Si ponga

$x \in \mathbb{R}$,

$r \in \mathbb{N}_0$

$$d_r(x;\underline{y}) = \begin{cases} \sqrt[r]{\frac{1}{N}\cdot\sum_{i=1}^{N}|x-y_i|^r}, & r>0, \\ \frac{1}{N}\cdot\sum_{i=1}^{N}|x-y_i|^0, & r=0. \end{cases}$$

DEFINIZIONE

Il centro di ordine r è quel numero reale ξ_r che rende minima la funzione $d_r(x,\underline{y})$:

$$\xi_r(\underline{y}) := \operatorname*{argmin}_{x\in\mathbb{R}} d_r(x,\underline{y}).$$

RISULTATI

- $\xi_0(\underline{y})$ è la moda di $\underline{y}$.

$$x \in \underline{y} \Rightarrow \exists J \in \{1, 2, \dots, N\} : x = y_J \Rightarrow$$

$$|x - y_J|^0 = 0^0 = 0$$

in quanto è nulla la distanza tra x e y_J.

- $\xi_1(\underline{y})$ è la mediana di $\underline{y}$.

TEOREMA

$\xi_2(\underline{y})$ è la media aritmetica di $\underline{y}$

ovvero

$$\xi_2(\underline{y}) = \bar{y} = \frac{1}{N} \sum_{i=1}^{N} y_i .$$

DIM

Bisogna dimostrare che:

$$\xi_2(\underline{y}) = \underset{x \in \mathbb{R}}{\operatorname{argmin}}\, d_2(\underline{x}; \underline{y})$$

$$= \underset{x \in \mathbb{R}}{\operatorname{argmin}} \sqrt{\frac{1}{N} \sum_{i=1}^{N} (x - y_i)^2}.$$

D'altre parte, posto

$$x \in \mathbb{R}, \quad f(x) := \frac{1}{N} \sum_{i=1}^{N} (x - y_i)^2$$

risulta

$$\xi_2(\underline{y}) = \operatorname{argmin} d_2(\underline{x}; \underline{y}) = \operatorname{argmin} f(x),$$

$\forall x \in \mathbb{R}$ $\forall x \in \mathbb{R}$

in quanto la funzione radice quadrata è strettamente crescente nel suo dominio.

La funzione $f(x)$

- è infinitamente grande nell'intorno di $+\infty$ e nell'intorno di $-\infty$:

$$\lim_{x \to -\infty} f(x) = +\infty \quad , \quad \lim_{x \to +\infty} f(x) = +\infty$$

- è derivabile in tutto $\mathbb{R}$.

Allora la ricerca del suo minimo assoluto è da eseguire tra i punti di minimo relativo.

A tale scopo, si ha:

$$f'(x) = D_x\left[\frac{1}{N}\sum_{i=1}^{N}(x-y_i)^2\right] = \frac{2}{N}\sum_{i=1}^{N}(x-y_i)$$

e

$$f''(x) = D_x\left[\frac{2}{N}\sum_{i=1}^{N}(x-y_i)\right] = \frac{2}{N}\sum_{i=1}^{N}1 = \frac{2N}{N} = 2>0.$$

<u>Punti stazionari</u>

$$f'(x)=0 \iff \frac{2}{N}\sum_{i=1}^{m}(x-y_i) = 0$$

$$\Leftrightarrow \sum_{i=1} x = \sum_{i=1} y_i$$

$$\Leftrightarrow N x = \sum_{i=1}^{N} y_i$$

$$\Leftrightarrow x = \frac{1}{N} \sum_{i=1}^{N} y_i = \bar{y}.$$

<u>Ricerca dei punti di minimo e massimo relativo</u>

$f''(\bar{y}) > 0 \Rightarrow \bar{y}$ è di minimo relativo e, quindi, per quanto osservato in precedenza $\bar{y}$ è il punto di minimo assoluto.

In definitiva:

$$\xi_2(\underline{y}) = \underset{x\in\mathbb{R}}{\arg\min}\, d_2(x,\underline{y}) = \underset{x\in\mathbb{R}}{\arg\min}\, f(x) = \overline{y}.$$

$\square$

TEOREMA

Si ponga

$$\xi_\infty(\underline{y}) := \lim_{r\to\infty} \xi_r(\underline{y}).$$

Risulta:

$$\xi_\infty(\underline{y}) = \frac{y_{(1)} + y_{(N)}}{2}.$$

dove $y_{(1)} = \min(y_1, y_2, \ldots, y_N)$

e

$$y_{(N)} = \max(y_1, y_2, \ldots, y_N).$$

$\square$

La quantità $\dfrac{y_{(1)} + y_{(N)}}{2}$ si dice "valore centrale" di $\underline{y}$.

