

Indici statistici di variabilità

Rendimento azioni Amazon e Microsoft a confronto

Anno	Microsoft	Amazon
1	1,6%	2,5%
2	5,4%	4,5%

Il rendimento medio e mediano sono uguali, ma Amazon ha rendimenti che vanno dal 2,5% al 4%, mentre Microsoft ha variazioni più ampie.

Indici di variabilità

L'utilizzo di indici di tendenza centrale non è sufficiente a discriminare situazioni molto differenti.

La **variabilità** rappresenta l'attitudine della variabile ad assumere diverse modalità.



In questo caso l'indicatore dovrà essere capace di graduare la variabile in termini di dispersione delle modalità rispetto ad un unico valore di sintesi (ad esempio una misura di posizione).

La concentrazione è in qualche modo una misura simmetrica della variabilità; quando un fenomeno è molto concentrato su un valore, si dice ragionevolmente che c'è poca variabilità.

Indici di variabilità

Un indice di variabilità:

- deve essere un numero positivo
- deve valere zero se calcolato su una distribuzione costante
- deve essere invariante se aggiungo una costante ad X

Gli indici di dispersione sono:

- l'**Intervallo di variazione (range)**
- lo **scarto interquartile**
- la **Varianza**
- la **Deviazione standard**
- il **Coefficiente di variazione**

Range

Il **range** (intervallo o campo di variazione) è la misura più semplice di variabilità ed è dato dalla differenza tra il valore più grande meno il valore più piccolo della distribuzione. Fornisce un'idea dello spazio all'interno del quale si muove il fenomeno, ma non dice nulla sulla variabilità all'interno dell'intervallo.

Range = Valore più grande – valore più piccolo
Esempio: calcolo il range dei seguenti stipendi

3310	3355	3450	3650	3730	3925
------	------	------	------	------	------

$$\text{Range} = 3925 - 3310 = 615$$

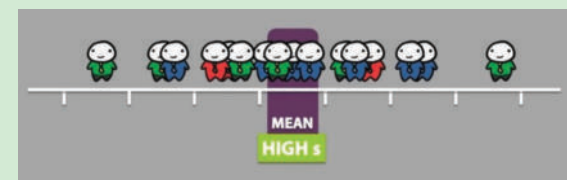
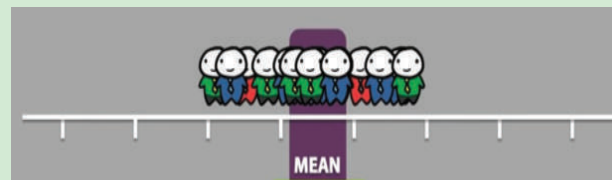


Limiti del range

Si base solo su due osservazioni e quindi è fortemente influenzato dai valori estremi.

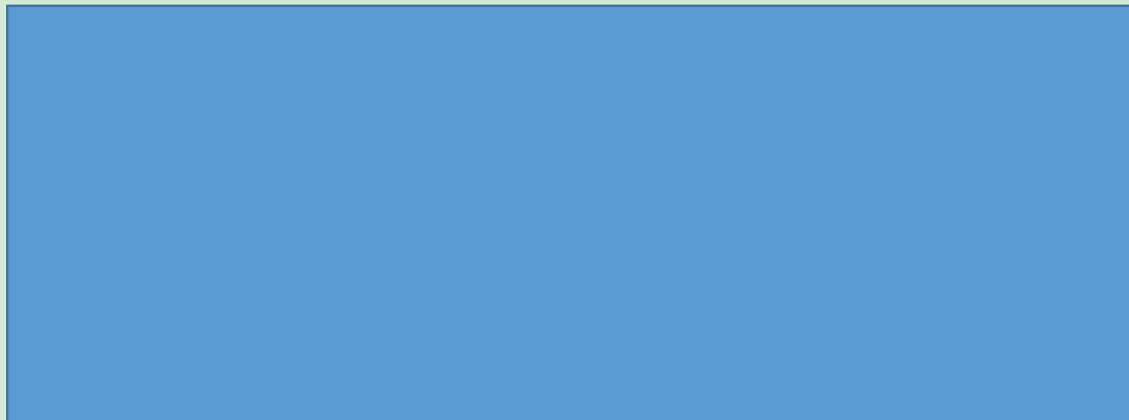
È una misura che non indica come, ma solo quanto sono dispersi i dati.

