MEDIE ALGEBRICHE
E
DI POSIZIONE

0. Introduzione

Tra le elaborazioni matematiche effettuate sui dati statistici rivestono particolare importanza quelle che hanno il compito di esprimere i diversi valori delle intensità di un fenomeno, mediante un solo numero che esprima sinteticamente quella successione. Ad esempio, per conoscere l'aumento del costo di certo prodotto nell'arco di un anno, più che la conoscenza dell'aumento mese per mese è utile conoscere l'*aumento medio*, cioè quell'aumento capace di sintetizzare tutti quelli che si sono verificati nei singoli mesi. Un numero capace di esprimere sinteticamente una distribuzione di intensità di un fenomeno collettivo viene detto *valore medio*.
In generale si definisce *valore medio* di un insieme di dati statistici quantitativi qualunque valore compreso fra il minimo ed il massimo di quelli dati. I valori medi, naturalmente, sono infiniti e ciascuno di essi può coincidere o no con un dato della successione. Fra questi infiniti valori medi assumono particolare importanza le cosiddette *medie algebriche* e le *medie di posizione* che ora andremo ad esaminare in dettaglio.

1. Media aritmetica

La media aritmetica viene applicata nei confronti di variabili che rappresentano grandezze aventi carattere additivo quali, ad esempio, redditi, consumi, produzioni, ecc. ed indica l'intensità che avrebbe ciascuna unità statistica nel caso in cui quella totale fosse ugualmente ripartita fra tutte le unità (reddito medio individuale, consumo medio individuale di un certo prodotto, ecc.).
Si chiama *media aritmetica semplice* di n numeri $x_1, x_2, \ldots, x_n$ aventi carattere quantitativo il numero M_a che si ottiene dividendo la loro somma per il numero n, cioè si ha:

$$M_a = \frac{x_1 + x_2 + \cdots + x_n}{n}$$

o, in forma compatta:

$$M_a = \frac{\sum_{k=1}^{n} x_k}{n} \qquad \text{o} \qquad M_a = \frac{1}{n}\sum_{k=1}^{n} x_k \; .$$

Osserviamo che la media aritmetica può essere determinata anche quando non si conoscono i vari termini $x_1, x_2, \ldots, x_n$ della successione: basta conoscere la loro somma ed il loro numero. Ad esempio si può calcolare il consumo medio annuale del gas per uso domestico di una famiglia senza conoscere il consumo in ciascun mese dell'anno: basta dividere il consumo totale nell'intero anno e dividerlo per 12.
Consideriamo, insieme alla successione $x_1, x_2, \ldots, x_n$ la somma degli n termini $M_a + M_a + \cdots + M_a$. Dalla definizione di media aritmetica si trae che:

$$x_1 + x_2 + \ldots + x_n = n \cdot M_a.$$

Ricordando che i numeri $x_1, x_2, \ldots, x_n$ rappresentano le intensità delle n modalità di quel carattere quantitativo, possiamo affermare che il numero M_a rappresenta il valore costante che dovrebbe avere ciascuno dei numeri $x_1, x_2, \ldots, x_n$ perché la somma complessiva dei loro valori resti invariata.
Possiamo verificare che M_a è effettivamente un valore medio compreso fra il minimo ed il massimo dei valori $x_1, x_2, \ldots, x_n$. Per fare ciò disponiamo i valori indicati in ordine non decrescente, cioè sia:

$$x_1 \le x_2 \le \cdots \le x_n.$$

Ovviamente vale la relazione:

$$\underbrace{x_1 + x_1 + \cdots + x_1}_{n\ volte} \le x_1 + x_2 + \cdots + x_n \le \underbrace{x_n + x_n + \cdots + x_n}_{n\ volte}$$

da cui:

$$n \cdot x_1 \le n \cdot M_a \le n \cdot x_n \quad \Rightarrow \quad x_1 \le M_a \le x_n.$$

L'aggettivo *aritmetica* è giustificato dal fatto che la media aritmetica di un numero dispari di valori che sono in progressione aritmetica è proprio il numero che occupa la posizione centrale. Se, ad esempio, si considerano i numeri 1; 3; 5; 7; 9; 11; 13 si vede che essi costituiscono una progressione aritmetica di ragione 2 e la media aritmetica di questi numeri è:

$$\mathrm{M}_a = \frac{1+3+5+7+9+11+13}{7} = \frac{49}{7} = 7$$

che coincide proprio con il termine centrale.

