

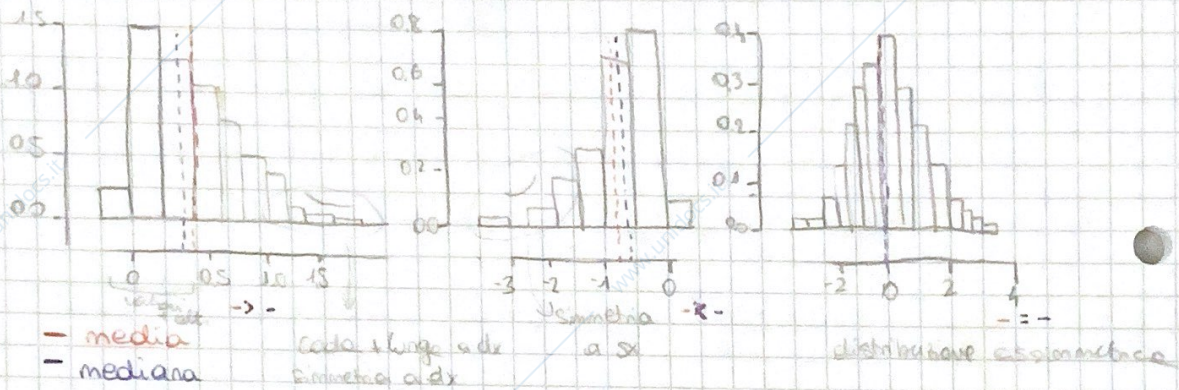
LEZIONE 10

INDICE di ASIMMETRIA e INDICE di FORNA

Si osserva cose accade nelle code delle distribuzioni

ASIMMETRIA

La direzione dell'asimmetria è data dalla coda più lunga



- ASIMMETRIA A DX : MEDIA > MEDIANA
- ASIMMETRIA A SX : MEDIA < MEDIANA
- SIMMETRIA : - MEDIA = MEDIANA ...

1° e 3° QUANTILE hanno la stessa distanza dalla mediana $q_3 - q_2 = q_2 - q_1$

Somma degli scarti della media aritmetica elevati a una potenza dispari è uguale a 0

$$\sum_{i=1}^N (x_i - \mu)^r = 0, r \text{ dispari}$$

1° INDICE di ASIMMETRIA che vediamo: INDICE di PEARSON

$$\alpha_1 = \frac{1}{N\sigma^3} \sum_{i=1}^K (x_i - \mu)^3 n_i$$

Somma di tutte le distanze tra le osservazioni e la media aritmetica prese al cubo (per la distr. assoluta)

deviazione standard al cubo $(\sqrt{\sigma^2})^3$

- indice di quantità
- $\alpha_1 > 0 \Rightarrow$ AS. a dx
 $\alpha_1 < 0 \Rightarrow$ AS. a sx
 $\alpha_1 = 0 \Rightarrow$ Simmetria

$\Rightarrow \alpha_1$ ha campo illimitato, non permette di coprire l'intera dell'asimmetria

2° INDICE che vediamo **INDICE di BOWLEY e JULE**

Si fa sui quartili

$$\alpha_2 = \frac{(q_3 - q_2) - (q_2 - q_1)}{q_3 - q_1} \quad -1 \leq \alpha_2 \leq 1$$

- Se:
- $\alpha_2 > 0 \Rightarrow$ As a dx
 - $\alpha_2 < 0 \Rightarrow$ As a sx
 - $\alpha_2 = 0 \Rightarrow$ Simmetrico

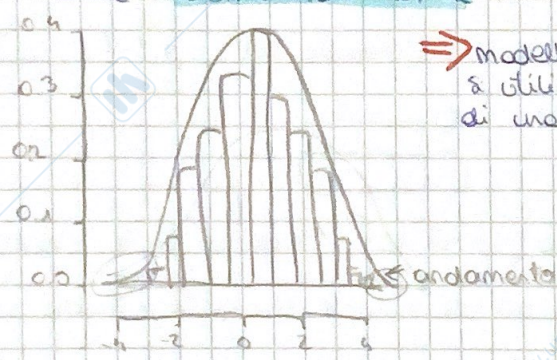
indice robusto perché basato sui quartili (non risente dei valori molto alti o molto bassi delle osservazioni) \neq media aritmetica

