MEDIE

- Passiamo ad un ulteriore elaborazione per ottenere una sintesi dei dati e consentire la comparazione del fenomeno cui si riferiscono i dati con altri fenomeni consimili.

- Spesso si rende necessario sintetizzare tutte queste cifre grezze in un unico valore che colga il sottofondo costante della molteplicità dei valori riguardanti il fenomeno collettivo

- Si può operare in modo diverso:
  - prendendo per base alcune quantità assunte come invarianti;
  - prendendo per base alcune condizioni tra gli scarti.

- Si ottengono così più valori che potranno essere assunti per esprimere sinteticamente l'intensità di un fenomeno collettivo: ***i valori medi***

- I valori medi sono dei valori caratteristici delle serie o delle distribuzioni statistiche

- Utilizzeremo il procedimento che assume per basi delle quantità come invarianti.

- Consideriamo i valori osservati x_1, x_2, …..x_N del carattere X in un collettivo di N unità.

- Per esprimere una grandezza che dipende dai valori x_i il modo migliore è scegliere una funzione matematica dei valori stessi.

- Si dice valore medio della X, al fine di lasciare immutata una determinata grandezza assunta come invariante ed espressa dalla funzione scelta f (x_1, x_2, …..x_N), quel valore $\bar{x}$ tale che, sostituito alle x_1, x_2, …..x_N soddisfi l'uguaglianza

$$f(x_1, x_2, \ldots x_N) = f(\bar{x}, \bar{x}, \ldots \ldots, \bar{x}) \text{ nel caso di una serie}$$

$$f(x_1, x_2, \ldots x_s) = f(\bar{x}, \bar{x}, \ldots \ldots, \bar{x}) \text{ nel caso di una distribuzione}$$

- Le medie si distinguono in:
  - medie effettive o reali, se i valori $\overline{x}$ sono uguali a una delle modalità osservate;
  - medie di conto, se non coincidono con alcuna modalità;
  - medie analitiche, prendono in considerazione tutti i valori della serie o della distribuzione;
  - medie lasche, prendono in considerazione solo alcuni valori che occupano determinate posizioni.

- Condizione di Cauchy: $x_1 \leq \overline{x} \leq x_N$

MEDIE ANALITICHE – Media aritmetica

- Se la funzione che si assume come invariante è di tipo additivo, vuol dire che si mira a lasciare costante l'ammontare totale del carattere, allora la media da utilizzare è la media aritmetica.

- Tale media risolve il problema della uniforme ripartizione del carattere fra le varie unità statistiche del collettivo

- Nel caso di una serie di valori, si avrà

$f(x_1, x_2, \ldots x_N) = x_1+x_2+ \ldots +x_N \Longrightarrow x_1+x_2+ \ldots +x_N=\bar{x}+\bar{x}+\ldots+\bar{x}$

$$\boxed{\bar{x}=\mu =\frac{\sum_{i=1}^{N} x_i}{N}}$$

Media Aritmetica Semplice

- Nel caso di una distribuzione di valori, si avrà

$f(x_1, x_2, \ldots x_s) = x_1n_1+x_2n_2+ \ldots +x_sn_s \Longrightarrow x_1n_1+x_2n_2+ \ldots +x_sn_s =\bar{x}n_1+\bar{x}\, n_2+\ldots+\bar{x}n_s$

$$\boxed{\bar{x}=\mu =\frac{\sum_{i=1}^{s} x_i n_i}{\sum_{i=1}^{s} n_i}=\frac{\sum_{i=1}^{s} x_i n_i}{N}}$$

Media Aritmetica ponderata

MEDIE ANALITICHE – Media aritmetica

- Es. 20 alunni di una classe liceale hanno conseguito i seguenti voti in matematica: 5,3,6,6,7,7,8,5,8,7,6,4,4,5,3,3,7,7,8,6. Si calcoli la media aritmetica

- Se lasciamo i valori così come li abbiamo rilevati siamo in presenza di una serie di valori, quindi calcoleremo la media aritmetica semplice:

$$\bar{x}=\mu = \frac{\sum_{i=1}^{N} x_i}{N} = \frac{5+3+6+6+7+7+8+5+8+7+6+4+4+5+3+3+7+7+8+6}{20} =5{,}75$$

