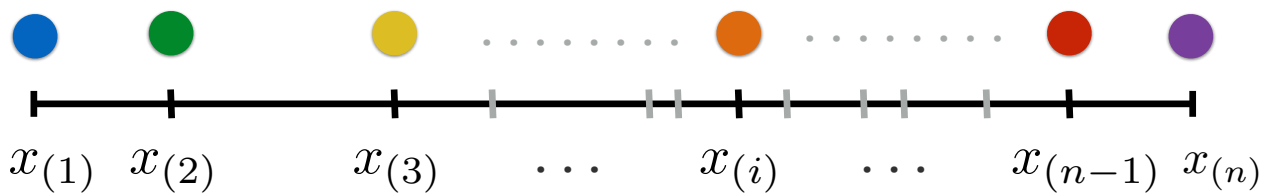
differenza
media

media delle $n \times (n - 1)$ distanze

senza
ripetizione

media delle $n \times n$ distanze
(considerando anche quelle "superflue")

con
ripetizione



Come misuro
le distanze
(o differenze)?

Proprietà desiderabili

non negatività

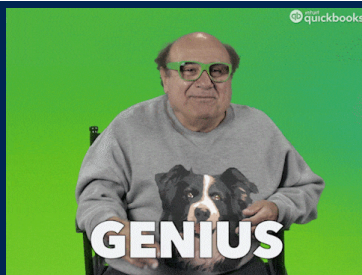
$$d(x_i, x_j) \geq 0$$

simmetria

$$d(x_i, x_j) = d(x_j, x_i)$$

riflessività

$$d(x_i, x_i) = 0$$



Come misuro
le distanze
(o differenze) ?

Variabilità come disuguaglianza

differenze semplici

$$d(x_i, x_j) = |x_i - x_j|$$

differenze quadratiche

$$d(x_i, x_j) = (x_i - x_j)^2$$

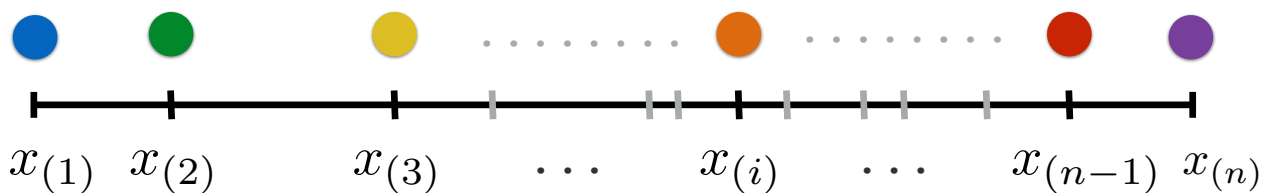
Variabilità come disuguaglianza

media delle $n \times (n - 1)$ distanze

differenza
semplice media

$$\Delta = \frac{\sum_{i=1}^n \sum_{j \neq i} |x_i - x_j|}{n(n-1)}$$

senza
ripetizione



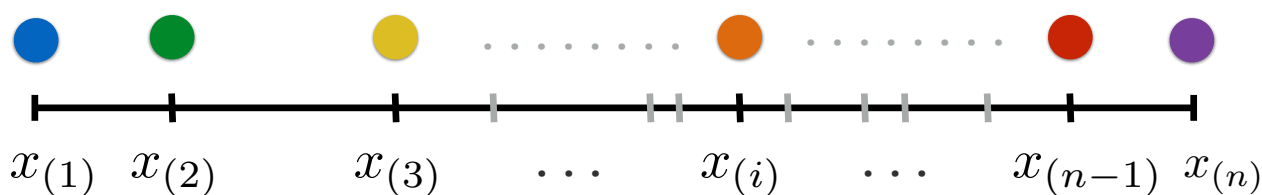
Variabilità come disuguaglianza

media delle $n \times n$ distanze

differenza
semplice media

$$\Delta = \frac{\sum_{i=1}^n \sum_{j=1}^n |x_i - x_j|}{n^2}$$

con
ripetizione



Variabilità come disuguaglianza

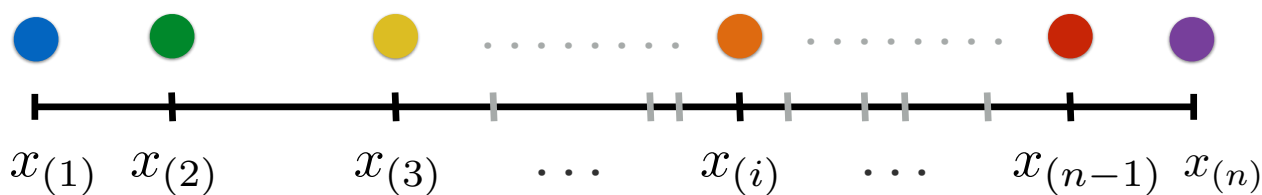
media delle $n \times (n - 1)$ distanze

differenza
quadratica media

$$(\Delta_2)^2 = \frac{\sum_{i=1}^n \sum_{j=1}^n (x_i - x_j)^2}{n(n-1)}$$

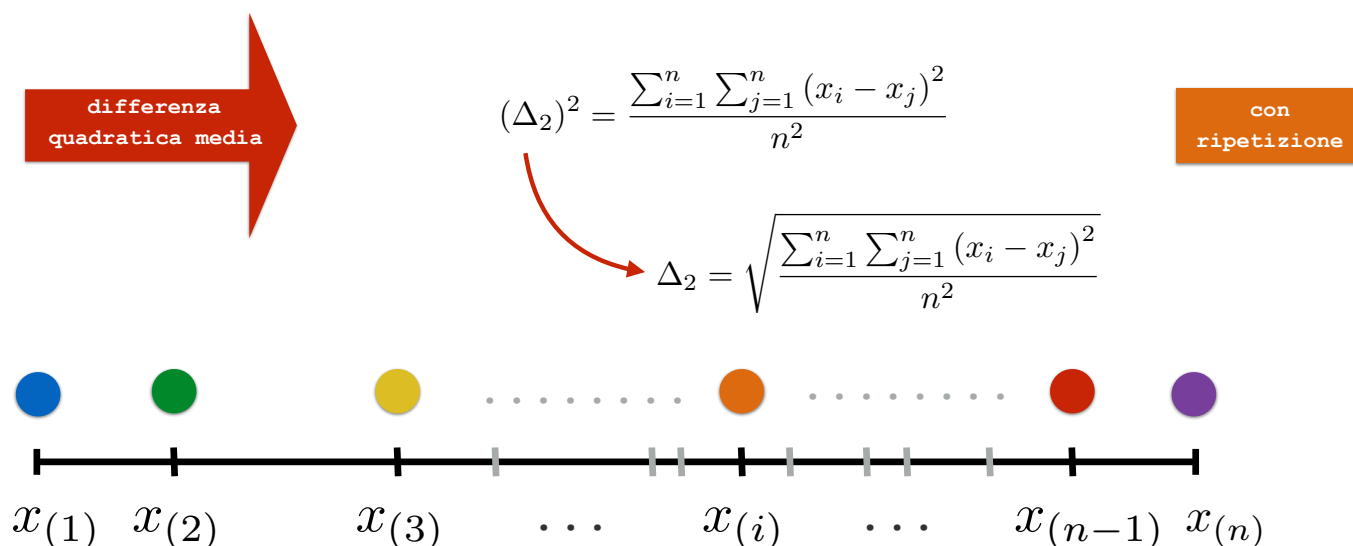
senza
ripetizione

$$\Delta_2 = \sqrt{\frac{\sum_{i=1}^n \sum_{j=1}^n (x_i - x_j)^2}{n(n-1)}}$$



Variabilità come disuguaglianza

media delle $n \times (n - 1)$ distanze



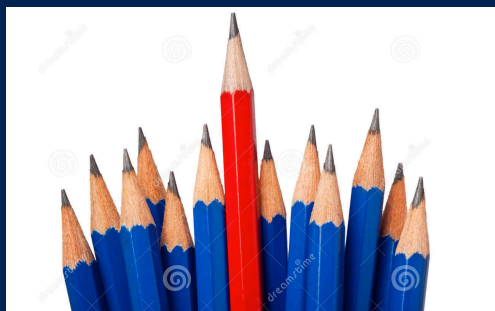
Variabilità come dispersione

		Criterio di misura della distanza	
		Valore assoluto	Quadrato
Senza ripetizione	$SR\Delta_1$	$SR\Delta_2$	
Con ripetizione	$SR\Delta_2$	$R\Delta_2$	

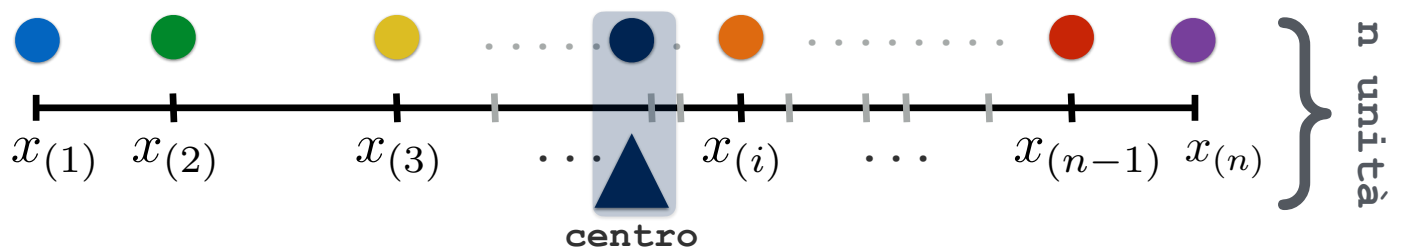
Variabilità come dispersione

Scelta comune Δ		Criterio di misura della distanza	
		Valore assoluto	Quadrato
Senza ripetizione	$SR\Delta_1$	$SR\Delta_2$	
Con ripetizione	$SR\Delta_2$	$R\Delta_2$	

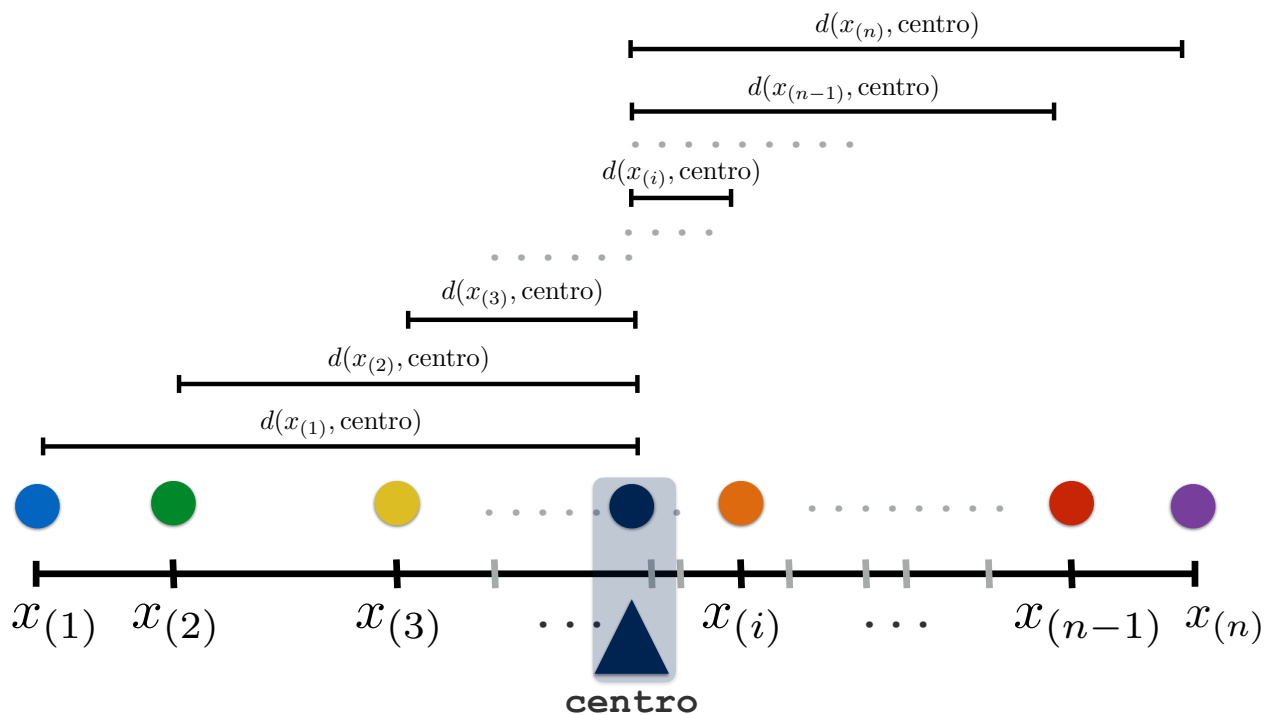
Lontani dal
"modello" tipico?



