



INTRODUZIONE ALLA STATISTICA

NOTE STORICHE SULLA STATISTICA

Riassumiamo brevemente le prime notizie storiche riguardo all'utilizzo della statistica.

- ✚ Cina - 2.300 a.c.: annotazioni e conteggi di abitanti, terreni
- ✚ Bibbia: conteggio degli Ebrei
- ✚ Impero Romano: "Census" ogni 5 anni
- ✚ Dal XV secolo viene definita la Statistica come "Scienza descrittiva degli Stati" (compaiono opere contenenti dati demografici ed economici)
- ✚ XVIII secolo
G.Achenwall - Università di Gottinga
Comparsa delle prime tabelle comparative delle risorse geografiche ed economiche dei vari Stati
- ✚ XIX secolo
Introduzione, nelle indagini Statistiche, del Calcolo delle Probabilità (Laplace, Pascal, Fermat)
- ✚ XX secolo La Statistica diventa una scienza autonoma. Sempre maggiore diffusione... e abuso!!! (esempio sondaggi commerciali, elezioni politiche,...)

CHE COSA E' LA STATISTICA

La statistica indaga su fenomeni collettivi, cioè su fenomeni che riguardano un insieme di individui, oggetti, beni raccogliendo informazioni relative ad essi traducendole poi in un modello numerico che possa essere analizzato semplicemente.

Il gruppo preso in esame è definito con il termine "popolazione". Il singolo elemento della popolazione è detto "individuo" o "unità statistica".

Facciamo alcuni esempi di popolazioni statistiche:

- I cittadini che hanno il diritto al voto nelle elezioni per il Parlamento.
- Gli abitanti della Francia.
- Le autovetture in circolazione attualmente in Italia.
- Le aziende tessili in Liguria.

- Gli studenti di una Istituzione Scolastica
- Le abitazioni esistenti in una cittadina
-

lo studio della statistica può essere suddiviso in due parti:

STATISTICA DESCRITTIVA

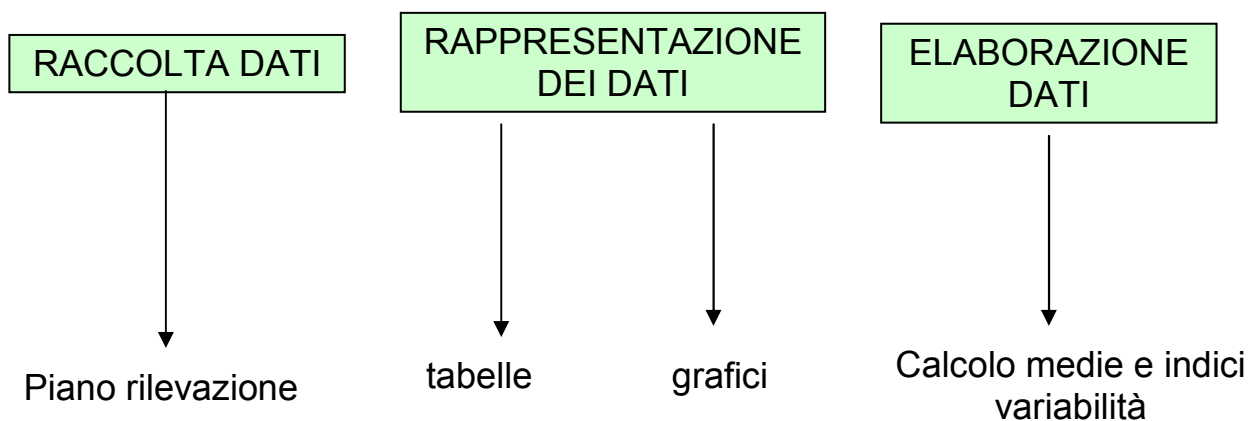
Raccoglie ed elabora i dati per studiare un fenomeno collettivo

STATISTICA INDUTTIVA

Si occupa dei metodi per stimare un fenomeno collettivo a partire da un campione ricavato da esso.

FASI DELLA STATISTICA DESCRITTIVA

Le principali fasi della statistica descrittiva le riportiamo di seguito:



LA RACCOLTA DEI DATI CHE COSA RILEVARE?

I caratteri che formano l'oggetto di una rilevazione statistica, possono essere:

- **Qualitativi**, espressi in forma verbale, sovente rappresentata da aggettivi.

Esempi: il colore; la nazionalità; lo stato civile; l'affidabilità; l'attitudine ai lavori manuali; l'attitudine agli studi...

- **Quantitativi**, espressi da numeri.

Esempi: la statura, il peso, il numero di stanze di un appartamento ...

Le modalità di un carattere quantitativo saranno, allora, espresse da numeri, che si chiamano anche "i valori" di quel carattere.

I caratteri qualitativi si possono classificare in:

- Sconnessi: relazioni di diversità (es. colore dei capelli)
- Ordinati: relazioni d'ordine (es. grado d'istruzione)

I caratteri quantitativi, invece, si possono classificare in:

- Discreti: insieme finito o infinito numerabile (es. numero componenti di una famiglia)
- Continui: carattere espresso con un numero reale (es. altezze di un certo gruppo di individui)

COME RACCOGLIERE I DATI?

La raccolta dei dati può avvenire attraverso l'utilizzo di:

- SCHEDE
- QUESTIONARI
- INTERVISTE (ANCHE TELEFONICHE)
- EXIT POLL

Tipo della rilevazione

Totale : viene effettuata su tutte le unità statistiche della popolazione (censimento) costi elevati e tempi elevati risultati affidabili

Parziale (o campionaria): viene effettuata su una parte (o campione) della popolazione costi contenuti rapidità risultati approssimati

Occasionale: fatta "quando serve" sondaggi d'opinione

Periodica: fatta regolarmente "ogni tanto tempo" censimento (10 anni)

COME RAPPRESENTARE I DATI?

