MEDIA E MEDIANA

Queste brevi note hanno lo scopo di raccogliere le definizioni di Media e di Mediana date in classe, assieme ad alcuni esempi ed esercizi.

In generale, la media e la mediana sono singoli valori numerici che si introducono per descrivere in maniera sintetica un insieme, nel nostro caso discreto, di dati numerici.

1. Media

Ci sono diversi tipi di media disponibili sul mercato. Noi useremo la cosiddetta media aritmetica, così definita

Definizione Dato un insieme D formato da numeri $\{x_i\}_{i=1}^{N}$, la media (aritmetica) è la somma degli x_i divisa per N, più precisamente

$$M := \frac{1}{N}\sum_{i=1}^{N} x_i = \frac{x_1 + x_2 + \ldots + x_N}{N}. \tag{1.1}$$

`eq:media`

È importante notare fin da subito che nella definizione precedente non si suppone assolutamente che i valori $\{x_i\}$ siano distinti e dunque nella somma in (1.1) gli x_i vengono sommati secondo la loro molteplicità. Ad esempio, supponiamo di voler calcolare la media del seguente insieme di numeri

$$D = \{1, 1, 1, 2, 3, 4, 4, 4\}.$$

Si ha $N = 8$ e dunque la definizione fornisce

$$M = \frac{1}{8}(1 + 1 + 1 + 2 + 3 + 4 + 4 + 4) = \frac{1}{8}(1 \times 3 + 2 + 3 + 4 \times 3) = \frac{20}{8} = \frac{5}{2}. \tag{1.2}$$

`eq:esempio1`

La molteplicità con cui un dato numerico compare nell'insieme D in questione è chiamata frequenza (assoluta)[1]. In particolare, nell'esempio precedente, 1 e 4 hanno frequenza 3, mentre 2 e 3 hanno frequenza 1. Si noti inoltre che N risulta essere la somma delle frequenze con cui compaiono i valori $1, 2, 3, 4$.

A seconda del contesto e del significato che si attribuisce ai dati numerici x_i, le frequenze con cui essi compaiono vengono chiamate pesi e la relativa media viene chiamata media pesata. Il seguente esempio dovrebbe chiarire la situazione. Supponiamo che un certo percorso universitario preveda il superamento di tre esami. Il primo esame, indicato con A, vale 9 crediti; il secondo esame, indicato con B, vale 6 crediti e l'ultimo esame, indicato con C, vale 3 crediti. Un certo studente, ha totalizzato 18/30 nel primo esame, 24/30 nel secondo esame e 30/30 nel terzo esame. Vogliamo calcolare la media dei voti dello studente in questione. Appare subito evidente che una buona descrizione deve tenere in considerazione il fatto che i singoli esami, A, B e C, contano

[1] Accanto alla frequenza assoluta, in statistica si parla anche di frequenza relativa. Detta f_i la frequenza assoluta dell'elemento i-esimo, la sua frequenza relativa è data da

$$fr_i := \frac{f_i}{\sum_{i=1}^{N} f_i}.$$

diversamente in base al numero di crediti che a loro viene attribuito. Il concetto di media da utilizzare è dunque quello di media pesata, vale a dire

$$MP := \frac{1}{\sum_{i=1}^{N} p_i} \sum_{i=1}^{N} x_i p_i. \tag{1.3}$$ eq:media_pesata

Operativamente, per calcolare MP si sommano tutti i valori x_i moltiplicati per i loro pesi e si divide per la somma dei pesi. Nel caso dell'esempio, abbiamo

$$D = \{18, 24, 30\},$$

mentre l'insieme dei pesi è dato da

$$P = \{9, 6, 3\}.$$

La media (pesata) è dunque

$$MP = \frac{1}{18}(18 \times 9 + 24 \times 6 + 30 \times 3) = \frac{396}{18} = 22. \tag{1.4}$$ eq:media_pesata_es2

È importante notare che avremmo ottenuto lo stesso risultato se invece di considerare distintamente l'insieme dei voti (D) e l'insieme dei pesi (P) avessimo considerato l'insieme dei voti contati con la loro frequenza (i pesi), vale a dire l'insieme

$$D = \{18, 18, 18, 18, 18, 18, 18, 18, 18, 24, 24, 24, 24, 24, 24, 30, 30, 30\}.$$

La media, secondo la definizione (1.1), è esattamente la media calcolata in (1.4) usando il concetto di media pesata. Si consiglia di risolvere i seguente problemi

prob:1 **Problema** In una famiglia composta da due adulti e tre adolescenti, la spesa media mensile per i cellulari è di 20 euro per ciascun adulto e di 30 euro per ciascun adolescente. Calcolare la spesa media mensile della famiglia per il cellulare. [*Risposta:* 26 *euro*]

prob:2 **Problema** Se alla fine della loro carriera universitaria i due studenti A e B hanno raggiunto la media di 27/30 e la media di 28/30, rispettivamente, si può affermare che lo studente A ha totalizzato voti più bassi dello studente B?

