Indici di dispersione

SAPIENZA
UNIVERSITÀ DI ROMA

Indici di disperione

- La dispersione è la seconda importante proprietà che descrive una distribuzione
- Gli indici di tendenza centrale ci dicono dove sono concentrati i valori della variabile, ovvero quali sono i valori più rappresentativi della distribuzione
- Gli indici di dispersione ci dicono quanto i valori si disperdono intorno alle misure di tendenza centrale, ovvero quanto si disperdono intorno al valore più rappresentativo della distribuzione
- Come per gli indici di tendenza centrale, esistono diversi indici di dispersione. La scelta dell'indice più adeguato dipende dal livello di misura della variabile

Indici di dispersione

Campione A

Punteggio	*f*
16	0
17	1
18	2
19	3
20	10
21	11
22	8
23	7
24	1
25	0
26	0

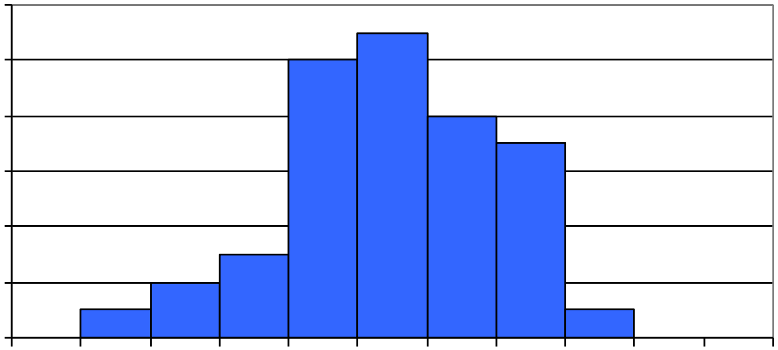
Media = 21

Campione B

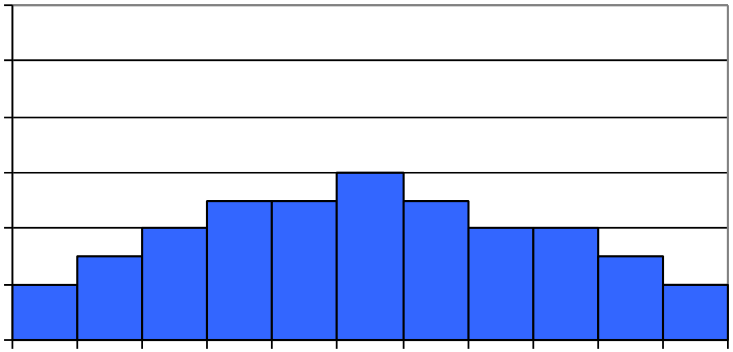
Punteggi	*f*
16	2
17	3
18	4
19	5
20	5
21	6
22	5
23	4
24	4
25	3
26	2

Media = 21

Campione A



Campione B



Indici di dispersione

Gli indici di tendenza centrale consentono di descrivere sinteticamente un fenomeno, ma non forniscono alcuna informazione sul modo in cui i dati si distribuiscono attorno al valore centrale

Esempio: quattro studenti hanno effettuato tre diverse prove di matematica, ottenendo le seguenti valutazioni:

	1° Prova	2° Prova	3° 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

In tutte e tre le prove la media è 6.25, ma i dati sono chiaramente distribuiti in modo diverso

- Se la variabile è misurata su scala a intervalli o a rapporti equivalenti si possono utilizzare diversi indici di dispersione:
- L'intervallo di variazione o range
- Lo scarto semplice medio
- La varianza
- La deviazione standard (scarto quadratico medio)

Indici di dispersione

- L’intervallo di variazione (o range) è la differenza tra il valore massimo e il valore minimo presenti nella distribuzione:
- Il campo di variazione misura quindi l’ampiezza dell’intervallo dei dati:

- Più il range è piccolo, più i valori sono concentrati attorno ai valori centrali

- Più il range è grande, più i dati sono dispersi attorno ai valori centrali

Confrontiamo tre distribuzioni

	1° Prova	2° Prova	3° 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*
range	*6*	*6*	*1*

L’intervallo di variazione è pari a:

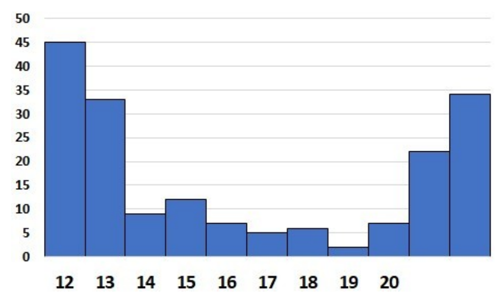
Intervallo di variazione o range

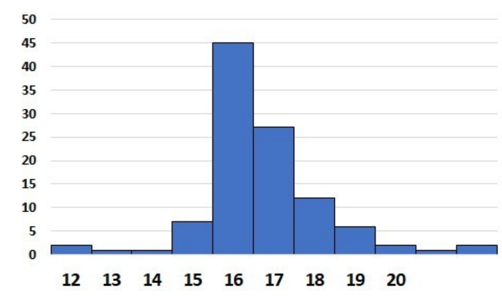
- Vantaggi:

- È facile da calcolare

- Limiti:

- Tiene conto di soli due valori della distribuzione, trascurando tutti glialtri
- È fortemente influenzato dai valoriestremi
- Aumenta con l'aumentare dclle osservazioni





Indici di dispersione

- È utile disporre di un indice di dispersione che consideri tutti i valori della distribuzione (e non solo quelli estremi)

- Un indice efficace di dispersione dovrebbe riflettere il grado in cui ciascun valore si discosta dalla tendenza centrale della distribuzione

Ovvero:

- Quanto, in media, i valori si discostano dalla media?

