

# Statistica

Ede Boanini

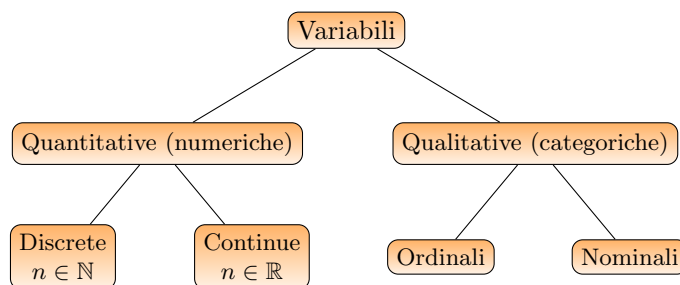
18 novembre 2025

## Indice

<b>1</b>	<b>Introduzione</b>	<b>2</b>
1.1	Classificazione delle Variabili . . . . .	2
1.2	Distribuzioni di Frequenza . . . . .	2
1.2.1	Tipi di Frequenza . . . . .	3
<b>2</b>	<b>Statistica Descrittiva</b>	<b>4</b>
2.1	Diagrammi a barre vs Istogrammi . . . . .	4
2.2	Media . . . . .	5
2.3	Moda . . . . .	5
2.4	Mediana . . . . .	5
2.5	Quartili . . . . .	6
2.6	Campo di Variazione / Range . . . . .	7
2.7	Differenza Interquartile . . . . .	7
2.8	Varianza . . . . .	7
2.9	Deviazione standard . . . . .	7
2.10	Coefficiente di variazione . . . . .	7
<b>3</b>	<b>Probabilità</b>	<b>8</b>
<b>4</b>	<b>Indipendenza e Probabilità Condizionata</b>	<b>8</b>
<b>5</b>	<b>Variabili Casuali</b>	<b>8</b>
5.1	Famiglie Parametriche . . . . .	8
<b>6</b>	<b>Inferenza Statistica</b>	<b>8</b>
6.1	Stima Puntuale . . . . .	8
6.2	Stima Intervallare . . . . .	8
6.3	Verifica delle Ipotesi . . . . .	8

# 1 Introduzione

## 1.1 Classificazione delle Variabili



Differenza tra ordinali e nominali:

- **Ordinali:** categorie che hanno un ordine, puoi solo dire se un valore è minore o maggiore rispetto ad un altro.
  - *Livello di istruzione:*  $elementare < media < \dots$
  - *Grado di soddisfazione:*  $nulla < basso < medio < \dots$
  - *Classifica di una gara:*  $quinto < quarto < \dots$
  - *Matricola:*  $17345 < 17346 < \dots$
- **Nominali:** categorie che non hanno un ordine.
  - *Colore occhi:*  $blu, verdi, marroni, \dots$
  - *Genere:*  $M, F$
  - *Marche auto:*  $Toyota, Ford, \dots$
  - *Nazionalità:*  $Giapponese, Italiano, \dots$

## 1.2 Distribuzioni di Frequenza

È una tabella che contiene modalità e frequenze.

Modalità di X ( $x_i$ )	Frequenze assolute $f_i$
$x_1$	$f_1$
$x_2$	$f_2$
$\dots$	$\dots$
$x_n$	$f_n$
	$N$

### 1.2.1 Tipi di Frequenza

1. **Frequenza assoluta:** numero di ripetizioni di una certa modalità (es: quanti studenti hanno preso 28 all'esame)

$$freq_{assoluta} = f_i$$

2. **Frequenza relativa:**

$$freq_{relativa} = \frac{f_i}{N}$$

3. **Frequenza percentuale:**

$$freq_{\%} = \frac{f_i}{N} \cdot 100$$

oppure

$$freq_{\%} = freq_{relativa} \cdot 100$$

4. **Frequenza cumulata:** somma progressiva delle frequenze assolute o relative.

$$freq_{cumulataAssoluta} = \sum_{i=1}^n f_i$$
$$freq_{cumulataRelativa} = \sum_{i=1}^n \frac{f_i}{N} = \sum_{i=1}^n freq_{relativa_i}$$

5. **Frequenza cumulata percentuale:**

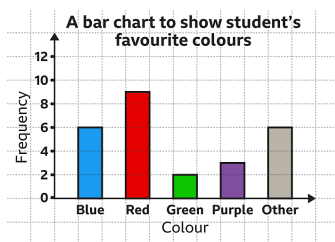
$$freq_{cumulataAssoluta\%} = \sum_{i=1}^n f_i \cdot 100$$
$$freq_{cumulataRelativa\%} = \sum_{i=1}^n \frac{f_i}{N} \cdot 100 = \sum_{i=1}^n freq_{relativa_i} \cdot 100$$

## 2 Statistica Descrittiva

### 2.1 Diagrammi a barre vs Istogrammi

**Definizione 2.1 (Diagrammi a barre).** Descrivono la distribuzione di frequenza di una o più variabili qualitative (categoriche). Le barre devono avere tutte la stessa base ed essere equi-spaziate (lasciare un pò di spazio tra una barra e l'altra).

