

Percentili

Monday, 20 March 2023

14:52

- K-esimo percentile campionario

Questo serve per praticamente togliere il fatto che la media è molto dipendente dai valori estremamenti.

Per farlo, noi prima di tutto dobbiamo comprendere delle terminologie:

Immaginate i percentili come se fossero una mediana che noi possiamo spostare

$$- \quad p = \frac{1}{4}, \quad k = 100 * p = 100 * \frac{1}{4} = 25 \text{ esimo percentile} = \text{primo quartile}$$

$$- \quad p = \frac{1}{2}, \quad k = 100 * p = 100 * \frac{1}{2} = 50 \text{ esimo percentile} = \text{secondo quartile} =$$

$$- \quad p = \frac{3}{4}, \quad k = 100 * p = 100 * \frac{3}{4} = 75 \text{ esimo percentile} = \text{terzo quartile}$$

Ora, reggete queste prossime linee che dopo con l'esempio pratico si capirà tutto
Per calcolarlo dobbiamo far finta che abbiamo una specie di mediana:

1) Ordiniamo i dati $x_1 \leq x_2 \leq \dots \leq x_n$

2) Ed ora come per la mediana, se:

i. N_p non è intero

$$t = x_i$$

ii. N_p è intero

$$t = \frac{x_{N_p} + x_{N_p+1}}{2}$$

Formule finite, ora esempio pratico

-6, -3, -3, 0, 0, 1, 10, 10, 10, 17

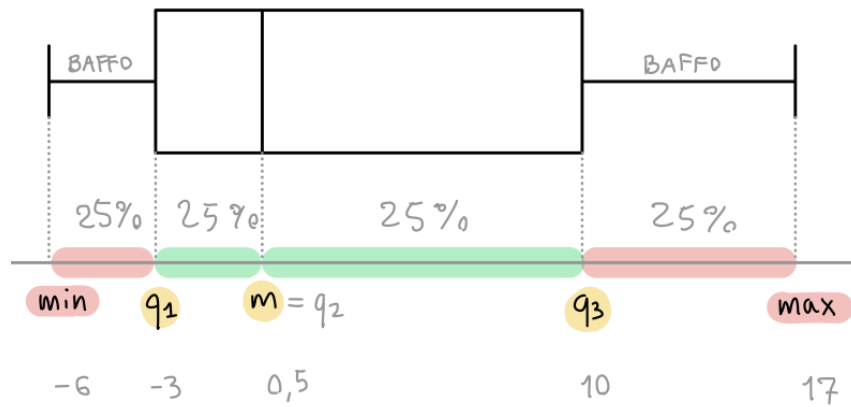
$$p = \frac{1}{2}: \left(N * \frac{1}{2} = 5 \right) \quad m = \frac{x_5 + x_6}{2} = \frac{0 + 1}{2} = 0,5$$

$$p = \frac{1}{4}: \left(N * \frac{1}{4} = 2,5 \right) \quad m = x_3 = -3$$

$$p = \frac{3}{4}: \left(N * \frac{3}{4} = 7,5 \right) \quad m = 10$$

Avendo calcolato $q_1, q_2 = m, q_3$ possiamo disegnare la box plot

Che è una rappresentazione grafica di come i nostri dati sono rappresentati in me



Certe persone tagliano gli outliers, cioè i valori rossi di sopra
R lo fa automaticamente, per non averlo bisogna:

> Boxplot(..., range=0)

Scarto interquartile: $\Delta := q_3 - q_1$

Questo costruisce almeno il 50% dei nostri dati ed è un indice di variabilità

Che misura la dispersione rispetto ad m

- Moda

La moda è il valore che compare con frequenza maggiore, e nota che non è sempre

Quando? 2 valori hanno frequenza massima