LA STATISTICA

Si occupa dei modi di *raccogliere e analizzare dati* relativi a un certo gruppo di persone o di oggetti per trarne conclusioni e fare previsioni.

La statistica è comunemente suddivisa in due branche principali:

Statistica descrittiva

Statistica inferenziale

Statistica Inferenziale
(o inferenza statistica)

Le conclusioni che ci permette di trarre sulla popolazione complessiva a partire dall'indagine sul campione non sono certezze, ma asserzioni fondate con i metodi, precisi e quantitativi, del **calcolo delle probabilità**.

Statistica Descrittiva

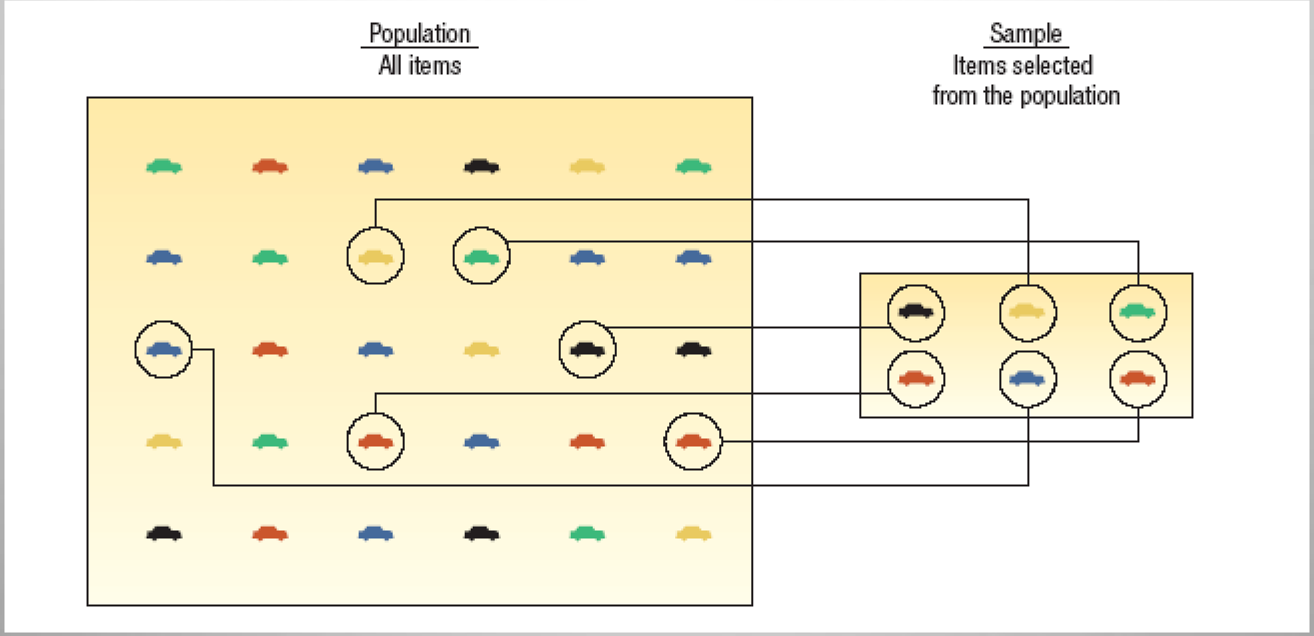
Si occupa dell’analisi dei dati osservati, prescindendo dal calcolo delle probabilità, con lo scopo di ridurre il volume degli stessi ed esprimendo le informazioni attraverso **grafici** ed **indicatori numerici**.

IL LINGUAGGIO DELLA STATISTICA

POPOLAZIONE : insieme di individui, oggetti, misure, oggetto di un’indagine statistica

UNITA’ STATISTICA: ciascun individuo, oggetto, misura che fa parte della popolazione

CAMPIONE : parte della popolazione



IL LINGUAGGIO DELLA STATISTICA

Le unità statistiche hanno diverse caratteristiche ed ognuna di esse rappresenta un **carattere**
della popolazione.