- Se ordiniamo i valori in una tabella avremo una distribuzione statistica e calcoleremo la media aritmetica ponderata:

$$\bar{x}=\mu = \frac{\sum_{i=1}^{s} x_i n_i}{\sum_{i=1}^{s} n_i} = \frac{115}{20} = 5{,}75$$

Voti x_i	**Alunnni** n_i	$x_i n_i$
3	3	9
4	2	8
5	3	15
6	4	24
7	5	35
8	3	24
Totale	20	115

MEDIE ANALITICHE – Media aritmetica

- Proprietà della media aritmetica:
  1. la somma algebrica degli scarti è uguale a 0: $\Sigma(x_i-\mu)=0$.
  2. la somma dei quadrati degli scarti è un minimo:

$$\left.\begin{array}{r}\Sigma(x_i-\mu)^2 \\ \\ \Sigma(x_i-\mu)^2 n_i\end{array}\right\} = \textit{minimo}$$

  3. la media aritmetica è associativa nel senso che possiamo sostituire ad un numero qualunque di valori diversi dalla X un ugual numero di valori tutti uguali alla loro media e il risultato non cambia.
  4. gode della proprietà traslativa ossia aggiungendo a tutti i valori x_i una costante la media risulta aumentata dello stesso valore.
  5. moltiplicando le x_i per una costante b diversa da 0 anche la media risulta moltiplicata per la stessa costante. La media aritmetica è dunque omogenea.
  6. la media è funzione crescente rispetto ad ognuna delle quantità singole x_i.
  7. se le modalità sono in progressione aritmetica e se N è un numero dispari, la media aritmetica coincide con il termine che occupa la posizione centrale nella graduatoria dei valori ordinati.

MEDIE ANALITICHE – Media geometrica

- Se la funzione che si assume come invariante è di tipo moltiplicativo la media da utilizzare è la media geometrica.

- La determinazione di tale media è fondata sulla invarianza del prodotto

- Nel caso di una serie di valori, si avrà

$f\,(x_1, x_2, \ldots x_N) = x_1 * x_2 * \ldots * x_N \qquad x_1 * x_2 * \ldots * x_N = \bar{x} * \bar{x} * \ldots * \bar{x} = \bar{x}^N$

$$\boxed{\bar{x} = M_g = \sqrt[N]{\prod_{i=1}^{N} x_i}}$$

Media geometrica semplice

- Nel caso di una distribuzione di valori, si avrà

$f\,(x_1, x_2, \ldots x_s) = x_1^{n1} * x_2^{n2} * \ldots * x_s^{ns}$

$x_1^{n1} * x_2^{n2} * \ldots * x_s^{ns} = \bar{x}^{n1} * \bar{x}^{n2} * \ldots * \bar{x}^{ns}$

$x_1^{n1} * x_2^{n2} * \ldots * x_s^{ns} = \bar{x}^{n1+n2+\ldots+ns}$

$$\boxed{\bar{x} = Mg = \sqrt[N]{\prod_{i=1}^{S} x_i^{ni}}}$$

Media geometrica ponderata

MEDIE ANALITICHE – Media geometrica

- Proprietà della media geometrica:
  1. La media geometrica di più rapporti è pari al rapporto tra la media geometrica dei termini a numeratore e la media geometrica dei termini a denominatore.

$$\underbrace{\sqrt[N]{\prod_{i=1}^{N} \frac{x_i}{y_i}} = \frac{\sqrt[N]{\prod_{i=1}^{N} x_i}}{\sqrt[N]{\prod_{i=1}^{N} y_i}}}_{\textit{serie}} \qquad \underbrace{\sqrt[N]{\prod_{i=1}^{S} \left(\frac{x_i}{y_i}\right)^{ni}} = \frac{\sqrt[N]{\prod_{i=1}^{S} x_i^{\mathrm{ni}}}}{\sqrt[N]{\prod_{i=1}^{S} y_i^{\mathrm{ni}}}}}_{\textit{distribuzione}}$$

  2. La media geometrica è omogenea, ossia moltiplicando le x_i per una costante b, diversa da 0, anche la media risulta moltiplicata per la stessa costante.