Una volta raccolti i dati, essi devono essere rappresentati. Il modo più semplice ed immediato per rappresentare i dati è costituito da tabelle.

La rappresentazione può avvenire nel modo seguente:

- SERIAZIONE STATISTICA: Tabella contenente frequenze di intensità di un carattere QUANTITATIVO
- SERIE STATISTICA: Tabella contenente frequenze di modalità di un carattere QUALITATIVO
- SERIE STORICHE: Tabella contenente intensità di un carattere QUANTITATIVO rilevato per modalità TEMPORALI(p.es. le vendite di automobili negli ultimi 10 anni)

TITOLO DI	N.PERS ONE
lic.elemen	2.000
lic. Media	6.000
diploma	10.000
laurea	2.000

Serie statistica

NUMERO	N.APPARTA
1	300
2	500
3	2.000
4	3.000
5	150
6	100
7	300

Seriazione
statistica

Peso (in grammi)	N° neonati
1.800-	10
2.200-	32
2.600-	120
3.000-	254
3.400-	134
3.800-	40
4.200-	10

Seriazione statistica con dati divisi in classi

Si suddividono i dati in classi quando il carattere analizzato è continuo

ANNO	VENDITE AUTO
1996	10.000
1997	15.000
1998	20.000
1999	18.000
2000	17.000
2001	22.000
2002	23.000
2003	10.000

Serie temporale

- Il numero che rappresenta quante unità statistiche presentano un certo carattere prende il nome di *frequenza assoluta o intensità* di quella modalità.
- L'insieme delle coppie ordinate(modalità, frequenza assoluta) si dice distribuzione di frequenze, cioè una funzione che può essere rappresentata elencandone tutte le coppie organizzate in una tabella.
- E' possibile calcolare anche le cosiddette frequenze relative che possono anche essere espresse in forma percentuale
- *Frequenza assoluta o intensità* è il numero di individui il cui carattere assume un determinato valore.
- *Frequenza relativa* è il rapporto tra la frequenza assoluta e la totalità della popolazione statistica su cui si sta svolgendo l'indagine.
- *Frequenza percentuale* è semplicemente la frequenza relativa espressa in termini percentuali.

Come esempio si riportano i risultati su un'indagine condotta su un gruppo di 50 persone relativa al numero di sigarette fumate ogni giorno:

Carattere	Frequenza assoluta	Frequenza relativa	Frequenza percentuale	Frequenza cumulata
0	20	0,4	40 %	40%
5	10	0,2	20 %	60%
10	10	0,2	20 %	80%
15	5	0,1	10 %	90%
20	5	0,1	10%	100%
TOTALE	50	1	100%	

$$F_{relativa} = F_{assoluta} / Totale$$

Talvolta può essere utile calcolare anche la frequenza cumulata (percentuale o relativa)

USO DEI GRAFICI

I dati raccolti in tabelle possono essere rappresentati anche attraverso grafici che offrono il vantaggio di una descrizione del fenomeno in forma visiva.

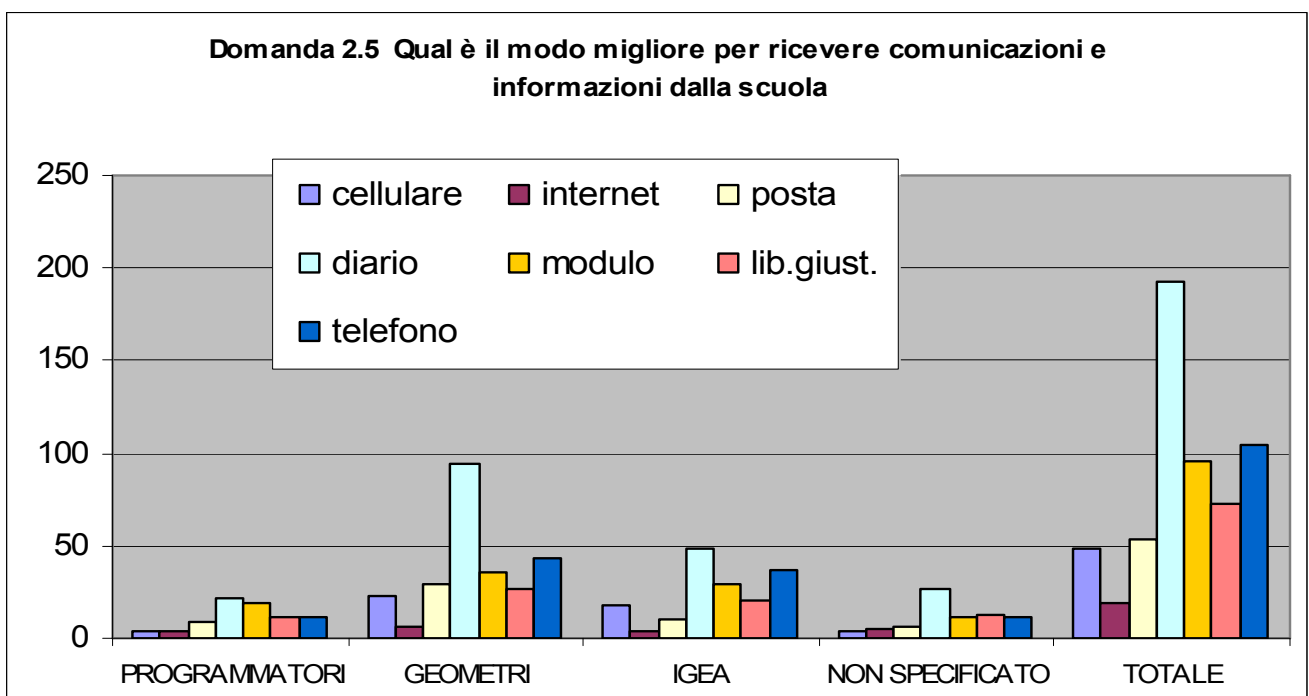
Per la rappresentazione grafica dei caratteri qualitativi vengono utilizzati i seguenti tipi di grafici:

- Ortogrammi o grafici a colonne
- Diagrammi circolari o areogrammi
- Ideogrammi
- Cartogrammi

Ortogrammi

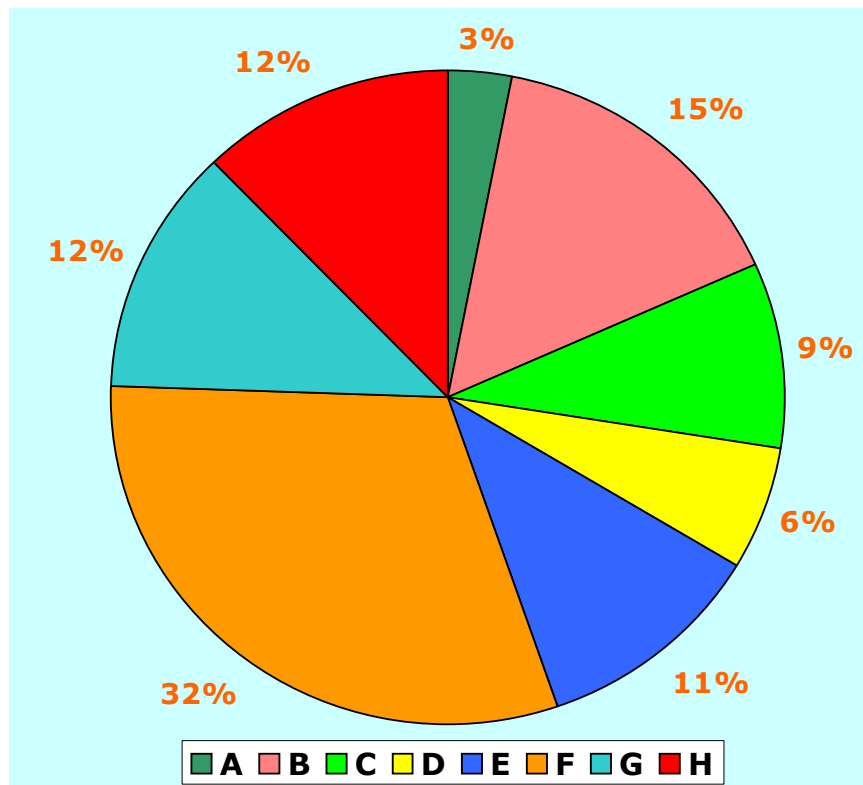
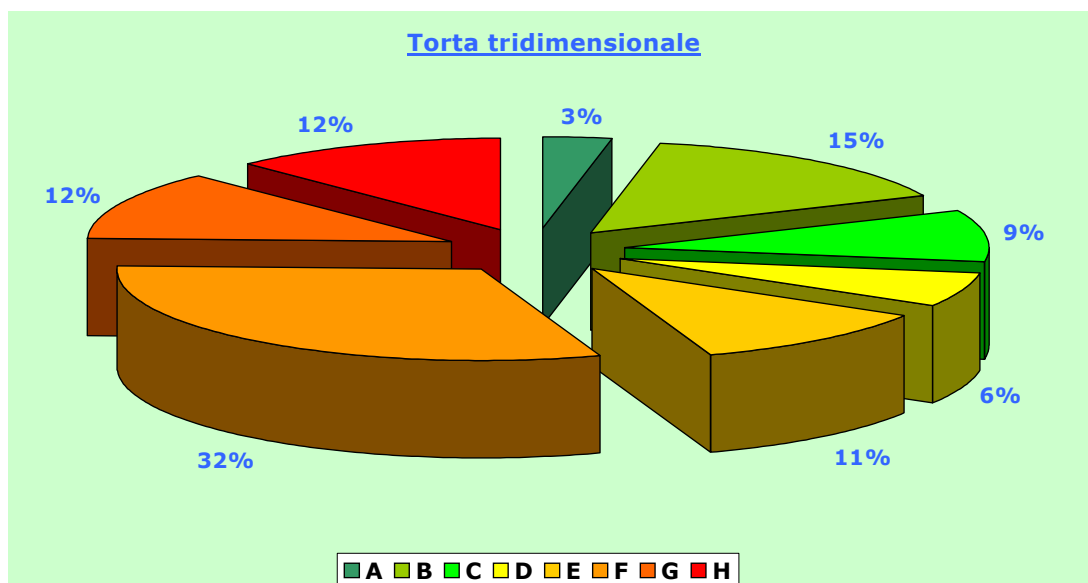
Questi grafici sono costituiti da rettangoli innalzati in corrispondenza ai diversi caratteri che si considerano e hanno:

basi uguali e altezze proporzionali alle frequenze.



