Statistica Descrittiva

Misure, indici (numerici) che descrivono le caratteristiche della distribuzione di una o più variabili in modo sintetico.

- indici di posizione o centralità:

  valore centrale, medie algebriche, mediana, moda
  (detti anche *misure di intensità, centri . . .)*

- indici di dispersione o variabilità:

  intervallo di variazione, varianza, varianza stimata, deviazione standard, deviazione standard stimata

- indici di simmetria o asimmetria: . . .

Valore Centrale

Dato l'insieme di valori $\{x_1, x_2, \ldots, x_n\}$, il valore centrale considera solo i due valori estremi (non tiene conto di tutti i valori):

$$\frac{x_{\text{max}} + x_{\text{min}}}{2}$$

dove $x_{\text{max}} = \max\{x_1, x_2, \ldots, x_n\}$ e $x_{\text{min}} = \min\{x_1, x_2, \ldots, x_n\}$.

Esempio: {3, 20, 27, 25, 30, 310}

$$\frac{x_{\text{max}} + x_{\text{min}}}{2} = \frac{310 + 3}{2} = 156.5$$

Media Aritmetica

MEDIA SEMPLICE: dato l'insieme di valori $\{x_1\,,\,x_2\,,\,\ldots\,,\,x_n\}$

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{x_1 + x_2 + \ldots + x_n}{n}$$

MEDIA PONDERATA (dati raggruppati): dato l'insieme di valori $\{x_1\,,\,x_2\,,\,\ldots\,,\,x_m\}$ con le rispettive frequenze assolute $\{f_1\,,\,f_2\,,\,\ldots\,,\,f_m\}$

$$\bar{x} = \frac{\sum_{i=1}^{m} f_i\,x_i}{\sum_{i=1}^{m} f_i} = \frac{1}{n}\cdot\sum_{i=1}^{m} f_i\,x_i = \frac{f_1\,x_1 + f_2\,x_2 + \ldots + f_m\,x_m}{n}$$

Media Aritmetica – Esercizi

Esercizio 1. Dato l'insieme di valori $\{12, 25, 37, 41, 0, 53\}$, calcolare la media aritmetica. [media aritmetica $= 28$]

Esercizio 2. Dato l'insieme di valori $\{28, 28, 28, 28, 28, 28\}$, calcolare la media aritmetica.

Esercizio 3. (dati raggruppati) In un campione di 200 persone si sa che 20 pesano 50kg, 30 pesano 55kg, 50 pesano 60kg, 70 pesano 65kg, 20 pesano 75Kg e 10 pesano 80kg. Calcolare il peso medio. [peso medio $= 62.5$Kg]

Media Aritmetica – Osservazioni

Alcune osservazioni:

- la media può non appartenere all'insieme dei dati
- insiemi di dati diversi possono avere la stessa media
- utilizza tutti i dati
- centro di gravità dei dati
- riduce l'effetto dei dati estremi *(outlier)*

Proprietà:

1) se applico una trasformazione lineare ai dati:

$$y_i = a\,x_i + b \quad \Rightarrow \quad \bar{y} = a\,\bar{x} + b$$

2) la somma degli scarti dalla media è nulla: $\displaystyle\sum_{i=1}^{n} (x_i - \bar{x}) = 0$

3) la somma dei quadrati degli scarti dalla media è minima:

$$\sum_{i=1}^{n} (x_i - x)^2 \text{ assume il valore minimo per } x = \bar{x}$$

Media Aritmetica – Osservazioni

- La somma degli scarti dalla media è nulla:

$$\sum_{i=1}^{n}(x_i - \bar{x}) = \sum_{i=1}^{n} x_i - \sum_{i=1}^{n} \bar{x} = \sum_{i=1}^{n} x_i - n\,\bar{x} = \sum_{i=1}^{n} x_i - n \cdot \frac{1}{n} \cdot \sum_{i=1}^{n} x_i = 0$$

- La somma dei quadrati degli scarti dalla media è minima:

poniamo $g(x) = \sum_{i=1}^{n}(x_i - x)^2$. Abbiamo che

$$g(x) = \sum_{i=1}^{n}(x_i)^2 - 2\sum_{i=1}^{n} x_i\, x + \sum_{i=1}^{n} x^2 = \sum_{i=1}^{n}(x_i)^2 - 2n\bar{x}\,x + nx^2$$

Quindi, g è un polinomio di secondo grado in x.
Pertanto, assume il suo valore minimo in $x = -\frac{-2n\bar{x}}{2n} = \bar{x}$.