Può capitare che in una distribuzione di frequenze ciascuno dei dati entri nel calcolo un numero diverso di volte: si ha in tal caso la *media aritmetica ponderata*.
Supponiamo che il valore x_1 compaia con una frequenza f_1, valore x_2 con frequenza f_2 ed infine il valore x_k con frequenza f_k. Si hanno così, complessivamente:

$$f_1 + f_2 + \cdots\cdots + f_k = n$$

numeri la cui media aritmetica è data dalla:

$$\mathrm{M}_a = \frac{\overbrace{(x_1 + x_1 + \cdots + x_1)}^{f_1\ volte} + \overbrace{(x_2 + x_2 + \cdots + x_2)}^{f_2\ volte} + \cdots + \overbrace{(x_k + x_k + \cdots + x_k)}^{f_k\ volte}}{f_1 + f_2 + \cdots\cdots + f_k}$$

che possiamo scrivere più semplicemente come:

$$\mathrm{M}_a = \frac{x_1 \cdot f_1 + x_2 \cdot f_2 + \cdots\cdots + x_k \cdot f_k}{f_1 + f_2 + \cdots\cdots + f_k}$$

o, in forma sintetica:

$$\mathrm{M}_a = \frac{\sum_{k=1}^{n} x_k \cdot f_k}{n} \qquad \Rightarrow \qquad \mathrm{M}_a = \frac{1}{n}\sum_{k=1}^{n} x_k \cdot f_k \ .$$

A queste espressioni si dà il nome di *media aritmetica ponderata* di x_1, x_2, …. , x_k ed i numeri f_1, f_2, ······ , f_k si dicono *pesi* o *frequenze*.

Possiamo perciò dare la seguente definizione:
Si chiama media aritmetica ponderata di più numeri il quoziente che si ottiene dividendo la somma dei prodotti dei singoli dati per le rispettive frequenze per la somma di tutte le frequenze.

Esempio
Si voglia determinare il salario medio mensile di un gruppo di lavoratori così composto:

n 25 con un salario mensile di 1200€
n 20 con un salario mensile di 1500€
n 5 con un salario mensile di 2000€

È evidente che la determinazione non può essere effettuata attraverso la media semplice dei tre valori perché ciascuno di essi si presenta con frequenza diversa. Si tratta quindi di una media ponderata ed abbiamo:

$$\mathrm{M}_a = \frac{1200 \cdot 25 + 1500 \cdot 20 + 2000 \cdot 5}{25 + 20 + 5} = 1400$$

cifra che rappresenta il salario medio per quel gruppo di lavoratori.
Quando si deve calcolare una media aritmetica ponderata su una distribuzione di frequenze di un carattere quantitativo i cui valori siano ripartiti in n classi di intensità, bisogna servirsi di particolari criteri. Si costruisce una tabella nel modo seguente:

Classi	*Frequenze* (n_i)	*Valori centrali* (x_i)	$x_i \cdot n_i$
$c_0 - c_1$	n_1	x_1	$x_1 \cdot n_1$
$c_1 - c_2$	n_2	x_2	$x_2 \cdot n_2$
…….	…	…	
$c_{k-1} - c_k$	n_k	x_k	$x_k \cdot n_k$
Totale	N	——	$\sum_{j=1}^{k} x_j \cdot n_j$

dove i valori centrali sono dati da $\frac{c_0 + c_1}{2}, \ldots, \frac{c_{k-1} + c_k}{2}$.

Esempio
Per la distribuzione dell'altezza di un collettivo di 50 persone si ha:

Altezza	*Frequenza* (n_i)	*Valori centrali* (x_i)	$x_i \cdot n_i$
150 - 160	1	155	155
160 - 170	10	165	1650
170 - 180	35	175	6125
180 - 200	4	190	760
Totale	50	——	8690

e quindi per l'altezza media si ha:

$$M_a = \frac{8690}{50} = 173{,}8\,.$$

2. Scarti e proprietà della media aritmetica

Dati n numeri $x_1, x_2, \ldots, x_n$ ed indicato con M_a il loro valore medio, si chiamano *scarti* (o *scostamenti*) dei numeri dati dal suo valore medio M_a, le differenze (positive, nulle o negative) fra ciascuno dei dati numeri ed il loro valore medio M_a. Gli scarti, pertanto, sono dati da:

$$x_1 - M_a, \quad x_2 - M_a, \quad \ldots\ldots, \quad x_n - M_a.$$

Prima proprietà

La somma algebrica degli scarti dalla media aritmetica è nulla.

Infatti si ha:

$$(x_1 - M_a) + (x_2 - M_a) + \ldots\ldots + (x_n - M_a) = x_1 + x_2 + \cdots\cdots + x_n - n \cdot M_a =$$

$$= x_1 + x_2 + \cdots\cdots + x_n - n \cdot \frac{x_1 + x_2 + \cdots + x_n}{n} =$$

$$= x_1 + x_2 + \cdots\cdots + x_n - (x_1 + x_2 + \cdots\cdots + x_n) = 0.$$

Questa proprietà vale anche nel caso della media aritmetica ponderata ed in tal caso risulta $\sum_{i=1}^{n} (x_i - M_a) \cdot f_i = 0$. Infatti si ha:

$$(x_1 - M_a) \cdot f_1 + (x_2 - M_a) \cdot f_2 + \ldots\ldots + (x_n - M_a) \cdot f_n =$$

$$= x_1 \cdot f_1 + x_2 \cdot f_2 + \cdots\cdots + x_n \cdot f_n - (f_1 + f_2 + \cdots\cdots + f_n) \cdot M_a =$$

$$= x_1 \cdot f_1 + \cdots\cdots + x_n \cdot f_n - (f_1 + f_2 + \cdots\cdots + f_n) \cdot \frac{x_1 \cdot f_1 + \cdots + x_k \cdot f_k}{f_1 + f_2 + \cdots + f_k} =$$

$$= x_1 \cdot f_1 + x_2 \cdot f_2 + \cdots\cdots + x_n \cdot f_n - (x_1 \cdot f_1 + x_2 \cdot f_2 + \cdots\cdots + x_n \cdot f_n) = 0.$$

Seconda proprietà

La somma dei quadrati degli scarti è minima quando gli scarti sono calcolati dalla media aritmetica M_a.

