

STATISTICA

La **Statistica** è la scienza che studia i fenomeni collettivi utilizzando metodi matematici.

Essa si occupa della tecnica per raccogliere ed elaborare “Dati” (studenti, abitanti, oggetti, ecc.) al fine di trarre da essi delle informazioni sintetiche.

L'insieme degli elementi su cui si svolge un'indagine statistica prende il nome di **POPOLAZIONE (STATISTICA)** o **FENOMENO COLLETTIVO** o **UNIVERSO**. Ogni singolo elemento di una popolazione statistica prende il nome di **UNITÀ STATISTICA**.

Punto di partenza di ogni indagine statistica è sempre la raccolta dei dati e la loro classificazione dopo aver definito:

- 1) Il **CARATTERE** cioè cosa si intende analizzare per lo studio di un fenomeno (Esempio: il colore degli occhi). Il CARATTERE può essere di tipo:
 - **QUALITATIVO** se è descritto con parole (Esempi: colore degli occhi; titolo di studio; sport praticato; ...)
 - **QUANTITATIVO** se è descritto con numeri (Esempi: l'età; il voto ottenuto in una materia; il numero di componenti di una famiglia; la statura; il peso; ...)
- 2) La **MODALITÀ** cioè il modo in cui si può presentare una caratteristica (Esempio: per il CARATTERE “Colore degli occhi”, le MODALITÀ possono essere: verdi, azzurri, castani, neri).

LA RACCOLTA DEI DATI

La raccolta dei dati si può effettuare con metodi diversi in base al tipo di **popolazione statistica** e al numero delle **unità statistiche** che la compongono. Un metodo per raccogliere i dati consiste nell'utilizzare le pubblicazioni periodiche dell'ISTAT, l'Istituto Nazionale di Statistica, oppure i numerosi dati contenuti negli uffici pubblici regionali e comunali, nelle Scuole e nelle Aziende Sanitarie Locali.

Ci sono due **METODI di raccolta dei dati** che si definiscono **DIRETTI** perché sono rivolti direttamente alla popolazione statistica: l'intervista e il questionario.

Il metodo dell'**intervista** consiste nel richiedere le informazioni direttamente agli interessati, ascoltarne le risposte e annotarle.

Il **metodo del questionario** consiste nel recapitare a ciascun interessato uno o più fogli con le domande e ritirarli con le risposte indicate.

La **RACCOLTA** dei dati con i metodi diretti può essere di due tipi: **completa** o **per campione**.

È **completa** se viene effettuata su tutta la popolazione statistica interessata all'indagine; è **per campione** se viene effettuata solo su una parte della popolazione statistica interessata, che viene appunto chiamata **campione**. È importante che il campione venga scelto in modo che sia rappresentativo della popolazione statistica su cui si svolge l'indagine.

Poi si procede alla elaborazione e classificazione che consiste nell'assegnare ciascuna osservazione alla corrispondente modalità della caratteristica osservata.

ORGANIZZAZIONE DEI DATI

Per ottenere una descrizione sintetica di un fenomeno, è utile associare a ciascuna modalità della caratteristica osservata, il numero di osservazioni che rientrano in tale modalità: questo numero prende il nome di **frequenza assoluta**.

FREQUENZA ASSOLUTA: Numero di volte in cui una certa modalità si manifesta nell'indagine.

N.B.: la somma delle frequenze assolute è uguale al numero di casi osservati, cioè al numero di componenti della popolazione statistica.

DISTRIBUZIONE STATISTICA O DI FREQUENZE (riferita ad un carattere osservato)

La tabella che riporta per ogni modalità le frequenze corrispondenti, prende il nome di **distribuzione statistica** o **distribuzione delle frequenze**. Dunque una distribuzione statistica si ottiene classificando le osservazioni in base alla **frequenza** con cui si presenta ciascuna modalità.

Per confrontare due o più distribuzioni di dati, è utile calcolare le **frequenze relative** e le **frequenze percentuali**.

