



UNIVERSITÀ  
DEGLI STUDI DI BARI  
ALDO MORO



DIPARTIMENTO DI SCIENZE POLITICHE

# Concetti e metodi per le analisi statistiche:

---

## introduzione alla statistica

Dott. M Dezio

# Info

---

Mail: [michele.dezio@uniba.it](mailto:michele.dezio@uniba.it)

Stanza: Centro di Calcolo, V piano, scala D

Codice teams: ciiuxnn

# Che cos'è la statistica?

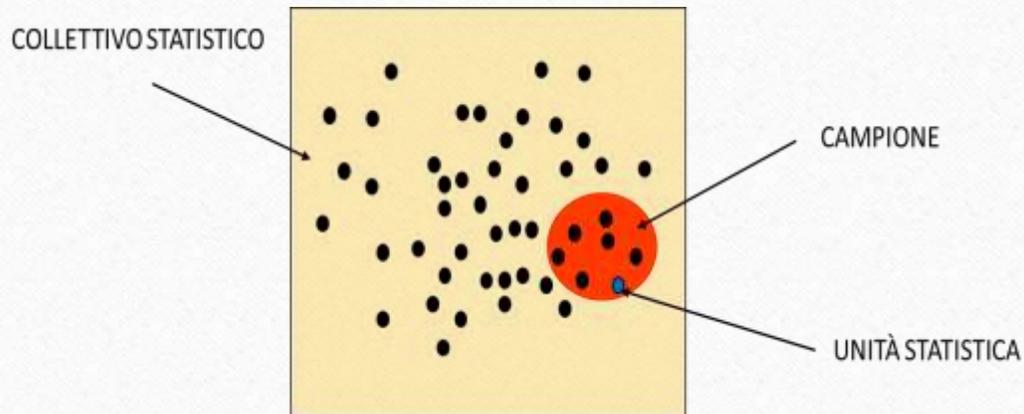
---

La **statistica** è una disciplina che si occupa dello studio dei **fenomeni collettivi** (qualitativi o quantitativi), attraverso la raccolta **dati** e l'elaborazione degli stessi, al fine di trarre **conoscenza**.



# Collettivo, unità e campione

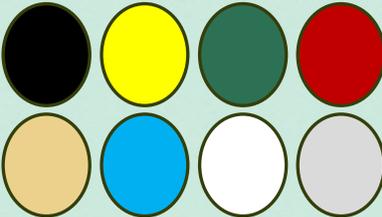
- **Collettivo**: un insieme di elementi di qualsiasi natura definito in modo tale che sia possibile determinare l'appartenenza di un elemento al gruppo.



- **Unità statistiche**: singoli elementi che costituiscono il collettivo. Possono essere semplici o composte.
- **Campione**: sottoinsieme estratto da una popolazione più ampia.

# Carattere e modalità

- Un **carattere** (o variabile) è un attributo, una proprietà o una misura osservabile all'interno delle unità di un collettivo. Un carattere è detto **quantitativo**, quando è indicato mediante un numero, **qualitativo** quando l'attributo associato alle unità è non numerico. La **modalità** è l'attributo o il numero che l'unità statistica manifesta.

CARATTERE QUANTITATIVO DISCRETO	CARATTERE QUANTITATIVO CONTINUO	CARATTERE DICOTOMICO	CARATTERE QUALITATIVO SCONNESSO	CARATTERE QUALITATIVO RETTILINEO
				

# Le fasi dell'indagine statistica (1)

## 1. Definizione degli obiettivi.

È la fase in cui vengono definiti gli obiettivi da raggiungere attraverso l'indagine.