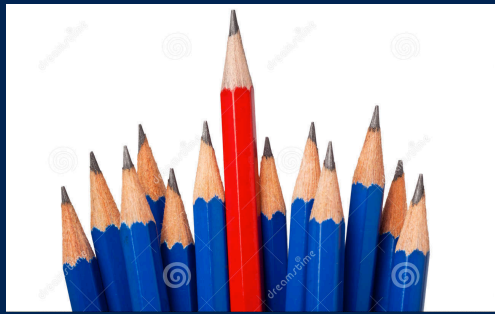
Variabilità come dispersione



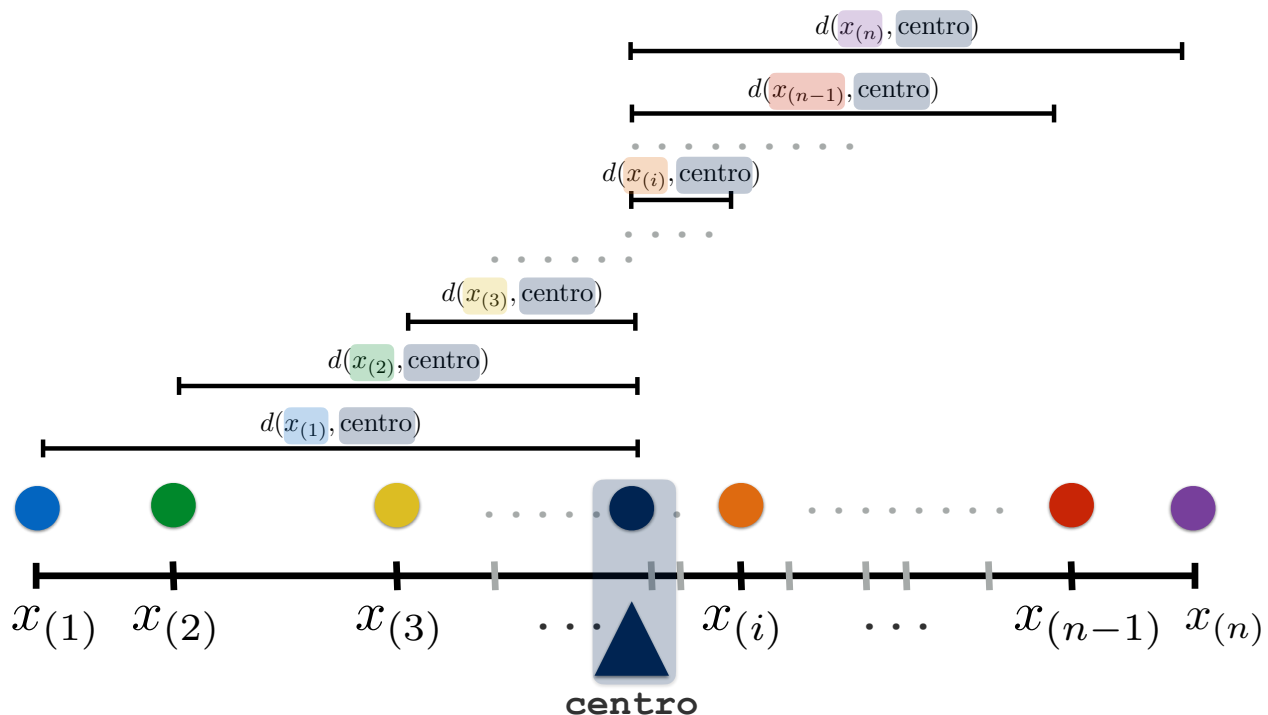
Variabilità come dispersione



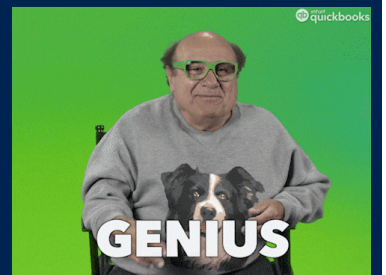
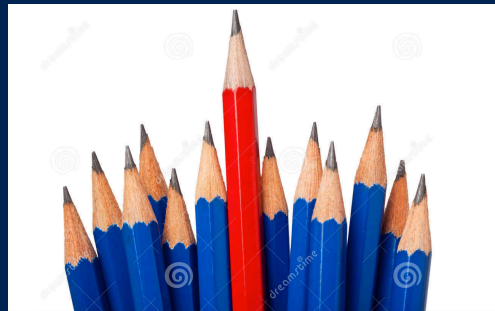
Quante distanze (o scarti o scostamenti)?



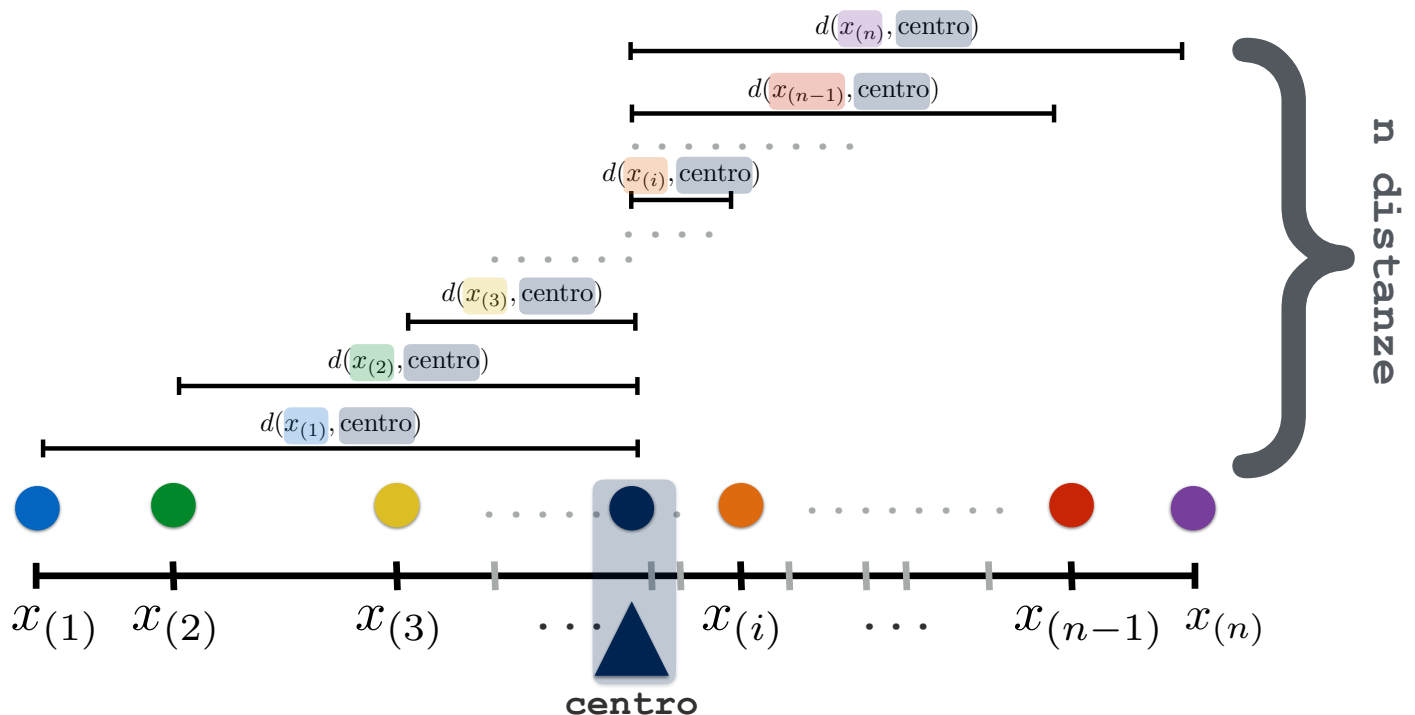
Variabilità come dispersione



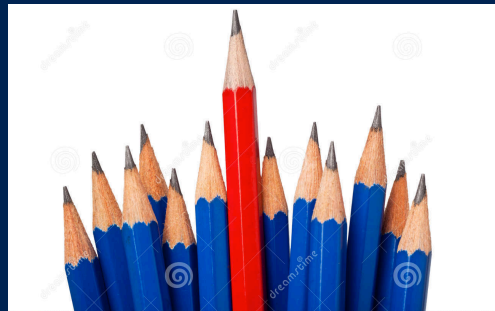
Quante distanze (o scarti o scostamenti)?



Variabilità come dispersione



Come misuro le distanze (o scarti o scostamenti)?



Variabilità come dispersione

scarti semplici

$$d(x_i, \text{centro}) = |x_i - \text{centro}|$$

scarti quadratici

$$d(x_i, \text{centro}) = (x_i - \text{centro})^2$$

Variabilità come dispersione

		Criterio di misura della distanza	
		Valore assoluto	Quadrato
Centro	Mediana	Scostamento semplice medio dalla mediana	Scostamento quadratico medio dalla mediana
	Media	Scostamento semplice medio dalla media	Scostamento quadratico medio dalla media

Variabilità come dispersione

		Criterio misura della distanza	
Criteri di ottimalità		Valore assoluto	Quadrato
Centro	Mediana	Scostamento semplice medio dalla mediana	Scostamento quadratico medio dalla mediana
	Media	Scostamento semplice medio dalla media	Scostamento quadratico medio dalla media

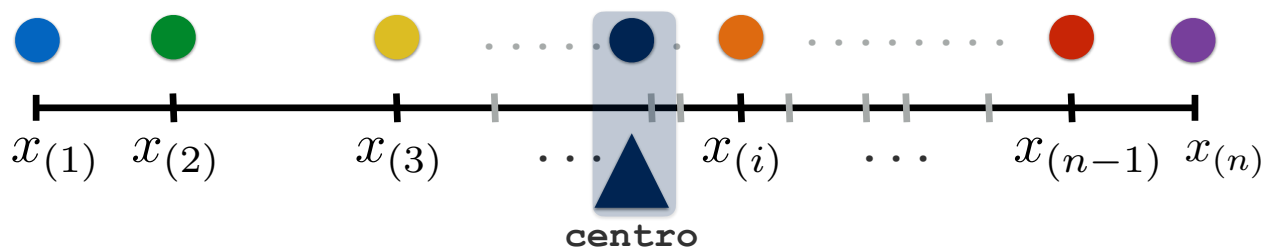
Variabilità come dispersione

media degli n scarti semplici

scarto
semplice medio

$$S_{Me} = \frac{\sum_{i=1}^n |x_i - Me|}{n}$$

dalla mediana



Variabilità come dispersione

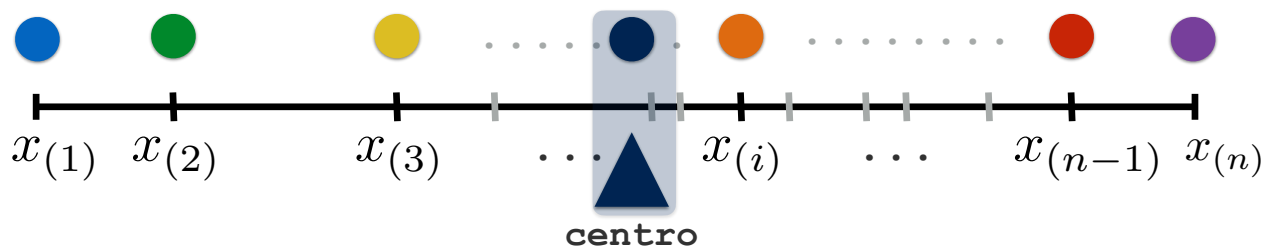
media degli n scarti semplici

scarto
quadratico medio

$$2S_{\bar{x}}^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n}$$

dalla media

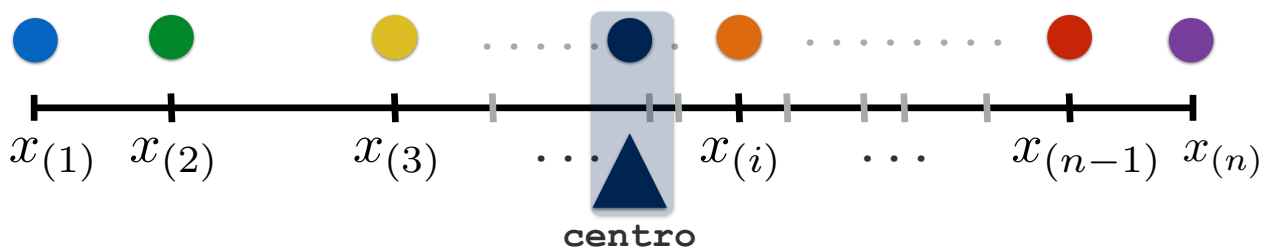
$$2S_{\bar{x}} = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n}}$$



Variabilità come dispersione

$$2S_{\bar{x}} = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n}}$$

varianza

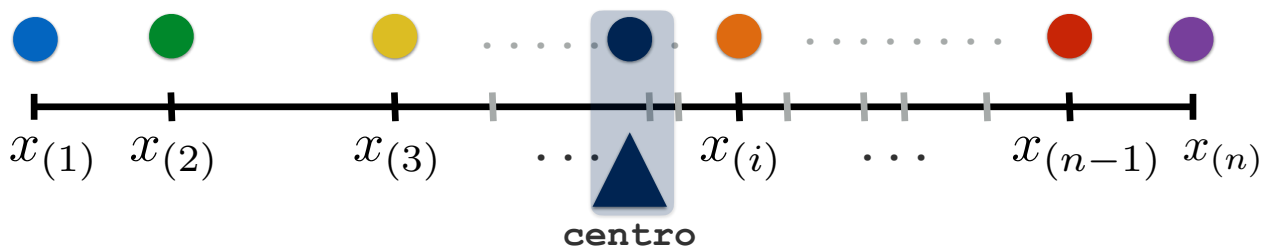


Variabilità come dispersione

$$2S_{\bar{x}} = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n}}$$

devianza

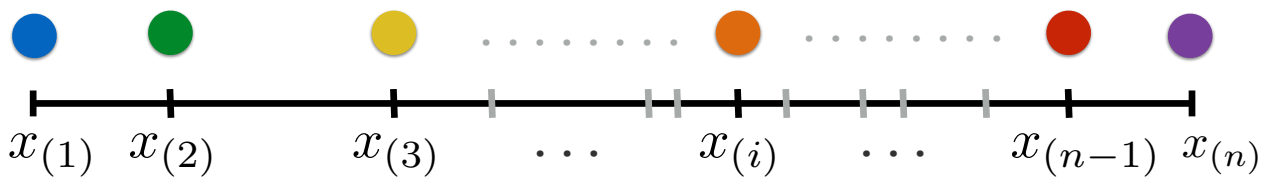
per alcuni "tecnicismi" si
rimanda alla prossima lezione



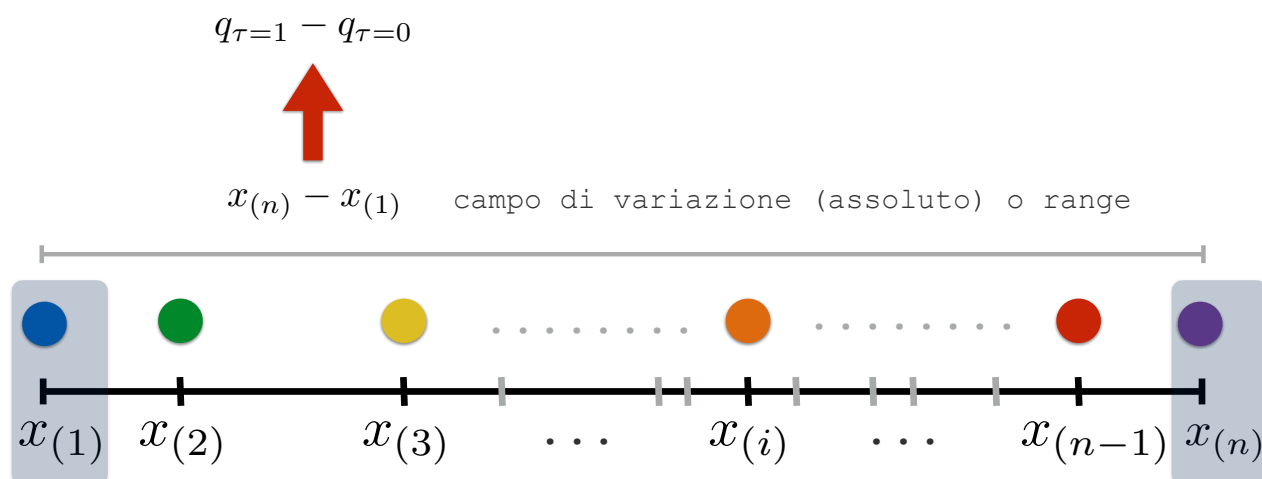
Quanto spazio c'è a disposizione?



