
STATISTICA

Corso A

Autore

Giuseppe Acocella

2024/25

Ultima Compilazione - February 24, 2025

Contents

1	Statistica Descrittiva	3
1.1	Frequenze e Campioni	3
1.2	Caratteri e Rappresentazioni Grafiche	3
1.2.1	Classi di grafici	5
1.3	Indici	6
1.3.1	Indici di Centralità - (Media, Mediana, Moda)	6
1.3.2	Indici di Dispersione - (Varianza, Deviazione Standard)	7
1.4	Funzione di Ripartizione Empirica (FDR/ECDF)	7
1.5	Prova	7

1 Statistica Descrittiva

Questo ramo della statistica cerca di raccogliere dati per descrivere degli oggetti. Elenchiamo delle definizioni standard:

1. **Popolazione:** Insieme di oggetti da studiare.
2. **Carattere:** Caratteristiche degli oggetti della popolazione.

(a) Colore di una biglia, altezza di un individuo.

Ricordiamo che un carattere può essere sia **qualitativo** (es. colore) sia **quantitativo** (es. altezza).

3. **Modalità:** Possibili valori che il carattere può assumere.
(a) Colore biglia istanziato: rosso, blu. Lancio moneta istanziato: testa/croce.
4. **Campione Statistico (Sample):** Sottoinsieme della popolazione scelto a rappresentarla.
5. **Dati:** Esiti delle misure del carattere sugli individui del campione.
(a) Lanci moneta: T, C, T, T, T, C, \dots
6. **Taglia Campione:** Numero di elementi nel campione.

1.1 Frequenze e Campioni

Abbiamo due tipi di frequenze:

1. **Frequenza Assoluta:** Corrisponde al numero di volte in cui la **modalità** appare nei **dati**:

$$\#\{ i \mid x_i = a \}$$

2. **Frequenza Relativa:** Corrisponde al numero di volte in cui la **modalità** appare nei dati fratto il numero dei dati stessi:

$$\text{frequenza relativa} = \frac{\text{frequenza assoluta di } a}{\text{taglia campione}}$$

1.2 Caratteri e Rappresentazioni Grafiche

La rappresentazione dei dati dipende fortemente dal tipo di **carattere**:

1. **Carattere Discreto:** Quantità piccola e finita di modalità assumibili.
(a) Lancio di un dado, esiti di un sondaggio.

In questo caso per le rappresentazioni si utilizzano **diagrammi a barre**.

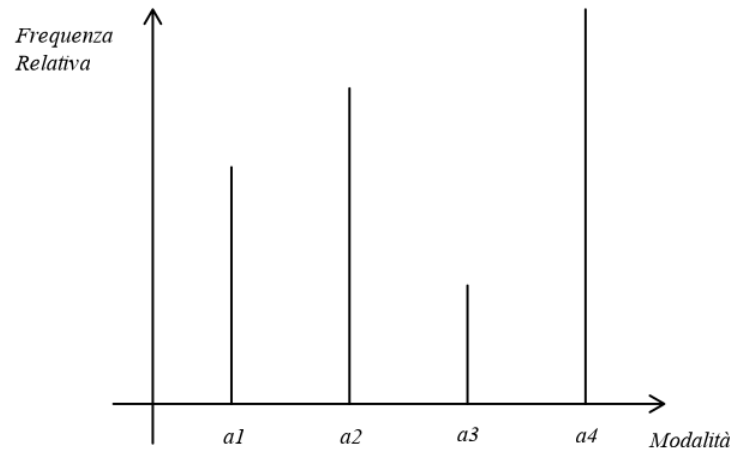


Figure 1: Esempio di diagramma a barre.

2. **Carattere Continuo:** Quantità assumibili in un intervallo continuo.

(a) Altezza della popolazione.

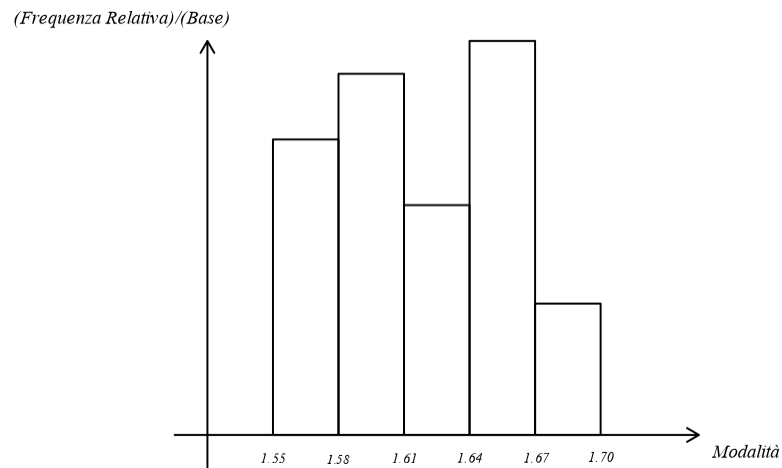


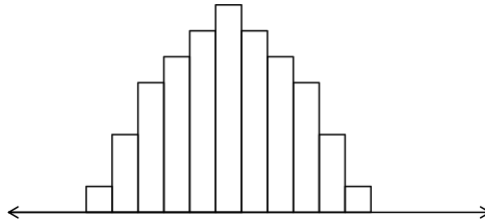
Figure 2: Esempio di istogramma.

