

VARIABILITÀ

(INDICI di VARIABILITÀ)

perché l'indice di posizione

- ⇒ attitudine del fenomeno a manifestarsi in modi diversi
- un fenomeno si manifesta diversamente in modi diversi
 - osservazione ripetuta di una stessa grandezza

⇒ **PROPRIETÀ** di un indice di variabilità:

- ① vale 0 se tutte le modalità sono uguali (= mode a μ)
- ② aumenta al crescere della variabilità
- ③ NON varia per trasformazioni di posizione (traslazioni)

NS

- Non vedremo:
- Scostamenti medi
 - Varianza
 - Intervalli di variazione
 - Indici relativi di variabilità
 - Variabilità per le variabili qualitative

SCOSTAMENTI MEDI ASSOLUTI d'ORDINE S

S_{μ}^s

Distribuzione disaggregata di X:

$$S_{\mu}^s = \left\{ \frac{1}{N} \sum_{i=1}^N |x_i - \mu|^s \right\}^{\frac{1}{s}}$$

indice di posizione

- Media potenziata delle differenze tra le modalità con cui si è manifestato il fenomeno X e μ

- $\forall s$ e $\forall \mu$, si ha che $S_{\mu}^s = 0$ SE e SOLO SE X ha variabilità nulla (1 sola modalità che coincide con μ)

• Distribuzione di freq. ass. per X con K modalità:

$$S_{\mu}^s = \left\{ \frac{1}{N} \sum_{i=1}^K |x_i - \mu|^s n_i \right\}^{\frac{1}{s}}$$

• Distribuzione di freq. relativa

$$S_{\mu}^s = \left\{ \sum_{i=1}^K |x_i - \mu|^s f_i \right\}^{\frac{1}{s}}$$

DISTRIBUZIONE DEGENERATA

VARIANZA

$$\Rightarrow \text{Var}(X) = \sum_{i=1}^N \frac{(x_i - \mu)^2}{N} = \sum_{i=1}^K \frac{(x_i - \mu)^2 n_i}{N} = \sum_{i=1}^K (x_i - \mu)^2 f_i \geq 0$$

Var(x)

 σ^2

Scostamenti medio assoluto di ordine 2 dalla media aritmetica, elevato al quadrato

- Più noto indice di variabilità

misura dell'attitudine di x ad assumere valori diversi

- $\text{Var}(X) \geq 0$

(sigma quadro)

- Si indica spesso con σ_x^2

- Quando X è degenere (tutti i valori osservati sono = tra loro) la varianza assume valore 0

Scosto quadratico medio

- $\sqrt{\text{Var}(X)} = \text{SD} = \sigma_x = \text{DEVIATION STANDARD}$ = scostamento medio assoluto di ordine 2

NB3

NON esiste un limite superiore alla variabilità

non si può stabilire se un fenomeno sia tanto o poco variabile in assoluto

- Si possono però fare confronti di variabilità tra fenomeni

FORMULA OPERATIVA DI CALCOLO per ottenere la varianza x distr. discog.

$$\begin{aligned} \text{Var}(X) &= \sum_{i=1}^N \frac{(x_i - \mu)^2}{N} = \sum_{i=1}^N \frac{(x_i^2 - 2x_i\mu + \mu^2)}{N} = \sum_{i=1}^N \frac{x_i^2}{N} + \sum_{i=1}^N \frac{\mu^2}{N} - \sum_{i=1}^N \frac{2x_i\mu}{N} = \\ &= \sum_{i=1}^N \frac{x_i^2}{N} + \mu^2 - 2\mu \sum_{i=1}^N \frac{x_i}{N} = \sum_{i=1}^N \frac{x_i^2}{N} - \mu^2 = \text{MEDIA DEI QUADRATI} - \text{QUADRATO DELLA MEDIA} \end{aligned}$$

VARIANZA di UNA TRASFORMAZIONE LINEARE: (IMPORTANTE in teoria spesso)

Si consideri $y = a + bx$, $a, b \in \mathbb{R}$. Ricordando la media di una trasformazione lineare

$$\begin{aligned} \text{Var}(y) &= \sum_{i=1}^N \frac{(y_i - \mu)^2}{N} = \sum_{i=1}^N \frac{(a + bx_i - a - b\mu)^2}{N} = \sum_{i=1}^N \frac{(bx_i - b\mu)^2}{N} = b^2 \sum_{i=1}^N \frac{(x_i - \mu)^2}{N} = \\ &= b^2 \text{Var}(x) \end{aligned}$$

NB3 LA VARIANZA NON È UN OPERATORE LINEARE

$$\otimes y = a + bx \rightarrow \text{Var}(y) = b^2 \text{Var}(x)$$

$$\mu_y = a + b\mu_x$$

COEFFICIENTE DI VARIAZIONE→ **CV(X)**

- è un indice relativo di **VARIABILITÀ** ⇒ elimina l'effetto dell'ordine di grandezza di un fenomeno e è **ADIMENSIONALE**
- Strumento per il confronto di variabilità tra due o più variabili
- Strumento per il confronto di variabilità tra fenomeni analoghi rilevati su gruppi distinti di dati statistiche
- Indice che elimina l'effetto dell'ordine di grandezza del fenomeno
- Il coefficiente di variazione è **ADIMENSIONALE**

$$CV(X) = \frac{\sqrt{Var(X)}}{|\mu|} = \frac{\text{Deviazione standard di } X}{|\mu|}$$