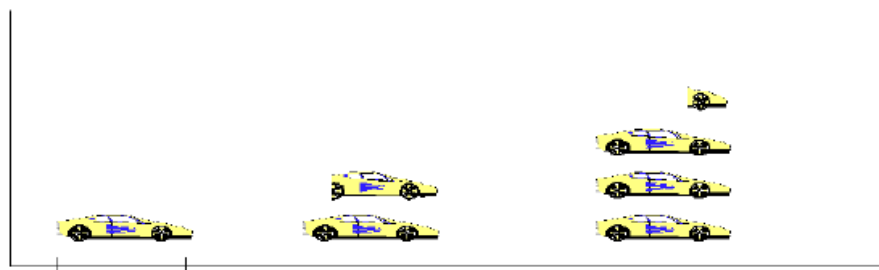
Diagrammi circolari o aerogrammi

Si ottengono dividendo un cerchio in settori circolari aventi angolo al centro proporzionale alle frequenze o intensità che rappresentano

Diagrammi circolari

Ideogrammi

Sono rappresentazioni mediante figure stilizzate che rappresentano il fenomeno studiato



Cartogrammi

Sono utilizzati per rappresentare serie territoriali



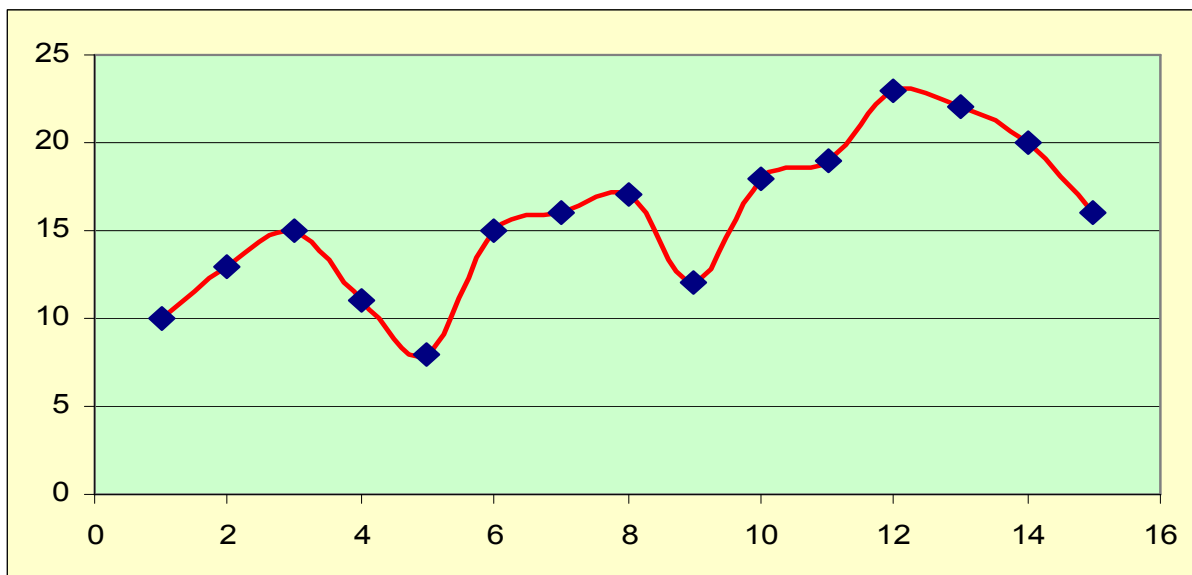
Invece per quanto riguarda le distribuzioni statistiche di carattere quantitativo si usano principalmente i seguenti grafici:

- Diagrammi cartesiani
- Istogrammi
- Diagrammi polari

Analizziamo ora in particolare i principali tipi di grafici:

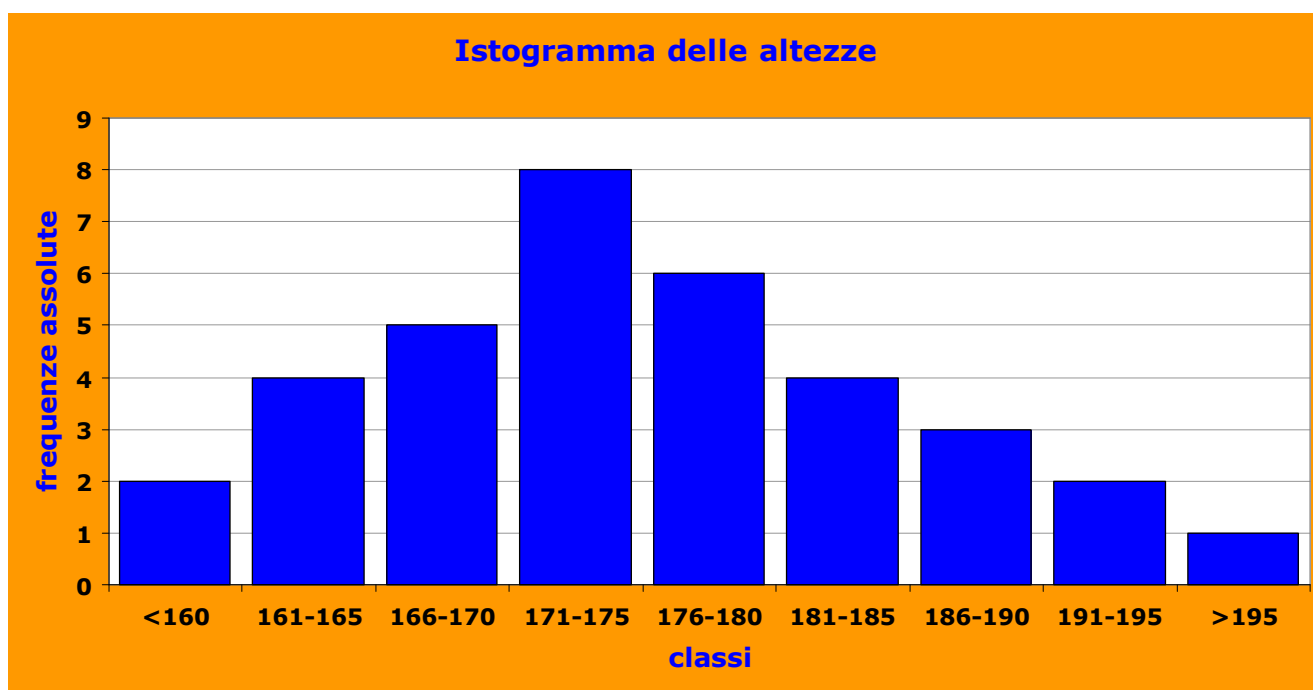
Diagrammi cartesiani

sono usati soprattutto per le serie temporali e possono essere sia discreti che continui.



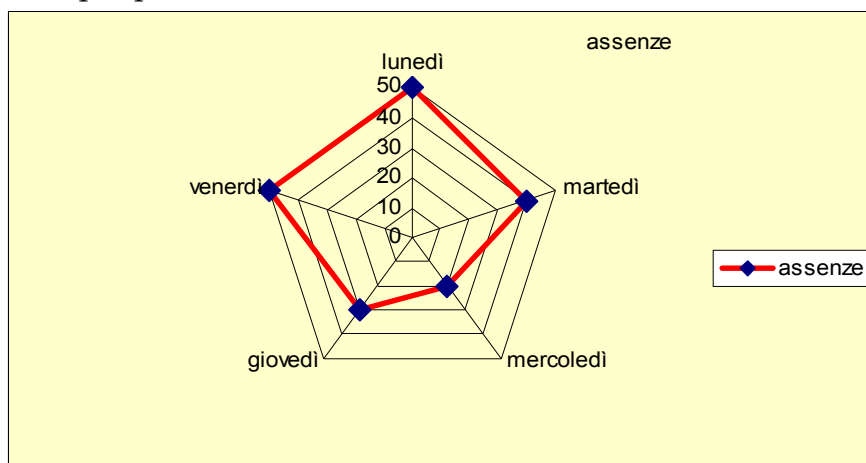
Istogrammi

Si usano per rappresentare seriazioni continue con i dati raggruppati in classi. Le altezze dei rettangoli sono proporzionali alle frequenze



Diagrammi polari

sono usati per particolari serie storiche con carattere di ciclicità



ELABORAZIONE DEI DATI

Questa fase serve ad esprimere in modo sintetico i risultati dell'indagine svolta calcolando degli indici:

- valori medi
- indici di variabilità

I valori medi

Il valore medio è un valore che esprime una tendenza centrale. Secondo *Cauchy* la media di un insieme è un valore compreso tra il minimo e il massimo.

Medie di calcolo: sono quelle che si calcolano tenendo conto di tutti i valori della distribuzione

- o **Media aritmetica e sue proprietà**
- o **Media Geometrica**
- o **Media quadratica**
- o **Media armonica**

Medie di posizione: sono quelle che si calcolano tenendo conto solo di alcuni valori della distribuzione

- o **Mediana**
- o **Moda o valore normale**

MEDIA ARITMETICA

$$M = \frac{x_1 + x_2 + \dots + x_n}{n}$$



Questa viene chiamata media semplice ed è utilizzata quando i valori rilevati non sono riportati con le frequenze

$$M = \frac{x_1 \cdot f_1 + x_2 \cdot f_2 + \dots + x_n \cdot f_n}{N} = \frac{\sum x \cdot f}{\sum f}$$



Questa viene chiamata media ponderata ed è utilizzata quando i valori rilevati hanno frequenze diverse

PROPRIETA' DELLA MEDIA ARITMETICA

Si definisce scarto della media aritmetica la differenza tra il valore osservato e la media stessa

$$x_i - M$$

La somma degli scarti positivi dalla media aritmetica è uguale, in valore assoluto, a quella degli scarti negativi, e quindi la somma algebrica di tutti gli scarti dalla media è uguale a zero.

Pertanto ne consegue che: $\sum (x_i - M) = 0$

Se i valori sono *tutti positivi e non nulli* si può calcolare la media geometrica.

MEDIA GEOMETRICA

Si definisce **media geometrica** dei valori $x_1, x_2, \dots, x_n$, quel numero G che sostituito ai valori x_i lascia invariato il loro prodotto:

$$G = \sqrt[n]{x_1 \cdot x_2 \cdot \dots \cdot x_n}$$

che è la *media geometrica semplice*.

Nel caso di valori x_i con frequenze o pesi y_i , si ha *media geometrica ponderata* :

$$G = \sqrt[N]{x_1^{f_1} \cdot x_2^{f_2} \cdot \dots \cdot x_n^{f_n}}$$

Dove f sono le frequenze

$$N = \sum_{i=1}^n f_i$$

QUANDO USARE LA MEDIA GEOMETRICA?

- La media geometrica viene utilizzata quando si vuole analizzare il variare di un fenomeno nel tempo, per esempio il tasso di variazione dei prezzi o i tassi di rendimento di capitali.

ESEMPIO

Un capitale è stato impiegato per 4 anni al tasso del 2%, per altri 3 anni al tasso del 3 % ed in fine per 2 anni al tasso del 5%.