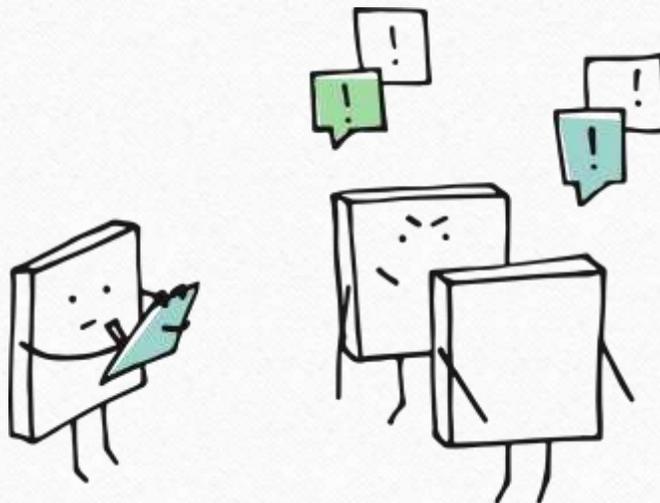


Qual è lo scopo della nostra indagine?  
Quale sarà la nostra popolazione di riferimento?  
Chi utilizzerà le nostre elaborazioni?  
Quali sono le limitazioni della nostra indagine?

# *Le fasi dell'indagine statistica (2)*

## 2. Rilevazione: raccolta dei dati relativi alle unità statistiche.

La rilevazione può essere completa (come nel caso dei censimenti) o parziale (attraverso campioni estratti dal collettivo statistico)



# *Le fasi dell'indagine statistica (3)*

## **3. Elaborazione dei dati.**

---

I dati rilevati vengono elaborati, classificati e sintetizzati per ottenere risultati più significativi.



# *Le fasi dell'indagine statistica (4)*

## 4. Presentazione dei dati.

Presentazione dei dati in forma tabulare o grafica e attraverso misure sintetiche.



# *Le fasi dell'indagine statistica (5)*

## 5. **Interpretazione dei dati.**

---

Interpretazione dei dati alla luce degli obiettivi prefissati nella fase di definizione.



# *Fonti Statistiche*

---

- **Dati primari**: ottenuti attraverso una raccolta dati diretta.
- **Dati secondari**: ottenuti attraverso l'utilizzo di dati raccolti da altre fonti.

In Italia, la principale fonte di dati secondari è l'**ISTAT** (Istituto nazionale di Statistica) che rappresenta il fulcro del **SISTAN** (Sistema Statistico Nazionale)

Fonti statistiche internazionali: **EUROSTAT, OCSE, ONU.**

# *Serie statistica*

Una **serie statistica** è una sequenza di dati disposti in ordine consecutivo, dove ogni dato rappresenta una modalità o un valore del carattere osservato su ciascuna delle unità statistiche considerate.

La serie statistica può essere rappresentata nel seguente modo:  $Y_1, Y_2, \dots, Y_N$

dove  $N$  rappresenta il numero di unità osservate

Una forma compatta per indicare la serie statistica oggetto d'interesse è la seguente:

$$\{Y_i\}, \quad i = 1, 2, \dots, N$$

Se il carattere è quantitativo o qualitativo rettilineo, la serie storica può essere ordinata. Una serie statistica ordinata può essere indicata come segue:

$$Y_{(1)}, Y_{(2)}, \dots, Y_{(N)}$$

$$\{Y_{(i)}\}, \quad i = 1, 2, \dots, N$$

## ***Serie statistica: esempio***

---

Consideriamo i voti presi al compito di Matematica da una classe di 20 studenti:

7, 5, 9, 4, 8, 6, 3, 10, 5, 7, 4, 6, 9, 8, 3, 10, 6, 4, 8, 5



**Serie statistica**

# ***Serie Storica***

---

Una tipologia fondamentale di serie statistica è la **serie storica**. La serie storica, mostra per ogni istante di tempo considerato  $t = 0, 1, 2, \dots, t, \dots T$ , l'intensità del fenomeno osservato

Tempo	$Y_t$
0	$Y_0$
1	$Y_1$
$\vdots$	$\vdots$
t	$Y_t$
$\vdots$	$\vdots$
T	$Y_T$

# *Serie territoriale o geografica*

---

Una **serie geografica** rappresenta l'intensità del fenomeno di interesse in diverse suddivisioni geografiche.

Questi dati sono utilizzati per analizzare e confrontare il fenomeno in ciascuna area geografica considerata.

Area	$Y_t$
1	$Y_0$
2	$Y_1$
⋮	⋮
k	$Y_t$
⋮	⋮
N	$Y_T$

# *Dalle serie alle tabelle di frequenza: la tabella di conteggio*

La serie statistica è molto scomoda e lo diventa di più al crescere di N.