FREQUENZA RELATIVA: rapporto tra la frequenza assoluta e il numero totale di unità statistiche.

N.B.: la somma delle frequenze relative è uguale a 1.

FREQUENZA (RELATIVA) PERCENTUALE: è la frequenza relativa espressa in percentuale.

N.B.: la somma delle frequenze percentuali è uguale a 100.

ESEMPIO:

POPOLAZIONE: Classe 28 alunni; **CARATTERISTICA:** mezzo di trasporto; **MODALITÀ:** automobile, a piedi, autobus, motorino, bicicletta

DISTRIBUZIONE STATISTICA

	MEZZO DI TRASPORTO	FREQUENZA ASSOLUTA	FREQUENZA RELATIVA	FREQUENZA %
MODALITÀ	Automobile	7	$7/28 = 0,25$	25%
	A piedi	3	$3/28 = 0,11$	11%
	Autobus	9	$9/28 = 0,32$	32%
	Motorino	4	$4/28 = 0,14$	14%
	Bicicletta	5	$5/28 = 0,18$	18%
	TOTALE		28	$28/28 = 1$

CLASSI DI FREQUENZA

Quando una modalità non si ripete più volte è utile raggrupparla con altre formando delle classi o intervalli, in tal caso si determina la frequenza di ogni singola classe o intervallo.

Per esempio se la **popolazione** è la **classe 1B** e il **carattere** è il **salto in lungo da fermo**, si verificherà che ogni alunno con il suo salto determinerà una modalità diversa (nessuno farà un salto esattamente uguale ad un altro!).

Quindi si possono formare delle classi o intervalli stabilendo che gli intervalli siano del tipo $(140 \div 160)$, $(160 \div 180)$, $(180 \div 200)$, Dove il valore minimo dell'intervallo si assume appartenente all'intervallo precedente, del quale ne rappresenta il massimo.

Ogni intervallo così definito si chiama **classe (di frequenza)** e vi appartengono tutte le modalità che ricadono nella classe stessa. Quindi, la frequenza assoluta di ciascuna classe è data dalla somma delle frequenze assolute delle modalità che vi appartengono.

ESEMPIO: un'indagine sulla statura di un gruppo di 40 coetanei ha dato i seguenti risultati (in centimetri):

140 137 145 158 164 148 162 144
136 143 142 153 153 152 139 161
153 152 164 160 155 164 153 163
164 160 138 135 144 160 138 154
162 135 161 148 149 138 158 145

I valori variano da un minimo di 135 cm a un massimo di 164 cm e alcuni di essi compaiono una sola volta; se si calcolasse la frequenza assoluta di tutte le modalità, la tabella di distribuzione delle frequenze sarebbe molto grande e poco significativa. Si possono invece raggruppare i valori in cinque gruppi distinti della stessa ampiezza di 6 cm, ovvero in 5 classi:

1^a classe: da 135 cm a 140 cm

2^a classe: da 141 cm a 146 cm

3^a classe: da 147 cm a 152 cm

4^a classe: da 153 cm a 158 cm

5^a classe: da 159 cm a 164 cm

e calcolare la frequenza assoluta di ciascuna classe, contando quanti dei dati rilevati sono compresi in essa. Si ottiene così la seguente tabella della distribuzione della frequenza assoluta per classi:

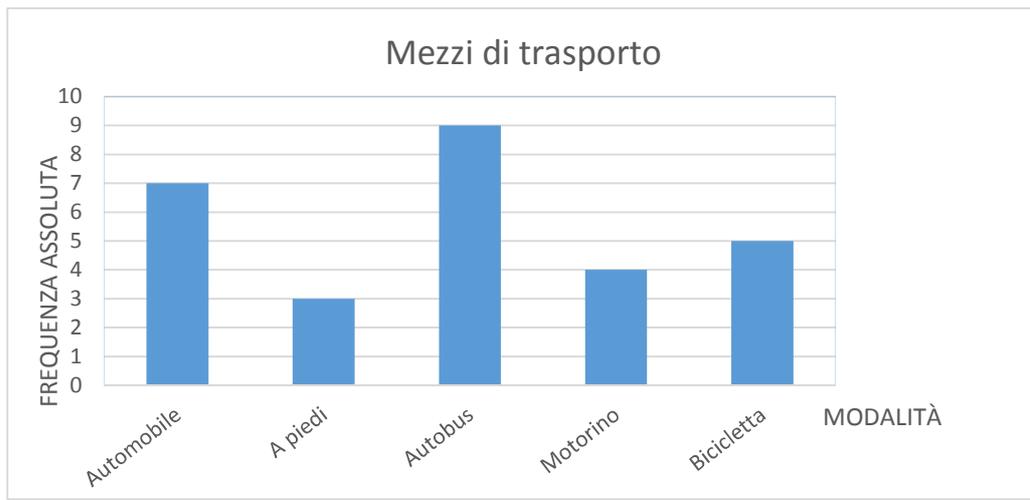
Classe	Intervallo (cm)	Frequenza assoluta
1 ^a	135 ÷ 140	9
2 ^a	141 ÷ 146	6
3 ^a	147 ÷ 152	5
4 ^a	153 ÷ 158	8
5 ^a	159 ÷ 164	12

RAPPRESENTAZIONE GRAFICA DEI DATI

La distribuzione delle frequenze oltre che mediante tabelle, può essere organizzata anche graficamente mediante:

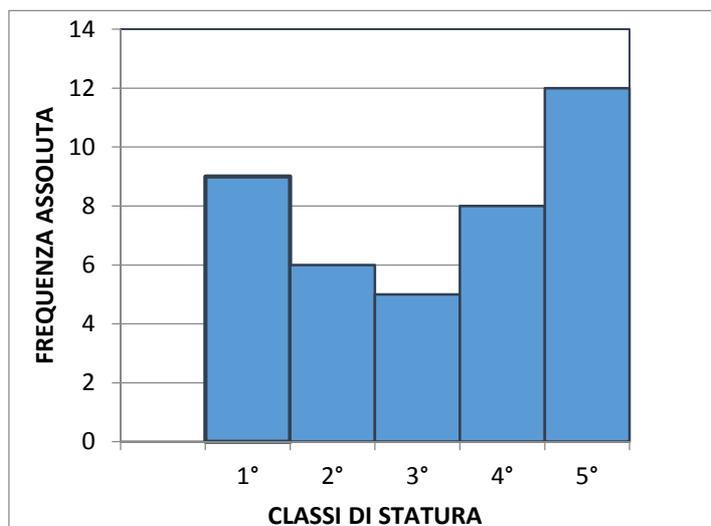
1) ORTOGRAMMA:

In esso, a ogni modalità, riportata sull'asse orizzontale, corrisponde un rettangolo che ha l'altezza proporzionata alla frequenza stessa e base costante al variare della modalità.



2) ISTOGRAMMA:

Un istogramma è costituito da rettangoli che hanno basi proporzionali alle ampiezze delle classi e altezze proporzionali alle corrispondenti frequenze.



POLIGONO DELLE FREQUENZE:

Se in un istogramma si congiungono i punti medi dei lati superiori dei rettangoli, di ottiene una spezzata, chiamata "poligono delle frequenze".

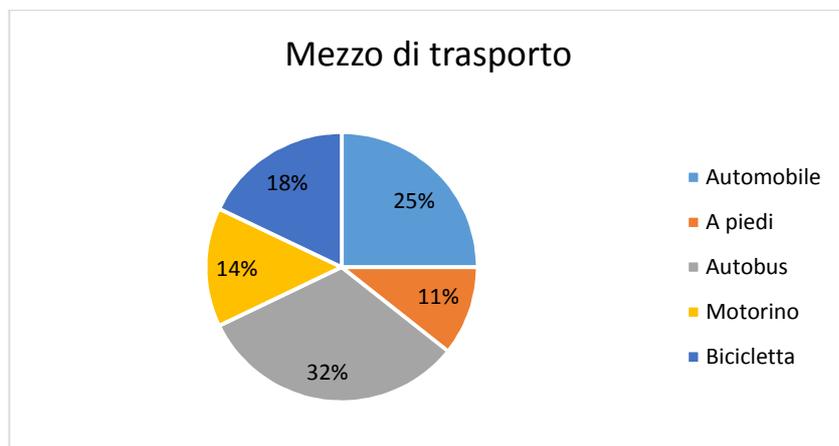
Ogni vertice del poligono delle frequenze corrisponde al valore centrale delle classi. La somma delle aree dei rettangoli di un istogramma è uguale all'area sottostante il poligono delle frequenze.

