

Elementi di Statistica

(a cura Prof.ssa M. Cola)

La raccolta d'informazioni e in particolare di dati numerici è lavoro quotidiano di chi si occupa di scienza. I dati, una volta raccolti, devono essere organizzati in tabelle e per essere studiati e compresi ci si avvale dei metodi della statistica.

La statistica è la scienza che studia i dati che si riferiscono a fenomeni di cui si desidera avere una comprensione e che bisogna sintetizzare ricorrendo a metodi matematici.

Le rappresentazioni grafiche accompagnano, quando non li sostituiscono, i dati raccolti in censimenti, indagini, rilevamenti e serie storiche d'informazioni economiche o ambientali (vedi il modulo "Rapporti e Proporzioni").

I dati raccolti e le tabelle, che sono alla base della rappresentazione grafica delle informazioni, non sempre sono d'immediata comprensione.

Il primo passo per ottenere un buon grafico è di organizzare le informazioni.

Lo strumento più utilizzato è una tabella a doppia entrata formata da un certo numero di righe e colonne. La prima riga e la prima colonna contengono i descrittori dei dati raccolti e nelle caselle, punto d'incontro di righe e colonne, si scrivono le occorrenze, la numerosità di quella particolare informazione.

Con il termine di **frequenza assoluta o frequenza**, s'indica quante volte un dato o una particolare informazione si presenta in un insieme di dati.

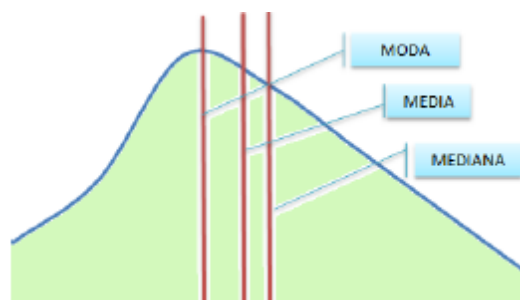
La **frequenza relativa** è la frequenza assoluta divisa per il numero totale di dati.

INDICI SINTETICI

Per sintetizzare i dati numerici sono disponibili degli indicatori sintetici.

Di un insieme di dati tabulati può essere interessante, ad esempio, individuare il valore minimo e massimo, l'intervallo in cui sono compresi e il valore medio.

Esistono vari tipi di medie. Quelle più usate sono la media aritmetica, la moda e la mediana.



La **media aritmetica** è il quoziente ottenuto dividendo la somma dei valori dei dati disponibili, per il numero dei dati. La formula della media aritmetica è:

$$\bar{x} = \frac{x_1 + x_2 + \dots + x_N}{N}$$

Esempio

Valori [7, 6, 8, 9, 7, 5], media = $7 + 6 + 8 + 9 + 7 + 5 : 6 = 42 : 6 = 7$

La **moda** è il dato che ha la massima frequenza.

In un istogramma la moda corrisponde al rettangolo di altezza massima ed è facilmente individuabile. La moda non è un dato calcolato ma di posizione e può anche mancare. Una distribuzione è detta **unimodale** se ha una sola moda. Ad esempio la serie [6, 7, 7, 7, 8, 9] ha moda 7 essendo tale valore ripetuto il più ricorrente. Una distribuzione è detta **bimodale** se ha due mode. Ad esempio la serie [6, 7, 7, 8, 8, 9] ha moda 7 e 8 essendo tali valori i più frequenti. Una distribuzione si dice plurimodale se ha più di due mode.

In una serie ordinata di dati in modo crescente, la **mediana** è il dato che occupa la posizione centrale.

Nel caso di serie formata da un numero pari di valori, la mediana è data dalla media aritmetica dei due valori centrali. La mediana è un indice di posizione (cioè dipende dai valori “centrali” della serie ordinata) e, come la moda, non è influenzata dai valori estremi. Per calcolare la mediana di un insieme di dati, occorre:

- 1 - disporre i valori in ordine crescente oppure decrescente e contare il numero totale di dati;
- 2 - se il numero di dati è dispari, la mediana corrisponde al valore numerico del dato centrale;
- 3 - se il numero di dati è pari, la mediana è calcolata come media aritmetica dei due valori centrali.

La serie [6, 7, 8, 5, 9] una volta riordinata ha mediana 7 essendo i valori in numero dispari e avendo tale valore tanti valori inferiori quanti superiori a esso. La serie [6, 7, 8, 10, 5, 9, 9, 6] una volta riordinata ha mediana 7,5. Essendo, infatti, i valori in numero pari occorre calcolare la media aritmetica dei due valori centrali 7 e 8.

Esercizi di statistica.

1. In una classe di venticinque studenti viene svolta un’indagine sul numero di ore passate giornalmente da ciascuno davanti alla televisione. Si è ottenuta la serie di dati seguente dove ogni valore indica la stima delle ore di televisione per ciascuno dei componenti della classe.

1 3 4 2 1 3 3 5 2 2 4 2 1 3 3 4 3 4 1 2 1 1 1 3 1

Costruisci una tabella delle frequenze e un istogramma dei dati. Calcola la media, la moda e la mediana della distribuzione.

2. In un corso post universitario su venticinque prove e con votazione in trentesimi Marco ha riportato i seguenti voti. Completa la tabella e costruisci un istogramma dei dati. Calcola la media e la moda della distribuzione.

voto	frequenza	frequenza %
20	5	20%
24	4	
25	8	
28	3	
30	5	
Totale		100%

3. Al termine di un corso di specializzazione viene richiesto ai partecipanti di esprimere, in modo anonimo, una valutazione del corso attribuendo un giudizio tra positivo (6) e negativo (1).

Completa la tabella con le frequenze relative e percentuali. Costruisci un istogramma dei dati. Calcola la media e la moda della distribuzione.

Giudizio	Numero	Frequenza relativa	Frequenza percentuale
1	1		5%
2	2	1/10	
3	3		
4	8		
5	4		
6	2		
Totale			100%

4. Di un campione di alunni estratto casualmente viene pesata la cartella. Si è ottenuta la serie di dati seguente dove ogni valore indica il peso in chilogrammi:

12 14 12 9 10 12 10 11 8 10 8 12 12 13 14 8 8 10 10 14 10 13 14 10.

Costruisci una tabella delle frequenze e un istogramma dei dati. Calcola la media, la moda e la mediana della distribuzione