Non ci dice se i dati sono concentrati attorno al valore centrale o attorno ad un altro valore o se sono distribuiti in modo omogeneo.



Esercizio

Calcolare il range e lo scarto interquartile della seguente distribuzione: 3, 3, 4, 5, 5, 6, 6, 6, 7, 8, 24.



Esercizio

Calcolare il range e lo scarto interquartile della seguente distribuzione: 3, 3, 4, 5, 5, 6, 6, 6, 7, 8, 24.

$$\text{Range} = 24 - 3 = \bullet 21$$

$Q1 = 0,25 * 11(N) = 2,75$ quindi la posizione 3, corrisponde al numero 4.

$Q3 = 0,75 * 11(N) = 8,25$ quindi la posizione 9, che corrisponde al numero 7.

$$\text{IQR} = 7 - 4 = 3$$

Varianza

La **varianza** è una misura della variabilità che utilizza tutti i dati.

Si basa sulla differenza tra il valore di ciascuna osservazione x_i e la media. Tale differenza viene definita scarto dalla media.

Nel calcolo gli scarti sono elevati al quadrato, quindi la varianza ha sempre valore positivo.

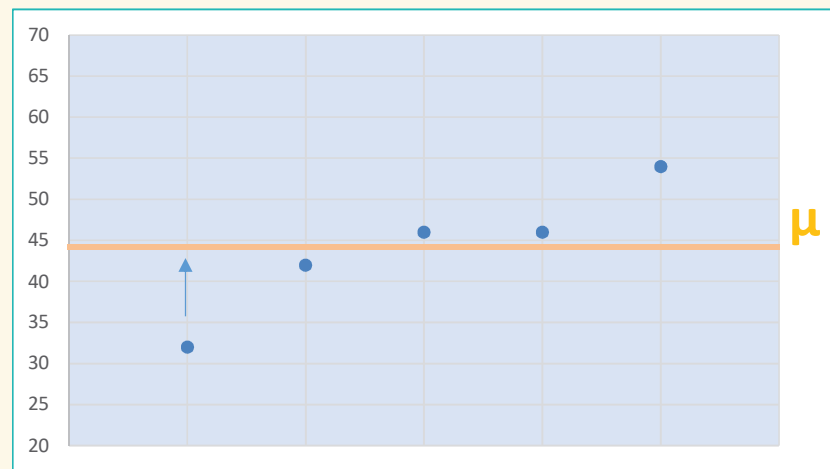
$$\sigma^2 = \frac{(x_1 - M)^2 + (x_2 - M)^2 + \dots + (x_n - M)^2}{n} = \frac{1}{n} \sum_{i=1}^n (x_i - M)^2$$



Calcolo della varianza

La **varianza** descrive e quantifica la dispersione dei dati intorno al valore centrale della popolazione.

Si calcola la distanza esistente tra ogni singola osservazione e la media, la cui somma per definizione è nulla, per cui bisogna elevare al quadrato le distanze rilevate.



Esercizio

N. di studenti	$x_i - \mu$	$x_i - \mu$	$(x_i - \mu)^2$
46	46-44	2	4
54	54-44	10	100
42	42-44	-2	4
46	46-44	2	4
32	32-44	-12	144
			Σ 256

$$\sigma^2 = 256/5 = 51,2$$

Proprietà della varianza

Il valore zero equivale alla non dispersione e può essere ottenuto solo se tutte le osservazioni sono identiche.

Es. la serie 3, 3, 3, 3 avrà media 3 e varianza 0.

Il caso di massima variabilità si ha quando una unità possiede tutto il fenomeno e le altre $n-1$ unità hanno la modalità pari a zero.