CARATTERE: proprietà oggetto dell'indagine statistica

MODALITA': ciascuna delle varianti con con cui può presentarsi un carattere. Le modalità osservate si dicono **DATI**

Tipi di caratteri

A. Qualitativo – le modalità non sono espresse da numeri

Esempio: carattere SESSO, modalità MASCHIO e FEMMINA

B. Quantitativo (variabile)- le modalità sono espresse da numeri

Esempio: carattere I GIORNI DI ASSENZA DI UNO STUDENTE, modalità espresse da numeri naturali

Caratteri quantitativi - Classificazione

Le variabili possono essere discrete o continue.

A. Variabili discrete: possono assumere solo un numero finito di valori

Esempio: Carattere: numero di figli (SI CONTA….)

B. Variabili continue: possono assumere ogni valore di un determinato intervallo.

Esempio: Carattere: peso di un bambino (SI MISURA…)

ESEMPI RIASSUNTIVI

Fenomeno studiato	Colore occhi degli Italiani	Altezza in metri degli studenti di una classe	Anno di nascita iscritti ad una palestra
Popolazione	Tutti gli italiani	Gli studenti di una classe	Gli iscritti alla palestra
Carattere	Il colore degli occhi	La misura dell’altezza	L’anno di nascita
Modalità	Verde, azzurro, marroni, ecc…	Per esempio: 1,60 m; 1,61 m; ecc..	…1970, 1971,…
Tipo di carattere	Qualitativo	Quantitativo continuo	Quantitativo discreto

LE FASI DI UN’INDAGINE STATISTICA

1. PIANIFICAZIONE DELL’INDAGINE STATISTICA : individuare carattere, popolazione/campione,…
2. RILEVAZIONED DEI DATI : attraverso questionari, interviste, misure,…
3. ELABORAZIONE DEI DATI : I dati sono tanti, bisogna riordinarli (tabelle) e sintetizzarli. Di questo si occupa proprio la statistica descrittiva. Se i dati sono stati rilevati solo su un campione, occorre anche estendere i risultati all’intera popolazione. Di questo si occupa la statistica inferenziale.
4. PRESENTAZIONE DEI RISULTATI : attraverso tabelle e/o grafici per facilitare la lettura.
5. INTERPRETAZIONE DEI RISULTATI: richiede attenzione!

ELABORAZIONE DEI DATI

La rilevazione fornisce DATI GREZZI (ad ogni unità statistica un dato, cioè la modalità del carattere osservato).

ESEMPIO: studiare I voti ottenuti in una verifica di matematica in una classe di 13 alunni

(i voti vanno dal 3 al 9)

Tabella dei dati grezzi:

Lucia	Maria	Carlo	Marco	Luca	Elisa	Lara	Giulia	Anna	Emma	Marta	Ale	Nico
6	4	5	8	6	8	5	6	7	5	6	6	8

Una prima elaborazione dei dati si può ottenere ordinandoli in una tabella (**tabella di sintesi**), dove riportare per ogni modalità osservata il numero di unità statistiche su cui è stata rilevata.

voto	n.alunni
3	0
4	1
5	3
6	5
7	1
8	3
9	0

ELABORAZIONE DEI DATI

E’ la tabella delle FREQUENZE. La tabella delle frequenze rappresenta la DISTRIBUZIONE DI FREQUENZE, ossia la funzione che associa ad ogni modalità di un carattere la rispettiva frequenza.

frequenza assoluta di una modalità è il numero di volte in cui si presenta tale modalità (nell’esempio precedente, il n. di alunni per voto)

frequenza relativa è il rapporto tra la frequenza e il numero totale delle unità statistiche (nell’esempio precedente il n. di alunni per voto diviso il numero totale di alunni). Può essere espressa anche come **frequenza percentuale** (la frequenza relativa trasformata in percentuale)

ESEMPI

Colore degli occhi	n. Studenti (FREQUENZA ASSOLUTA)	FREQUENZA RELATIVA	FREQUENZA %
Neri	2	1/10	10%
Marroni	10	½	50%
Azzurri	4	1/5	20%
Verdi	4	1/5	20%
TOTALE	**20**	**1**	**100%**

ESEMPI

n. di figli	**n. Famiglie intervistate (FREQUENZA ASSOLUTA)**	**FREQUENZA RELATIVA**	**FREQUENZA %**
0	10	1/20	5%
1	55	11/40	27,5%
2	95	19/40	47,5%
3	30	3/20	15%
4	4	1/50	2%
5	6	3/100	3%
TOTALE	**200**	**1**	**100%**