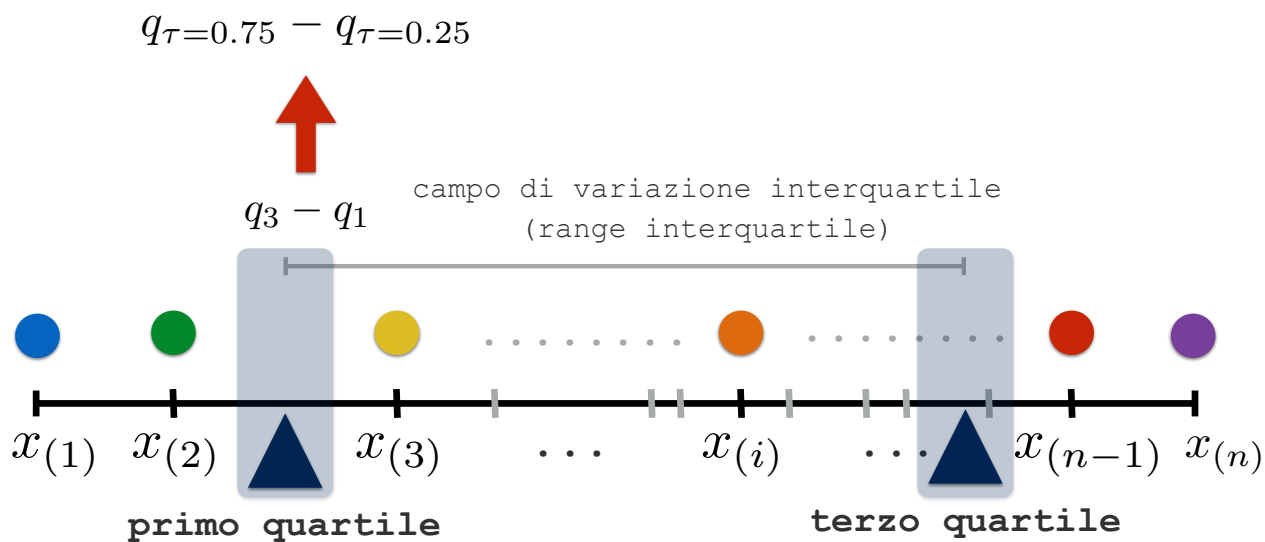
Variabilità come "range"



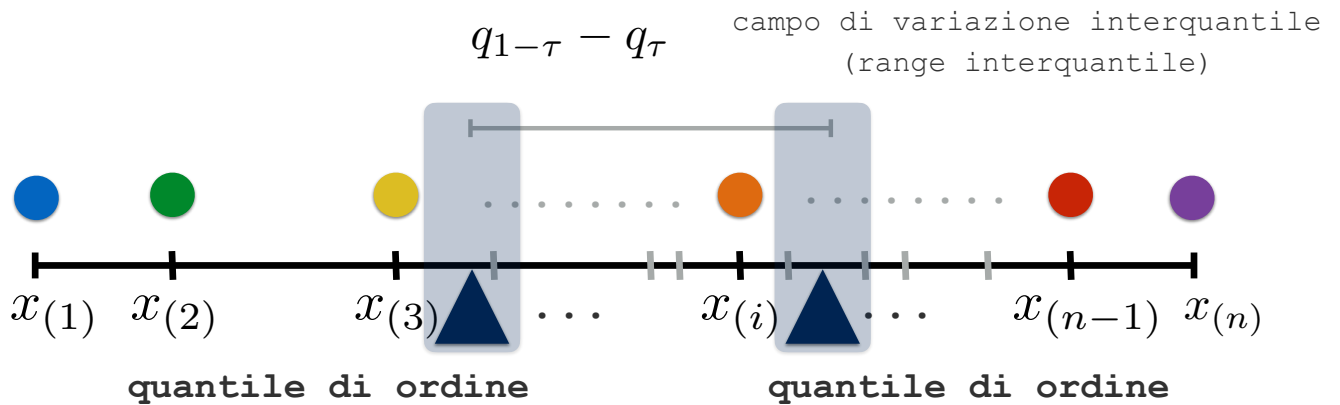
Variabilità come "range"



Variabilità come "range"



Variabilità come "range"



L'angolo dei
"tecnicismi"

La differenza semplice media

Serie grezza:

$$\Delta = \frac{\sum_{i=1}^n \sum_{j \neq i} |x_i - x_j|}{n(n-1)}$$

Tabella di frequenza
per singole modalità:

$$\Delta = \frac{\sum_{i=1}^k \sum_{j \neq i} |x_i - x_j| n_i n_j}{n}$$

Tabella di frequenza
per classi di modalità:

$$\Delta = \frac{\sum_{i=1}^k \sum_{j \neq i} |c_i - c_j| n_i n_j}{n}$$

$$c_i = \hat{x}_i = \frac{x_{i-1} + x_i}{2}$$

Lo scostamento semplice medio

Serie grezza:

$${}_1S_{Me} = \frac{\sum_{i=1}^n |x_i - Me|}{n} = S_{Me}$$

Tabella di frequenza
per singole modalità:

$${}_1S_{Me} = \frac{\sum_{i=1}^k |x_i - Me| n_i}{n}$$

Tabella di frequenza
per classi di modalità:

$${}_1S_{Me} = \frac{\sum_{i=1}^k |c_i - Me| n_i}{n}$$

$$c_i = \hat{x}_i = \frac{x_{i-1} + x_i}{2}$$

Lo scostamento quadratico medio

Serie grezza:

$${}_2S_{\bar{x}} = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n}}$$

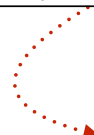
Tabella di frequenza
per singole modalità:

$${}_2S_{\bar{x}} = \sqrt{\frac{\sum_{i=1}^k (x_i - \bar{x})^2 n_i}{n}}$$

Tabella di frequenza
per classi di modalità:

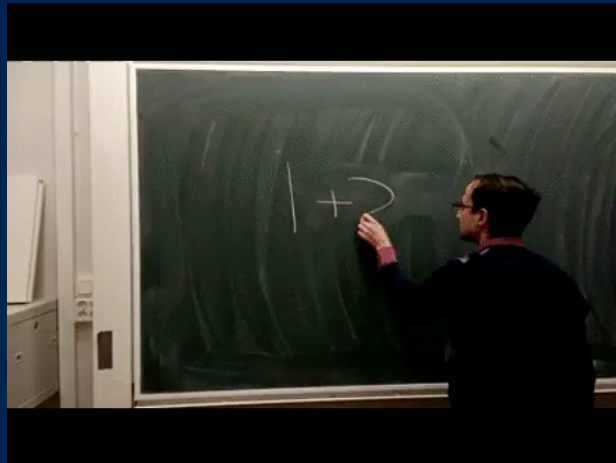
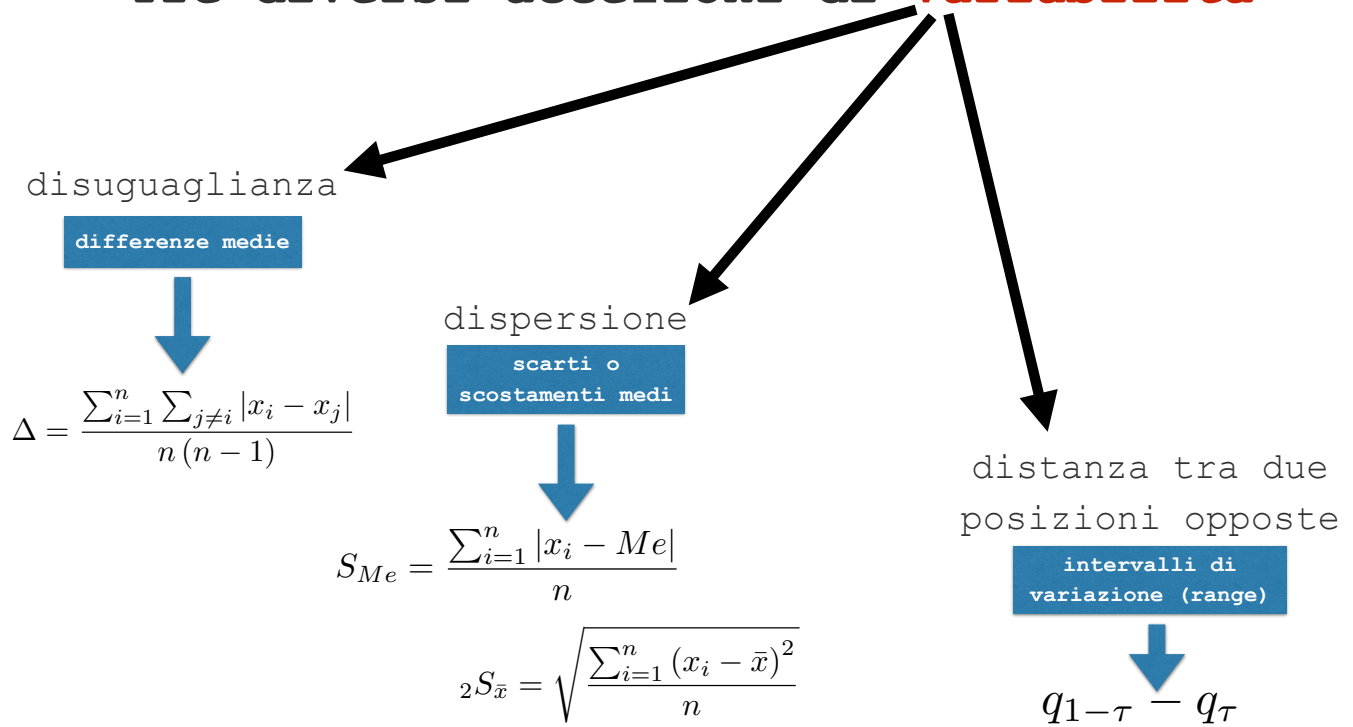
$${}_2S_{\bar{x}} = \sqrt{\frac{\sum_{i=1}^k (c_i - \bar{x})^2 n_i}{n}}$$

per alcuni "tecnicismi" si
rimanda alla prossima lezione


$$c_i = \hat{x}_i = \frac{x_{i-1} + x_i}{2}$$

Tiriamo le somme?

Tre diversi accezioni di **variabilità**



Si riparte?

"Trucchi" di calcolo...

- Suggerimenti di calcolo per le differenze semplice medie (caso di serie grezze)
- Suggerimenti di calcolo per le differenze semplice medie (caso di tabelle di frequenza)
- Suggerimenti di calcolo per gli scostamenti medi (caso di serie grezze)
- Suggerimenti di calcolo per gli scostamenti medi (caso di tabelle di frequenza)

Un indicatore robusto...



Un indicatore robusto di variabilità

$$S_{Me} = \frac{\sum_{i=1}^n |x_i - Me|}{n}$$

la "tirannia" della media

mediana come centro di riferimento

scarti in valore assoluto (criterio di ottimalità)