  3. Se le modalità sono in progressione geometrica e se N è un numero dispari, la media geometrica coincide con il termine che occupa la posizione centrale nella graduatoria dei valori ordinati.

MEDIE ANALITICHE – Media armonica

- Qualora il problema richieda di utilizzare la proporzionalità inversa al singolo termine, e siamo nel caso di una distribuzione, la funzione che si assume come invariante è la seguente:

$$f\left(\frac{n_1}{x_1}, \frac{n_2}{x_2}, \ldots\ldots, \frac{n_s}{x_s}\right) = \frac{n_1}{x_1} + \frac{n_2}{x_2} + \ldots\ldots + \frac{n_s}{x_s}$$

Se cerchiamo il valore unico $\bar{x}$ da sostituire alle x_i osservate in modo da lasciare invariata la funzione avremo la seguente uguaglianza

$$\frac{n_1}{x_1} + \frac{n_2}{x_2} + \ldots\ldots + \frac{n_s}{x_s} = \frac{n_1}{\bar{x}} + \frac{n_2}{\bar{x}} + \ldots\ldots + \frac{n_s}{\bar{x}}$$

$$\sum_{i=1}^{s} \frac{n_i}{x_i} = \frac{1}{\bar{x}} \sum_{i=1}^{s} n_i \longrightarrow \boxed{\bar{x} = M_{ar} = \frac{N}{\sum_{i=1}^{s} \frac{n_i}{x_i}}}$$

Media armonica ponderata

MEDIE ANALITICHE – Media armonica

- Se siamo nel caso di una serie di valori la funzione diventa la seguente:

$$f\left(\frac{1}{x_1}, \frac{1}{x_2}, \ldots\ldots\ldots, \frac{1}{x_s}\right) = \frac{1}{x_1} + \frac{1}{x_2} + \ldots\ldots\ldots + \frac{1}{x_s}$$

Se cerchiamo il valore unico $\bar{x}$ da sostituire alle x_i osservate in modo da lasciare invariata la funzione avremo la seguente uguaglianza

$$\frac{1}{x_1} + \frac{1}{x_2} + \ldots\ldots\ldots + \frac{1}{x_s} = \frac{1}{\bar{x}} + \frac{1}{\bar{x}} + \ldots\ldots\ldots + \frac{1}{\bar{x}}$$

$$\sum_{i=1}^{N} \frac{1}{x_i} = \frac{1}{\bar{x}} N \longrightarrow \boxed{\bar{x} = M_{ar} = \frac{N}{\sum_{i=1}^{N} \frac{1}{x_i}}}$$

Media armonica semplice

MEDIE ANALITICHE – Media di potenze

- Una quantità caratteristica più generale che considerale x_i elevate alla potenza t è la seguente

$$x_1^t n_1 + x_2^t n_2 + \ldots . x_s^t n_s = \overline{x}^t n_1 + \overline{x}^t n_2 + \ldots . \overline{x}^t n_s$$

Se cerchiamo il valore unico $\overline{x}$ da sostituire alle x_i osservate in modo da lasciare invariata la funzione avremo la seguente uguaglianza

$$\sum_{i=1}^{s} x_i^t n_i = \overline{x}^t \sum_{i=1}^{s} n_i$$

$$\boxed{\overline{x} = M_t = \sqrt[t]{\frac{\sum_{i=1}^{s} x_i^t n_i}{N}}}$$

Media di potenze ponderata

$$\boxed{\overline{x} = M_t = \sqrt[t]{\frac{\sum_{i=1}^{N} x_i^t}{N}}}$$

Media di potenze semplice

Per le medie vale la seguente relazione:

$$x_1 \leq \ldots . \leq M_{ar} \leq M_g \leq \mu \leq M_q \leq \ldots . \leq x_N$$