2. Mediana

Viene organizzata una lotteria in un gruppo di 100 persone. Ognuno compra un singolo biglietto al costo di 10 euro e l'intero ricavato viene investito nel monte premi che dunque è di 1000 euro. Vengono preparati 100 biglietti di cui solo uno risulta vincente e ogni membro del gruppo estrae un biglietto. Vogliamo calcolare la media delle vincite. Per modellizzare la lotteria, indichiamo con x_i la vincita del giocatore i-esimo e supponiamo che il primo giocatore sia quello vincente, cioè $x_1 = 1000$. Abbiamo dunque la seguente situazione

$$x_i = \begin{cases} 1000, & \text{se } i = 1, \\ 0 & \text{se } i \neq 1. \end{cases}$$

La vincita media è dunque la seguente

$$M = \frac{1}{100} \sum_{i=1}^{100} x_i = \frac{1}{100}(1000 + 0 + \ldots + 0) = 10. \tag{2.1}$$ eq:vincita_media

La domanda che ci poniamo è se la media delle vincite riesce a fornire una buona descrizione della lotteria. Si noti a tal proposito che nel gruppo di 100 giocatori ci sono 99 perdenti (investono 10 euro per il biglietto e vincono 0 euro) ed uno solo vincente che a fronte di un investimento di 10 euro per il biglietto vince 1000 euro. La singola vincita di 1000 euro, a fronte di 99 vincite di 0 euro costituisce un esempio di *outlier*, cioè un elemento dell'insieme di dati molto distante da tutti gli altri. In presenza di *outliers*, al posto della media risulta essere più significativo usare il concetto di mediana. Per prima cosa introduciamo il concetto di insieme ordinato di dati.

def:insieme_ord **Definizione** Un insieme $D = \{x_i\}_{i\in\mathbb{N}}$ con $x_i \in \mathbb{R}$ è detto ordinato in maniera crescente se vale

$$x_i \leq x_{i+1}, \quad \text{per } i \in \mathbb{N}. \tag{2.2}$$ eq:crescente

Se vale invece

$$x_{i+1} \leq x_i, \quad \text{per } i \in \mathbb{N}, \tag{2.3}$$ eq:decrescente

l'insieme D è detto ordinato in maniera decrescente.

La mediana risulta così definita.

Definizione Dato un insieme finito e ordinato di dati $D = \{x_i\}_{i=1}^N$ con $x_i \in \mathbb{R}$ per ogni i, La mediana è

$$Me := \begin{cases} x_{\frac{N+1}{2}} & \text{se } N \text{ è dispari }, \\ \frac{1}{2}(x_{\frac{N}{2}} + x_{\frac{N}{2}+1}) & \text{se } N \text{ è pari.} \end{cases} \tag{2.4}$$ eq:calcolo_mediana

Si noti che nel caso in cui N sia dispari allora la mediana è il valore, appartenente a D che occupa la posizione centrale. Se invece N è pari si può pensare di dividere l'insieme di dati in esattamente due insiemi ordinati contenenti lo stesso numero di elementi, più precisamente l'insieme

$$D_1 = \left\{x_1, \ldots, x_{N/2}\right\}$$

e l'insieme

$$D_2 = \left\{x_{N/2+1}, \ldots, x_N\right\}.$$

La mediana è la media tra $x_{N/2}$ ed $x_{N/2+1}$.
Calcoliamo la mediana delle vincite nel caso della lotteria. Per prima cosa ordiniamo in maniera, ad esempio decrescente, l'insieme delle vincite D. Abbiamo cosí

$$D = \{1000, 0, 0, 0, \ldots, 0, 0\}.$$

Visto che $N = 100$ è un numero pari la mediana risulta essere

$$Me = \frac{1}{2}(x_{50} + x_{51}) = \frac{1}{2}(0 + 0) = 0.$$

prob:perdite **Problema** Calcolare la media delle perdite della lotteria precedente.