Un indicatore robusto di variabilità



mediana
"per ogni dove"

$$S_{Me} = \frac{\sum_{i=1}^n |x_i - Me|}{n}$$

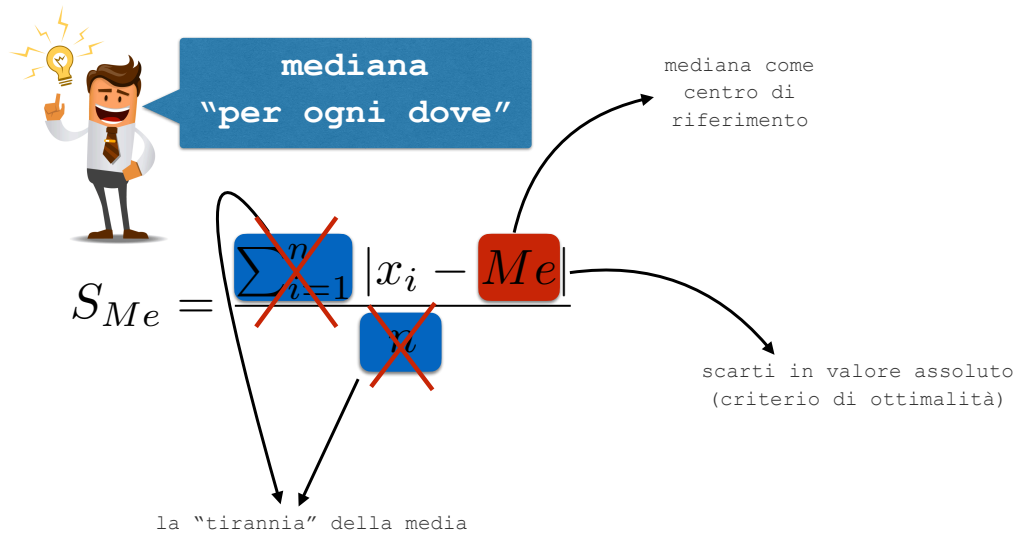
la "tirannia" della media

mediana come centro di riferimento

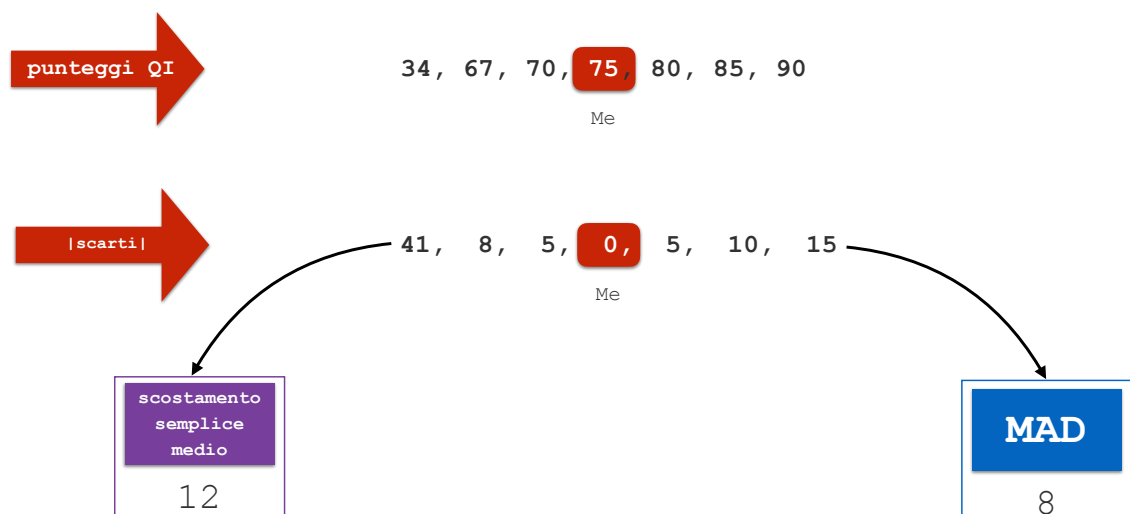
scarti in valore assoluto (criterio di ottimalità)

$MAD = |x - Me|_{(\frac{n+1}{2})}$

Un indicatore robusto di variabilità



Un indicatore robusto di variabilità



E tu sei più variabile?



Variabilità "relativa"

X (peso in kg)	$\Delta(X)$	${}_1S_{Me}(X)$	${}_2S_{\bar{x}}(X)$	$MAD(X)$	$q_{1-\tau}(X) - q_{\tau}(X)$
Y (altezza in cm)	$\Delta(Y)$	${}_1S_{Me}(Y)$	${}_2S_{\bar{y}}(Y)$	$MAD(Y)$	$q_{1-\tau}(Y) - q_{\tau}(Y)$
Z (punteggio Q.I.)	$\Delta(Z)$	${}_1S_{Me}(Z)$	${}_2S_{\bar{z}}(Z)$	$MAD(Z)$	$q_{1-\tau}(Z) - q_{\tau}(Z)$
W (punteggio ansia)	$\Delta(W)$	${}_1S_{Me}(W)$	${}_2S_{\bar{w}}(W)$	$MAD(W)$	$q_{1-\tau}(W) - q_{\tau}(W)$
K (ore di studio)	$\Delta(K)$	${}_1S_{Me}(K)$	${}_2S_{\bar{k}}(K)$	$MAD(K)$	$q_{1-\tau}(K) - q_{\tau}(K)$
L (voto esame)	$\Delta(L)$	${}_1S_{Me}(L)$	${}_2S_{\bar{l}}(L)$	$MAD(L)$	$q_{1-\tau}(L) - q_{\tau}(L)$

Variabilità "relativa"

X (peso in kg)	$\Delta(X)$	${}_1S_{Me}(X)$	${}_2S_{\bar{x}}(X)$	$MAD(X)$	$q_{1-\tau}(X) - q_{\tau}(X)$
Y (altezza in cm)	$\Delta(Y)$	${}_1S_{Me}(Y)$	${}_2S_{\bar{y}}(Y)$	$MAD(Y)$	$q_{1-\tau}(Y) - q_{\tau}(Y)$
Z (punteggio Q.I.)	$\Delta(Z)$	${}_1S_{Me}(Z)$	${}_2S_{\bar{z}}(Z)$	$MAD(Z)$	$q_{1-\tau}(Z) - q_{\tau}(Z)$
W (punteggio ansia)	$\Delta(W)$	${}_1S_{Me}(W)$	${}_2S_{\bar{w}}(W)$	$MAD(W)$	$q_{1-\tau}(W) - q_{\tau}(W)$
K (ore di studio)	$\Delta(K)$	${}_1S_{Me}(K)$	${}_2S_{\bar{k}}(K)$	$MAD(K)$	$q_{1-\tau}(K)$
L (voto esame)	$\Delta(L)$	${}_1S_{Me}(L)$	${}_2S_{\bar{l}}(L)$	$MAD(L)$	$q_{1-\tau}(L)$



Soluzione



Variabilità "relativa"

Amarcord

Il rapporto tra due grandezze omogenee (stessa unità di misura) è un numero puro, adimensionale (privo di unità di misura)

esempio

Il rapporto tra il prezzo libro del cellulare aiCel (290€) e il prezzo del libro di Statistica (29€) è uguale a **10**



un cellulare aiCel (290€) costa come **10** libri di Statistica (29€)



le due unità di misura dell'antecedente e del conseguente si sono "semplificate"

Variabilità "relativa"

Amarcord

Il rapporto tra due grandezze omogenee (stessa unità adimensionale (privo di unità di misura)

Cosa scegliereste?



esempio

Il rapporto tra il prezzo libro del cellulare aiCel (290€) e il prezzo del libro di Statistica (29€) è uguale a **10**



un cellulare aiCel (290€) costa come **10** libri di Statistica (29€)



le due unità di misura dell'antecedente e del conseguente si sono "semplificate"

Coefficienti di variazione

Variabilità relativa: i coefficienti di variazione

X (peso in kg)	$\frac{\Delta(X)}{2\bar{x}}$	$\frac{{}_1S_{Me}(X)}{Me(X)}$	$\frac{{}_2S_{\bar{x}}(X)}{\bar{x}}$	$\frac{MAD(X)}{Me(X)}$	$\frac{q_{1-\tau}(X) - q_{\tau}(X)}{Me(X)}$
Y (altezza in cm)	$\frac{\Delta(Y)}{2\bar{y}}$	$\frac{{}_1S_{Me}(Y)}{Me(Y)}$	$\frac{{}_2S_{\bar{y}}(Y)}{\bar{y}}$	$\frac{MAD(Y)}{Me(Y)}$	$\frac{q_{1-\tau}(Y) - q_{\tau}(Y)}{Me(Y)}$
Z (punteggio Q.I.)	$\frac{\Delta(Z)}{2\bar{z}}$	$\frac{{}_1S_{Me}(Z)}{Me(Z)}$	$\frac{{}_2S_{\bar{z}}(Z)}{\bar{z}}$	$\frac{MAD(Z)}{Me(Z)}$	$\frac{q_{1-\tau}(Z) - q_{\tau}(Z)}{Me(Z)}$
W (punteggio ansia)	$\frac{\Delta(W)}{2\bar{w}}$	$\frac{{}_1S_{Me}(W)}{Me(W)}$	$\frac{{}_2S_{\bar{w}}(W)}{\bar{w}}$	$\frac{MAD(W)}{Me(W)}$	$\frac{q_{1-\tau}(W) - q_{\tau}(W)}{Me(W)}$
K (ore di studio)	$\frac{\Delta(K)}{2\bar{k}}$	$\frac{{}_1S_{Me}(K)}{Me(K)}$	$\frac{{}_2S_{\bar{k}}(K)}{\bar{k}}$	$\frac{MAD(K)}{Me(K)}$	$\frac{q_{1-\tau}(K) - q_{\tau}(K)}{Me(K)}$
L (voto esame)	$\frac{\Delta(L)}{2\bar{l}}$	$\frac{{}_1S_{Me}(L)}{Me(L)}$	$\frac{{}_2S_{\bar{l}}(L)}{\bar{l}}$	$\frac{MAD(L)}{Me(L)}$	$\frac{q_{1-\tau}(L) - q_{\tau}(L)}{Me(L)}$

Variabilità relativa: i coefficienti di variazione

X (peso in kg)	$\frac{\Delta(X)}{2\bar{x}}$	$\frac{{}_1S_{Me}(X)}{Me(X)}$	$\frac{{}_2S_{\bar{x}}(X)}{\bar{x}}$	$\frac{MAD(X)}{Me(X)}$	$\frac{q_{1-\tau}(X) - q_{\tau}(X)}{Me(X)}$
Y (altezza in cm)	$\frac{\Delta(Y)}{2\bar{y}}$	$\frac{{}_1S_{Me}(Y)}{Me(Y)}$	$\frac{{}_2S_{\bar{y}}(Y)}{\bar{y}}$	$\frac{MAD(Y)}{Me(Y)}$	$\frac{q_{1-\tau}(Y) - q_{\tau}(Y)}{Me(Y)}$
Z (punteggio Q.I.)	$\frac{\Delta(Z)}{2\bar{z}}$	$\frac{{}_1S_{Me}(Z)}{Me(Z)}$	$\frac{{}_2S_{\bar{z}}(Z)}{\bar{z}}$	$\frac{MAD(Z)}{Me(Z)}$	$\frac{q_{1-\tau}(Z) - q_{\tau}(Z)}{Me(Z)}$
W (punteggio ansia)	$\frac{\Delta(W)}{2\bar{w}}$	$\frac{{}_1S_{Me}(W)}{Me(W)}$	$\frac{{}_2S_{\bar{w}}(W)}{\bar{w}}$	$\frac{MAD(W)}{Me(W)}$	$\frac{q_{1-\tau}(W) - q_{\tau}(W)}{Me(W)}$
K (ore di studio)	$\frac{\Delta(K)}{2\bar{k}}$	$\frac{{}_1S_{Me}(K)}{Me(K)}$	$\frac{{}_2S_{\bar{k}}(K)}{\bar{k}}$	$\frac{MAD(K)}{Me(K)}$	$\frac{q_{1-\tau}(K) - q_{\tau}(K)}{Me(K)}$
L (voto esame)	$\frac{\Delta(L)}{2\bar{l}}$	$\frac{{}_1S_{Me}(L)}{Me(L)}$	$\frac{{}_2S_{\bar{l}}(L)}{\bar{l}}$	$\frac{MAD(L)}{Me(L)}$	$\frac{q_{1-\tau}(L) - q_{\tau}(L)}{Me(L)}$

Variabilità relativa: i coefficienti di variazione



utili anche se le variabili sono espresse nella stessa unità di misura ma ci sono forti differenza in media

peso adulti (kg)

peso neonati (kg)

prezzo libri (€)

prezzo riviste (€)

45.50, 46.40, 46.70, 47.50, 48.00

2.0, 2.9, 3.2, 4.0, 4.5

media
46.82

sqm
0.76

=

sqm
0.76

media
3.32



0.02 (2%) < 0.23 (23%)

Questioni di omonimia? (tecnicissimi)

Lo scostamento (scarto) quadratico medio

The diagram illustrates the relationship between different measures of standard deviation. On the left, three labels are positioned: s at the top, s_X at the bottom left, and $SQM(X)$ at the bottom right. Three curved arrows originate from these labels and point towards a central red circle containing the expression $2s_{\bar{x}}$. To the right of this circle is an equals sign followed by a square root symbol. Inside the square root, the numerator is the sum of squared deviations from the mean, $\sum_{i=1}^n (x_i - \bar{x})^2$, and the denominator is the sample size, n . The entire fraction is enclosed in a light blue rectangular box.

$$2s_{\bar{x}} = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n}}$$

Lo scostamento (scarto) quadratico medio

The diagram illustrates the relationship between the standard deviation and its square. It features the formula for the sample standard deviation, $s = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n}}$. The term s is highlighted in a red circle, and the term s^2 is highlighted in a blue box. Curved arrows connect s to s^2 and s^2 to s . Labels s , s_X , and $SQM(X)$ are placed near the formula. Below the formula, the text "deviazione standard" and "standard deviation" are written, with a small circle between them.

$$s = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n}}$$

s
 s_X
 $SQM(X)$

deviazione standard
o
standard deviation

Lo scostamento (scarto) quadratico medio

The diagram illustrates the relationship between the standard deviation, its square, and the variance. It features the formula for the sample standard deviation, $s = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n}}$. The term s is highlighted in a red circle, and the term s^2 is highlighted in a blue box. Curved arrows connect s to s^2 and s^2 to s . Labels s , s_X , $SQM(X)$, s^2 , s_X^2 , and $Var(X)$ are placed near the formula. Below the formula, the text "deviazione standard" and "standard deviation" are written, with a small circle between them.

$$s = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n}}$$

s
 s_X
 $SQM(X)$
 s^2
 s_X^2
 $Var(X)$

deviazione standard
o
standard deviation

Lo scostamento (scarto) quadratico medio

$$s = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n}} = \sqrt{\frac{SQ(X)}{n}} = \sqrt{\frac{SS(X)}{n}}$$

Diagram illustrating the relationship between the standard deviation (s) and the quadratic mean deviation ($SQM(X)$), and their connection to the sum of squares ($SS(X)$) and the sum of squared deviations ($SQ(X)$).

The formula shows that the standard deviation (s) is the square root of the quadratic mean deviation ($SQM(X)$), which is the square root of the sum of squared deviations ($SQ(X)$) divided by the sample size (n). The sum of squared deviations ($SQ(X)$) is also equal to the sum of squares ($SS(X)$).

Labels in the diagram:

- s : Standard deviation
- $SQM(X)$: Quadratic mean deviation
- $SS(X)$: Sum of squares
- $SQ(X)$: Sum of squared deviations
- $Var(X)$: Variance
- s_X^2 : Sample variance
- $Dev(X)$: Deviation
- SQ_X : Sum of squared deviations (alternative notation)
- SS_X : Sum of squares (alternative notation)

Popolazione o campione?



Lo scostamento (scarto) quadratico medio

Popolazione o campione?