MEDIE ANALITICHE – Media di potenze

- Se nella media di potenze la potenza t=2 si avrà la media quadratica:

$$\boxed{\bar{x} = M_q = \sqrt[2]{\frac{\sum_{i=1}^{S} {x_i}^2 n_i}{N}}}$$

Media quadratica ponderata

$$\boxed{\bar{x} = M_t = \sqrt[2]{\frac{\sum_{i=1}^{N} {x_i}^2}{N}}}$$

Media quadratica semplice

Per le medie vale la seguente relazione:

$$x_1 \leq \ldots\ldots \leq M_{ar} \leq M_g \leq \mu \leq M_q \leq \ldots\ldots \leq x_N$$

MEDIE LASCHE

- Si dicono medie lasche quei valori medi che si basano solo su alcuni valori della intera distribuzione ordinati dal più piccolo al più grande

- I principali tipi di medie lasche sono:
  - valore centrale (v.c.)
  - mediana
  - quartili
  - moda

MEDIE LASCHE – Valore Centrale

- Il valore centrale, che esprime il centro di variazione della v.s., si calcola facendo la semisomma dei valori estremi di un collettivo, dopo aver ordinato le osservazioni in ordine non decrescente.

$$V.C. = \frac{x_{(1)}+x_{(N)}}{2}$$

$x_{(1)}$ = il più piccolo valore tra quelli osservati

$x_{(N)}$ = il più grande valore tra quelli osservati

- Inconveniente: risente dell'influenza dell'inclusione o meno di valori eccezionalmente bassi o alti.
- Il v.c. va sempre calcolato quando trattiamo v.s. divise per intervalli e viene utilizzato quale modalità in tutte le formule che riguardano i caratteri quantitativi continui. Si indica con x_i'

Classi	n_i	x_i'
50 ⊢ 80	50	65
80 ⊢ 180	30	130
180 ⊢ 280	70	230
Totale	150	

MEDIE LASCHE – Mediana

- Ordinati i valori osservati in ordine non decrescente, si dice valore mediano o mediana quel valore che bipartisce la graduatoria, lasciando un ugual numero di termine da una parte e dell'altra. Si indica con Me.
- La mediana gode della seguente proprietà: la somma dei valori assoluti degli scarti dalla mediana è un minimo: $\sum_{i=1}^{N}|x_i - Me|$= minimo
- Esaminiamo dapprima il caso di una serie e distinguiamo 2 casi:

  - N è pari
  - N è dispari
- Se N è pari, la mediana si ottiene dalla media aritmetica dei 2 termini che occupano le posizioni centrali, ossia rispettivamente la posizione $\left(\frac{\mathrm{N}}{2}\right)$ ed $\left(\frac{\mathrm{N}}{2}+1\right)$:

$$Me = \frac{x_{\left(\frac{N}{2}\right)} + x_{\left(\frac{N}{2}+1\right)}}{2}$$

- Se N è dispari, la mediana coincide con il termine che occupa la posizione $\frac{\mathrm{N}+1}{2}$:

$$Me = x_{\left(\frac{N+1}{2}\right)}$$

MEDIE LASCHE – Mediana

- Esempi

1. Serie dispari di valori: 2, 3, 4, 5, 6, 7, 8 N=7

   prima ricerchiamo la posizione occupata dalla modalità mediana: $\frac{\mathrm{N}+1}{2} = \frac{7+1}{2} = 4$

   poi individuiamo la modalità che occupa la posizione mediana: Me = $x_{\left(\frac{N+1}{2}\right)} = x_{(4)} = 5$

2. Serie pari di valori: 2, 3, 4, 5, 6, 7 N=6

   prima cerchiamo la posizione occupata dalle modalità centrali:

   $\left(\frac{\mathrm{N}}{2}\right) = \left(\frac{6}{2}\right) = 3$

   $\left(\frac{\mathrm{N}}{2} + 1\right) = \left(\frac{6}{2} + 1\right) = 4$

   poi individuiamo le modalità che occupano le posizioni centrali:

   $x_{\left(\frac{N}{2}\right)} = x_{(3)} = 4$

   $x_{\left(\frac{N}{2}+1\right)} = x_{(3+1)} = x_{(4)} = 5$ e poi calcoliamo il valore mediano

   Me = $\frac{x_{\left(\frac{N}{2}\right)} + x_{\left(\frac{N}{2}+1\right)}}{2}$ = Me = $\frac{x_{(3)} + x_{(4)}}{2}$ = Me = $\frac{4+5}{2} = 4{,}5$

MEDIE LASCHE – Mediana

- Esaminiamo ora il caso di una distribuzione e distinguiamo il caso in cui il carattere rilevato sia discreto o continuo.
- In entrambi i casi si calcoleranno dapprima le frequenze accumulate che forniscono informazioni circa il posto occupato dalle varie modalità nella distribuzione.