EVITA che vi siano **EFFETTO DI COMPENSAZIONE** per i quali la media assume valori prossimi allo 0 rendendo falsamente ampio il coefficiente di variazione

INTERVALLI DI VARIAZIONE→ Δ_c

• Indicatore ottenuto come **DIFFERENZA** tra valori che occupano particolari posizioni nella distribuzione

→ Δ_q • **Lampo di variazione**

$$\Delta_c = x_{(n)} - x_{(1)} = \text{valore massimo} - \text{valore minimo}$$

⊗ Validità limitata poiché non considera gli altri $N-2$ termini della distribuzione

• **Scarto interquartile**

$$\Delta_q = q_3 - q_1$$

⊗ Validità limitata poiché può essere **NULLO** anche se non è nulla la variabilità

MUTABILITÀ⇒ Variabilità per variabili **QUALITATIVE****MINIMA**

si ha quando una sola modalità presenta la max frequenza (pari al tot. delle osservazioni)

MASSIMA

si ha quando tutte le modalità presentano la stessa frequenza (= numero di u.s. per ciascuna modalità)

non è possibile misurare la distanza tra le modalità (rk non sono numeriche)

si valuta la distribuzione delle frequenze sul supporto della variabile

• l'indice di **MUTABILITÀ** deve rappresentare alcune caratteristiche

- * valore minimo in caso di minime mutabilità
- * valori crescenti al crescere della mutabilità
- * valore massimo quando la mutabilità è massima

Distribuzione di frequenza relative $(x_1, f_1) \dots (x_k, f_k)$

2 indici di mutualità

INDICE DI GINI

$$G = 1 - \sum_{i=1}^k f_i^2$$

MIN
MUTUAB

$$G = 1 - (0^2 + \dots + 0^2 + 1^2 + 0^2 + \dots + 0^2) = 0$$

Solo un valore
 $f_k \neq 0$

$$G = 1 - \left(\frac{1}{k^2} + \dots + \frac{1}{k^2} \right) = 1 - k \cdot \frac{1}{k^2} = \frac{k-1}{k}$$

$$f_k = \frac{1}{k} \quad \forall k$$

MAX
MUTUAB

INDICE di ENTROPIA di SHANNON

$$H = - \sum_{i=1}^k f_i \ln f_i$$

(NB) C'è il - perché il $\ln$ di f_i è un valore compreso tra 0 e 1, quindi è un valore negativo. Così cambia il segno

MIN
MUTUAB

$$H = -(0 \ln 0 + \dots + 0 \ln 0 + 1 \ln 1 + 0 \ln 0 + \dots + 0 \ln 0) = 0$$

Solo un valore $f_k \neq 0$

MAX
MUTUAB

$$f_k = \frac{1}{k} \quad \forall k$$

$$H = - \left(\frac{1}{k} \ln \frac{1}{k} + \dots + \frac{1}{k} \ln \frac{1}{k} \right) = -k \cdot \frac{1}{k} \ln \frac{1}{k} = - \ln \frac{1}{k} = \ln k$$

Quindi

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

$$0 \leq H \leq \ln k$$

(NB) INDICI di MUTUAB. NORMALIZZATI

Normalizzazione: trasformazione di un indice I con un proprio campo di variazione $[I_{\min}, I_{\max}]$ in un indice definito sull'intervallo $[0, 1]$

$$\tilde{I} = \frac{I - I_{\min}}{I_{\max} - I_{\min}}$$

IND di GINI NORM: $\tilde{G} = \frac{G}{\frac{k-1}{k}} = \frac{k}{k-1} G$

ENTROPIA di SHANNON NORM: $\tilde{H} = \frac{H}{\ln k}$

(NB) $\tilde{G} = 0 \circ \tilde{H} = 0 \Rightarrow$ MIN MUTUABILITÀ

$\tilde{G} = 1 \circ \tilde{H} = 1 \Rightarrow$ MAX MUTUABILITÀ

STANDARDIZZAZIONE

• Strumento per rendere i dati confrontabili tra loro
(usata molto)

↓
Sia X variabile di media μ e varianza σ^2 :

- **VARIABILE STANDARDIZZATA**: $z = \frac{X - \mu}{\sigma}$ da cui: $X = \mu + z\sigma$

- **MEDIA di Z** : $\mu_z = \frac{\mu - \mu}{\sigma} = 0 \rightarrow \mu_z = -\frac{\mu}{\sigma} + \frac{1}{\sigma} X \cdot \mu = -1 + 1 = 0$

- **VARIANZA di Z** : $Var(z) = Var\left(\frac{X - \mu}{\sigma}\right) = \frac{Var(X)}{\sigma^2} = 1$
 $= Var\left(\frac{X}{\sigma}\right) = \frac{1}{\sigma^2} Var(X) = \frac{\sigma^2}{\sigma^2} = 1$

SCORE

• la quantità: $z_i = \frac{x_i - \mu}{\sigma}$ detta score misura il NUMERO di SCARTI quadratici medi che separano una osservazione x_i dal valore medio μ

↓
• $z_i = k, k > 0$: l'osservazione è sopra la media μ di k volte lo scarto quadratico medio $\Rightarrow \frac{x_i - \mu}{\sigma} = k \Rightarrow x_i = \mu + k\sigma$

• $z_i = k, k < 0$: l'osservazione è sotto la media μ di k volte lo scarto quadratico medio $\Rightarrow \frac{x_i - \mu}{\sigma} = -k \Rightarrow x_i = \mu - k\sigma$

• $z_i = 0, k = 0$: l'osservazione coincide con la media