Scarto quadratico medio

$$s_x^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})}{n}$$

Scarto quadratico medio "corretto"
(campionario)

$$\hat{s}_x^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})}{n - 1}$$

Lo scostamento (scarto) quadratico medio

Popolazione o campione?

$$s_x^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})}{n}$$

σ_X^2
 σ^2

μ_X
 μ

N

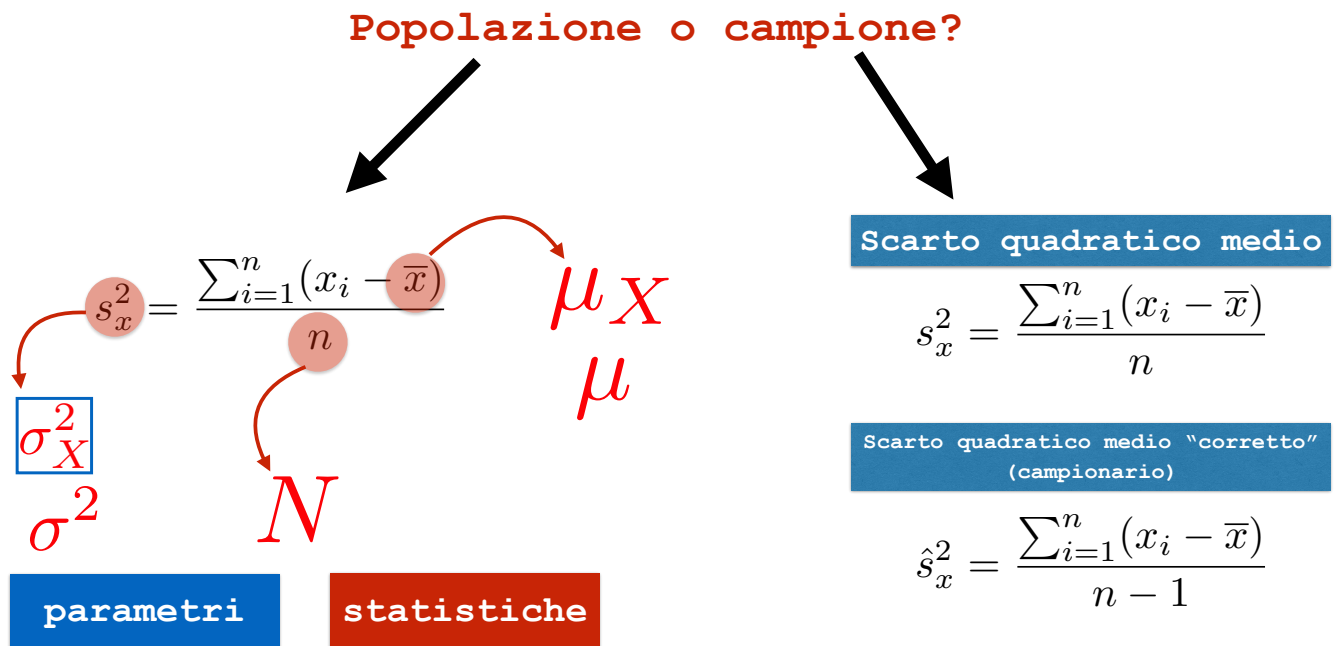
Scarto quadratico medio

$$s_x^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})}{n}$$

Scarto quadratico medio "corretto"
(campionario)

$$\hat{s}_x^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})}{n - 1}$$

Lo scostamento (scarto) quadratico medio



Proprietà desiderabili

EXPECTATION



Proprietà desiderabili di un indice di variabilità

1

non deve mai essere negativo

differenze
medie



scarti o
scostamenti medi



intervalli di
variazione (range)



2

deve annullarsi se e solo se tutte le osservazioni sono uguali (assenza variabilità)

differenze
medie



scarti o
scostamenti medi



intervalli di
variazione (range)



3

deve crescere al crescere della disuguaglianza tra i valori osservati

differenze
medie



scarti o
scostamenti medi



intervalli di
variazione (range)



Proprietà desiderabili di un indice di variabilità

2

deve annullarsi se e solo se tutte le osservazioni sono uguali (assenza variabilità)

differenze
medie



scarti o
scostamenti medi



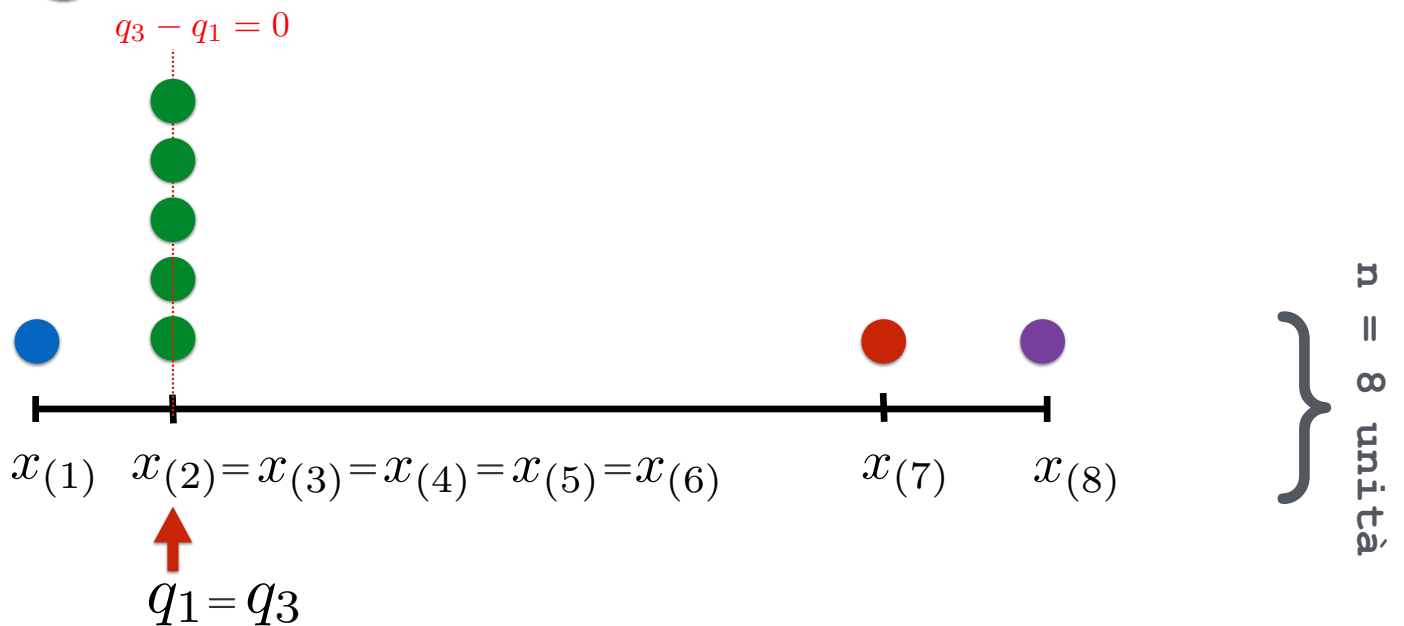
intervalli di
variazione (range)



Proprietà desiderabili di un indice di variabilità

2

deve annullarsi se e solo se tutte le osservazioni sono uguali (assenza di variabilità)



Proprietà desiderabili di un indice di variabilità

3

deve crescere al crescere della disuguaglianza tra i valori osservati

differenze
medie

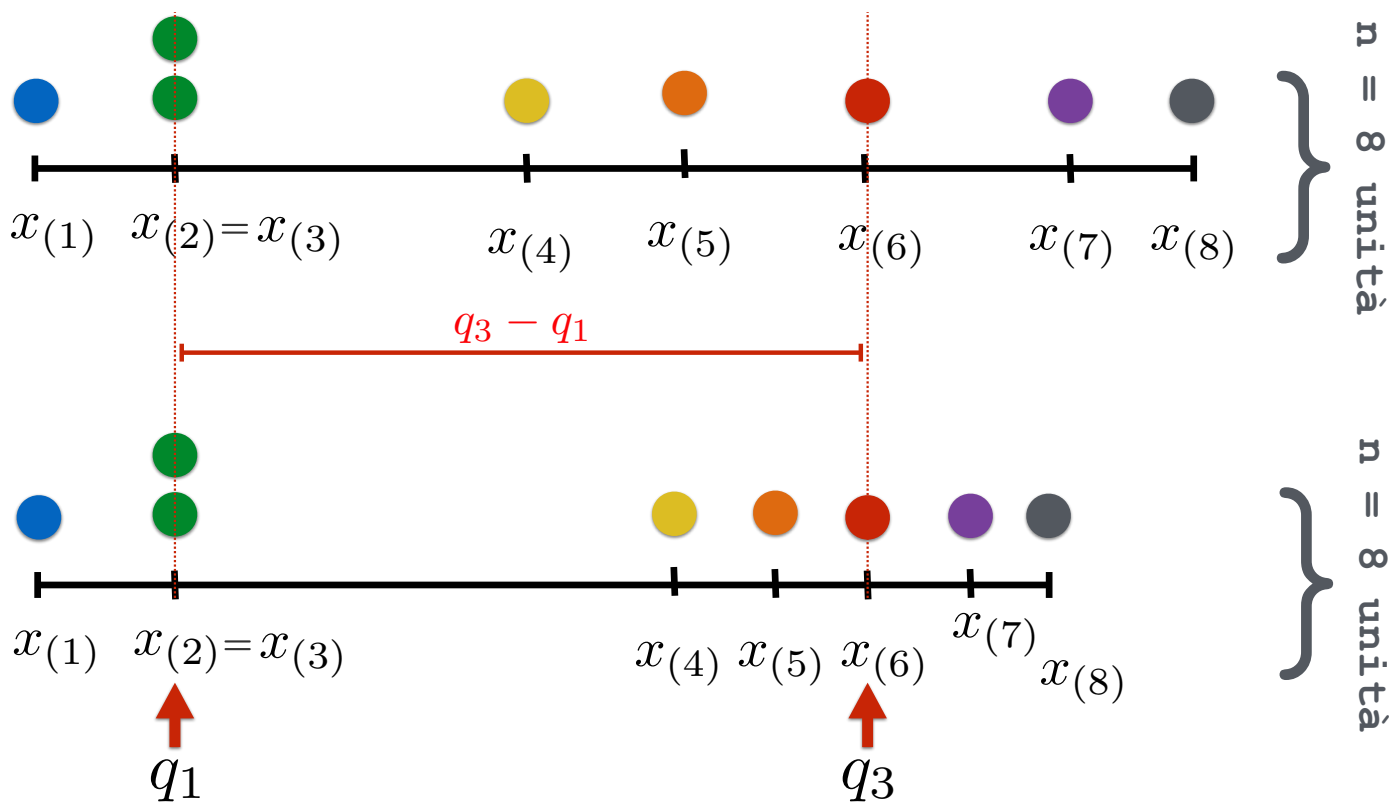
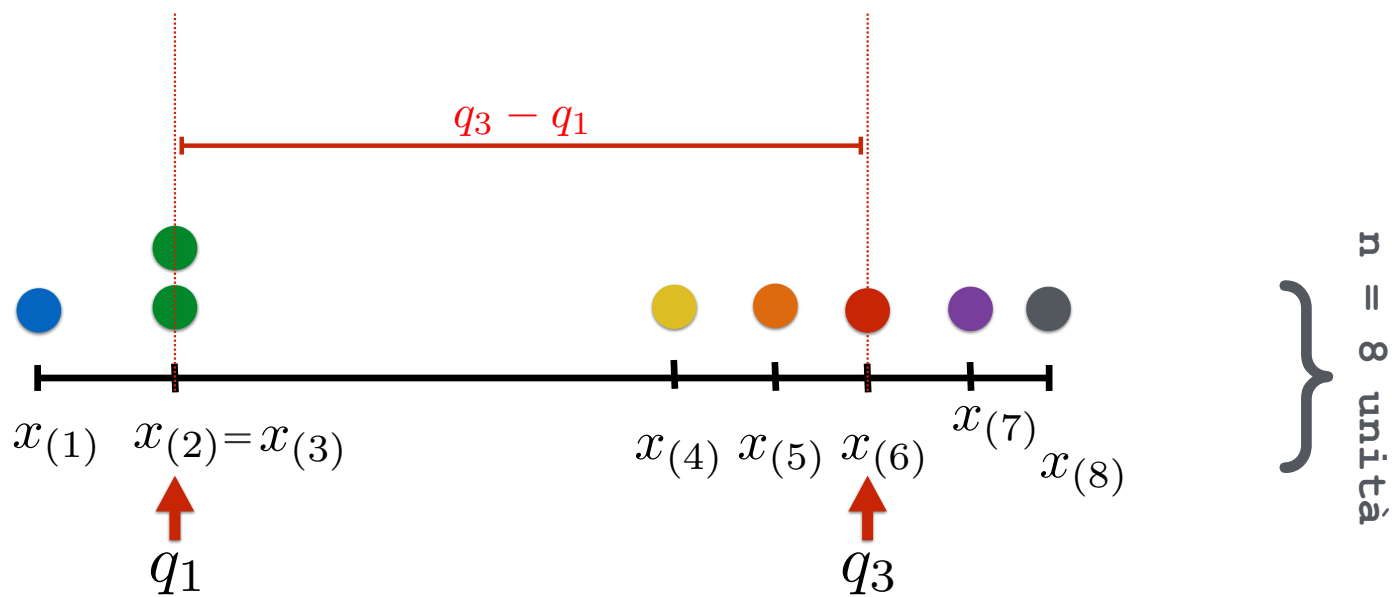


scarti o
scostamenti medi



intervalli di
variazione (range)



