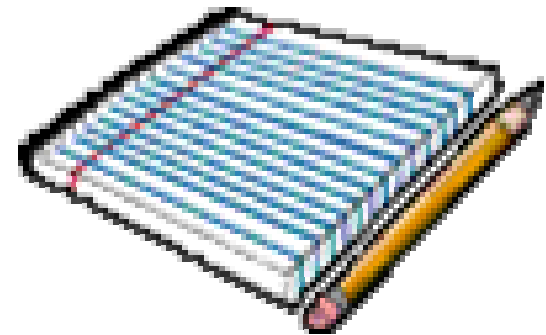


Lezione 4

La Variabilità





Definizione

Un valore medio, comunque calcolato, non è sufficiente a rappresentare l'insieme delle osservazioni effettuate (o l'insieme dei valori assunti dalla variabile statistica); è necessario quindi affiancare ad esso altri indici che siano in grado di fornire delle informazioni sulla **dispersione**, in pratica sulla distanza delle varie osservazioni dal valore medio che rappresenta il centro della distribuzione.

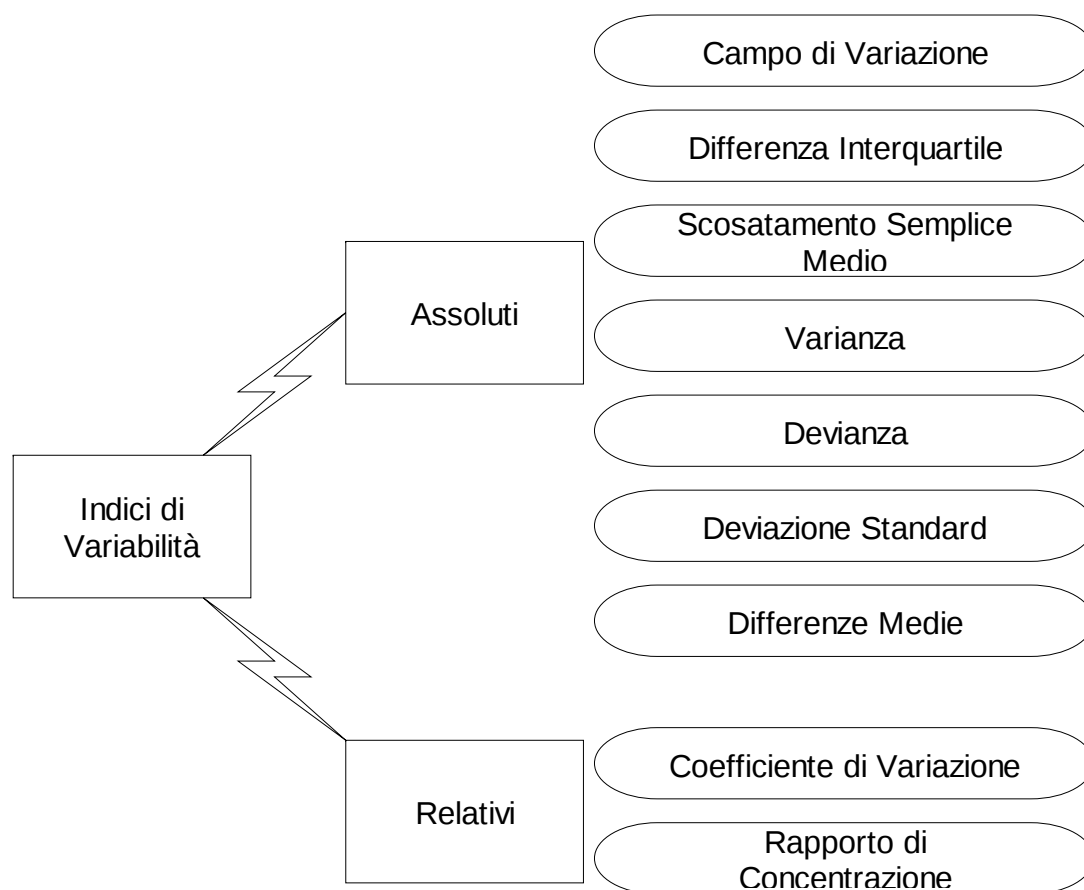
Caratteristiche



Indicatore di variabilità è una misura che:

- **deve annullarsi** quando, e solo quando, tutte le unità osservate presentano il medesimo stato di grandezza del carattere;
- **deve assumere valori crescenti** all'aumentare della variabilità.

Classificazione degli Indici di Variabilità



Indici di Variabilità Assoluti

Campo di Variazione



È il più semplice da calcolare ed è dato dalla *differenza fra il maggiore e il minore dei valori rilevati*. Talvolta il campo di variazione si esprime indicando, invece della differenza fra il maggiore e il minore dei valori rilevati, gli estremi dell'intervallo. Il campo di variazione è un indice molto semplice da calcolare, ma di scarsa importanza perché tiene conto solo dei valori estremi e non degli altri.

$$CV = x_{\max} - x_{\min}$$

Limiti: Troppo sensibile ai valori estremi

Esempio



CAMPO DI VARIAZIONE

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

$$CV = 57 - 23 = 34$$



Differenza Interquartile

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

$$Q = Q_3 - Q_1$$

E' una misura di variabilità 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

DIFFERENZA INTERQUARTILE

ESEMPIO

15 soggetti hanno espresso il loro grado di soddisfazione (punteggio da 1 a 7) su un determinato prodotto.

