Prof. Anna Paola Ercolani (Università di Roma)

Lez. 12 - ***Indicatori di dispersione***

Consentono di descrivere la variabilità all'interno della distribuzione di frequenza tramite un unico valore che ne sintetizza le caratteristiche

- **CAMPO DI VARIAZIONE**
- **DIFFERENZA INTERQUARTILE**
- **SCOSTAMENTO SEMPLICE MEDIO**
- **VARIANZA**
- **SCARTO QUADRATICO MEDIO o DEVIAZIONE STANDARD**
- **COEFFICIENTE DI VARIAZIONE**

CAMPO DI VARIAZIONE

Questo indice richiede una scala di misura metrica cioè almeno a intervalli o a rapporti

Esempio

11 ragazzi di 8 anni hanno ottenuto ad un test la seguente serie di punteggi:

23 45 34 57 23 57 48 38 38 54 48

Definire il campo di variazione

23 23 34 38 38 45 48 48 54 57 57

CV=57-23 =34

Se la distribuzione di frequenza è in classi si calcola la differenza tra i punti centrali delle classi estreme (superiore e inferiore)

$$\boxed{CV = X_{c(classe_{max})} - X_{c(classe_{min})}}$$

Per definire il campo di variazione è necessario innanzi tutto raccogliere i dati in una distribuzione di frequenza in classi

Classi	f
61-70	8
71-80	17
81-90	14
91-100	11
101-110	7

-punto medio classe inferiore:
(61+70)/2 = 65.5
-punto medio classe superiore:
(101+110)/2=105.5

CV= 105.5-65.5 =40

Campo di variazione totale: calcolato sull'intera distribuzione di frequenza

Campo di variazione parziale: calcolato su parte della distribuzione di frequenza
(ad es. togliendo i casi estremi se anomali rispetto al resto della distribuzione, outliers).

Campo di variazione totale:
2 10 10 10 10 11 11 11 12 12 12 12 13 13 14 14 14 14 15 15 16 16 16 30

CV = 30-2 =28

Campo di variazione parziale:
2 10 10 10 10 11 11 11 12 12 12 12 13 13 14 14 14 14 15 15 16 16 16 30

CV = 16-10 = 6

Limiti

Troppo sensibile ai valori estremi

Dà poche indicazioni

Viene usato solo in modo generico

ESEMPIO:

Gruppo A: 1 2 2 4 5 5 7 9 9 10

Gruppo B: 1 1 1 1 1 10 10 10 10 10

Gruppo C: 1 5 5 5 5 5 5 5 5 10

CV = 10-1 = 9

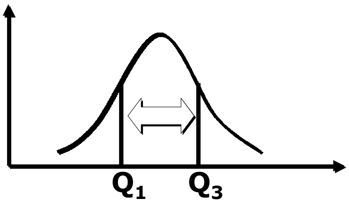
Il campo di variazione è sempre lo stesso ma le distribuzioni hanno caratteristiche estremamente diverse che non vengono evidenziate da questo indice di dispersione.

DIFFERENZA INTERQUARTILE

La differenza interquartile è data dalla differenza tra il terzo e il primo quartile

DI (o Q) = $Q_3 - Q_1$

E' analoga al campo di variazione ma tiene conto soltanto dei valori che cadono tra il 1° e 3° quartile
(cioè del 50% della distribuzione)



Limiti

E' un indice che non tiene conto di cosa accade all'interno della distribuzione (casi centrali) e agli estremi distribuzione

Esempio
15 soggetti hanno espresso il loro grado di adesione (punteggio da 1 a 7) alla seguente affermazione:
"Meglio cento anni da pecora che un giorno da leone"
1 5 4 6 7 2 5 6 3 1 2 4 4 7 7

Trovare il 1°, e il 3° quartile

Punteggi	f	f_{cum}
1	2	2
2	2	4
3	1	5
4	3	8
5	2	10
6	2	12
7	3	15

$$posQ_1 = \left(\frac{15+1}{4}\right).1 = 4$$

$$Q_1 = 2$$

$$posQ_3 = \left(\frac{15+1}{4}\right).3 = 12$$

$$Q_3 = 6$$

DI (o Q) = $Q_3 - Q_1 = 6 - 2 = 4$

INDICI DI DISPERSIONE

Come si può ottenere un indice di dispersione che tenga conto del contributo dei singoli casi?