Per il calcolo del tasso medio si può impiegare la Media Geometrica Ponderata:

$$i = \sqrt[9]{0,02^4 \cdot 0,03^3 \cdot 0,05^2}$$

LA MEDIA QUADRATICA

$$Q = \sqrt{\frac{x_1^2 + x_2^2 + \dots + x_n^2}{n}}$$

Media quadratica semplice

$$Q = \sqrt{\frac{x_1^2 \cdot f_1 + x_2^2 \cdot f_2 + \dots + x_n^2 \cdot f_n}{N}}$$

Media quadratica ponderata

Dove f sono le frequenze

dove

$$N = \sum f_i$$

La media quadratica è quella con valore maggiore e viene usata per mettere in evidenza i valori che si scostano molto dai valori centrali

LA MEDIA ARMONICA

$$A = \frac{n}{\frac{1}{x_1} + \frac{1}{x_2} + \dots + \frac{1}{x_n}} \quad \text{Media armonica semplice}$$

$$\text{Media armonica ponderata} \quad A = \frac{N}{\frac{f_1}{x_1} + \frac{f_2}{x_2} + \dots + \frac{f_n}{x_n}}$$

Dove f sono le frequenze e dove $N = \sum f_i$

Questa media è utilizzata quando ha significato il calcolo del **reciproco** di una certa grandezza.

Per esempio si definisce potere di acquisto di una moneta la quantità di beni acquistabile con una unità di tale moneta. Se, per esempio un prodotto costa 0,85 euro il potere di acquisto riferito a tale bene è: $1/0,85$.

Ora supponiamo che uno stesso prodotto in 4 città differenti abbia i seguenti prezzi:

Milano 0,85 euro Torino 0,90 Roma 0,75 Bari 0,70

$$\text{Il potere di acquisto medio sarà} \quad \frac{4}{\frac{1}{0,85} + \frac{1}{0,90} + \frac{1}{0,75} + \frac{1}{0,70}}$$

LA MEDIANA

Dati un insieme di valori $x_1, x_2, x_3, \dots, x_n$ ordinati si definisce Mediana il valore che bipartisce la successione.

Ovvero il valore centrale se il numero dei termini è dispari

La media aritmetica dei due valori centrali se il numero dei termini è pari.

Esempio:

Calcolare la Mediana dei seguenti valori:

4	7	8	9	10	12	55
---	---	---	---	----	----	----

Il Valore 9 è quello che sta in centro ed è pertanto la mediana

Esempio:

Calcolare la Mediana dei seguenti valori:

8	22	33	36	38	41	44	46
---	----	----	----	----	----	----	----

La mediana è la media aritmetica dei due valori centrali $(36+38)/2 = 37$

Se si ha invece una distribuzione di frequenze, occorre calcolare le frequenze cumulate. Indicando con N la somma delle frequenze, la Mediana è il valore corrispondente

- a $N/2$ se N è pari
- a $(N+1)/2$ se N è dispari

Ad esempio se abbiamo la seguente indagine

N.Figli	Frequenze assolute	Frequenze cumulate
0	3	3
1	8	11
2	7	18
3	4	22
4	1	23
5	1	24
6	1	25
	Tot 25	Tot 25

$N = 25$ quindi la mediana è il valore corrispondente a $(N+1)/2 = (25+1)/2 = 13$ cioè il valore 2

Se i dati sono raggruppati in classi si determina la classe mediana mediante le frequenze cumulate quindi si procede con un'interpolazione lineare.

Classi cilindrata(cc)	F.ass.	F. cum.
0-800	100	100
800-1200	840	940
1200-1600	955	1895
1600-2000	305	2200
2000-2500	200	2400
2500-3000	100	2500

	Tot	2500
	2500	

Poiché N è pari $N/2 = 1250$ che corrisponde alla Classe Mediana 1200-1600

Per giungere ad un valore preciso della Mediana si deve impostare una proporzione:

$$\begin{array}{ccc} 1200 & & 940 \\ x & 1250 & \\ 1600 & & 1895 \end{array}$$

$$(x-1200) : (1600-1200) = (1250-940) : (1895-940)$$

LA MODA

Si dice Moda o Valore Modale di una distribuzione di frequenze il valore corrispondente alla massima frequenza.

Esempio

Voti matematica	F.ass.
3	2
4	1
5	7
6	5
7	4
8	2
9	1
10	1

La Moda è il valore 5 poiché è il valore corrispondente alla frequenza più alta.

Se i valori sono raggruppati in classi si determina la Classe Modale:

Se l'ampiezza delle classi è costante si dirà classe modale quella con frequenza maggiore;

Se l'ampiezza delle classi non è costante si divide ogni frequenza per l'ampiezza della rispettiva classe calcolando così la densità di frequenza; il valore più alto di densità definisce la classe modale

N.B. esistono delle distribuzioni di frequenza che presentano più valori della moda (distribuzioni plurimodali)

COME SI SCEGLIE UNA MEDIA

Si possono solo fare considerazioni a carattere generale:

- La media aritmetica è più utilizzata per esprimere un concetto di equidistribuzione
- La media geometrica è utilizzata quando i dati si susseguono in progressione geometrica
- La media quadratica quando si vuole eliminare l'influenza dei segni e quando si vuole dare importanza a valori piuttosto grande
- La moda si usa quando è importante rilevare il valore che ha maggior probabilità di presentarsi
- La mediana è caratterizzata dal fatto di non essere influenzata dalla rilevante differenza tra i dati

