

INDICI DI VARIABILITÀ

Valori che sono calcolati per esprimere sinteticamente la variabilità di un fenomeno, o meglio la sua attitudine ad assumere valori differenti tra loro

Proprietà essenziali

1. NON NEGATIVITÀ

Una qualsiasi misura di variabilità deve presentare sempre segno positivo

2. NULLITÀ

Una qualsiasi misura di variabilità deve essere nulla se e solo se tutti i termini della distribuzione sono uguali tra loro

3. INVARIANZA PER TRASLAZIONE

Se a ciascun termine della distribuzione si aggiunge una costante (c) l'indice di variabilità della nuova distribuzione assume lo stesso valore numerico

$$V(X+c) = V(X)$$

4. NON-INVARIANZA PER PRODOTTO

Se a ciascun termine della distribuzione si moltiplica per una costante (c), l'indice di variabilità della nuova distribuzione ha come valore quello precedente moltiplicato per $(c)^2$

$$V(X \cdot c) = c^2 V(X)$$

5. Una qualsiasi misura di variabilità deve crescere all'aumentare della disuguaglianza dei termini

Distinguiamo:

➤ **Indici assoluti di variabilità**

Sono direttamente collegati all'unità di misura nella quale sono espressi.

Si distinguono a loro volta in:

1. **Indici di dispersione:** misurano la dispersione delle modalità attorno ad un centro
 - **scarto quadratico medio**
 - **devianza**
 - **varianza**
2. **Indici di disuguaglianza:** misurano la disuguaglianza tra le differenti modalità della distribuzione
 - **campo di variazione**

➤ **Indici relativi di variabilità**

Sono indipendenti dall'unità di misura e possono essere utilizzati per il confronto di distribuzioni differenti. Si ottengono rapportando gli indici di variabilità assoluta o al loro valore medio o al valore massimo

- **coefficiente di variazione** (rapporta il valore assoluto al valore medio)

CAMPO DI VARIAZIONE

È l'indice più semplice tra le misure di variabilità e fornisce una prima indicazione della variabilità di un fenomeno pur conoscendo solo il valore più grande e il valore più piccolo delle modalità della distribuzione

SUCCESSIONE, DISTRIBUZIONE DI FREQUENZA , DISTRIBUZIONE DI FREQUENZA IN CLASSI

$$W = X_{\max} - X_{\min}$$

n.b. il campo di variazione è un indice che non tiene conto della numerosità ed inoltre risente della presenza di valori anomali nella distribuzione

PROBLEMI DI CONTROLLO DI QUALITÀ

In un'azienda produttrice di pasta la lunghezza degli spaghetti appropriata all'inscatolamento è compresa tra 15 e 19 cm.

Il macchinario che taglia la pasta è tarato in modo da produrre spaghetti lunghi (in media aritmetica) 18,6; la variabilità della lunghezza degli spaghetti prodotti dipende, invece, dalle condizioni atmosferiche, e deve quindi essere costantemente sotto controllo in modo da bloccare in tempo il processo di produzione quando questo genera spaghetti non idonei all'inscatolamento.

A quale misura di variabilità occorre ricorrere?

$$W=19-15=4$$

VARIANZA

▫ SUCCESIONE

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

▫ VARIABILE STATISTICA

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

$$\text{dove } N = \sum_{i=1}^s n_i$$

▫ VARIABILE STATISTICA IN CLASSI

$$\sigma^2 = \frac{\sum_{i=1}^s (x_i^c - M)^2 n_i}{N}$$

dove x_i^c rappresenta il valore centrale di ogni classe

SCARTO QUADRATICO MEDIO

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

N.B. è la media quadratica degli scarti di ciascun termine dalla media

DEVIANZA

$$\text{Dev} = s\sigma^2$$

N.B. è il numeratore della varianza

FORMULA ALTERNATIVA PER IL CALCOLO DELLA VARIANZA

$$\begin{aligned}\sigma^2 &= \frac{\sum_{i=1}^s (x_i - M)^2}{s} = \frac{\sum_{i=1}^s x_i^2}{s} + M^2 - 2M \frac{\sum_{i=1}^s x_i}{s} \\ &= \frac{\sum_{i=1}^s x_i^2}{s} - M^2\end{aligned}$$

la varianza può essere calcolata come differenza tra la media aritmetica del quadrato dei termini e il quadrato della media aritmetica