Se il **carattere è quantitativo le modalità vanno ordinate**, in senso crescente o descrescente

ESEMPI

Altezza studenti	n. studenti (FREQUENZA ASSOLUTA)	FREQUENZA RELATIVA
(160,165]	1	1/18
(165,170]	3	1/6
(170,175]	5	5/18
(175,180]	2	1/9
(180,185]	5	5/18
(185,190]	2	1/9
TOTALE	**18**	**1**

DISTRIBUZIONE DI FREQUENZE PER CLASSI : utilizzate sia nel caso di caratteri quantitativi discreti, che presentano numerose modalità distinte tra loro, sia nel caso di caratteri quantitativi continui.

ESEMPI

DISTRIBUZIONE DI FREQUENZE PER CLASSI :
Nell'esempio precedente il carattere oggetto di studio si presenta in molte modalità. Per costruire una tabella di sintesi è conveniente accorpare le modalità in intervalli tra loro disgiunti (le CLASSI).

Data una classe definita da un intervallo *a* e *b*, con *a*<*b*, si definisce l'ampiezza della classe il valore *b-a*, il valore centrale il numero *(a+b)/2.*

Le ampiezze delle classi non devono essere necessariamente tutte uguali.

FREQUENZE CUMULATE:

Frequenze assolute cumulate: si ottengono attraverso la progressiva somma delle frequenze assolute

Frequenze relative cumulate: si ottengono attraverso la progressiva somma delle frequenze relative

Se il carattere è quantitativo , ordinando le varie modalità, è possibile rispondere alla domanda: qual è la frequenza per cui il carattere è minore di una certa modalità?

ESEMPI

Altezza studenti	n. studenti (FREQUENZA ASSOLUTA)	FREQUENZA COMULATA
(160,165]	1	1
(165,170]	3	1+3=4
(170,175]	5	4+5=9
(175,180]	2	9+2=11
(180,185]	5	11+5=16
(185,190]	2	16+2=18
TOTALE	**18**	

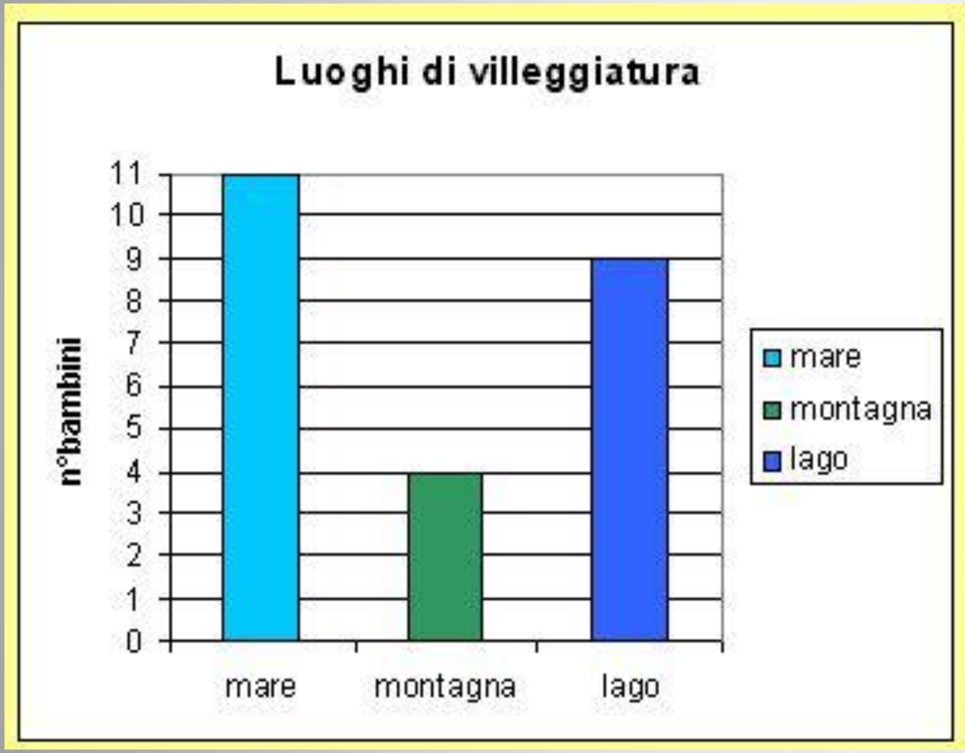
Quanti sono gli studenti con altezza minore o uguale a 1,75 m? Sono 7!

