Esercizio sulla giustificazione della deviazione standard del campione

Si dimostri la disuguaglianza

$$\sum_{i=1}^{N}(x_i-\bar{x})^2 \le \sum_{i=1}^{N}(x_i-x)^2$$

per qualsiasi x.

Con $\bar{x}=\frac{\sum_{i=1}^{N}x_i}{N}$, ovvero la media aritmetica.

(suggerimento si calcoli la seguente uguaglianza e si confronti con quanto sopra: $\sum_{i=1}^{N}(x_i-x)^2=\sum_{i=1}^{N}(x_i-\bar{x}+\bar{x}-x)^2$)

Se vale per ogni x varrà anche X valore centrale della distribuzione di Gauss.

Premessa all'esercizio

Mediante il principio di massima verosimiglianza si osserva che per la distribuzione di Gauss $G_{X,\sigma}(x)=\frac{1}{\sigma\cdot\sqrt{2\pi}}e^{-\frac{(x-X)^2}{2\sigma^2}}$, il parametro σ utilizzato come parametro standard per la distanza dal valore centrale X risulterebbe che è la ***deviazione standard della popolazione*** ovvero per la varianza (σ^2) si ottiene

$$\sigma^2=\frac{\sum_{i=1}^{N}(x_i-X)^2}{N}$$

Si osservi X , il valore di aspettazione centrale della curva di Gauss, è un'aspettativa che abbiamo per dati casuali. Dagli esperimenti ripetuti noi possiamo fornire una stima di tale valore ovvero la media aritmetica $\bar{x}$.

Pertanto questo giustifica, per la stima dell'errore, l'utilizzo della ***deviazione standard del campione,*** mediante la media aritmetica come miglior stima di X ovvero

$$\sigma_x^2=\frac{\sum_{i=1}^{N}(x_i-\bar{x})^2}{N-1}$$

Questo in quanto si osserva che per qualsiasi valore x :

$$\sum_{i=1}^{N}(x_i-\bar{x})^2 \le \sum_{i=1}^{N}(x_i-x)^2 \qquad \textbf{Equ. 1}$$

Espresso in formule sarebbe affermare che la media è quel valore che rende minima la media degli scarti.

Se l'equazione 1 vale per qualsiasi x varrà anche per X il valore di aspettazione.

Pertanto il valore che noi possiamo fornire del parametro σ èsottostimato

$$\frac{\sum_{i=1}^{N}(x_i-\bar{x})^2}{N} \le \frac{\sum_{i=1}^{N}(x_i-X)^2}{N}$$

percui per "riaggiustarlo" lo moltiplichiamo per un numero maggiore (*N/(N-1))*

$$\sigma_x^2=\sigma^2\cdot\frac{N}{N-1}=\sum_{i=1}^{N}(x_i-\bar{x})^2\Big/_{N-1}$$

Per questo motivo utilizzeremo la deviazione standard del campione come miglior stima della distribuzione delle incertezze nella nostra misura.