- altezza barre: frequenza

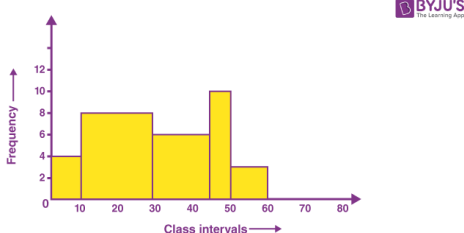


**Definizione 2.2 (Istogrammi).** Descrivono la distribuzione di frequenza di una o più variabili quantitative. Ogni barra rappresenta una classe e la sua frequenza.

- altezza barre: densità di frequenza

$$densita_{freq} = \frac{\text{Frequenza}}{\text{Ampiezza classe}}$$

- base barre: ampiezza delle classi



**Osservazione:** Definire  $k$  classi di uguale ampiezza

$$\text{Ampiezza classe} = \frac{\max - \min}{k}$$

I dati sulla statura di 48 adulti vanno da un minimo di 160 a 180 cm. Come fare  $k$  classi di uguale ampiezza?

1. Scelgo  $k$  (es:  $k = 5$ )
2. Uso formula  $\text{Ampiezza classe} = \frac{\max - \min}{k}$  (es:  $\text{Ampiezza classe} = \frac{180-160}{5} = 4$ )

cm); quindi ogni classe avrà ampiezza 4.

3. Gli estremi inferiore della classe sono (contando ampiezza 4):

- 160
- 164
- 168
- 172
- 176

Conclusione: le  $k = 5$  classi di ugual ampiezza sono:

[160, 164), [164, 168), [168, 172), [172, 176), [176, 180]

## 2.2 Media

Qual è il centro dei dati? **valore tipico** attorno a cui si concentrano i dati.

$f_i$  indica la frequenza assoluta.

- **Formula della media per distribuzione di frequenze:** (variabili discrete)

$$\bar{x} = \frac{\sum_{i=1}^n (x_i \cdot f_i)}{N} = \sum_{i=1}^n (x_i \cdot freq_{relativa_i})$$

- **Formula della media per distribuzione di frequenze:** (variabili continue)

$$\bar{x} = \frac{\sum_{i=1}^n (m_i \cdot f_i)}{N} = \sum_{i=1}^n (m_i \cdot freq_{relativa_i})$$

dove  $a, b$  estremi dell'intervallo e  $m_i = \frac{a+b}{2}$  il valore centrale della classe.

## 2.3 Moda

Qual è il centro dei dati? **valore tipico** attorno a cui si concentrano i dati.

La Moda è il valore che si ripete più spesso nei dati.

- **Formula della moda per distribuzione di frequenze:** (variabili discrete)

$$Moda = x_i \text{ con maggior frequenza}$$

- **Formula della media per distribuzione di frequenze:** (variabili continue)

$$Moda = \frac{a+b}{2}$$

### Esempio Moda

Per esempio, per l'esame di analisi 2 ci sono stati tanti studenti che hanno preso tra il 20 e il 25 (classe), allora [20-25] è la classe modale. Pertanto, nel nostro esempio  $Moda = \frac{20+25}{2} = 22.5$

## 2.4 Mediana

Qual è il centro dei dati? **valore tipico** attorno a cui si concentrano i dati.

La Mediana è il valore che è più grande (o uguale) della prima metà dei dati e allo stesso tempo, più piccolo (o uguale) della seconda metà dei dati.