INDICI di FORMA o curtosi

Si guardano sempre le code di distribuzione

Si valuta il **PESO** delle code di distribuzione

Si confronta rispetto al modello teorico rappresentato dalle **distribuzione normale**



\Rightarrow modello matematico che si utilizza per l'approssimazione di una distribuzione reale

IND di CURTOSI STANDARDIZZATO di PEARSON

\Rightarrow indice usato per la simmetria. Si basa sulle potenze sparse

$$\text{gamma} \Rightarrow \gamma = \frac{\sum_{i=1}^K (x_i - \mu)^4 n_i}{N \sigma^4} - 3 \Rightarrow \text{logli } 3 \text{ confronto ideale}$$

rapporto tra due indici di variabilità

indice costruito sui nostri dati

momento di confronto con le code della distri normale

- SE**:
- $\gamma > 0 \Rightarrow$ in caso di code + **PESANTI** della normalità: **IPERNORMALITÀ**
 - $\gamma < 0 \Rightarrow$ in caso di code + **LEGGERE** della normalità: **IPONORMALITÀ**
 - $\gamma = 0 \Rightarrow$ distribuzione normale

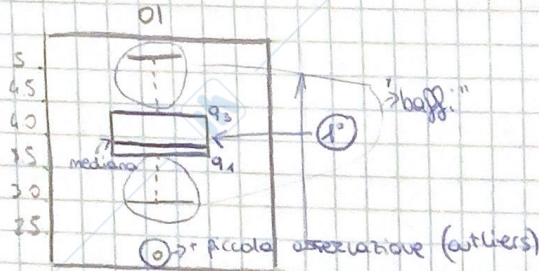
BOXPLOT

per descrivere una variabile

condiziona l'info relativa all'asimmetria

= diagramma a scatola e baffi

➊ Rapp. grafica di variabilità e asimmetria di una distribuzione



• costruito sulle basi dei tre quartili si inizia a fare la scatola partendo da q_1 a q_3 in mezzo c'è la mediana

• I baffi si estendono fino ai DATI PIÙ ESTREMI (non + distanti di $k \times \Delta q$)
quote dalle scatole
 $k=1,5$ a sotto

➋ osservazioni del fenomeno $n = \dots$

le oss. devono essere messe in ordine crescente x fare i quartili!

➌ trovo i valori di q_1 , m e $q_3 \Rightarrow$ ampiezza scatola: $\Delta q = q_3 - q_1$

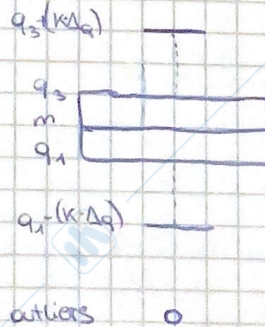
➍ trovo la quota delle scatole (lunghezza baffi): $1,5 \cdot (q_3 - q_1)$

➎ Disegno la scatola:

➏ BAFFO INF: arriva fino all'oss. + piccola tra quelle + grandi di $q_1 - (k \times \Delta q)$

➐ BAFFO SUP: fino all'oss. + grande tra quelle + piccole di $q_3 + (k \times \Delta q)$

➑ tra le osservazioni si trovano quelle che non rientrano tra i baffi e diventano OUTLIERS. Si disegnano fuori



OUTLIERS

valori che si discostano dall'andamento generale delle distrib.

- POSSIBILI CAUSE:
- * Errori di rilevazione osservazione
 - * Circostanze eccezionali dovrebbero esse eliminate se possibile
 - * Contaminazione il cui anomalo proviene da un gruppo specifico all'interno della pop.

- n^* fortemente ANOMALA se

$$|n^* - \mu_{0.25}| > 3(\mu_{0.75} - \mu_{0.25}) \text{ oppure } |n^* - \mu_{0.75}| > 3(\mu_{0.75} - \mu_{0.25})$$

- n^* ANOMALA se

$$|n^* - \mu_{0.25}| > 1.5(\mu_{0.75} - \mu_{0.25}) \text{ oppure } |n^* - \mu_{0.75}| > 1.5(\mu_{0.75} - \mu_{0.25})$$