$\square$

QUARTILI

La mediana di una rilevazione dati è il quartile di ordine 2: Q_2.

La mediana suddivide $\underline{y}$ in due parti: i dati "minori" di essa e i dati "maggiori" di essa.

La mediana dei dati "minori" di Q_2

si indica con il simbolo Q_1: il "quantile di ordine 1".

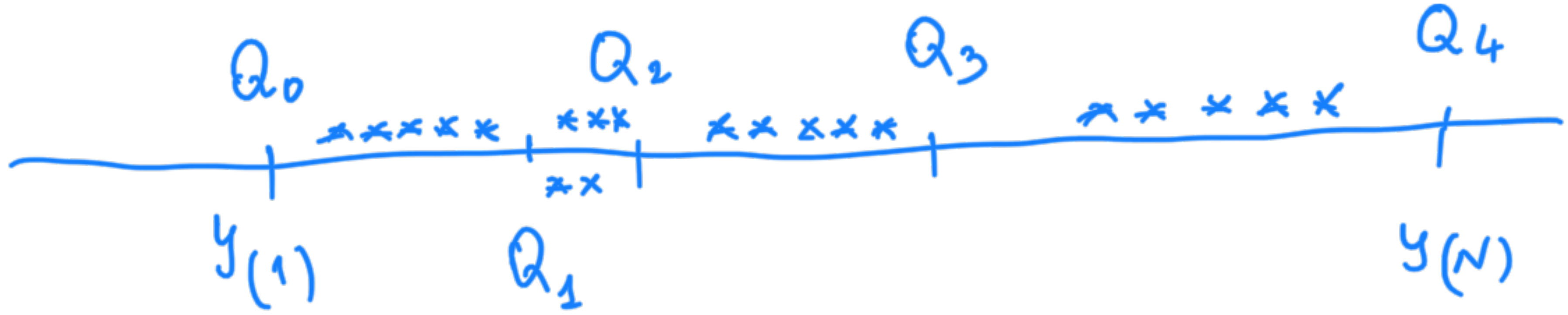
La mediana dei dati maggiori di Q_2 si indica con il simbolo Q_3: il quantile di ordine 3.

Si completa chiamando $y_{(1)}$ quantile di ordine 0

$$Q_0 = y_{(1)},$$

e chiamando $y_{(N)}$ quantile di ordine 4

$$Q_4 = y_{(N)}.$$



INDICI DI DISPERSIONE

Nel caso di un carattere qualitativo quello che viene maggiormente utilizzato come indice di dispersione è l'"indice di ricchezza" ovvero il numero delle modalità con le quali il carattere si manifesta.

L' indice di ricchezza è un esempio del-la famiglia degli "indici di diversità".

Per i caratteri quantitativi ce ne sono invece vari:

1) campo o "intervallo di variazione"

(riferito al valore centrale $\frac{y_{(1)} + y_{(N)}}{2}$)

$$\Gamma := y_{(N)} - y_{(1)} = Q_4 - Q_0 .$$

2) differenza tra il terzo e il primo quartile: "differenza interquartilica"

(riferito alla mediana Q_2)

$$\gamma := Q_3 - Q_1 .$$

3) "scarto mediano assoluto"

(riferito alla mediana Q_2)

$$S_{Q_2} := \frac{1}{N} \sum_{i=1}^{N} |y_i - Q_2| .$$

4) " scarto medio assoluto

(riferito alla media aritmetica $\bar{y}$)