Media Geometrica

MEDIA SEMPLICE: dato l'insieme di valori $\{x_1, x_2, \ldots, x_n\}$ con la condizione che siano *tutti positivi*

$$x_g = \sqrt[n]{\prod_{i=1}^{n} x_i} = \sqrt[n]{x_1 \cdot x_2 \cdots x_n} \quad \Rightarrow \quad \log x_g = \frac{1}{n} \cdot \sum_{i=1}^{n} \log x_i$$

MEDIA PONDERATA: dato l'insieme di valori $\{x_1, x_2, \ldots, x_m\}$, tutti positivi, con le rispettive frequenza assolute $\{f_1, f_2, \ldots, f_m\}$

$$\log x_g = \frac{1}{n} \cdot \sum_{i=1}^{m} f_i \log x_i$$

Mediana

Dato l'insieme di valori ordinati $x_1 \leq x_2 \leq \cdots \leq x_{n-1} \leq x_n$, si chiama mediana (o valore mediano) il valore M_e che occupa la posizione centrale:

- se n è dispari, c'è un unico termine mediano di posto $\frac{n+1}{2}$

$$M_e = x_{\frac{n+1}{2}}$$

- se n è pari ci sono due termini mediani di posti $\frac{n}{2}$ e $\frac{n}{2}+1$

$$M_e = \frac{1}{2}\left(x_{\frac{n}{2}} + x_{\frac{n}{2}+1}\right)$$

Utilizza tutti i valori ma si basa soltanto sull'ordinamento degli stessi.

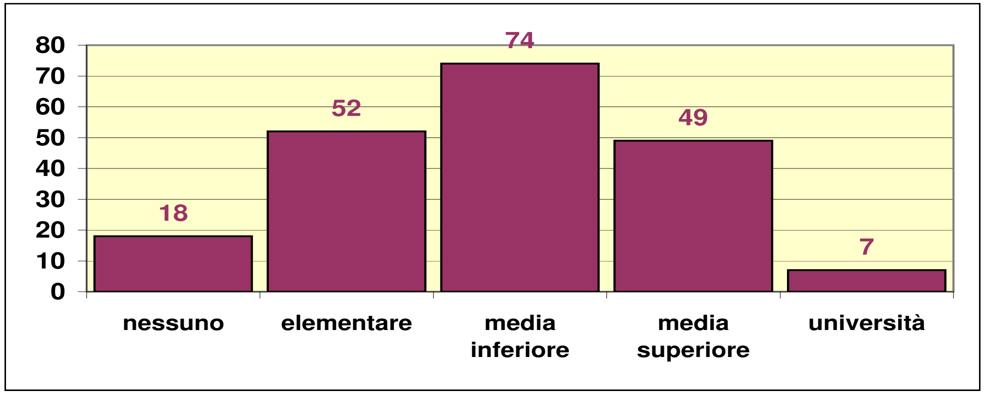
Esempio 1. $\{0, 13, 25, 81, 503\} \quad \Rightarrow \quad M_e = 25$

Esempio 2. $\{1, 2, 81, 93, 327, 503\} \quad \Rightarrow \quad M_e = 87$

Moda

MODA: valore (o classe) al quale è associata la frequenza più alta

titolo di studio	fi	fi / n
nessuno	18	0,09
elementare	52	0,260
media inferiore	74	0,370
media superiore	49	0,245
università	7	0,035
	200	1,000



Si può applicare anche a dati qualitativi espressi su scala nominale.