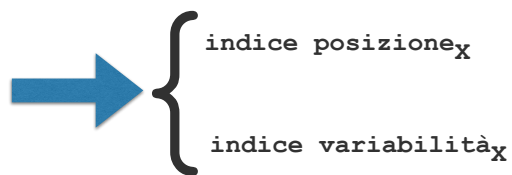


Trasformazioni lineari



Trasformazioni lineari

$$X$$
$$(x_1, \dots, x_i, \dots, x_n)$$

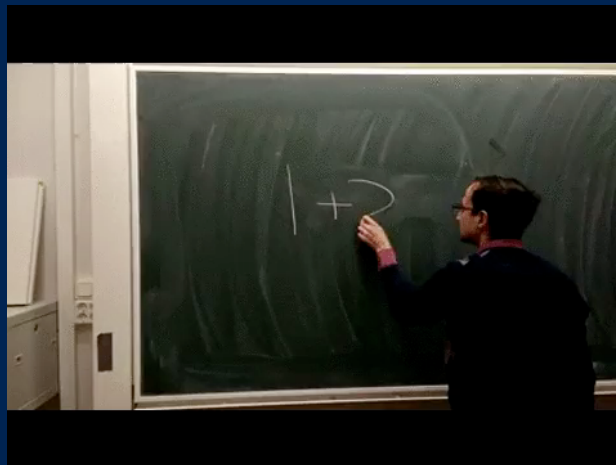
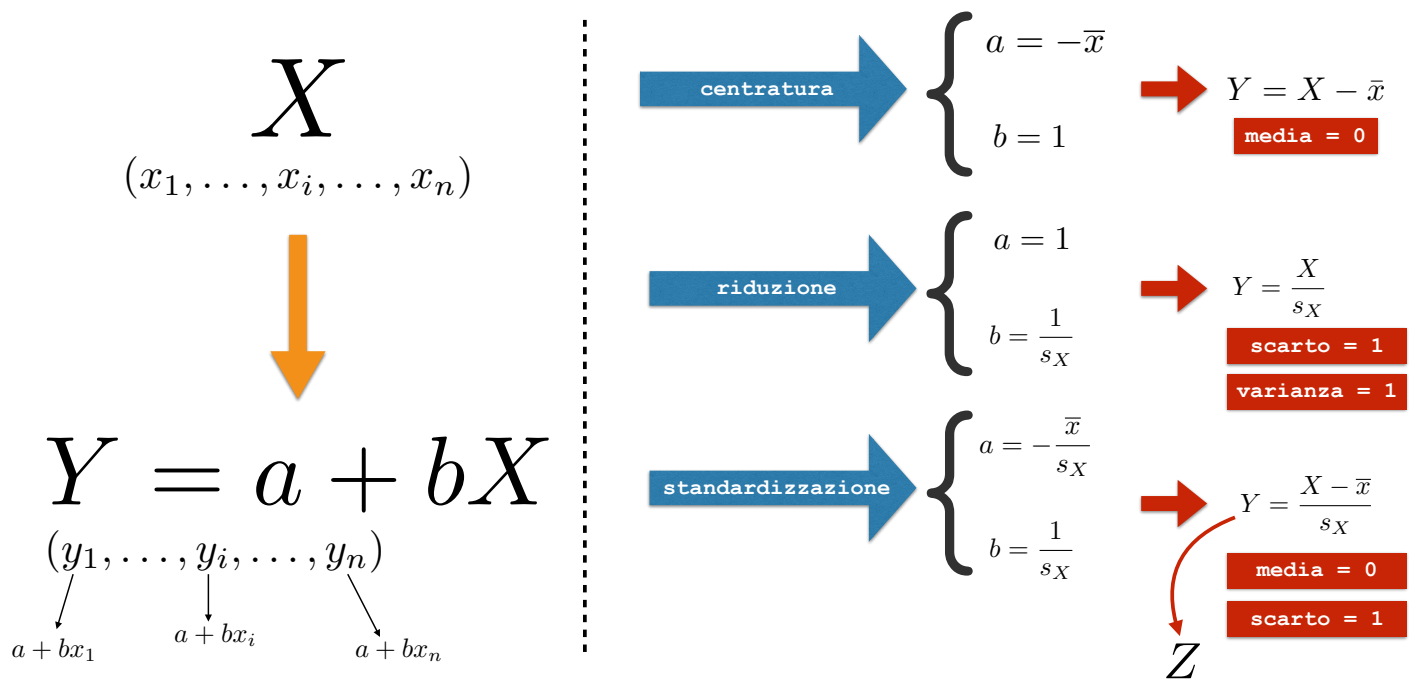


$$Y = a + bX$$
$$(y_1, \dots, y_i, \dots, y_n)$$

$a + bx_1$ $a + bx_i$ $a + bx_n$

$$\left\{ \begin{array}{l} \text{indice posizione}_Y = a + b \text{ indice posizione}_X \\ \text{indice variabilità}_Y = |b| \text{ indice variabilità}_X \end{array} \right.$$

Trasformazioni lineari



Si riparte?

Dati divisi in gruppi: esempi

- Studenti divisi in base al sesso
- Studenti divisi in base al diploma di provenienza
- Soggetti divisi al livello di istruzione
- Soggetti divisi in base al loro orientamento politico
- Pazienti divisi in base alla gravità di una data tipologia
- Pazienti divisi in base ai diversi livelli di una scala (ansia, competenze, estroversione, ecc.)

Dati divisi in gruppi: struttura

Gruppo	Unità	Cardinalità
1	$x_{1,1}, x_{1,2}, \dots, x_{1,j}, \dots, x_{1,n_1}$	n_1
2	$x_{2,1}, x_{2,2}, \dots, x_{2,j}, \dots, x_{2,n_2}$	n_2
$\vdots$	$\vdots$	$\vdots$
i	$x_{i,1}, x_{i,2}, \dots, x_{i,j}, \dots, x_{i,n_i}$	n_i
$\vdots$	$\vdots$	$\vdots$
G	$x_{G,1}, x_{G,2}, \dots, x_{G,j}, \dots, x_{G,n_G}$	n_G

$$n = \sum_{i=1}^G n_i$$

Dati divisi in gruppi: sintesi (1)

Gruppo	Unità	Cardinalità	Medie
1	$x_{1,1}, \dots, x_{1,j}, \dots, x_{1,n_1}$	n_1	$\bar{x}_1$
2	$x_{2,1}, \dots, x_{2,j}, \dots, x_{2,n_2}$	n_2	$\bar{x}_2$
$\vdots$	$\vdots$	$\vdots$	
i	$x_{i,1}, \dots, x_{i,j}, \dots, x_{i,n_i}$	n_i	$\bar{x}_i = \frac{\sum_{j=1}^{n_i} x_{i,j}}{n_i}$
$\vdots$	$\vdots$	$\vdots$	
G	$x_{G,1}, \dots, x_{G,j}, \dots, x_{G,n_G}$	n_G	$\bar{x}_G$

$$n = \sum_{i=1}^G n_i$$

Proprietà di associatività della media

La media complessiva:

$$\bar{x} = \frac{\sum_{i=1}^G \sum_{j=1}^{n_i} x_{i,j}}{n}$$

Può essere ottenuta come media ponderata delle medie di gruppo con pesi dati dalle numerosità di gruppo:

$$\bar{x} = \frac{\sum_{i=1}^G \bar{x}_i n_i}{n}$$

Dati divisi in gruppi: sintesi (2)

Gruppo	Unità	Cardinalità	Medie	Devianze
1	$x_{1,1}, \dots, x_{1,j}, \dots, x_{1,n_1}$	n_1	$\bar{x}_1$	Dev_1
2	$x_{2,1}, \dots, x_{2,j}, \dots, x_{2,n_2}$	n_2	$\bar{x}_2$	Dev_2
$\vdots$	$\vdots$	$\vdots$		
i	$x_{i,1}, \dots, x_{i,j}, \dots, x_{i,n_i}$	n_i	$\bar{x}_i$	$\text{Dev}_i = \sum_{j=1}^{n_i} (x_{i,j} - \bar{x}_i)^2$
$\vdots$	$\vdots$	$\vdots$		
G	$x_{G,1}, \dots, x_{G,j}, \dots, x_{G,n_G}$	n_G	$\bar{x}_G$	Dev_G

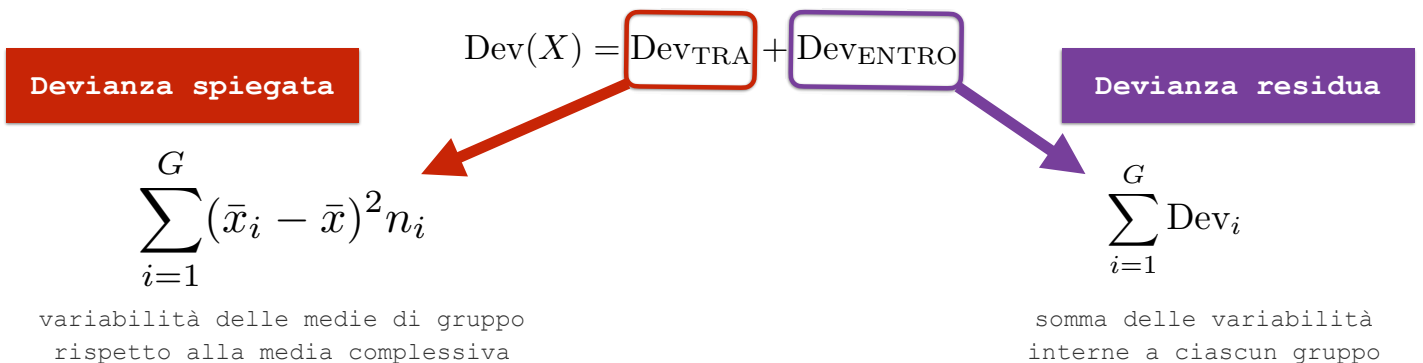
$$\overline{n = \sum_{i=1}^G n_i}$$

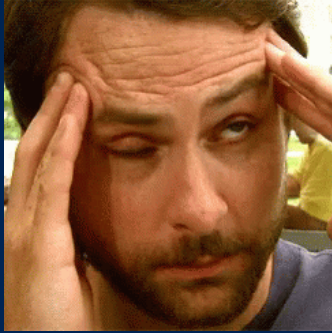
Proprietà di decomposizione della devianza

La devianza complessiva:

$$\text{Dev}(X) = \sum_{i=1}^G \sum_{j=1}^{n_j} (x_{i,j} - \bar{x})^2$$

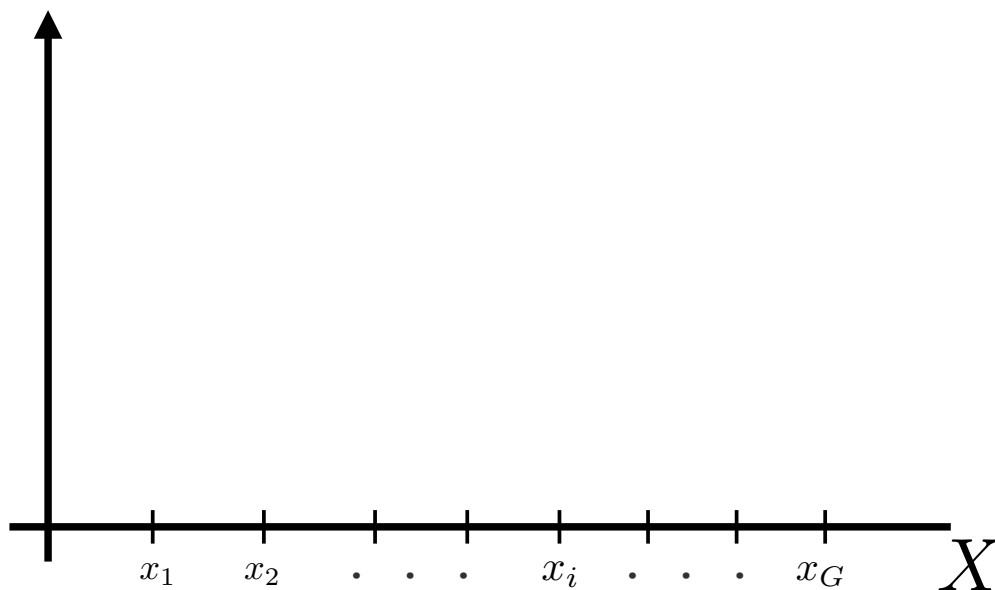
Può essere scomposta in devianza esterna (tra i gruppi o among) e devianza interna (entro i gruppi o within):



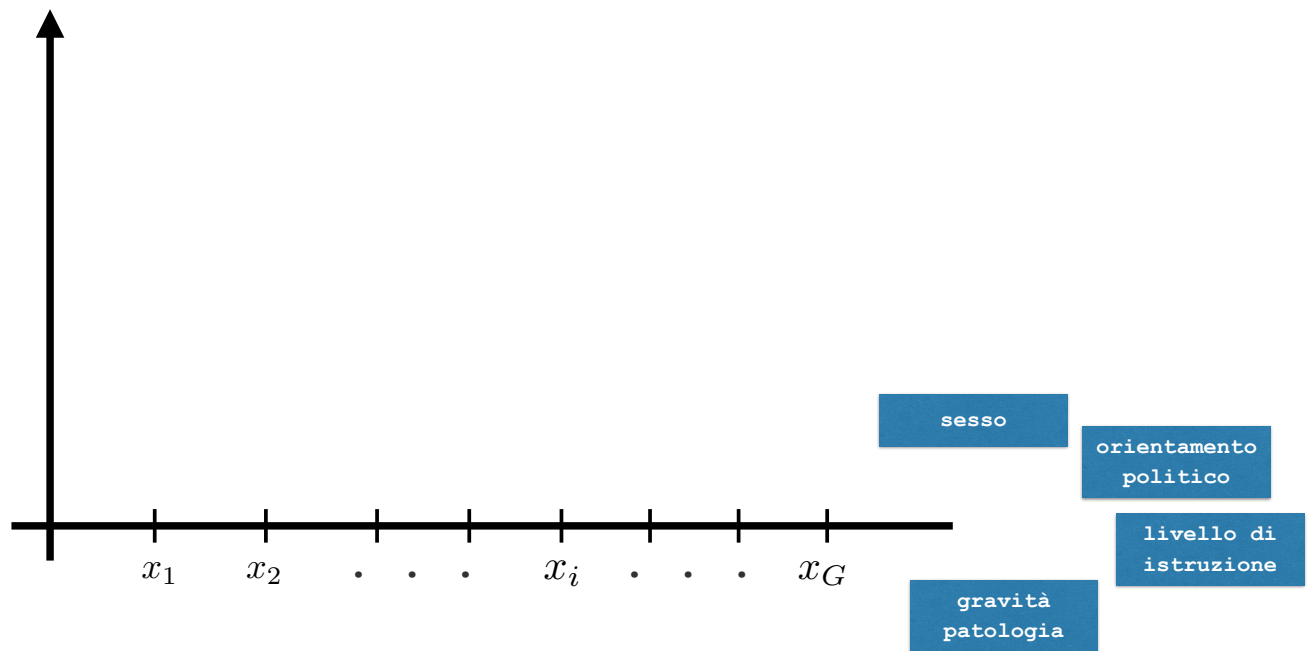


A cosa serve tutto questo?