ELABORAZIONE DEI DATI

Esistono vari tipi di grafici per rappresentare i dati statistici e le loro frequenze, fra i quali:

- DIAGRAMMA A BARRE
- DIAGRAMMA CIRCOLARE (A TORTA)
- ISTOGRAMMA
- DIAGRAMMA CARTESIANO (A PUNTI)

DIAGRAMMA A BARRE



L'asse verticale è graduato e serve per indicare la frequenza (assoluta o relativa) con cui le modalità si presentano. L'asse orizzontale serve soltanto come base di appoggio dell'elemento grafico (le due barre).

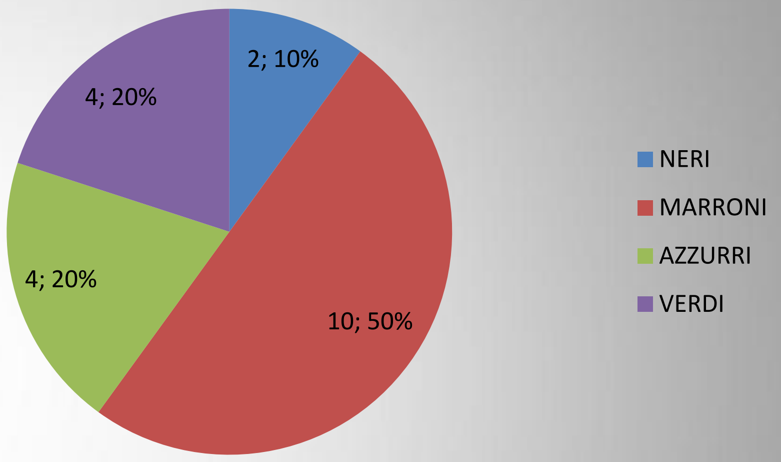
La distribuzione di frequenza è rappresentata da rettangoli, distanziati tra loro, aventi tutti la stessa base.

Sono molto utilizzati per rappresentare la frequenza con cui si presentano le modalità di un carattere qualitativo.

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

DIAGRAMMA CIRCOLARE

- Un diagramma circolare viene costruito dividendo un cerchio in spicchi le cui ampiezza circolari sono proporzionali alle classi di frequenza. Le aree di ogni spicchio sono proporzionali alle frequenze.
- Il diagramma circolare evidenzia le proporzioni tra le sue vari “parti”.
- E’ utile per rappresentare frequenze relative e percentuali.
- E’ utile per rappresentare caratteri qualitativi e di caratteri quantitativi, dove però non importa l’ordine delle modalità.



COLORE DEGLI OCCHI	N.ALUNNI
NERI	2
MARRONI	10
AZZURRI	4
VERDI	4

ISTOGRAMMA

- Gli istogrammi vengono utilizzati per rappresentare graficamente distribuzioni di caratteri quantitativi suddivisi in classi.
- L'istogramma è un grafico costituito da rettangoli non distanziati, ciascuno dei quali ha un'area proporzionale alla frequenza della classe che rappresenta
- Ogni frequenza è quindi rappresentata dall'area di un rettangolo, la cui base è uguale all'ampiezza della classe e l'altezza è pari alla ***densità di frequenza***, cioè al rapporto tra la frequenza della classe e l'ampiezza della classe stessa.
- Se l'ampiezza di ogni classe è uguale, l'altezza è proporzionale alla frequenza di ogni classe.