Se moltiplico X per qualsiasi costante b , $Y=bX$ allora la varianza di Y sarà moltiplicata per b^2 ovvero $\text{var}Y=b^2 \sigma^2$

Varianza per le distribuzioni di frequenza

$$\sigma^2 = \frac{(x_1 - M)^2 n_1 + (x_2 - M)^2 n_2 + \dots + (x_k - M)^2 n_k}{n_1 + n_2 + \dots + n_k} = \frac{1}{n} \sum_{i=1}^k (x_i - M)^2 n_i = \sum_{i=1}^k (x_i - M)^2 f_i$$

xi (voti)	ni (studenti)	xi - μ	(xi - μ) ²	(xi - μ) ² *ni
24	18	-2,5	6,25	112,5
25	12	-1,5	2,25	27
26	16	-0,5	0,25	4
27	17	0,5	0,25	4,25
28	10	1,5	2,25	22,5
29	22	2,5	6,25	137,5
	Σ 95	0		Σ 307,75

μ=26,5

$$\sigma^2 = 307,75/95 = 3,239$$

Deviazione standard

La **deviazione standard** (detta anche **errore standard**, **scarto quadratico medio**, **scarto tipo**) è definita come la radice quadrata della varianza.

Oltre a fornire informazioni su come il fenomeno sia disperso intorno al valore medio, dà un risultato nella stessa unità di misura della media.

$$\sigma = \sqrt{\sigma^2}$$

$$\sigma = \sqrt{\frac{\sum (x_i - \mu)^2}{n}}$$



Indici di variabilità

La **varianza** e la **deviazione standard** appartengono al gruppo degli scostamenti medi, indicatori che misurano la tendenza delle varie modalità del carattere a disperdersi attorno a un valore medio (dispersione), solitamente la media aritmetica.

In assenza di dispersione (distribuzione costante), gli indici assumono valore nullo.

Varianza

espressa al quadrato dell'unità di misura del carattere



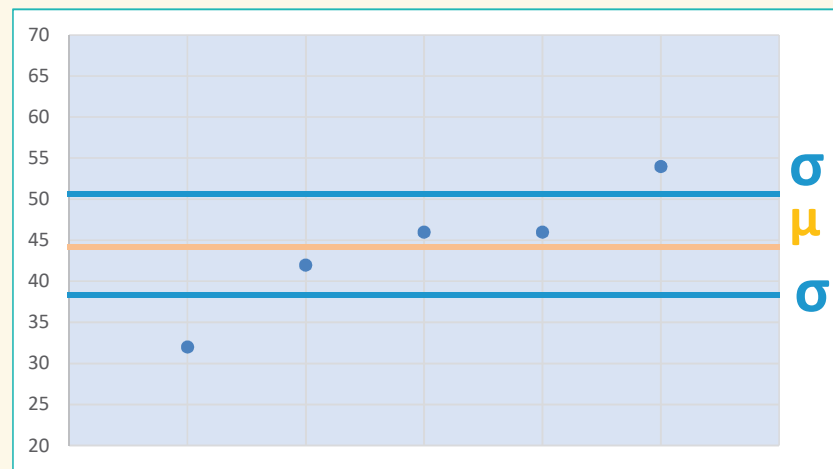
Deviazione standard

espressa su scala lineare, di migliore interpretazione



Calcolo

$$\sigma^2 = 256/5 = 51,2 \quad \sigma = 51,2^{(1/2)} = 7,1$$



Esercizio

Calcolare la varianza e la deviazione standard per la seguente distribuzione: 4, 9, 9, 5, 8, 5, 6, 1.

$$\mu = 5,875$$

x_i	
4	
9	
9	
5	
8	
5	
6	
1	

Esercizio

Calcolare la varianza e la deviazione standard per la seguente distribuzione: 4, 9, 9, 5, 8, 5, 6, 1.

x_i	$x_i - \mu$	$(x_i - \mu)^2$
4	-1,875	3,52
9	3,125	9,77
9	3,125	9,77
5	-0,875	0,77
8	2,125	4,52
5	-0,875	0,77
6	0,125	0,02
1	-4,875	23,77
		Σ 52,87

$$\mu = 5,875$$

$$\sigma^2 = 52,87/8 = 6,61$$

$$\sigma = \sqrt{6,61} = 2,57$$

Esempio

Abbiamo 2 farmaci per il trattamento della pressione arteriosa e vediamo che effetto hanno dopo la somministrazione:

Farmaco	Prima		Dopo	
	Media	DV	Media	DV
1	90	15	70	10
2	90	15	70	15

Esempio

I 2 farmaci hanno prodotto gli stessi risultati medi?