a. si calcolano gli scarti dei valori osservati dalla media

b. si fa una media di questi scarti

Per ottenere un indice unico e sintetico di dispersione dei dati è necessario che i dati siano misurati su scale metriche a intervalli equivalenti o a rapporti equivalenti

SCOSTAMENTO SEMPLICE MEDIO

Poiché la somma degli scarti dalla media è zero, sommo gli scarti in valore assoluto:

$$SSM = \frac{\sum_{i=1}^{N} |X_i - M|}{N}$$

Esempio

Ad un test di personalità, 10 adolescenti hanno ottenuto i seguenti punteggi:

8 9 5 4 7 8 9 7 4 3

Calcolare lo scostamento semplice medio

Calcolo la media:

$$M = \frac{8+9+5+4+7+8+9+4+3}{10} = 6.4$$

$$SSM = \frac{|8-6.4|+|9-6.4|+|5-6.4|+\ldots+|3-6.4|}{10} = \frac{|1.6|+|2.6|+|-1.4|+\ldots+|-3.4|}{10} = 1.95$$

VARIANZA

Poiché la somma degli scarti dalla media è zero, sommo gli scarti elevati al quadrato:

$$s^2 = \frac{\sum_{i=1}^{N}(X_i - M)^2}{N}$$

Se i dati sono raggruppati in classi, la formula diventa

$$s^2 = \frac{\sum_{j=1}^{k} f_j \left(X_{c_j} - M\right)^2}{N}$$

Si indica con s^2 se si tratta di dati osservati su campioni

Si indica con σ^2 se si tratta di distribuzioni teoriche

Esempio

Ad un test di personalità, 10 adolescenti hanno ottenuto i seguenti punteggi:

8 9 5 4 7 8 9 7 4 3

Calcolare la varianza

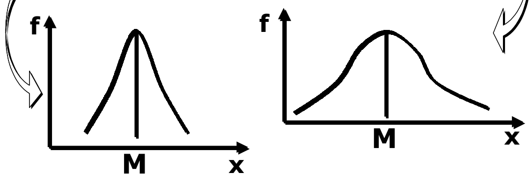
M = 6.4 (vedi es. precedente)

$$s^2 = \frac{(8-6.4)^2+(9-6.4)^2+(5-6.4)^2+\ldots+(3-6.4)^2}{10} = \frac{2.56+6.76+1.96+\ldots+11.56}{10} = 4.44$$

La varianza non è mai negativa

Maggiore è la varianza più i casi sono dispersi attorno alla media

Minore è la varianza più i casi sono concentrati attorno alla media



SCARTO QUADRATICO MEDIO (DEVIAZIONE STANDARD)

Radice quadrata della Varianza: indice di dispersione con unità di misura uguale alla media. Indica di quanto, mediamente, i dati osservati si discostano dalla loro media.

$$s = \sqrt{\frac{\sum_{i=1}^{N}(X_i - M)^2}{N}}$$

Se i dati sono raggruppati in classi la formula diventa:

$$s = \sqrt{\frac{\sum_{j=1}^{k} f_j \left(X_{c_j} - M\right)^2}{N}}$$

Esempio

Ad un test di personalità, 10 adolescenti hanno ottenuto i seguenti punteggi:

8 9 5 4 7 8 9 7 4 3

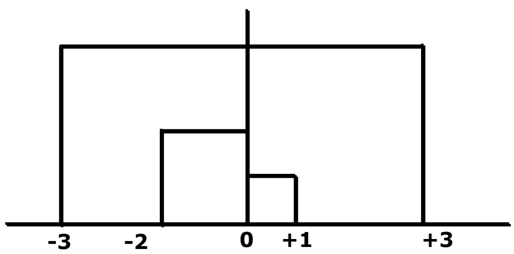
Calcolare la deviazione standard

M = 6.4 (vedi es. precedente)