  Es. la prima frequenza accumulata indica che i primi N_1 posti sono occupati dalla modalità x_1; la seconda indica che i posti dall'N_{1+1} sino all'N_2 sono occupati dalla modalità x_2 e così via.
- Nel caso discreto si va poi a considerare la numerosità del collettivo:

  - se N è pari, si andranno a calcolare $\left(\frac{N}{2}\right)$ e $\left(\frac{N}{2}+1\right)$ che individuano i posti occupati dai valori che interessano per il calcolo, le rispettive modalità di riferimento si individueranno confrontando detti valori con le frequenze accumulate dopo di chè si procede come nel caso della serie;

  - se N è dispari, si andrà a calcolare il valore $\frac{N+1}{2}$ confrontandolo con le frequenze accumulate e quindi verrà individuata la modalità che occupa tale posto come visto nel caso della serie.

MEDIE LASCHE – Mediana

N pari

Ampiezza	n_i	N_i
1	1100	1100
2	1250	2350
3	1340	3690
Totale	3690	

$\frac{N}{2} = \frac{3690}{2} = 1845$

$\left(\frac{N}{2}+1\right) = \left(\frac{3690}{2}+1\right) = 1846$

confrontando questi valori con le freq. accum. scopriamo che i posti dal 1101 al 2350 sono occupati dalla modalità 2 quindi:

$$Me = \frac{x_{\left(\frac{N}{2}\right)} + x_{\left(\frac{N}{2}+1\right)}}{2} = \frac{x_{(1845)} + x_{(1846)}}{2} = \frac{2+2}{2} = 2$$

N dispari

Ampiezza	n_i	N_i
1	1100	1100
2	1250	2350
3	1341	3691
Totale	3691	

$\frac{N+1}{2} = \frac{3691+1}{2} = 1846$

confrontando questio valori con le freq. accum. scopriamo che i posti dal 1101 al 2350 sono occupati dalla modalità 2 quindi:

$Me = x_{\left(\frac{N+1}{2}\right)} = x_{(1846)} = 2$

MEDIE LASCHE – Mediana

- Nel caso delle v.s. divise per intervalli non importa la numerosità del collettivo.
- Si calcola $\frac{N}{2}$ e lo si confronta con la frequenza accumulata: $\frac{N}{2} \leq N_i$, si individua così la classe mediana $x_i \vdash x_{i+1}$
- La mediana si calcolerà con la seguente formula:

$$Me = x_i + \frac{x_{i+1} - x_i}{n_i} (N/2 - N_{i-1})$$

$$\frac{150}{2} = 75 = \frac{N}{2} \leq N_i = 80$$

$x_i \vdash x_{i+1}$ ⟶ $80 \vdash 180$

$$Me = 80 + \frac{180-80}{30}(150/2-50) = 163{,}33$$

Classi	n_i	N_i
$50 \vdash 80$	50	50
$80 \vdash 180$	30	80
$180 \vdash 280$	70	150
Totale	150	