3) AREOGRAMMA o DIAGRAMMA A TORTA:

Un cerchio viene suddiviso in tanti settori circolari, ognuno dei quali corrisponde a una classe o modalità.

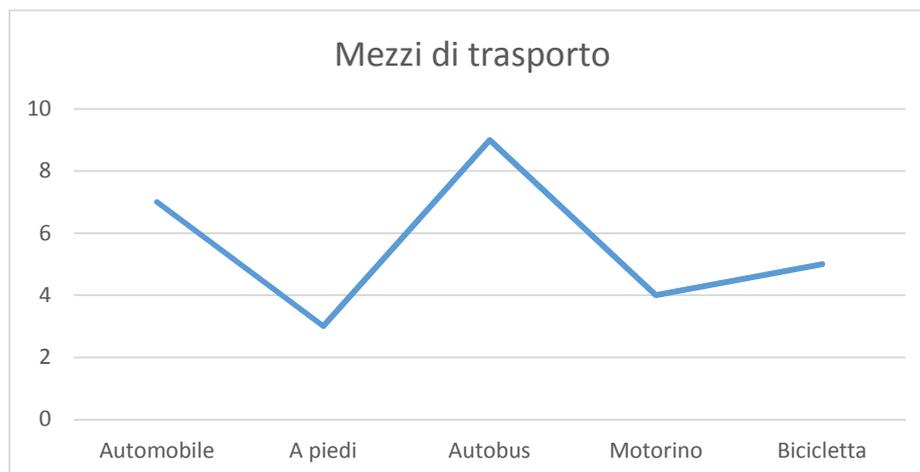
Ogni settore circolare ha un angolo al centro di ampiezza proporzionale alla frequenza percentuale di quella modalità.

$$X^\circ : 360^\circ = f.\text{relativa} : 100$$



4) DIAGRAMMI CARTESIANI:

Si tratta di una spezzata che descrive la forma della distribuzione delle frequenze.



5) IDEOGRAMMI:

Gli ideogrammi utilizzano figure che richiamano il contenuto del fenomeno e ne danno una visione immediata.

Le figure hanno dimensioni proporzionali al valore del dato che rappresentano.

Programmi per ragazzi	          
Intrattenimento	    
Telegiornali	  
Sport	       
Documentari	    
Film	      
Fiction	    

 = 5 bambini

INDICI DI POSIZIONE CENTRALE

Come già detto, la statistica si pone l'obiettivo di trarre dai dati delle informazioni sintetiche.

La distribuzione delle frequenze rappresenta già una forma di sintesi dei dati, ma spesso è necessario disporre di un unico valore numerico che sintetizzi tutti i dati della distribuzione rappresentandone in qualche modo il "centro".

Esiste più di un "indice di posizione centrale" che può essere utilizzato a tal fine.

I più importanti indici di posizione centrale sono: la **MODA**, la **MEDIANA** e la **MEDIA DELLA DISTRIBUZIONE**.

- La **MODA** è il valore o la modalità a cui corrisponde la frequenza assoluta massima

ESEMPLI:

- 1) **CARATTERE:** colore preferito. **MODALITÀ:** rosso; verde; giallo; nero; blu. **POPOLAZIONE:** gruppo di studenti di un istituto superiore.