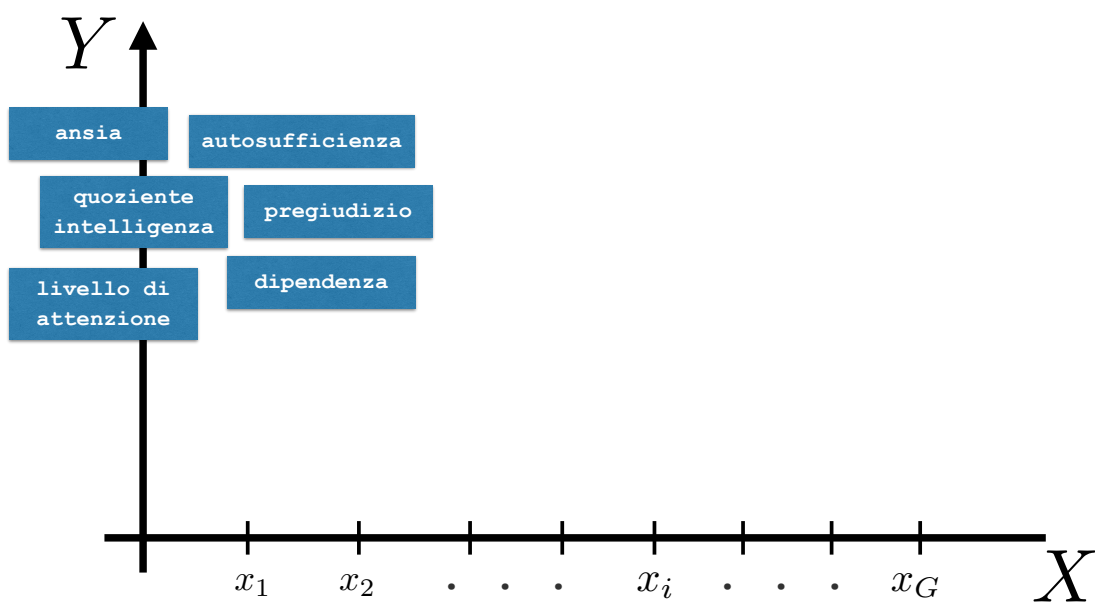
Proprietà di decomposizione della devianza



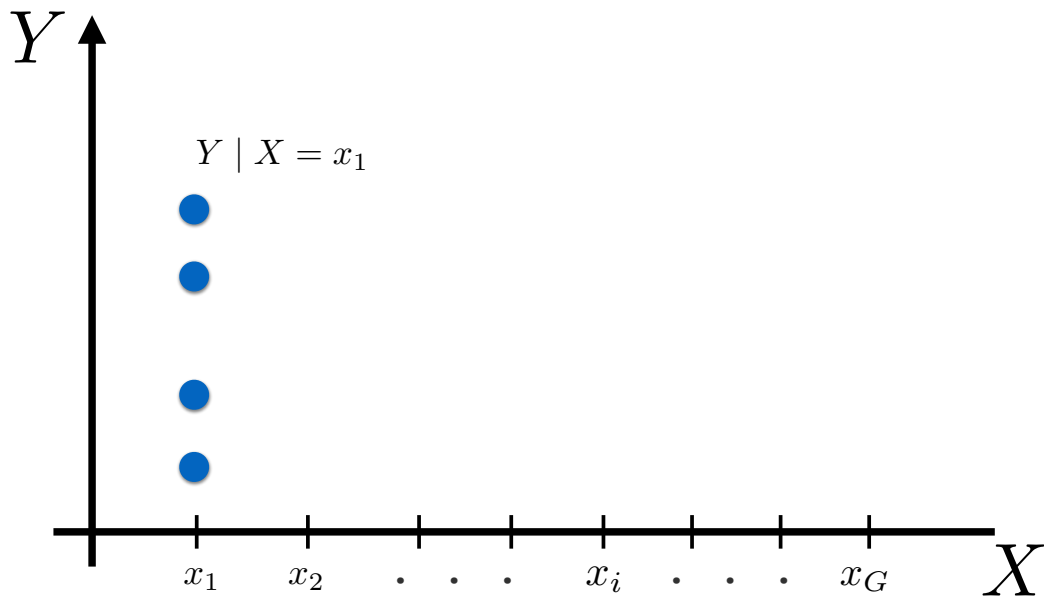
Proprietà di decomposizione della devianza



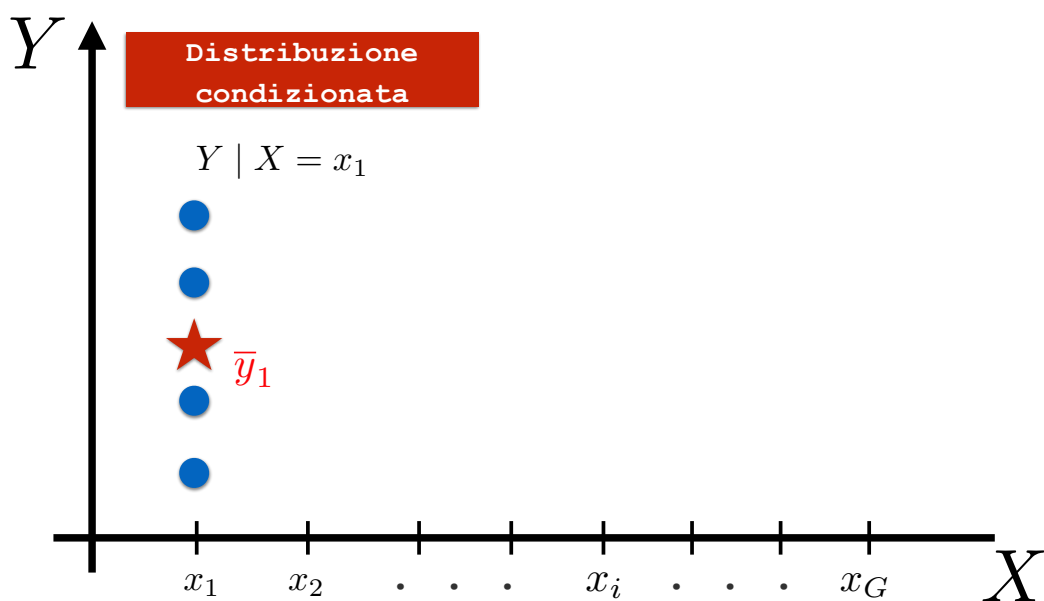
Proprietà di decomposizione della devianza



Proprietà di decomposizione della devianza



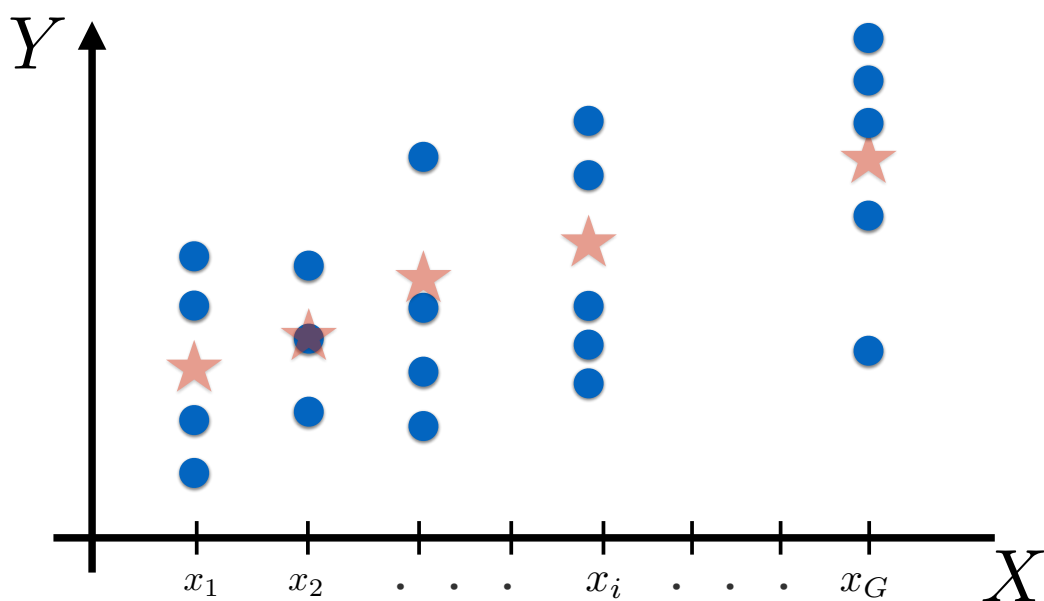
Proprietà di decomposizione della devianza



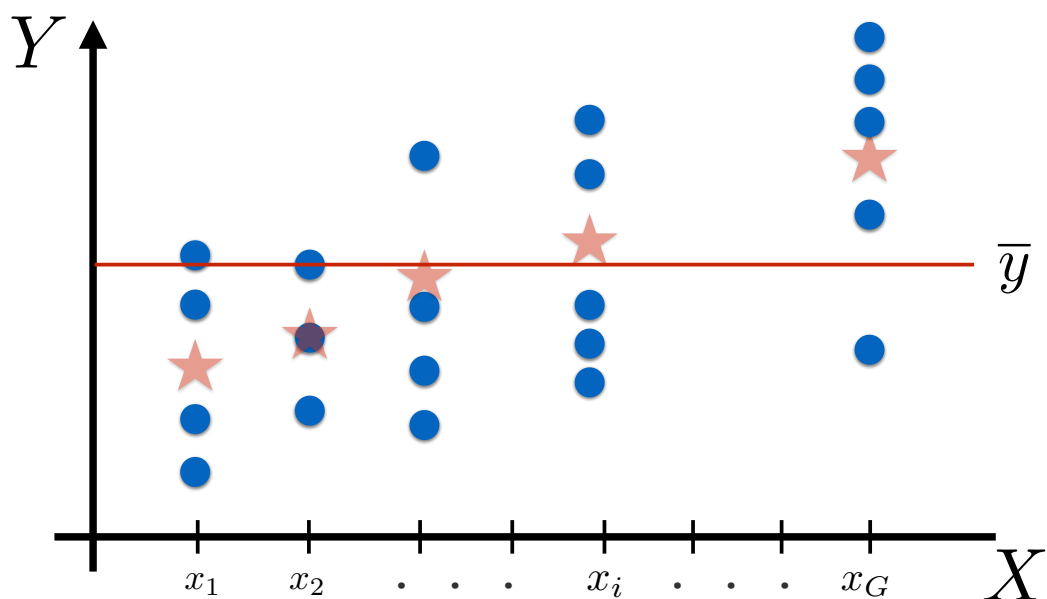
Attenzione al cambio di notazione

Cosa esprimono le due componenti della variabilità?

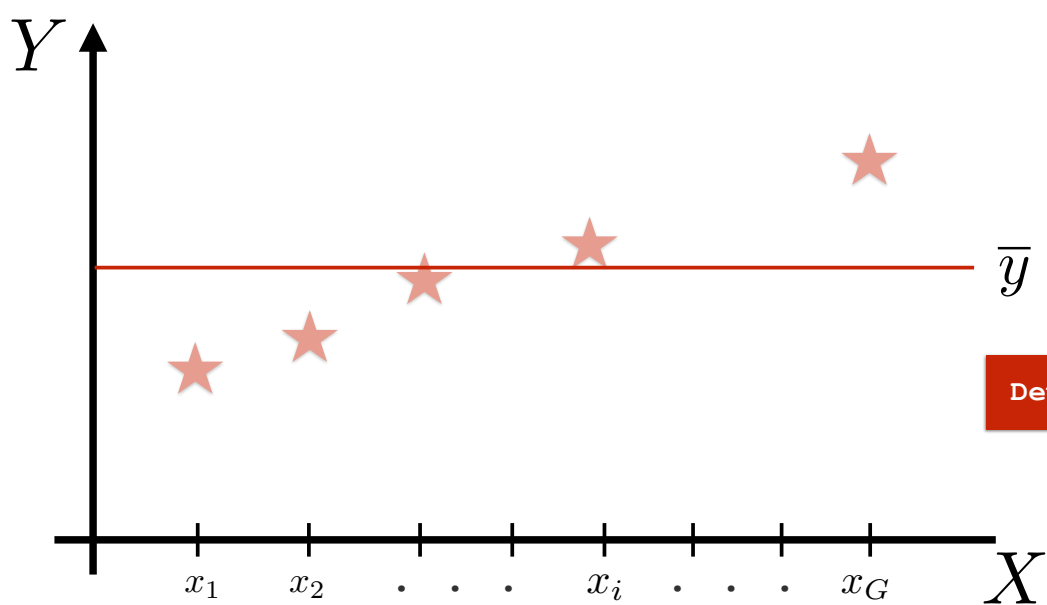
Proprietà di decomposizione della devianza