LA VARIABILITA'

I valori medi sono indici importanti per la descrizione sintetica di un fenomeno statistico però non forniscono informazioni sulla dispersione dei dati cioè sulla loro variabilità

Esempio

- Due gruppi di studenti hanno sostenuto una prova di matematica con i seguenti risultati:

1° Gruppo	3	4	5	8	10
2° gruppo	5	6	6	7	6

La media dei voti per entrambi i gruppi vale 6, ma è evidente che i dati del 1° gruppo sono molto più dispersi

Per stimare la variabilità si ricorre ad alcuni indici:

- Campo di variazione
- Scarto quadratico medio
- Varianza
- Coefficiente di variazione
- Scostamento semplice medio

CAMPO DI VARIAZIONE

E' il più semplice degli indici di variazione:

Si calcola facendo la differenza tra il dato più grande e il dato più piccolo

$$\text{Campo variazione } R = x_{\max} - x_{\min}$$

Il C.V. è molto semplice da calcolare però è poco significativo poiché tiene conto solo del valore più piccolo e di quello più grande trascurando tutti gli altri valori.

Il campo di variazione dà informazioni sulla distribuzione dei dati:

più R è piccolo più i dati sono concentrati;

più R è grande più i dati sono dispersi

SCARTO QUADRATICO MEDIO E VARIANZA

Sono gli indici di variabilità più utilizzati, e tengono conto della distribuzione di tutti i dati.

Scarto quadratico medio (o deviazione standard)

Rappresenta la media quadratica, semplice o ponderata degli scarti dei dati dalla media aritmetica M

$$\text{Scarto quadratico medio} = \sigma = \sqrt{\frac{(x_1 - M)^2 + (x_2 - M)^2 + \dots + (x_n - M)^2}{n}} = \sqrt{\frac{\sum_{i=1}^n (x_i - M)^2}{n}}$$

Con

$$\sigma = \sqrt{\frac{\sum_{i=1}^x (x_i - M)^2 \cdot f_i}{N}}$$

Dove $N = \sum f_i$

E' definita come il quadrato dello scarto quadratico medio e viene indicata con σ^2

Per il calcolo della varianza e dello scarto quadratico medio si preferisce adottare la seguente formula facilmente dimostrabile:

che se i dati sono senza frequenza vale la prima formula mentre se i dati hanno frequenza vale la seconda formula

$$\sigma^2 = \frac{\sum_{i=1}^n x_i^2}{n} - M^2$$

$$\sigma^2 = \frac{\sum_{i=1}^n x_i^2 \cdot f_i}{N} - M^2$$

dove $N = \sum f_i$

La varianza σ^2 e lo scarto quadratico medio σ danno informazioni sulla distribuzione dei dati:

- più sono piccoli più i dati sono concentrati;
- più sono grandi più i dati sono dispersi.

Vediamo un esempio

valore	frequenza
2	3
4	1
8	2
11	4

Calcolare varianza e scarto quadratico medio

$$M = \frac{2 \cdot 3 + 4 \cdot 1 + 8 \cdot 2 + 11 \cdot 4}{10} = 7$$

$$\sigma^2 = \frac{\sum_{i=1}^n x_i^2 \cdot f_i}{N} - M^2 = \frac{2^2 \cdot 3 + 4^2 \cdot 1 + 8^2 \cdot 2 + 11^2 \cdot 4}{10} - 7^2 = 15 \quad \sigma = \sqrt{15} = 3,87$$

IL COEFFICIENTE DI VARIAZIONE CV

Il CV è una misura relativa di dispersione (le precedenti sono misure assolute) ed è una grandezza adimensionale.

E' particolarmente utile quando si devono confrontare le distribuzioni di due gruppi con medie molto diverse o con dati espressi in scale differenti (es. confronto tra variazione del peso e variazione dell'altezza).

$$CV = \left(\frac{\sigma}{M} \cdot 100 \right) \%$$

SCOSTAMENTO SEMPLICE MEDIO

Un altro modo per calcolare la variabilità dei dati (tenendo conto di tutti i dati) consiste nel calcolare la distanza di tutti i dati dalla media e fare la media aritmetica di tali distanze

Scostamento semplice medio = Distanza media dei dati dalla media

Se i dati sono senza frequenza vale la prima formula, la seconda in caso contrario

$$S_M = \frac{\sum_{i=1}^n |x_i - M|}{n}$$

$$S_M = \frac{\sum_{i=1}^n |x_i - M| \cdot f_i}{N}$$

Dove $N = \sum f_i$

In alcuni casi in tali formule è possibile sostituire alla Media Aritmetica M la Mediana Me

Lo scostamento semplice medio dalla media dà informazioni sulla distribuzione dei dati:

- più SM è piccolo più i dati sono concentrati;
- più SM è grande più i dati sono dispersi

LA CONCENTRAZIONE

Un particolare aspetto della variabilità di un fenomeno è la **concentrazione**. Lo studio della concentrazione è utile per vedere se il fenomeno è equamente distribuito fra tutte le unità statistiche oppure è concentrato in poche unità.

Ad esempio, si può affermare che la distribuzione delle scuole elementari è poco concentrata sul territorio italiano perché quasi tutti i Comuni ne possiedono almeno una, mentre la distribuzione delle Università è molto concentrata perché pochi Comuni sono sedi di Università.