Tabella16 - Impiegati di un ufficio per classi di età

CLASSI DI ETÀ	Impiegati Frequenza	Ampiezza della classe	Densità di frequenza
20-29	12	10	1,2
30-49	20	20	1,0
50-59	26	10	2,6
60-64	4	5	0,8
Totale	**62**		

Figura F - Impiegati di un ufficio per classi di età

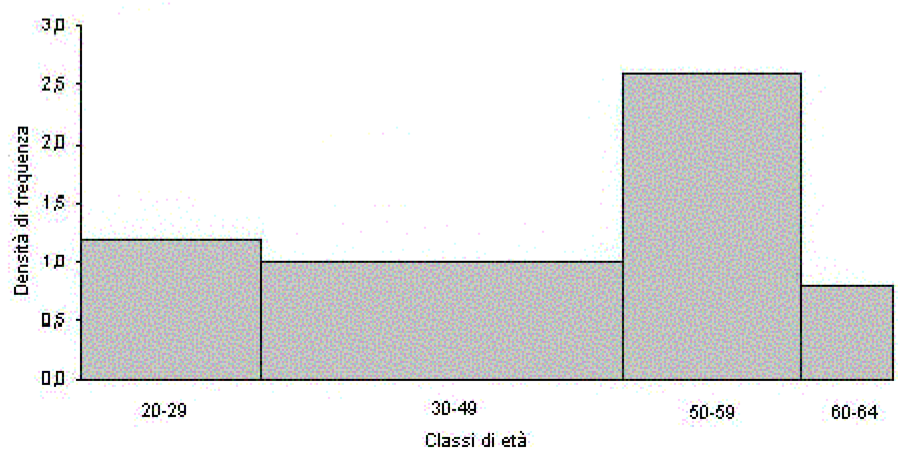
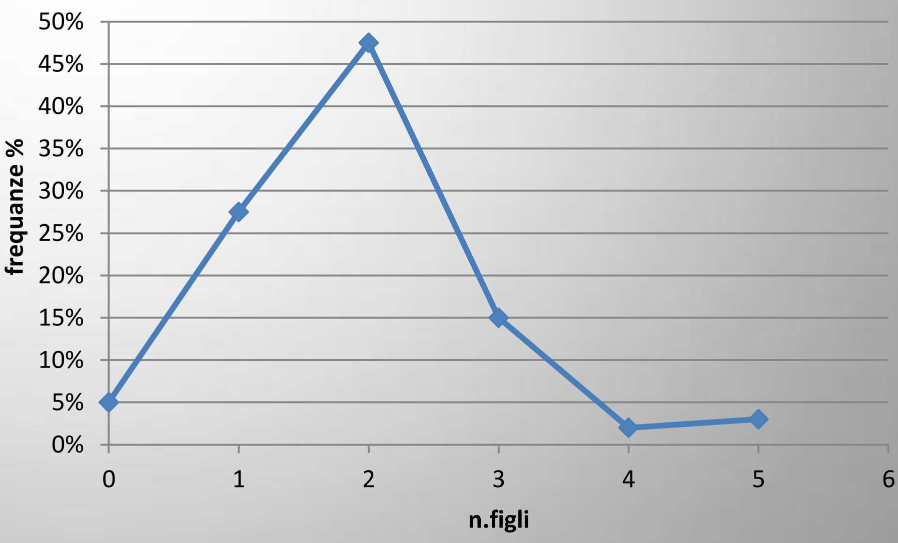


DIAGRAMMA CARTESIANO

- I diagrammi cartesiani si possono utilizzare per rappresentare le distribuzioni di frequenze di un carattere quantitativo.
- Per costruirli, si rappresentano i punti che hanno come ascisse i valori osservati e come ordinate le corrispondenti frequenze.
- Si possono collegare i punti ottenendo una poligonale

n. di figli	n. Famiglie intervistate (FREQUENZA ASSOLUTA)	FREQUENZA RELATIVA	FREQUENZA %
0	10	1/20	5%
1	55	11/40	27,5%
2	95	19/40	47,5%
3	30	3/20	15%
4	4	1/50	2%
5	6	3/100	3%



ELABORAZIONE DEI DATI

Un ulteriore elaborazione dei dati è cercare di sintetizzare con pochi numeri significativi i dati grezzi rilevati in un'indagine statistica

Gli <u>indici di posizione centrale</u> : la media, la mediana e la moda.