I risultati sono:

1 5 4 6 7 2 5 6 3 1 2 4 4 7 7

Trovare il 1°, e il 3° quartile

Lo Scostamento Semplice Medio dalla Media Aritmetica



Un altro indice di variabilità è lo scostamento semplice medio, che è la media aritmetica dei valori assoluti degli scarti da un valore medio.

Esistono due tipi di scostamenti:

1. Scostamento semplice medio dalla media aritmetica:

$$\frac{\sum_{i=1}^{i=n} |x_i - \bar{x}| n_i}{\sum_{i=1}^{i=n} n_i}$$

2. Scostamento semplice medio dalla mediana

$$\frac{\sum_{i=l}^{i=n} |x_i - Me| n_i}{\sum_{i=l}^{i=n} n_i}$$



Esempio

SCOSTAMENTO SEMPLICE MEDIO

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

Varianza



VARIANZA

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

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

Proprietà della Varianza



- proprietà:
 - ✓ aumenta all'aumentare del numero di dati
 - ✓ ha come unità di misura il quadrato della variabile x (molto scomodo)
 - ✓ sono confrontabili solo devianze che derivano dallo stesso numero di dati



Esempio

VARIANZA

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

Formula Alternativa per il calcolo della Varianza



La Varianza può essere inoltre calcolata nel seguente modo:

$$S^2 = M_q^2 - M^2 = \frac{\sum_{i=1}^n x_i^2}{n} - \left(\frac{\sum_{i=1}^n x_i}{n} \right)^2$$

Devianza



La quantità

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

Viene definita devianza

Scarto Quadratico Medio (Deviazione Standard)



La *deviazione standard* è rappresentata dal *radice quadrata della varianza*

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

La *Deviazione standard* viene espressa nella *stessa unità di misura dei dati originari*.

Esempio



- Varianza : esercizio di calcolo

calcolare la varianza del seguente campione di 6 dati: 5, 6, 7, 7, 8, 10

1) con la formula classica euristica
$$s^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n}$$

2) con la formula alternativa
$$s^2 = \frac{\sum_{j=1}^n x_j^2}{n} - \left(\frac{\sum_{j=1}^n x_j}{n} \right)^2$$



Indici di Variabilità Relativi

Tutti gli indici di variabilità sono definiti *indici di variabilità assoluta* e sono espressi nella stessa unità di misura del fenomeno considerato; nel caso occorra confrontare più distribuzioni che siano espresse con diverse unità di misura, si ricorre agli *indici di variabilità relativa*.

Gli *indici di variabilità relativa* hanno quindi la *caratteristica di essere dei numeri puri*, indipendenti cioè dall'unità di misura prescelta, e permettono di confrontare più distribuzioni.



Il Coefficiente di Variazione

COEFFICIENTE DI VARIAZIONE

- Il coefficiente di variazione sintetizza il rapporto tra **Media** e **Deviazione Standard**
- Determina la dispersione dei dati osservati mediante l'uso della **Media** come *unità di misura*
- E' un indicatore di *variabilità relativa*

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

Indici di Mutabilità



Mutabilità: Attitudine di un carattere qualitativo ad assumere differenti modalità.

Insieme Omogeneo: Tutte le unità rispetto ad un determinato carattere presentano le stesse modalità

Insieme Eterogeneo: Tutte le unità statistiche tendono a distribuirsi in maniera uniforme.

Indice di Eterogeneità di Gini



$$IE = 1 - \sum_{i=1}^k \left(\frac{n_i}{n} \right)^2$$

Proprietà dell'indice di Gini



$$0 \leq IE \leq \frac{k-1}{k}$$

IE= 0 quando il collettivo è omogeneo

IE= (k-1)/k quando ciascuna unità statistica possiede n/k unità del carattere

Indice di Gini Normalizzato



$$IE_n = \frac{IE}{\max(IE)} = \frac{IE}{\frac{k-1}{k}}$$



ESEMPIO: CORSO DI LAUREA

xi	ni	fi	fi ²
SAM	137	18,69%	0,034933
SPO	251	34,24%	0,117257
ORU	186	25,38%	0,06439
IES	159	21,69%	0,047053
	733	100,00%	0,263633
Indice di eterogeneità di Gini			
E1	0,736367	k = 4	
E1*	0,981823		
INDICE NORMALIZZATO PROSSIMO A 1: C'è quasi massima eterogeneità			

$K = n^{\circ}$ delle modalità = 4

$$E_1 = 1 - \sum_{i=1}^k p_i^2 = 1 - (0.1869^2 + 0.3424^2 + 0.2538^2 + 0.2169^2) = 1 - 0.2636633 = 0.7363$$

$$E_1^* = E_1 \frac{k}{k-1} = 0.7363 \cdot \frac{4}{3} = 0.9818$$

L'indice di Mutabilità di Frosini



$$d = \sqrt{\sum_{i=1}^k \left(f_i - \frac{1}{k} \right)^2}$$

$$0 \leq d \leq \sqrt{\frac{k-1}{k}}$$

Indice Normalizzato di Frosini

$$d_n = \frac{d}{\max(d)} = \frac{d}{\sqrt{\frac{k-1}{k}}}$$