Per semplificare la visione dei dati, si può ricorrere alle tabelle di frequenza.

Un modo per passare dalle serie statistiche alle tabelle di frequenza è la **tabella di conteggio**

7, 5, 9, 4, 8, 6, 7, 7, 5, 7, 4, 6, 9, 8, 3, 10, 6, 4, 7, 5

1	
2	
3	I
4	III
5	III
6	III
7	HHH
8	II
9	II
10	I

# Tabella di Frequenza

CARATTERE      FREQUENZE



$$N = n_1 + n_2 + \dots + n_i + \dots + n_s = \sum_{i=1}^s n_i$$

MODALITÀ →

$Y_i$	$n_i$
$Y_1$	$n_1$
$Y_2$	$n_2$
⋮	⋮
$Y_i$	$n_i$
⋮	⋮
$Y_s$	$n_s$
TOTALE	N

Voto	Numero studenti
3	1
4	3
5	3
6	3
7	5
8	2
9	2
10	1

# Le Frequenze

1. **Frequenze assolute:** indicano il numero di volte in cui una determinata modalità del carattere oggetto di interesse si presenta nel collettivo analizzato. (Sono le frequenze che abbiamo visto e utilizzato finora).

Se in un gruppo di 100 persone, 5 hanno la licenza elementare, 13 la licenza media, 51 il diploma, 27 la laurea e 5 il dottorato, i numeri (5, 13, 51, 27, 5) sono le frequenze assolute associate alle modalità (Licenza Elementare, Licenza Media, Diploma, Laurea, Dottorato) del carattere «Titolo di Istruzione» raccolti nel nostro collettivo

Titolo di Studio ( $Y_i$ )	Frequenza assoluta ( $n_i$ )
Licenza Elementare	5
Licenza Media	13
Diploma	51
Laurea	27
Dottorato	4
Totale	100

# Le Frequenze

2. **Frequenze cumulate**: associano ad ogni valore  $i$ -esimo del carattere discreto (quantitativo o rettilineo) o continuo, la somma delle frequenze assolute dalla prima fino all' $i$ -esima.

Nel nostro esempio, la frequenza cumulata associata alla modalità Diploma, ci dice quanti dei nostri intervistati ha il diploma o un titolo di studio inferiore.

Titolo di Studio ( $Y_i$ )	Frequenze assolute ( $n_i$ )	Frequenze cumulate ( $N_i$ )
Licenza Elementare	5	5
Licenza Media	13	18
Diploma	51	69
Laurea	27	86
Dottorato	4	100
Totale	100	

$$N_i = n_1 + n_2 + \dots + n_{i-1} + n_i$$

$$N_i = \sum_{j=1}^i n_j$$

# Le Frequenze

3. **Frequenze relative:** si ottengono rapportando le frequenze assolute al numero totale di unità.

In questo modo, se analizziamo un secondo collettivo e rileviamo lo stesso carattere per ciascuna unità, possiamo effettuare dei confronti

Titolo di Studio ( $Y_i$ )	Frequenze assolute ( $n_i$ )	Frequenze cumulate ( $N_i$ )	Frequenze relative ( $f_i$ )
Licenza Elementare	5	5	0,05
Licenza Media	13	18	0,13
Diploma	51	69	0,51
Laurea	27	86	0,27
Dottorato	4	100	0,04
Totale	100		1

$$f_i = \frac{n_i}{N} = \frac{n_i}{\sum_{i=1}^s n_i}$$

# Le Frequenze

4. **Frequenze percentuali**: si ottengono moltiplicando le frequenze relative per 100.

Le frequenze percentuali sono molto utilizzate, poiché di facile comprensione e consentono il confronto.

Titolo di Studio ( $Y_i$ )	Frequenze assolute ( $n_i$ )	Frequenze cumulate ( $N_i$ )	Frequenze relative ( $f_i$ )	Frequenze percentuali ( $f_i$ )
Licenza Elementare	5	5	0,05	5%
Licenza Media	13	18	0,13	13%
Diploma	51	69	0,51	51%
Laurea	27	86	0,27	27%
Dottorato	4	100	0,04	4%
Totale	100		1	1

$$p_i = f_i \times 100$$