Indici di dispersione

- Un modo per calcolare la variabilità dei dati tenendo conto di tutti i valori potrebbe consistere nel:

1. Calcolare la distanza (o scarto) di ciascun valore dalla media
2. Fare la media aritmetica di tali distanze

- Un modo per calcolare la variabilità dei dati tenendo conto di tutti i valori potrebbe consistere nel:
- Calcolare la distanza (o scarto) di ciascun valore dalla media
- Fare la media aritmetica di tali distanze

$$\frac{(x_1 - \bar{x}) + (x_2 - \bar{x}) + \ldots.. + (x_n - \bar{x})}{n} = \frac{\sum_{i=1}^{n} (x_i - \bar{x})}{n}$$

Per una proprietà della media, tuttavia, la somma degli scarti di tutti i valori dalla loro media è pari a:

$$\sum_{i=1}^{n} (x_i - \bar{x}) = 0$$

Indici di dispersione. Esempio

	1° Prova
1° studente	3
2° studente	5
3° studente	8
4° studente	9
media	*6.25*
range	*6*

Punteggio	$x_i - \bar{\bar{x}}$
3	-3.25
5	-1.25
8	1.75
9	2.75
	$\Sigma = 0$

Quest'approccio <u>non</u> ci consente pertanto di ottenere una misura della variabilità dei dati

- Il problema si può risolvere facilmente considerando gli scarti dalla media in valore assoluto

- Lo scarto semplice medio (SSM) è la media degli scarti (in valore assoluto) tra tutti i valori della distribuzione e la loro media:

$$\mathrm{SSM} = \frac{\left|x_1 - \bar{x}\right| + \left|x_2 - \bar{x}\right| + \ldots\ldots + \left|x_n - \bar{x}\right|}{n} = \frac{\sum_{i=1}^{n} \left|x_i - \bar{x}\right|}{n}$$

SSM = media della distanza di tutti i valori dalla media

Scarto semplice medio. Esempio

	1° Prova
1° studente	3
2° studente	5
3° studente	8
4° studente	9
media	6.25
range	*6*
SSM	*2.25*

Punteggio	$X_i-\bar{X}$	$\lvert X_i-\bar{X}\rvert$
3	-3.25	3.25
5	-1.25	1.25
8	1.75	1.75
9	2.75	2.75
		$\Sigma = 9$

$$SSM = \frac{9.00}{4} = 2.25$$

Esercitazione

- Eseguire lo stesso calcolo per la seconda e la terza prova

Scarto semplice medio

- Un limite dello SSM è che la somma degli scarti in valore assoluto dalla media può essere uguale alla somma degli scarti in valore assoluto da un altro valore della distribuzione (diverso dalla media)

Varianza

- Un indice analogo si calcola elevando al quadrato le deviazioni dalla media
- Se sommiamo i quadrati delle deviazioni (o "scarti") dalla media e dividiamo questa somma per il numero delle osservazioni otteniamo la varianza:

$$\text{Varianza} = \frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \ldots\ldots + (x_n - \bar{x})^2}{n} = \frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n}$$

- La varianza rappresenta quindi la media dei quadrati degli scarti dalla media

Varianza

- Quando la varianza viene calcolata su un campione si indica con s2
- In questo caso il denominatore n è sostituito da n-1 in modo da ottenere una stima corretta della dispersione della variabile nella popolazione da cui il campione in esame è stato estratto

Varianza. esempio

	1° Prova
1° studente	3
2° studente	5
3° studente	8
4° studente	9
media	*6.25*
range	*6*
SSM	*2.25*
S^2	*7.58*

Punteggio	$X_i-\bar{X}$	$(X_i-\bar{X})^2$
3	-3.25	10.56
5	-1.25	1.56
8	1.75	3.06
9	2.75	7.56

$\Sigma = 22.75$

$$S^2 = \frac{22.75}{3} = 7.58$$

Varianza. Esercitazione

- Calcolare la varianza per la seconda e terza prova

Varianza

- Un limite della varianza come misura di dispersione è quella di avere una unità di misura espressa al quadrato rispetto alla media e all'unità di misura originale
- Essa indica la media dei quadrati degli scarti fra i valori osservati e la loro media
- Per facilitare l'interpretazione dell'indice, esprimendolo in un'unità di misura più facilmente comprensibile, si ricorre spesso al calcolo della deviazione standard

Deviazione standard

- La deviazione standard è l’indice di dispersione della distribuzione più largamente utilizzato
- È pari alla radice quadrata della varianza
- Quando viene calcolata sulla popolazione si indica con il simbolo greco σ
- $\sigma = \sqrt{\frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n}}$
- Quando invece viene calcolata su di un campione si indica con la lettera s minuscola e la sua formula è

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

Deviazione standard

- La deviazione standard ha la stessa unità di misura dei valori osservati
- Essa indica quanto, in media, ciascun valore si discosta dalla media
- Può essere interpretata anche come una misura della accuratezza della media come indice di tendenza centrale: più è elevata, meno la media è efficace nel fornire un valore che sia rappresentativo dell’intera distribuzione

Deviazione standard. esempio

	1° Prova
1° studente	3
2° studente	5
3° studente	8
4° studente	9
media	*6.25*
range	*6*
SSM	*2.25*
S^2	*7.58*
S	*2.75*

Punteggio	$X_i-\bar{X}$	$(X_i-\bar{X})^2$
3	-3.25	10.56
5	-1.25	1.56
8	1.75	3.06
9	2.75	7.56

$\Sigma = 22.75$

$$S = \sqrt{7.58} = 2.75$$