COLORE PREFERITO	FREQUENZA ASSOLUTA
Rosso	19
Verde	38
Giallo	12
Nero	4
Blu	15
TOTALE	88

MODA = colore verde (è la modalità corrispondente alla massima frequenza).

2) **CARATTERE:** peso (Kg). **MODALITÀ:** 50; 60; 70; 80; 90. **POPOLAZIONE:** gruppo di persone adulte.

PESO (Kg)	FREQUENZA ASSOLUTA
50	119
60	97
70	85
80	143
90	20
TOTALE	464

MODA = 80 kg (è la modalità corrispondente alla massima frequenza)

- La **MEDIANA** di N dati o osservazioni ordinati in senso crescente o decrescente è:
 - il valore centrale se N è dispari.

ESEMPIO:

Sono dati i seguenti valori: 7, 12, 2, 3, 9.

Per calcolare la MEDIANA, bisogna prima metterli in ordine crescente (o decrescente):

2 3 **7** 9 12

MEDIANA = 7

↑

MEDIANA

- la media aritmetica dei due valori centrali se N è pari.

ESEMPIO:

Sono dati i seguenti valori: 1, 7, 12, 2, 3, 9.

Per calcolare la MEDIANA, bisogna prima metterli in ordine crescente (o decrescente):

1 2 **3 7** 9 12

$$\text{MEDIANA} = \frac{3+7}{2} = 5$$

- Si dice **MEDIA ARITMETICA (SEMPLICE)** di n numeri $a_1, a_2, a_3, \dots, a_n$ il quoziente tra la loro somma e il loro numero:

$$M = \frac{a_1 + a_2 + \dots + a_n}{n}$$

ESEMPIO: uno studente ha ottenuto in inglese i seguenti otto voti:

6 4 6,5 8 5,5 7 5 6

La media aritmetica è:

$$M = \frac{6 + 4 + 6,5 + 8 + 5,5 + 7 + 5 + 6}{8} = 6$$

Si dice che 6 è il voto medio dello studente in inglese.

- La **MEDIA PONDERATA** di n osservazioni è il quoziente tra le somme dei prodotti di ciascun valore per il proprio peso (frequenza) e la somma dei pesi (somma delle frequenze).

ESEMPIO:

Nel caso di un'indagine sull'età dei ragazzi iscritti a un corso di nuoto, i cui risultati sono riportati nella tabella seguente di distribuzione delle frequenze:

ETÀ (anni)	FREQUENZA ASSOLUTA
8	140
9	98
10	102
11	34
12	42
TOTALE	416

la MEDIA ARITMETICA PONDERATA risulta:

$$Mp = \frac{8 \cdot 140 + 9 \cdot 98 + 10 \cdot 102 + 11 \cdot 34 + 12 \cdot 42}{416} \cong 9,38$$

(N.B.: \cong si legge "circa uguale a")

Ciò significa che l'età media degli iscritti è di poco superiore ai 9 anni.