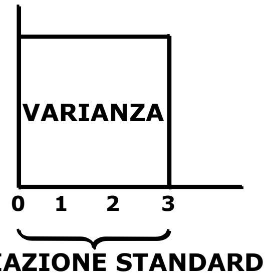
$$s = \sqrt{\frac{(8-6.4)^2+(9-6.4)^2+(5-6.4)^2+\ldots+(3-6.4)^2}{10}} =$$

$$= \sqrt{\frac{2.56+6.76+1.96+\ldots+11.56}{10}} = \sqrt{4.44} = 2.11$$

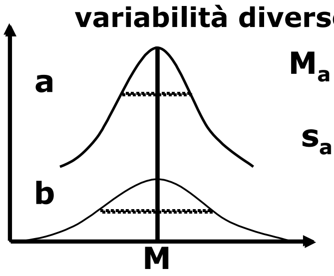
QUADRATI e SCARTI DALLA MEDIA



VARIANZA e DEVIAZIONE STANDARD



Queste due distribuzioni hanno la stessa media ma variabilità diverse

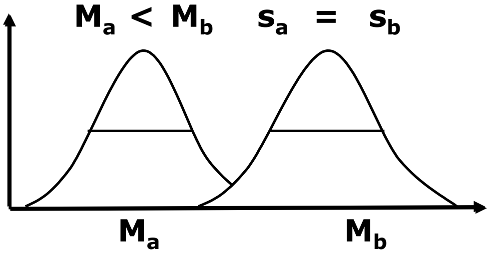


$M_a = M_b$

$s_a < s_b$

Queste due distribuzioni hanno la medie diverse ma La stessa variabilità

$M_a < M_b \quad s_a = s_b$



Nel riportare le statistiche descrittive di un gruppo di soggetti si scrive M ± s

Esempio

Si può dire che i 10 adolescenti al test di personalità ottengono una media di 6.4 ± 2.11

VARIANZA e DEVIAZIONE STANDARD

Esistono formule che consentono il calcolo partendo direttamente dai dati grezzi:

non occorre calcolare la media e i singoli scarti

$$s = \sqrt{\frac{\sum_{i=1}^{N} X_i^2}{N} - \left(\frac{\sum_{i=1}^{N} X_i}{N}\right)^2}$$

Deviazione standard

$$s^2 = \frac{\sum_{i=1}^{N} X_i^2}{N} - \left(\frac{\sum_{i=1}^{N} X_i}{N}\right)^2$$

Varianza

Con semplici passaggi:

$$s=\sqrt{\frac{\sum X_i^2}{N}-\frac{(\sum X)^2}{N^2}}=$$
$$=\sqrt{\frac{N\sum X^2-(\sum X)^2}{N^2}}=\frac{1}{N}\sqrt{N\sum X_i^2-(\sum X)^2}$$

Analoghe alle precedenti:

$$s=\frac{1}{N}\sqrt{N\sum_{i=1}^{N}x_i^2-\left(\sum_{i=1}^{N}x_i\right)^2}$$

Deviazione standard

$$s^2=\frac{1}{N^2}\left[N\sum_{1}^{N}x_i^2-\left(\sum_{1}^{N}x_i\right)^2\right]$$

Varianza

Esempio

Ad un test di personalità, 10 adolescenti hanno ottenuto i seguenti punteggi:

8 9 5 4 7 8 9 7 4 3

Calcolare la varianza e la deviazione standard utilizzando le formule con i dati grezzi

DEVIAZIONE STANDARD

1

$$s=\sqrt{\frac{8^2+9^2+5^2+\ldots+3^2}{10}-\left(\frac{8+9+5+\ldots+3}{10}\right)^2}=$$
$$=\sqrt{45.4-40.96}=\sqrt{4.44}=2.11$$

2

$$s=\frac{1}{10}\sqrt{\left[10\cdot(8^2+9^2+5^2+\ldots+3^2)\right]-(9+8+5+..+3)^2}=$$
$$=\frac{1}{10}\sqrt{4540-4096}=\frac{1}{10}\sqrt{444}=\frac{1}{10}21.1=2.11$$

VARIANZA

1

$$s^2=\frac{8^2+9^2+5^2+\ldots+3^2}{10}-\left(\frac{8+9+5+\ldots+3}{10}\right)^2=$$
$$=45.4-40.96=4.44$$

2

$$s^2=\frac{1}{100}\left[10\cdot(8^2+9^2+5^2+\ldots+3^2)\right]-(9+8+5+..+3)^2=$$
$$=\frac{1}{100}[4540-4096]=\frac{1}{100}x444=4.44$$

ESEMPIO

x_i	x_i^2	$(x_i - M)$	$(x_i - M)^2$
1	1	(1-3)=-2	4
2	4	(2-3)=-1	1
3	9	(3-3)=0	0
4	16	(4-3)=1	1
5	25	(5-3)=2	4
$\Sigma x_i=15$	$\Sigma x_i^2=55$	$\Sigma(x_i-M)=0$	$\Sigma(x_i-M)^2=10$

