

Indici di posizione

Leonardo Bizzoni

April 6, 2024

1 Media campionaria

Per descrivere il "*centro*" dell'insieme di dati, definiamo la **media** per un insieme x_1, x_2, \dots, x_N .

$$\bar{x} := \frac{1}{N} \sum_{i=1}^N x_i.$$

Siano X_1, X_2, \dots variabili aleatorie indipendenti e identicamente distribuite. La media \bar{X}_N è un'altra variabile aleatoria $\bar{X}_N = \frac{1}{N} * \sum_{i=1}^N X_i$, dove:

- Valore medio: $E[\bar{X}_N] = \frac{E[X_1]+E[X_2]+\dots+E[X_N]}{N} = \frac{N * E[X_i]}{N} = E[X_i]$
- Varianza: $Var[\bar{X}_N] = \frac{Var[X_1]+Var[X_2]+\dots+Var[X_N]}{N^2} = \frac{N * Var[X_i]}{N^2} = \frac{Var[X_i]}{N}$

2 Mediana campionaria

Un'altra misura del "*centro*" dell'insieme di dati è la **mediana** campionaria.

Per calcolare la mediana, bisogna ordinare i dati x_1, x_2, \dots, x_n e la mediana sarà l'elemento in *posizione centrale*.

- Quando il numero dei dati è **dispari** la mediana è $m := x_{(\frac{N+1}{2})}$.
- Quando il numero dei dati è **pari** la mediana è la media tra i 2 valori centrali: $m = \frac{x_{(\frac{N}{2})} + x_{\frac{N}{2}+1}}{2}$.

3 Percentile campionario (quartili)

Fissiamo un numero $k \in 0, 100$.

Il **percentile** campionario è un valore t per cui:

- almeno il *almeno il* k mo percentile ($100 - k$)

I valori di k più importanti sono: 25, 50, 75. È possibile riscrivere questi valori come $p \in [0, 1] = \frac{k}{100}$.

- $p = \frac{1}{4} : k = 25$ detto **primo** quartile q_1 .
- $p = \frac{1}{2} : k = 50$ detto **secondo** quartile q_2 (*coincide con la mediana*)
- $p = \frac{3}{4} : k = 75$ detto **terzo** quartile q_3 .

3.1 Calcolo del percentile

1. Ordina l'insieme dei dati x_1, x_2, \dots, x_N .
2. Se $N * p$ **non** è intero allora $t = x_{(i)}$ dove $i = \text{round}(p)$.
3. Se invece è intero allora $t = \frac{x_{(N*p)} + x_{(N*p+1)}}{2}$, ovvero la media dell'elemento in posizione $N * p$ e del dato successivo.

4 Moda

Se tra i dati c'è un valore che compare con frequenza maggiore esso prende il nome di moda.