MEDIE LASCHE – Quartili

- Conosciamo 3 quartili:
  - primo quartile, Q_1, è quel valore al di sotto del quale ci sono 1/4 dei valori della X e al di sopra si trovano i 3/4 dei valori della X;
  - secondo quartile, Q_2=Me;
  - terzo quartile, Q_3, è quel valore al di sotto del quale si trovano i 3/4 dei valori della X e al di sopra 1/4.
- Q_1 è la mediana della prima metà della distribuzione; Q_3 della seconda metà, quindi tra Q_1 e Q_3 è compreso il 50% dei casi.
- Quanto detto per la mediana si applica anche ai quartili con l'unica differenza che si utilizzeranno quali rapporti di riferimento $\frac{\mathrm{N}}{4}$ e $\frac{3}{4}N$. Con una particolarità.
- Nel caso di una serie o una distribuzione discreta in cui N non risulta multiplo di 4, quindi il risultato del rapporto sarebbe un numero con la virgola, il procedimento da seguire muta se ***N è dispari***, ossia anziché considerare $x_{\left(\frac{N+1}{4}\right)}$ o $x_{\left(\frac{3N+1}{4}\right)}$ si utilizzeranno $x_{\left(\frac{N}{4}+1\right)}$ e $x_{\left(\frac{3N}{4}+1\right)}$ prendendo come risultato del rapporto la parte intera del risultato.

  Mentre se ***N è pari*** il procedimento rimane identico, $x_{\left(\frac{N}{4}\right)}$ e $x_{\left(\frac{N}{4}+1\right)}$ o $x_{\left(\frac{3N}{4}\right)}$ e $x_{\left(\frac{3N}{4}+1\right)}$ anche qui se il numero non è multiplo di 4 si considera sempre la parte intera del risultato del rapporto.

MEDIE LASCHE – Quartili

Esempi.

1. Siamo in presenza della seguente serie:

5,6 6,2 2,3 4,2 3,1 7,8 9,3 8,7 1,2 N=9 serie dispari

Prima si ordinano i valori

1,2 2,3 3,1 4,2 5,6 6.2 7,8 8,7 9,3

$Q_1 = x_{\left(\frac{N}{4}+1\right)} = x_{(2,25+1)} = x_{(2+1)} = x_{(3)} = 3,1$

$Q_3 = x_{\left(\frac{3N}{4}+1\right)} = x_{(6,75+1)} = x_{(6+1)} = x_{(7)} = 7,8$

2. Siamo in presenza della seguente serie:

5,6 6,2 2,3 4,2 3,1 7,8 9,3 8,7 1,2 10,3 N=10 serie pari

Prima si ordinano i valori

1,2 2,3 3,1 4,2 5,6 6.2 7,8 8,7 9,3 10,3

$$Q_1 = \frac{x_{\left(\frac{N}{4}\right)} + x_{\left(\frac{N}{4}+1\right)}}{2} = \frac{x_{\left(\frac{10}{4}\right)} + x_{\left(\frac{10}{4}+1\right)}}{2} = \frac{x_{(2,5)} + x_{(2,5+1)}}{2} = \frac{x_{(2)} + x_{(3)}}{2} = \frac{2,3+3,1}{2} = 2,7$$

$$Q_3 = \frac{x_{\left(\frac{3N}{4}\right)} + x_{\left(\frac{3N}{4}+1\right)}}{2} = \frac{x_{\left(\frac{3*10}{4}\right)} + x_{\left(\frac{3*10}{4}+1\right)}}{2} = \frac{x_{\left(\frac{30}{4}\right)} + x_{\left(\frac{30}{4}+1\right)}}{2} = \frac{x_{(7,5)} + x_{(7,5+1)}}{2} = \frac{x_{(7)} + x_{(7+1)}}{2} = \frac{7,8+8,7}{2} = 8,25$$

Stesso procedimento nel caso di una distribuzione

MEDIE LASCHE – Moda

- Si dice <u>*valore normale* </u>o <u>*valore modale* </u>o <u>*norma*</u> o <u>moda</u> il valore x_i che si presenta con la massima frequenza all'interno del collettivo considerato. Si indica con M_o.

  Se la v.s. è divisa per intervalli si individuerà la classe modale attraverso la frequenza o la densità di frequenza a seconda che le classi abbiano o non la stessa ampiezza.
- Le v.s. possono essere unimodali o plurimodali. Se la moda è negli estremi si dicono zeromodali
- Per v.s. simmetriche unimodali si avrà μ=Me=Mo
- Per v.s. moderatamente asimmetriche i 3 valori sono approssimativamente legati dalla seguente relazione: Mo ~ μ-3(μ-Me)

  cioè la mediana si trova tra la moda e la media aritmetica approssimativamente a 2/3 di distanza dalla moda e a 1/3 dalla media aritmetica