Proprietà della devianza

Siano dati N valori numerici reali

$$x_1, x_2, ..., x_N$$

Sussistono le seguenti definizioni

Definizione_1

Si definisce media aritmetica semplice dei numeri x_i il valore numerico

$$x_m = \frac{x_1 + x_2 + ...x_N}{N}$$

Definizione_2

Preso un qualsiasi valore reale *m*, generalmente compreso tra il valore minimo ed il valore massimo dei valori x_i si definisce **scarto semplice** del valore x_i dal numero *m* la differenza x_i-*m*.

In particolare si possono considerare gli scarti semplici dei valori x_i dalla media aritmetica.

***

Proprietà degli scarti semplici dalla media aritmetica

La somma degli scarti semplici dalla media aritmetica dei dati numerici $x_1, x_2, ..., x_N$ è sempre nulla.

Dim.

Osserviamo che per le definizioni precedenti di media aritmetica e di scarto possiamo scrivere

$$\sum_{i=1}^{N}(x_i - x_m) = \sum_{i=1}^{N} x_i - Nx_m = \sum_{i=1}^{N} x_i - \cancel{N} \cdot \frac{x_1 + x_2 + ...x_N}{\cancel{N}} = \sum_{i=1}^{N} x_i - (x_1 + x_2 + ...x_N) = 0$$

***

Definizione_3

Si chiama **devianza** dei dati numerici reali $x_1, x_2, ..., x_N$ la somma dei quadrati degli scarti dei singoli valori numerici dalla media aritmetica. Dunque

$$Dev = \sum_{i=1}^{N}(x_i - x_m)^2$$

***

Spesso si è interessati a valutare gli scarti dei valori numerici $x_1, x_2, ..., x_N$ rispetto ad un indice particolare x_0, considerandone gli scarti semplici $(x_i - x_0)$, o i valori assoluti $|x_i - x_0|$, oppure i loro quadrati

$(x_i - x_0)^2$, per ottenere delle informazioni sulla dispersione dei valori numerici $x_1, x_2, ..., x_N$ rispetto al valore di riferimento x_0. Ebbene, sussiste il seguente

Teorema

La somma dei quadrati degli scarti di un insieme di dati numerici $x_1, x_2, ..., x_N$ da un indice particolare x_0 assunto come riferimento risulta minima solo quando x_0 coincide con la media aritmetica dei valori x_i.

Dim.

Sia x_0 un numero reale qualsiasi ed x_m la media aritmetica dei valori $x_1, x_2, ..., x_N$. Poniamo $x_0 - x_m = d$, dunque $x_0 = x_m + d$.

Per la somma dei quadrati degli scarti dei valori x_i da x_0 possiamo scrivere

$$\sum_{i=1}^{N}(x_i - x_0)^2 = \sum_{i=1}^{N}(x_i - x_m - d)^2 = \sum_{i=1}^{N}\left[(x_i - x_m)^2 - 2(x_i - x_m)d + d^2\right] =$$

$$\sum_{i=1}^{N}(x_i - x_m)^2 - 2d\sum_{i=1}^{N}(x_i - x_m) + Nd^2 = \sum_{i=1}^{N}(x_i - x_m)^2 - 2d \cdot 0 + Nd^2 =$$

$$\sum_{i=1}^{N}(x_i - x_m)^2 + Nd^2 \geq \sum_{i=1}^{N}(x_i - x_m)^2 .$$

L'uguaglianza vale solo se d=0, mentre con d≠0 vale la disuguaglianza stretta.

Concludiamo che la somma dei quadrati degli scarti dalla media aritmetica x_m rappresenta il valore minimo che può assumere la somma dei quadrati degli scarti dei valori numerici considerati valutati rispetto ad ogni altro numero reale x_0.