**È il valore che sta in mezzo a dati ordinati;** quindi per poter stimare la  $Me$  è necessario ordinare i dati:

- **Formula della mediana per distribuzione di frequenze:** (variabili discrete)

1. Ordina i dati
2. Trova indice  $i$ :
  - se  $N$  pari:  $i_1 = \frac{N}{2}$ ,  $i_2 = \frac{N}{2} + 1$
  - se  $N$  dispari:  $i = \frac{N}{2}$
3. La mediana è il valore associato all'indice trovato ( $i = x_i$ ):
  - Se ho due indici  $i_1, i_2$ , allora  $Me = \frac{x_1 + x_2}{2}$
  - Se ho un solo indice  $i$ , allora  $Me = x_i$

- **Formula della mediana per distribuzione di frequenze:** (variabili continue)

1. Calcola frequenza cumulata di ogni classe

$$freq_{cumulataAssoluta} = \sum_{i=1}^n f_i$$

2. Trova indice  $i$ :
  - se  $N$  pari:  $i_1 = \frac{N}{2}$ ,  $i_2 = \frac{N}{2} + 1$
  - se  $N$  dispari:  $i = \frac{N}{2}$
3. Osserva  $i$  in che classe cade (vedi frequenza cumulata), allora  $Me = classe$ .  
Oppure, se abbiamo due indici  $i_1, i_2$  con valori  $x_1, x_2$ , allora  $Me = \frac{x_1 + x_2}{2}$

## 2.5 Quartili

Il  $p$ -esimo percentile è il valore che ha  $\%p$  dei dati sotto/dietro di sé.

- $Q_1$  = 25-esimo percentile  
(25% dei dati sotto questo valore)
- $Q_2$  = 50-esimo percentile = Mediana  
(50% dei dati sotto questo valore)
- $Q_3$  = 75-esimo percentile  
(75% dei dati sotto questo valore)

Divido la distribuzione in 4 parti uguali, per questo si chiamano "quartili".

- **Come trovare il  $Q_k$  per distribuzione di frequenze:** (variabili discrete)

1. Ordina i dati
2. Trova indice:  $i = \frac{N+1}{4} \cdot k$
3.  $Q_k$  è il valore associato all'indice:
  - Se  $i \in \mathbb{N}$ , allora  $Q_k = x_i$
  - Se  $i \in \mathbb{Q}$ , allora  $Q_k = \frac{\text{somma dei valori associati}}{2}$

Esempio: se  $i = 6.75$  allora  $i_1 = 6, i_2 = 7$ , e i valori associati a  $i_1 = 20, i_2 = 25$ , allora  $Q_k = \frac{x_1 + x_2}{2} = \frac{20 + 25}{2} = 22.5$

- **Come trovare il  $Q_k$  per distribuzione di frequenze:** (variabili continue)

1. Calcola frequenza cumulata di ogni classe

$$freq_{cumulataAssoluta} = \sum_{i=1}^n f_i$$

2. Trova indice:  $i = \frac{N+1}{4} \cdot k$

3. Osserva  $i$  in che classe cade (vedi frequenza cumulata)

4. Allora avremo:

$$Q_k = L + \frac{i - f_{cumulata}}{f_i} \cdot h$$

dove:

- $L$ : estremo inferiore della classe attuale (dell'indice)
- $i$ : indice
- $f_{cumulata}$ : frequenza cumulata classe precedente
- $f_i$ : frequenza assoluta classe attuale
- $h$ : ampiezza classe attuale

## 2.6 Campo di Variazione / Range

Distanza tra min e max.

$$Range = max - min$$

## 2.7 Differenza Interquartile

Si usano i quartili per capire la variabilità centrale.

$$IQR = Q_3 - Q_1$$

## 2.8 Varianza

Quanto sono variabili i dati? i dati sono **vicini o molto sparsi**.

Quanto i dati si allontanano dalla loro media ( $\mu$  oppure  $\bar{x}$ ).

- **Formula varianza per distribuzione di frequenze:** (variabili discrete)

- Varianza della popolazione ( $P$ ):

$$\sigma^2 = \frac{\sum_{i=1}^n (x_i - \mu)^2}{N}$$

- Varianza campionaria ( $C \subseteq P$ ):

$$s^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{N - 1}$$

## 2.9 Deviazione standard

Quanto sono variabili i dati? i dati sono **vicini o molto sparsi**.

$$\sigma = \sqrt{\sigma^2}$$

dove  $\sigma^2$  è la varianza.

## 2.10 Coefficiente di variazione

Si calcola quando  $\mu, \bar{x}, \sigma$  sono positivi e si esprime in percentuale.

$$CV = \frac{\sigma}{\mu} \quad \text{oppure} \quad CV = \frac{\sigma}{\bar{x}}$$

### **3 Probabilità**

### **4 Indipendenza e Probabilità Condizionata**

### **5 Variabili Casuali**

#### **5.1 Famiglie Parametriche**

### **6 Inferenza Statistica**

#### **6.1 Stima Puntuale**

#### **6.2 Stima Intervallare**

#### **6.3 Verifica delle Ipotesi**