PROBLEMI SULL'UTILIZZO DI DIFFERENTI INDICI DI VARIABILITÀ

Date le seguenti due distribuzioni:

| a) Modalità | Frequenze | b) Modalità | Frequenze |
|-------------|-----------|-------------|-----------|
| 0 | 5 | 0 | 1 |
| 1 | 5 | 1 | 4 |
| 2 | 5 | 2 | 15 |
| 3 | 5 | 3 | 4 |
| 4 | 5 | 4 | 1 |

Senza effettuare i calcoli essendo le due distribuzioni simmetriche si può affermare che la media aritmetica di ambedue è 2 ed inoltre il campo di variazione di ambedue è 4 .

a) le due distribuzioni sono identiche?

b) quale delle due distribuzioni mostra minore variabilità?

SOLUZIONE

Le due distribuzioni pur avendo la stessa media non presentano la stessa dispersione dei caratteri attorno ad essa, in particolare la seconda distribuzione è più concentrata attorno al valore medio rispetto alla prima. Infatti:

a)

$$\begin{aligned}\sigma &= \sqrt{\frac{\sum_{i=1}^s (x_i - M)^2 n_i}{N}} = \\ &= \frac{(0-2)^2 5 + (1-2)^2 5 + (2-2)^2 5 + (3-2)^2 5 + (4-2)^2 5}{25} = \\ &= \sqrt{\frac{50}{25}} = 1,4\end{aligned}$$

b)

$$\begin{aligned}\sigma &= \sqrt{\frac{\sum_{i=1}^s (x_i - M)^2 n_i}{N}} = \\ &= \frac{(0-2)^2 1 + (1-2)^2 4 + (2-2)^2 15 + (3-2)^2 4 + (4-2)^2 1}{25} = \\ &= \sqrt{\frac{16}{25}} = 0,8\end{aligned}$$

COEFFICIENTE DI VARIAZIONE

SUCCESSIONE, DISTRIBUZIONE DI FREQUENZA, DISTRIBUZIONE DI FREQUENZA IN CLASSI

$$CV = \frac{\sigma}{M} \cdot 100$$

n.b. il coefficiente di variazione è un indice che non è definito per $M=0$, tende ad assumere valori sempre più grandi al tendere di $M \rightarrow 0$, se M è negativa occorre considerarla in modulo

PROPRIETA' DEL CV

1. è uguale a 0 se e solo se la variabilità è nulla
2. può assumere anche valori superiori a 100, e precisamente il suo valore massimo è $\max(CV) = 100\sqrt{s-1}$ che si riscontra quando $(s-1)$ valori sono uguali a 0 ed uno è uguale all'ammontare complessivo del carattere, e cioè pari a sM (massima variabilità)
3. è invariante anche per trasformazioni che riguardano il prodotto

PROBLEMI DI CONFRONTO DI VARIABILITÀ

Ogni 2 anni in Francia viene effettuata dall'Istituto XXX per la ricerca contro le malattie genetiche una raccolta di fondi che coinvolge tutti i mezzi di comunicazione (internet, stampa, tv, ecc.) Sono state calcolate le seguenti statistiche che riguardano 3 differenti periodi:

| | <i>donazione media</i> | <i>scarto q.medio</i> |
|------------|------------------------|-----------------------|
| 1° periodo | 78 | 21 |
| 2° periodo | 90 | 38 |
| 3° periodo | 150 | 45 |

In quale dei tre periodi si è presentata una maggiore variabilità nelle donazioni?

SOLUZIONE

$$1^{\circ} \text{ periodo } CV=21/78=0,27$$

$$2^{\circ} \text{ periodo } CV=38/90=0,42$$

$$3^{\circ} \text{ periodo } CV=45/150=0,30$$