Gli <u>indici di dispersione o variabilità</u>: campo di variazione, scarto quadratico medio

ELABORAZIONE DEI DATI

MEDIA

È un indice di posizione centrale per caratteri quantitativi

Tra i tipi di media ricordiamo la **MEDIA ARITMETICA**

Si consideri una serie di n termini $x_1, x_2, ..., x_n$, la media aritmetica semplice, $\bar{x}$, è data dalla somma dei termini diviso il loro numero; in simboli:

$$\bar{x} = \frac{x_1 + x_2 + \ldots + x_n}{n} = \frac{\sum_{i=1}^{n} x_i}{n}$$

MEDIA ARITMETICA SEMPLICE

Esempio 1: si vuole valutare il profitto medio in matematica di uno studente che in 5 prove abbia riportato i seguenti voti:

u.s.	mod.
PROVA	**VOTO**
prima	7
seconda	8
terza	6
quarta	2
quinta	8

7+8+6+2+8 =

$= \bar{x} + \bar{x} + \bar{x} + \bar{x} + \bar{x}$

voto medio

$$\bar{x} = \frac{7+8+6+2+8}{5} =$$

6,2

MEDIA ARITMETICA PONDERATA

Si consideri una serie di k numeri x_1, x_2, ..., x_k, con frequenze assolute rispettivamente f_1, f_2, ..., f_k , la media aritmetica ponderata, $\bar{x}$, è data da:

$$\bar{x} = \frac{x_1 f_1 + x_2 f_2 + \cdots + x_k f_k}{f_1 + f_2 + \cdots + f_k} = \sum_{i=1}^{k} \frac{x_i f_i}{f_i}$$

Se le frequenze relative sono rispettivamente f_{r1}, f_{r2}, ..., f_{rk} , la media aritmetica ponderata, $\bar{x}$, è data da:

$$\bar{x} = x_1 f_{r1} + x_2 f_{r2} + \cdots + x_k f_{rk} = \sum_{i=1}^{k} x_i f_{ri}$$

MEDIA ARITMETICA PONDERATA

ESEMPIO: numero medi di figli

n. di figli	n. Famiglie intervistate (FREQUENZA ASSOLUTA)	FREQUENZA RELATIVA	FREQUENZA %
0	10	1/20	5%
1	55	11/40	27,5%
2	95	19/40	47,5%
3	30	3/20	15%
4	4	1/50	2%
5	6	3/100	3%
TOTALE	**200**	**1**	**100%**

$$\bar{x} = \frac{0 \cdot 10 + 1 \cdot 55 + 2 \cdot 95 + 3 \cdot 30 + 4 \cdot 4 + 5 \cdot 6}{200} = 1{,}905$$

MEDIA ARITMETICA PONDERATA

ESEMPIO: media aritmetica di un carattere suddiviso per classi

Altezza studenti	n. studenti (FREQUENZA ASSOLUTA)	Valore centrale
(160,165]	1	162,5
(165,170]	3	167,5
(170,175]	5	172,5
(175,180]	2	177,5
(180,185]	5	182,5
(185,190]	2	187,5
TOTALE	**18**	

Se un carattere è suddiviso per classi non è possibile calcolare il valore esatto della media aritmetica, perché non si conoscono esattamente i valori osservati all'interno di ogni classe. In questi casi conviene assumere come media aritmetica il valore che si calcola utilizzando il valore centrale di ogni classe.

$$\bar{x} = \frac{162{,}5 \cdot 1 + 167{,}5 \cdot 3 + 172{,}5 \cdot 5 + 177{,}5 \cdot 2 + 182{,}5 \cdot 5 + 187{,}5 \cdot 2}{18} = 176{,}1$$

ELABORAZIONE DEI DATI

MEDIANA

È un indice di posizione centrale per caratteri quantitativi (e qualitativi se si manifestano con modalità *ordinabili)*

Data una successione di n dati, disposti in ordine crescente (o decrescente), la mediana è il valore che occupa la posizione di mezzo.