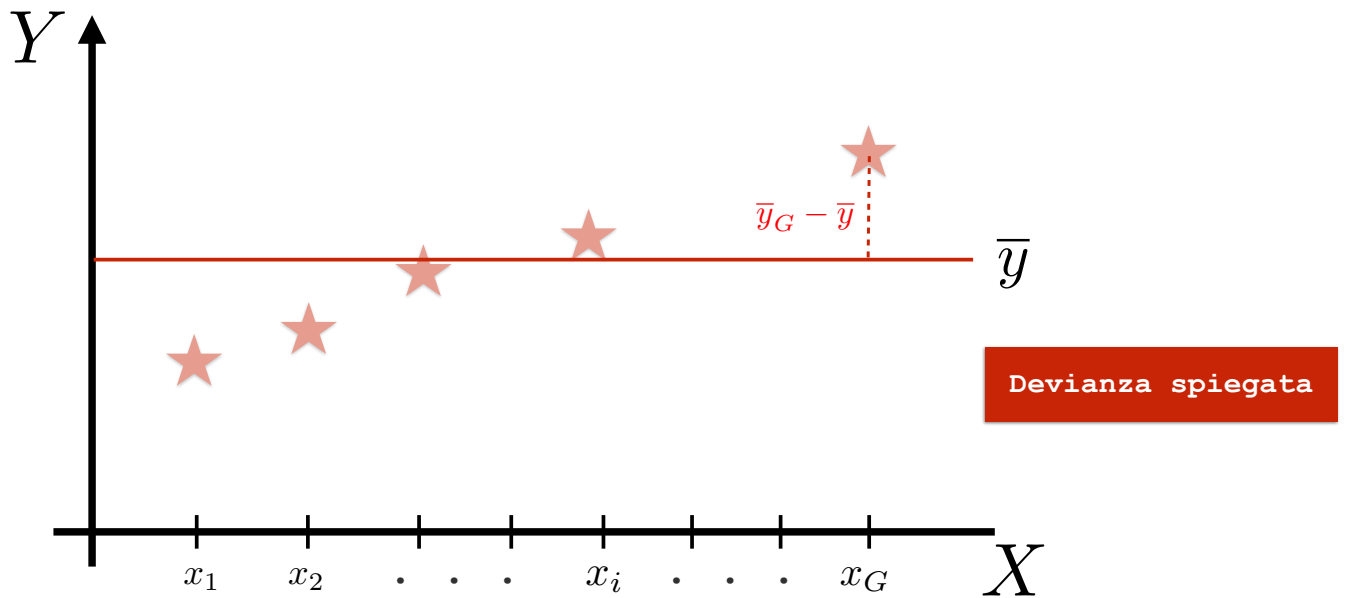
Proprietà di decomposizione della devianza



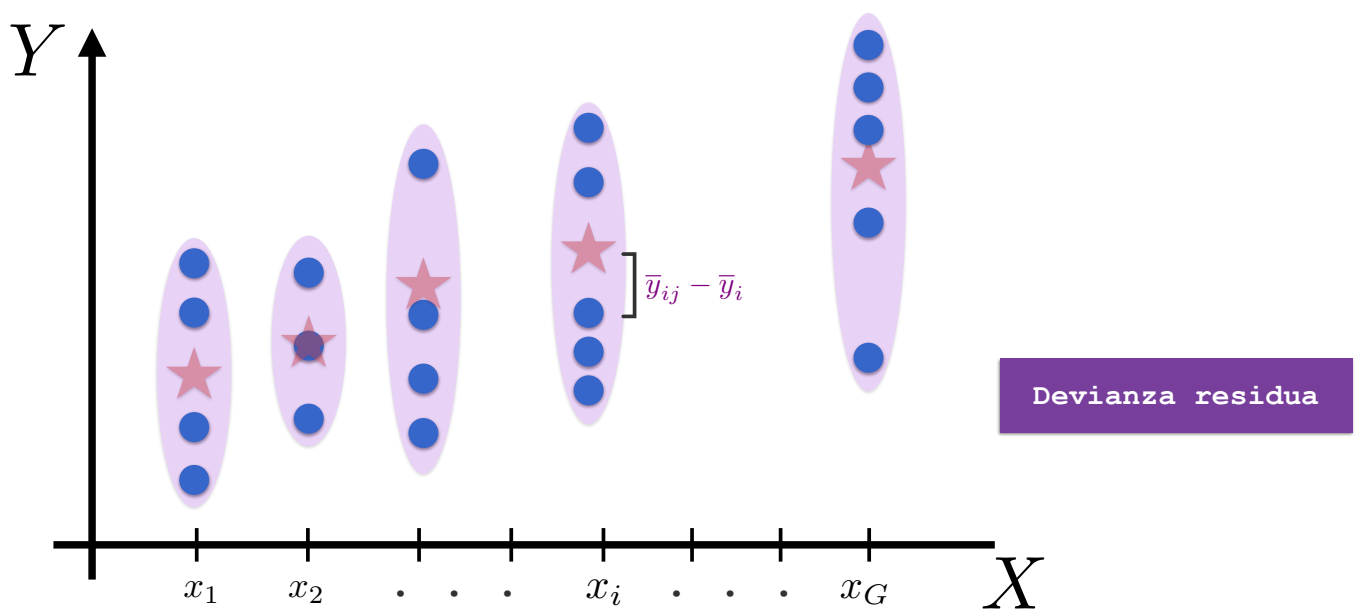
Proprietà di decomposizione della devianza



Proprietà di decomposizione della devianza

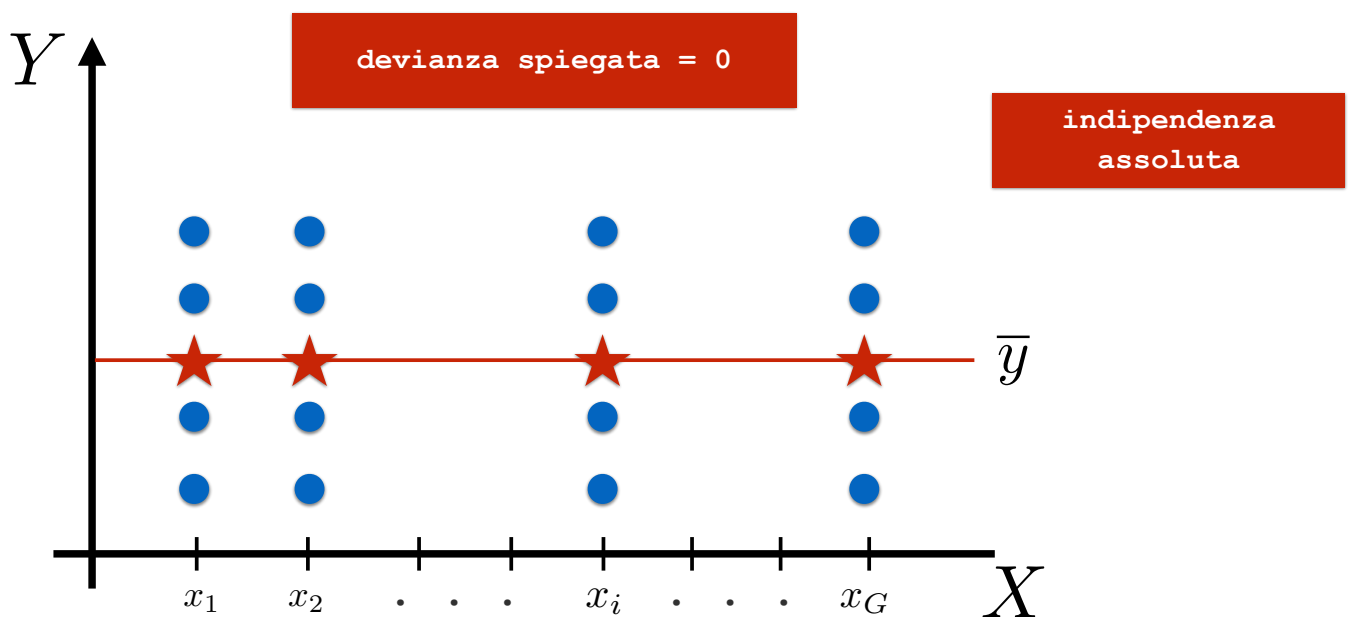


Proprietà di decomposizione della devianza

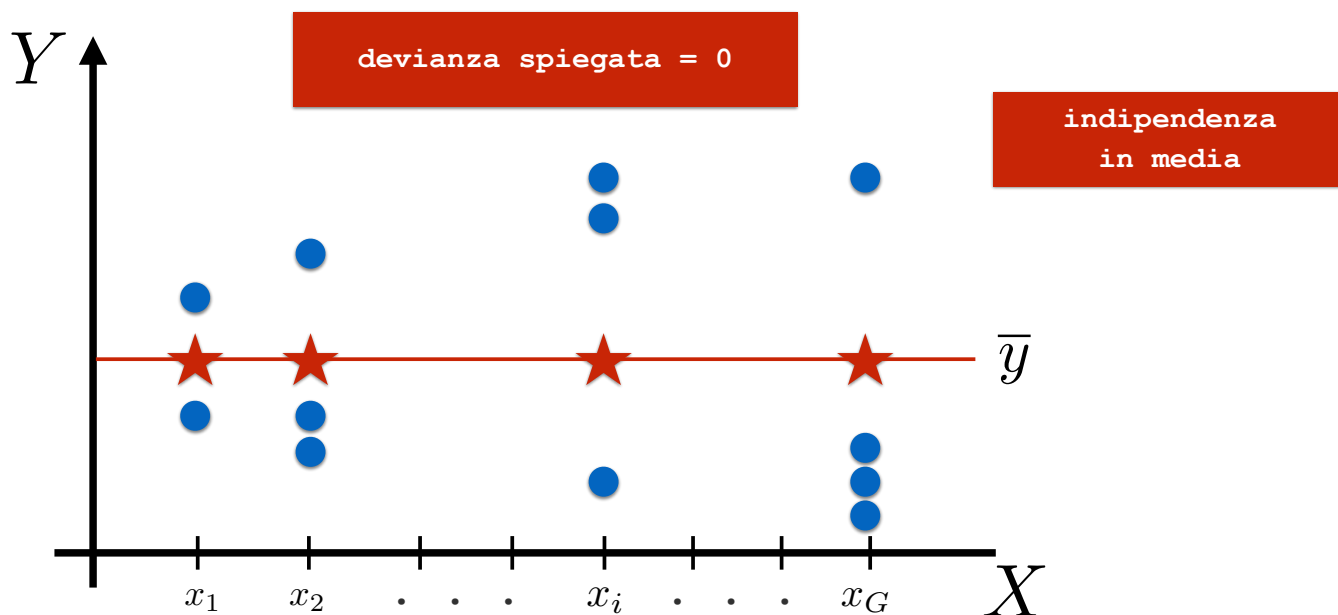


Alcuni casi "caratteristici"

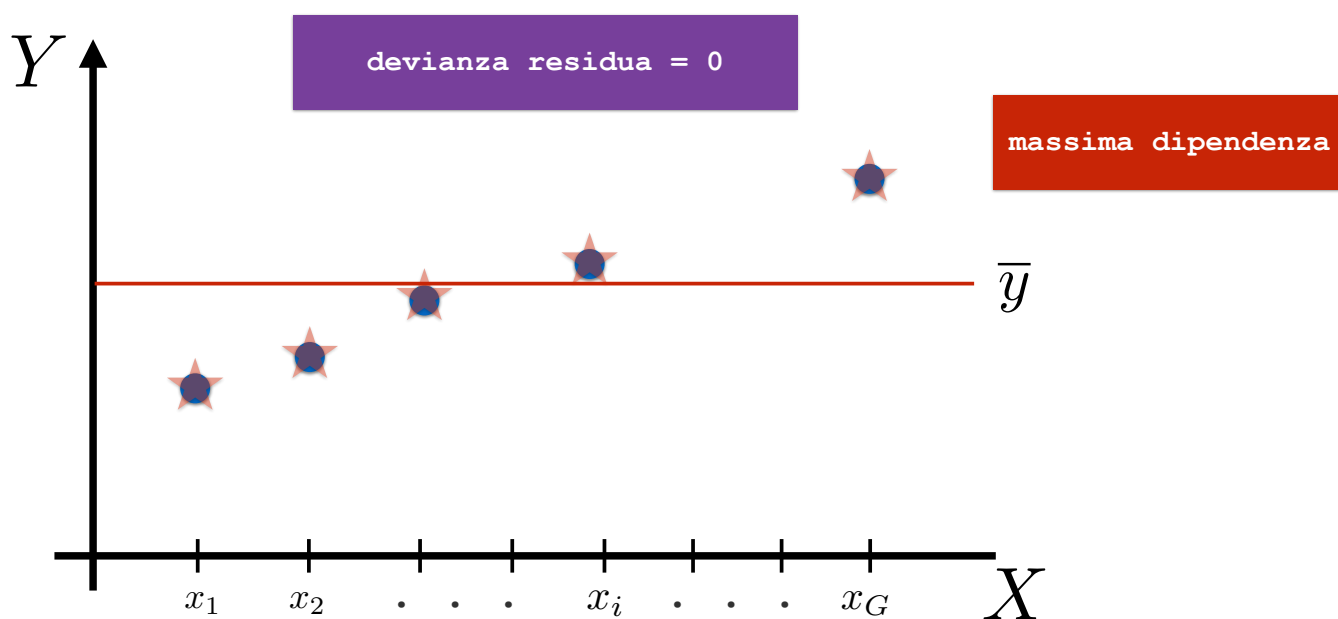
Proprietà di decomposizione della devianza



Proprietà di decomposizione della devianza



Proprietà di decomposizione della devianza



Statistiche di sintesi

Proprietà di decomposizione della devianza

$$F = \frac{\text{Devianza spiegata} / G - 1}{\text{Devianza residua} / n - G}$$

$$\eta^2 = \frac{\text{Devianza spiegata}}{\underbrace{\text{Devianza residua} + \text{Devianza spiegata}}_{\text{Devianza totale}}}$$

Proprietà di decomposizione della devianza

$$\text{Devianza totale} = \text{Devianza residua} + \text{Devianza spiegata}$$

$$\frac{\text{Devianza totale}}{n} = \frac{\text{Devianza residua}}{n} + \frac{\text{Devianza spiegata}}{n}$$

Formula alternativa di calcolo della devianza / varianza

$$\text{Devianza totale} = \text{Devianza residua} + \text{Devianza spiegata}$$

$$\frac{\text{Devianza totale}}{n - 1} = \frac{\text{Devianza residua}}{n - 1} + \frac{\text{Devianza spiegata}}{n - 1}$$

L'angolo dei tecnicismi

Formula alternativa di calcolo della
varianza / devianza

varianza →
$$\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n} = \frac{\sum_{i=1}^n x_i^2}{n} - \bar{x}^2$$

devianza →
$$\sum_{i=1}^n (x_i - \bar{x})^2 = \sum_{i=1}^n x_i^2 - n \bar{x}^2$$