I pazienti del farmaco 1 hanno avuto una reazione più omogenea?

Ogni paziente del farmaco 1 dopo il trattamento ha registrato livello di pressione inferiore rispetto ai pazienti del trattamento 2?

I pazienti del farmaco 2 hanno avuto reazioni molto diverse al trattamento?

Esempio

La deviazione standard è una misura comunemente utilizzata del rischio associato all'investimento in azioni o in fondi azionari.

Essa fornisce una misura di come i rendimenti mensili fluttuano attorno al rendimento medio di lungo periodo.

Azioni	Microsoft	Amazon
Rendimento	5% negli ultimi 5 anni	15% negli ultimi 5 anni
Deviazione standard	10	20

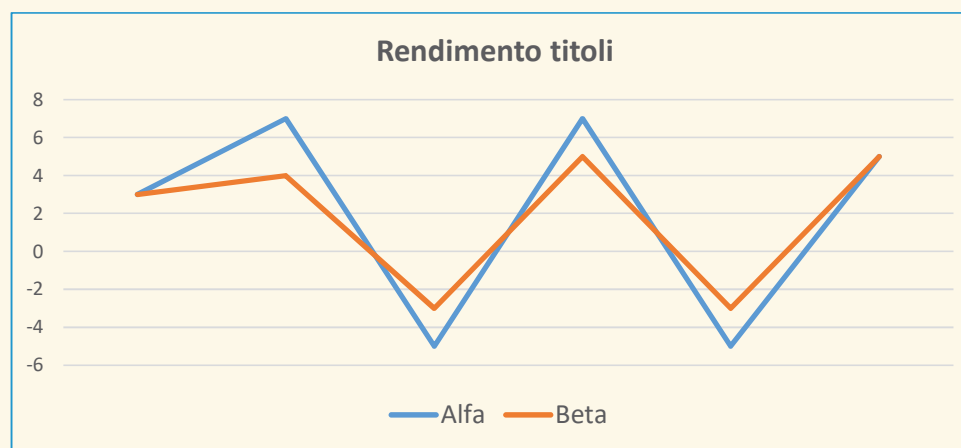
Minore è la deviazione standard, maggiore è la probabilità di ottenere un rendimento vicino a quello medio.

Esempio

La deviazione standard in ambito finanziario indica la **volatilità**, ossia le fluttuazioni/oscillazioni dei prezzi (o dei tassi di rendimento).

La volatilità finanziaria indica il grado di incertezza di rendimento di un determinato asset di mercato.

La deviazione standard in questo ambito misura la dispersione dei prezzi (o rendimenti) rispetto alla media degli stessi.



Esempio

Andamento del petrolio fine 2018 - inizi 2020



σ

In rosso i 3 momenti in cui il mercato ha mostrato un cambiamento di trend.

In questi casi la deviazione standard indica che vi è un'elevata volatilità e può dare indicazioni utili per anticipare un cambio di trend.

Coefficiente di variazione

Permette di confrontare fenomeni riferiti a unità di misura differenti o un ordine di misura diversa.

È una misura della variabilità relativa.

Ad esempio non possiamo confrontare la varianza di due diverse valute oppure il peso dei bambini con quello degli adulti.

In questi casi, per confrontare la variabilità di due distribuzioni per il carattere X con media M positiva, può essere utilizzato il **coefficiente di variazione**:

CV ha un campo di variazione positivo.

$$CV = \frac{\sigma}{|\mu|} * 100$$

Valore assoluto della media con $\mu \neq 0$

Esempio

	Altezza (m)	Peso (kg)
Media	165,8	60,64
Varianza	123,49	23,13



DEV	11,11	4,81
CV	6,70%	7,93%

CV del peso > CV dell'altezza