MEDIANA

Esempio: dopo una gara tra 5 concorrenti viene compilata la classifica

u.s.	*mod.*
giocat.	**punti**
Barbara	8
Enza	8
Andrea	7
Carlo	6
Diego	2

Ordiniamo i dati: 8,8,7,6,2

informazione da conservare: punteggio centrale

MEDIANA

7

MEDIANA

Esempio : tra i 25 dipendenti di una ditta, il titolo di studio risulta così distribuito:

modalità ordinate	frequenza
Lic. elemen.	1
Lic. media	10
Diploma	8
Laurea	6

modalità con posizione centrale (la 13ª)

MEDIANA

Diploma

MEDIANA

<u>Esempio :</u> In un esame universitario sono stati assegnati

voto	frequenza
20	3
24	2
28	3
30	2

Ordiniamo i dati in ordine crescente:

Voti: 20,20,20,24,<u>24,28</u>,28,28,30,30

I dati sono in numero pari, perciò le posizioni centrali sono la quinta e la sesta, cioè il voto 24 e il voto 28.
In questo caso (i dati sono in numero pari) la mediana è data dalla media aritmetica (se il carattere è quantitativo!) dei due dati che occupano la posizione centrale:

$$MEDIANA = \frac{24+28}{2} = 26$$

MEDIANA

Per trovare con più facilità la posizione centrale conviene utilizzare la frequenza comulata

Esempio : In un esame universitario sono stati assegnati

voto	frequenza	Frequenza cumulata
20	3	3
24	2	5
28	3	8
30	2	10

I dati sono in numero pari (n=10). Le posizione centrali sono la n° 5 (n/2)e la n°6 (n/2 +1), che corrispondono al voto 24 e al voto 28. Perciò la mediana è:

$$MEDIANA = \frac{24+28}{2} = 26$$

MEDIANA

Esempio :

n. di figli	n. Famiglie intervistate (FREQUENZA ASSOLUTA)	FRE QUENZA CUMULATA
0	10	10
1	55	65
2	95	160
3	30	190
4	5	195
5	6	201
TOTALE	**201**	

I dati sono in numero dispari (201). La posizione centrali è la n° 101 ((n+1)/2), che corrisponde la dato 2 figli. Perciò la mediana è proprio 2!

MEDIANA

<u>Esempio :</u> mediana di un carattere suddiviso in classi

Altezza studenti	Valore centrale	n. studenti (FREQUENZA ASSOLUTA)	Frequenza cumulata
(160,165]	162,5	1	1
(165,170]	167,5	3	4
(170,175]	172,5	5	9
(175,180]	177,5	2	11
(180,185]	182,5	5	16
(185,190]	187,5	2	18
TOTALE		**18**	

I dati sono in numero pari (18). Le posizione centrali sono la n° 9 e la n°10, che appartengono rispettivamente alla classe (170,175] e (175,180].
La mediana sarà la media aritmetica dei valori centrali di queste due classi:

$$MEDIANA = \frac{172{,}5 + 177{,}5}{2} = 175$$

ELABORAZIONE DEI DATI

MODA

È un indice di posizione centrale per caratteri qualitativi e quantitativi .

Si chiama MODA la modalità che si presenta con la massima frequenza (assoluta, relativa o percentuale)

MODA

Esempio : tra i 4 corsi offerti da un centro sportivo, 25 ragazzi hanno optato per uno di essi secondo la seguente tabella:

modalità	frequenza
tennis	7
danza	4
calcio	12
atletica	2

modalità con frequenza max

MODA
calcio

MODA

Esempio :

La maggior parte dei dipendenti guadagna 1.700 euro

euro/mese	frequenza
23.000	1
9.400	1
6.500	2
2.600	3
2.200	19
1.700	22
1.300	2

moda = 1.700

MODA

Nel caso di caratteri quantitativi suddivisi in classi, al posto della moda si introduce il concetto di CLASSE MODALE:

- se le classi hanno la stessa ampiezza, la classe modale è quella di maggior frequenza
- se le classi hanno ampiezze diverse, la classe modale è quella con maggiore densità di frequenza.

Esempio :

Altezza studenti	n. studenti (FREQUENZA ASSOLUTA)
(160,165]	1
(165,170]	3
(170,175]	5
(175,180]	2
(180,185]	5
(185,190]	2

In questo caso sono due le classi con frequenza maggiore. E' una distribuzione BIMODALE. Le due classi modali sono (170,175] e (180.185].

OSS: vi sono caso in cui la moda non è un indice centrale significativo, in quanto non esiste un dato che sia notevolmente più frequente degli altri.