Sono state introdotte varie misure della concentrazione. Interessante è il **metodo grafico di Lorenz** applicato allo studio della distribuzione della ricchezza; esso è assai utile se i dati sono numerosi e raggruppati in classi. Data una variabile statistica:

X_1	Y_1
X_2	Y_2
...	...
X_n	Y_n

Dove x_i sono i valori nel discreto o i centri delle classi nel caso di raggruppamenti in classi e le y_i sono le frequenze, si considerino i prodotti $x_i y_i$ (essi rappresentano l'intensità del carattere della classe i -esima) e la loro somma:

$$\sum_{i=1}^n x_i y_i$$

(rappresenta l'intensità globale del fenomeno). Si calcolano le frequenze cumulate e le intensità cumulate, quindi le frequenze relative cumulate e le intensità relative cumulate, che si ottengono, le prime dividendo le frequenze cumulate per la somma delle frequenze e le seconde dividendo le intensità cumulate per l'intensità globale. Se le intensità relative cumulate sono *eguali* alle frequenze relative cumulate, significa che il fenomeno è equidistribuito poiché a una certa frazione delle unità statistiche corrisponde una frazione uguale dell'intensità del fenomeno; se, invece, sono *minori*, il fenomeno è tanto più concentrato quanto più le intensità relative cumulate differiscono dalle frequenze relative cumulate. La concentrazione è massima quando la curva di concentrazione coincide con i cateti del triangolo OAB, in pratica quando l'intensità globale è concentrata in una sola unità statistica. Se l'intensità è equamente distribuita, la curva di concentrazione coincide con la retta di equidistribuzione OB. Quanto maggiori sono le differenze $p_i - q_i$, altrettanto maggiore è la concentrazione. Un **rapporto di concentrazione** si ottiene dividendo la misura dell'area di concentrazione per la misura dell'area del triangolo OAB;

$$R = \frac{\text{area di concentrazione}}{\text{area del triangolo OAB}}$$

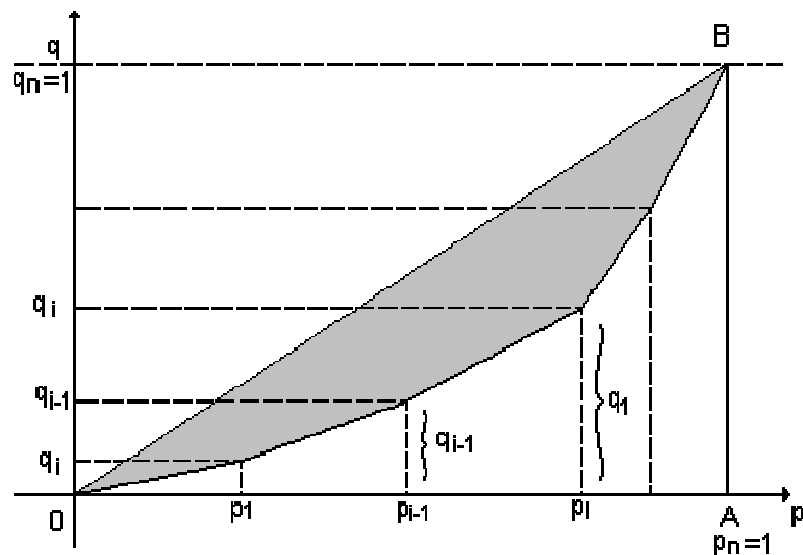
Risulta:

$$0 \leq R \leq 1$$

Se l'area di concentrazione è zero, ossia se la curva di concentrazione coincide con la retta di equidistribuzione, allora $R = 0$ e si afferma che non esiste concentrazione. Se l'area di concentrazione coincide con quella del triangolo OAB, allora $R = 1$ e la concentrazione è massima. In tutti gli altri casi si ha $0 < R < 1$. Quindi si ha:

$$R = \frac{\text{area triangolo OAB} - \text{somma area dei trapezi}}{\text{area triangolo OAB}}$$

Graficamente avremo:



Considerando che:

L'area del triangolo OAB equivale a $\frac{1}{2}$,

L'area di ogni trapezio si ottiene con la formula:

$$A = \frac{1}{2} \cdot h \cdot (B + b)$$

e nel nostro caso:

$$A = \frac{1}{2} \cdot (p_i - p_{i-1}) \cdot (q_i + q_{i-1})$$

Il primo trapezio in realtà è un triangolo rettangolo,

la formula per il calcolo di R diventa:

$$R = \frac{\frac{1}{2} - \frac{1}{2} \cdot \sum_{i=1}^{i=n} (p_i - p_{i-1}) \cdot (q_i + q_{i-1})}{\frac{1}{2}}$$

La formula di R si può semplificare dividendo il numeratore e il denominatore per $\frac{1}{2}$:

$$R = 1 - \sum_{i=1}^{i=n} (p_i - p_{i-1}) \cdot (q_i + q_{i-1})$$

Effettuando il prodotto otteniamo:

$$R = 1 - \sum (p_i q_i + p_i q_{i-1} - p_{i-1} q_i - p_{i-1} q_{i-1}) = 1 - \sum p_i q_i - \sum p_i q_{i-1} + \sum p_{i-1} q_i + \sum p_{i-1} q_{i-1}$$

Poiché:

$$\sum p_i q_i - \sum p_{i-1} q_{i-1} = p_n q_n$$

e $p_n q_n$ è uguale a 1, la precedente uguaglianza diventa:

$$R = 1 - 1 - \sum p_i q_{i-1} + \sum p_{i-1} q_i$$

In definitiva il rapporto di concentrazione risulta:

$$R = \sum_{i=1}^{i=n} p_{i-1} q_i - \sum_{i=1}^{i=n} p_i q_{i-1}$$