$$S_{\bar{y}} := \frac{1}{N} \sum_{i=1}^{N} |y_i - \bar{y}| .$$

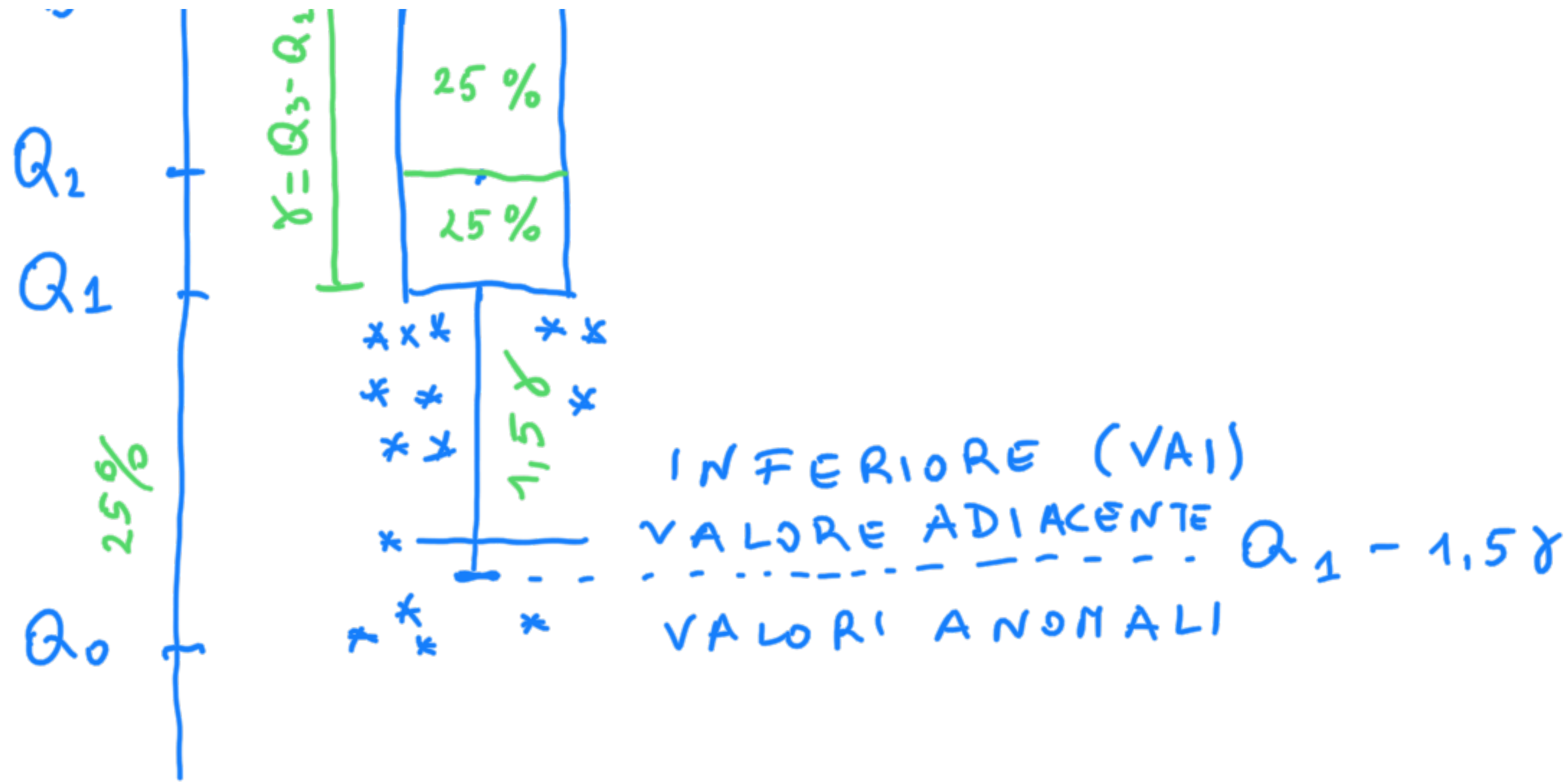
5) " varianza

(riferito alla media aritmetica $\bar{y}$)

$$S^2_{\bar{y}} := \frac{1}{N} \sum_{i=1}^{N} (y_i - \bar{y})^2$$

6) $$\sqrt{S^2_{\bar{y}}} := \sqrt{\frac{1}{N} \sum_{i=1}^{N} (y_i - \bar{y})^2}.$$

DIAGRAMMA SCATOLA CON BAFFI



Q2
Q1
Q0
γ = Q3 - Q1
25 %
25 %
25 %
1,5 γ
INFERIORE (VAI)
VALORE ADIACENTE
Q1 - 1,5 γ
VALORI ANOMALI