La scelta di mettere sull'asse y il rapporto tra freq. relativa e base non è casuale, infatti se scegliessimo intervalli di ampiezza diversa si andrebbe in contro ad un errore di rappresentazione.

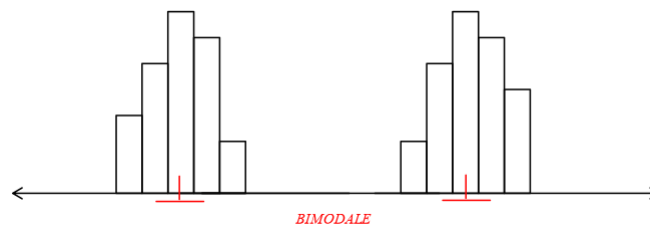
1.2.1 Classi di grafici

Elenchiamo qualche classificazione di rappresentazioni grafiche:

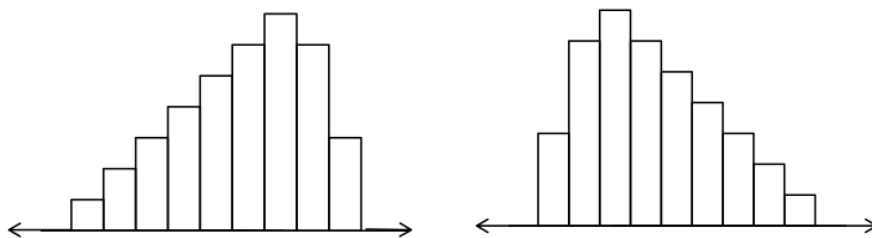
1. **Normale:** Simile ad una campana simmetrica:



2. **Uni/Bi/Tri Modale:** Si concentra attorno ad un numero k di colonne più alte:



- (a) **Modale Asimmetrico S_x/D_x :** Si concentrano attorno ad una colonna più alta in maniera asimmetrica:



1.3 Indici

Gli indici statistici sono quantità numeriche che riassumono proprietà significative sulla **distribuzione dei dati**.

1.3.1 Indici di Centralità - (Media, Mediana, Moda)

Descriviamo tre tipi di indici di centralità:

1. **Media Campionaria:** Descriviamo questo indice:

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$$

ossia semplicemente la media aritmetica dei dati. Un modo **alternativo** è rappresentarlo così:

$$\frac{\text{modalità} * \text{frequenza ass. della modalità}}{\text{quantità di dati}}$$

che formalmente si esprime così:

$$\bar{x} = \sum_{j=1}^M a_j p(a_j)$$

dove a_j sta per modalità e $p(a_j)$ sta per frequenza relativa della modalità.

Sensibilità ai Valori Estremi Una delle caratteristiche della media campionaria è quella di essere molto sensibile ai valori estremi del campione.

Caratteristiche Riesce a vedere tutti i dati del campione e gode di alcune proprietà matematiche come la linearità.

2. **Mediana Campionaria:** Il dato x_i sarà **centrale**, dunque avrà metà dei dati a sinistra e metà a destra. La calcoliamo dunque in due modi:

- (a) **Numero dispari di modalità:** Dato centrale.

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

- (b) **Numero pari di modalità:** Media tra i due dati centrali.

$$\text{mediana} = \frac{1}{2}(x_{\frac{n}{2}} + x_{\frac{n}{2}+1})$$

Caratteristiche Più robusta rispetto ai valori estremi.

3. **Moda:** Modalità più frequente tra i dati.

1.3.2 Indici di Dispersione - (Varianza, Deviazione Standard)

Gli indici di dispersione ci permettono di stabilire quanto i valori della distribuzione si allontanano da un valore centrale scelto come riferimento. Elenchiamoli:

1. **Varianza Campionaria/Empirica:** Permette di

$$\text{CAMPIONARIA: } Var(x) = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2$$

$$\text{EMPIRICA: } Var_e(x) = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 = \left(\frac{1}{n} \sum_{i=1}^n x_i^2 \right) - \bar{x}^2$$

E' possibile calcolare la varianza anche con le frequenze relative:

$$Var_e(x) = \left(\sum_{j=1}^M a_j^2 * p(a_j) \right) - \bar{x}^2$$

2. **Scarto Quadratico Medio:** Indice basato sulla varianza.

$$\sigma(x) = \sqrt{Var(x)}$$

3. **Indice Campionario di Asimmetria:** Un indice che permette di stabilire se una distribuzione sia o meno asimmetrica:

$$b = \frac{1}{b^3} \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^3$$

(a) $b > 0$: Distribuzione Asimmetrica a destra.

(b) $b < 0$: Distribuzione Asimmetrica a sinistra.

1.4 Funzione di Ripartizione Empirica (FDR/ECDF)

Dati x_1, x_2, \dots, x_n dati quantitativi definiamo $F_e : \mathbb{R} \rightarrow \mathbb{R}$ data da

$$F_e(t) = \frac{\#\{i \mid x_i \leq t\}}{n}$$

1.5 Prova