MODA

Esempio : classi di diversa ampiezza

Tabella16 - Impiegati di un ufficio per classi di età

CLASSI DI ETÀ	Impiegati Frequenza	Ampiezza della classe	Densità di frequenza
20-29	12	10	1,2
30-49	20	20	1,0
50-59	26	10	2,6
60-64	4	5	0,8
Totale	**62**		

La classe modale è 50-59, perché a tale classe corrisponde la densità di frequenza maggiore

ELABORAZIONE DEI DATI

INDICI DI VARIABILITA’

Un carattere importante dei dati statistici è la variabilità.

Per analizzare una distribuzione, dopo aver calcolato uno o più valori medi, si cerca di mettere in evidenza la dispersione dei dati, dispersione che caratterizza la variabilità del fenomeno.

Sono presenti vari indici che misurano la variabilità di un fenomeno (per caratteri quantitativi).

Campo di variazione

Scarto quadratico medio (deviazione standard)

ELABORAZIONE DEI DATI

CAMPO DI VARIAZIONE

Il campo di variazione è dato dalla differenza fra il maggiore ed il minore tra i dati rilevati. Tiene conto solo dei valori estremi e non degli altri.

Esempio: Considerando le 5 velocità rilevate dall' autovelox: 59,67,80,98,110.

Il campo di variazione è la differenza tra 110 e 59 ossia 51.

ELABORAZIONE DEI DATI

SCARTO QUADRATICO MEDIO

Consideriamo n dati numerici $x_1, x_2, \dots, x_n$, *sia* $\bar{x}$ *la media aritmetica.*
Per ogni dato è possibile calcolare lo scarto dalla media:

$$x_1 - \bar{x}, \quad x_2 - \bar{x}, \dots, \quad x_n - \bar{x},$$

Chiamiamo varianza V *di media aritmetica* $\bar{x}$,
la media aritmetica dei quadrati degli scarti

$$V = \frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \cdots + (x_n - \bar{x})^2}{n}$$

LO SCARTO QUADRATICO è la radice quadrata della varianza

$$\sigma = \sqrt{\frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \cdots + (x_n - \bar{x})^2}{n}}$$

ELABORAZIONE DEI DATI

SCARTO QUADRATICO MEDIO

Considerando la distribuzione di frequenza

Consideriamo k dati numerici $x_1, x_2, \dots, x_k$, con frequenze $f_1, f_2, \dots, f_k$, sia $\bar{x}$ la media aritmetica.

Per ogni dato è possibile calcolare lo scarto dalla media:

$$x_1 - \bar{x}, \qquad x_2 - \bar{x}, \dots, \qquad x_k - \bar{x},$$

Chiamiamo varianza V di media aritmetica $\bar{x}$, la media aritmetica dei quadrati degli scarti

$$V = \frac{(x_1 - \bar{x})^2 \cdot f_1 + (x_2 - \bar{x})^2 \cdot f_2 + \dots + (x_k - \bar{x})^2 \cdot f_k}{f_1 + f_2 + \dots + f_k}$$

LO SCARTO QUADRATICO è la radice quadrata della varianza

$$\sigma = \sqrt{\frac{(x_1 - \bar{x})^2 \cdot f_1 + (x_2 - \bar{x})^2 \cdot f_2 + \dots + (x_k - \bar{x})^2 \cdot f_k}{f_1 + f_2 + \dots + f_k}}$$

SCARTO QUADRATICO MEDIO

ESEMPIO

Tre studenti, hanno riportato le seguenti successioni di voti nelle prove scritte di matematica:

STUDENTI	VOTI
Anna	5,6,6,7
Gianni	4,5,7,8
Elena	3,4,8,9

Calcoliamo le medie aritmetiche dei voti di ogni studente, il campo di variazione e lo scarto quadratico medio.

SCARTO QUADRATICO MEDIO

STUDENTI	VOTI	media	Campo di variazione	Scarto quadratico medio
Anna	5,6,6,7	6	7-5=2	$\sqrt{\frac{(5-6)^2+(6-6)^2+(6-6)^2+(7-6)^2}{4}} = 0{,}71$
Gianni	4,5,7,8	6	8-4=4	$\sqrt{\frac{(4-6)^2+(5-6)^2+(7-6)^2+(8-6)^2}{4}} = 1{,}58$
Elena	3,4,8,9	6	9-3=6	$\sqrt{\frac{(3-6)^2+(4-6)^2+(8-6)^2+(9-6)^2}{4}} = 2{,}50$