CALCOLO DI VARIANZA E DEVIAZIONE STANDARD

Con la formula

$$s = \sqrt{\frac{\sum_{i=1}^{N}(X_i - M)^2}{N}}$$

$$s = \sqrt{\frac{10}{5}} = 1.41$$ **Deviazione standard**

$$s^2 = (1.41)^2 = 2$$ **Varianza**

Con la formula

$$s = \sqrt{\frac{\sum_{i=1}^{N} x_i^2}{N} - \left(\frac{\sum_{i=1}^{N} x_i}{N}\right)^2}$$

$$s = \sqrt{\frac{55}{5} - \left(\frac{15}{5}\right)^2} = 1.41$$ **Deviazione standard**

$$s^2 = \frac{55}{5} - \left(\frac{15}{5}\right)^2 = 11 - 9 = 2$$ **Varianza**

Con la formula

$$s = \frac{1}{N}\sqrt{N\sum_{i=1}^{N} x_i^2 - \left(\sum_{i=1}^{N} x_i\right)^2}$$

$$s = \frac{1}{5}\sqrt{5(55) - (15)^2} = 1.41$$ **Deviazione standard**

$$s^2 = \frac{1}{5^2}\left[5(55) - (15)^2\right] = \frac{1}{25}(50) = 2$$ **Varianza**

La deviazione standard è l'indicatore più usato per descrivere la variabilità di una distribuzione

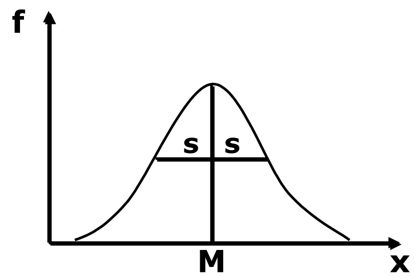
Usa la stessa unità di misura della media

E' direttamente confrontabile con la stessa media

Gli indicatori

Media e deviazione standard

vengono sempre citati per descrivere le distribuzioni

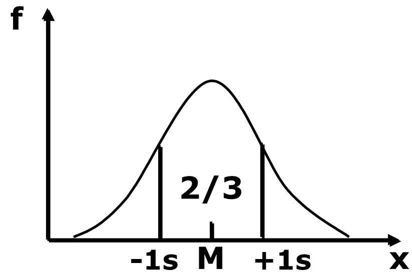


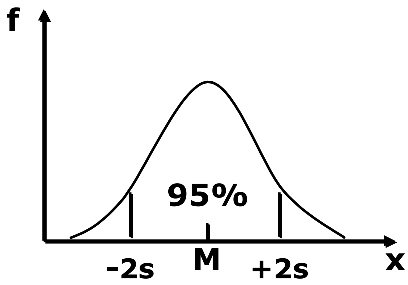
Nelle distribuzioni simmetriche e unimodali circa 2/3 delle osservazioni cadono nell'intervallo

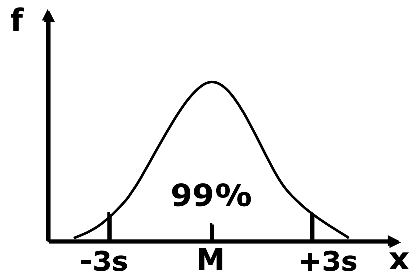
$M \pm 1s$

circa 95% → $M \pm 2s$

circa 99% → $M \pm 3s$







COEFFICIENTE DI VARIAZIONE

- **Il coefficiente di variazione sintetizza il rapporto tra Media e Deviazione Standard**

$$V = \frac{s}{M} \cdot 100$$

- **Determina la dispersione dei dati osservati mediante l'uso della Media come unità di misura**
- **E' un indicatore di variabilità relativa**

Esempio
Le medie e le deviazioni standard ad un test di motivazione al lavoro dei lavoratori di due aziende sono rispettivamente:

$M_1=84 \pm 7$ e $M_2=68 \pm 6$

Quale è l'azienda con maggior variabilità assoluta?
E maggior variabilità relativa?

$M_1=84 \pm 7$ e $M_2=68 \pm 6$

$s_1= 7 > s_2= 6$

La 1° azienda ha una maggiore variabilità assoluta

$$V_1 = \frac{7}{84} \cdot 100 \cong 8 \qquad V_2 = \frac{6}{68} \cdot 100 \cong 9$$

$V_2=9 > V_1=8$

La 2° azienda ha una maggiore variabilità relativa