LA TEORIA IN SINTESI

<p>1. La statistica</p>	<p>Statistica</p>	<p>■ È la scienza che studia i fenomeni collettivi mediante indagini statistiche, con cui si raccolgono informazioni, si organizzano, si analizzano e si interpretano i dati ottenuti, usando metodi matematici.</p>
<p>2. Le indagini statistiche</p>	<p>Popolazione statistica</p> <p>Unità statistica</p> <p>Carattere o variabile statistica</p> <p>Fasi di un'indagine statistica</p>	<p>■ Si dice popolazione statistica l'insieme degli elementi su cui si svolge un'indagine statistica.</p> <p>■ Ciascun elemento di una popolazione statistica si dice unità statistica.</p> <p>■ Si dice carattere o variabile statistica di una popolazione statistica ogni aspetto o caratteristica che è oggetto dell'indagine; ciascun carattere si manifesta con modalità diverse. I caratteri di una popolazione statistica possono essere:</p> <ul style="list-style-type: none"> - qualitativi, se sono espressi da aggettivi o sostantivi (come il titolo di studio, il colore preferito, la religione); - quantitativi, se sono espressi da numeri (il peso, l'età, la statura). <p>■ Le fasi di un'indagine statistica sono tre: la raccolta dei dati, la loro organizzazione, e la loro elaborazione.</p>
<p>3. La raccolta dei dati</p>	<p>Raccolta dei dati</p>	<p>■ La raccolta dei dati si effettua mediante intervista o mediante questionario e può essere completa (se si effettua su tutta la popolazione statistica) oppure a campione (se si effettua solo su una parte della popolazione statistica, detta campione).</p>
<p>4. L'organizzazione dei dati</p>	<p>Organizzazione dei dati</p>	<p>■ L'organizzazione dei dati consiste nello spoglio e nella successiva tabulazione dei dati (in tabelle semplici o multiple o a doppia entrata); spesso le tabelle sono completate con rappresentazioni grafiche: istogrammi, ortogrammi, areogrammi, diagrammi cartesiani.</p>
<p>5. Il calcolo della frequenza</p>	<p>Frequenza assoluta</p> <p>Frequenza relativa</p>	<p>■ Si dice frequenza assoluta di una modalità di un carattere il numero di volte in cui la modalità si è manifestata nell'indagine.</p> <p>■ Si dice frequenza relativa di una modalità il rapporto fra la frequenza assoluta e il numero di dati raccolti, cioè il numero delle unità statistiche.</p>

ESERCIZI

1. E' stata effettuata una indagine statistica sulle materie teoriche preferite dagli studenti delle classi prime di un Istituto Alberghiero e le risposte al questionario sono le seguenti:

Materia preferita: italiano; numero studenti: 142

Materia preferita: matematica; numero studenti: 62

Materia preferita: storia; numero studenti: 78

Materia preferita: inglese; numero studenti: 67

Materia preferita: scienze; numero studenti: 99

Dopo aver definito il carattere e le modalità, costruisci una tabella con le frequenze assolute e calcola, rappresentandole nella stessa tabella, le frequenze relative e percentuali.

Infine, rappresenta con un diagramma a barre (= ortogramma) i dati e determina quindi la moda.

2. È stata effettuata una indagine statistica sulle 5.429 famiglie residenti in un Comune della Lombardia per conoscere il numero dei componenti delle stesse famiglie ed il risultato è il seguente:

N. componenti	N. famiglie
1	892
2	1.310
3	1.089
4	1.210
> 4	928

Dopo aver definito il carattere e le modalità dell'indagine statistica, rappresenta i dati in una tabella indicando le frequenze assolute e calcolare, inserendo i dati nella stessa tabella, le frequenze relative ed assolute. Rappresenta quindi i dati sia con un grafico a barre che con un diagramma circolare (a torta).

3. Tieni nota per una settimana del tempo e la tipologia di programmi televisivi che guardi ogni giorno. Segna quanto rilevato distinguendo tra film, film ad episodi, documentari, eventi sportivi, musicali, ecc. e costruisci una tabella indicando il carattere, le modalità, le frequenze assolute e calcola le frequenze relative e percentuali. Rappresenta quindi i dati sia con un diagramma a barre che con un diagramma circolare (a torta) e determina la moda.

4. Scrivi tutti i tuoi voti che hai riportato nelle materie scolastiche dell'anno formativo appena terminato. Calcola la media e determina la moda e la mediana dei voti per ogni materia. Esegui gli stessi calcoli e determina la moda e la mediana su tutti i voti di tutte le materie, in modo tale da avere la media generale, la moda e la mediana.

(N.B.: se non ricordi tutti i voti li puoi inventare)

5. Tieni nota per 10 giorni di quanti soldi spendi in un giorno e calcola quindi la media della tua spesa giornaliera. Determina infine la mediana.

(Puoi anche inventare i dati)

6. Con riferimento all'esercizio precedente, calcola anche quanto risparmi ogni giorno rispetto alle disponibilità di soldi che hai e calcola la media del tuo risparmio giornaliero e determina infine la mediana.

(Puoi anche inventarti i dati)