Ciò vuol dire che se si calcolano gli scarti anziché dalla media aritmetica M_a da un altro numero qualunque α, la somma dei quadrati di tali scarti risulta maggiore rispetto a quella degli scarti dalla media aritmetica.

Consideriamo gli scarti dalla media aritmetica M_a e gli scarti da un numero qualsiasi α e supponiamo che sia $\alpha = M_a - \delta$. Possiamo scrivere:

$$\begin{gathered} x_1 - \alpha = x_1 - (M_a - \delta) = (x_1 - M_a) + \delta \\ x_2 - \alpha = x_2 - (M_a - \delta) = (x_2 - M_a) + \delta \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ x_n - \alpha = x_n - (M_a - \delta) = (x_n - M_a) + \delta \end{gathered}$$

La somma dei quadrati degli scarti è:

$$\begin{gathered} (x_1 - \alpha)^2 + (x_2 - \alpha)^2 + \cdots\cdots + (x_n - \alpha)^2 = \\ = [(x_1 - M_a) + \delta]^2 + [(x_2 - M_a) + \delta]^2 + \cdots\cdots + [(x_n - M_a) + \delta]^2 = \\ = [(x_1 - M_a)^2 + 2\delta(x_1 - M_a) + \delta^2] + \cdots\cdots + [(x_n - M_a)^2 + 2\delta(x_n - M_a) + \delta^2] = \\ = [(x_1 - M_a)^2 + (x_2 - M_a)^2 + \cdots\cdots + (x_n - M_a)^2] + \\ + 2\delta[(x_1 - M_a) + \cdots\cdots + (x_n - M_a)] + n\delta^2. \end{gathered}$$

La quantità dentro la seconda parentesi quadra essendo la somma degli scarti da M_a vale zero e quindi si ha:

$$\begin{gathered} (x_1 - \alpha)^2 + (x_2 - \alpha)^2 + \cdots\cdots + (x_n - \alpha)^2 = \\ = [(x_1 - M_a)^2 + (x_2 - M_a)^2 + \cdots\cdots + (x_n - M_a)^2] + n\delta^2. \end{gathered}$$

Essendo $n\delta^2$ sicuramente positivo risulta evidente che:

$$(x_1 - \alpha)^2 + \cdots\cdots + (x_n - \alpha)^2 > (x_1 - M_a)^2 + (x_2 - M_a)^2 + \cdots\cdots + (x_n - M_a)^2.$$

La proprietà vale anche nel caso della media ponderata, cioè risulta

$$\sum_{k=1}^{n} (x_k - M_a)^2 \cdot f_k < \sum_{k=1}^{n} (x_k - \alpha)^2 \cdot f_k .$$

3. Media geometrica e sue proprietà

La media geometrica viene usata quando si ha a che fare con valori il cui prodotto ha un significato logico poiché la media geometrica è quel numero che sostituito a ciascuno dei valori dati ne conserva inalterato il prodotto. Essa viene applicata, ad esempio, nei problemi di capitalizzazione in cui si tratta di sostituire ad una serie di tassi variabili nel tempo un tasso unico costante equivalente.
Si chiama *media geometrica semplice*, e si indica con G, di n numeri x_1, x_2, …. , x_n la radice n-esima del loro prodotto, cioè:

$$\mathrm{G} = \sqrt[n]{x_1 \cdot x_2 \cdot ..... \cdot x_n} \; .$$

In particolare la media geometrica di due numeri, α e β, è la radice quadrata del loro prodotto, cioè:

$$\mathrm{G} = \sqrt{\alpha \cdot \beta} \; .$$

Quando i dati che compongono la successione hanno una diversa frequenza, si ha la *media geometrica ponderata* che si può scrivere nella forma:

$$\mathrm{G} = \sqrt[f_1+f_2+\cdots+f_n]{x_1^{f_1} \cdot x_2^{f_2} \cdot ..... \cdot x_n^{f_n}}$$

il cui calcolo è possibile con l'uso dei logaritmi.
La media geometrica non può essere determinata se qualcuno dei termini è uguale a zero perché in tal caso si annulla il prodotto contenuto nel radicando.
Passiamo ora ad esaminare le proprietà della media geometrica.

Prima proprietà
Il numero G rappresenta il valore che dovrebbe possedere ciascuno degli n numeri $x_1, x_2, \ldots, x_n$ affinché il prodotto rimanga immutato.
Infatti, volendo che sia:

$$\underbrace{\mathrm{G} \cdot \mathrm{G} \cdot \, ... \, \cdot \mathrm{G}}_{n \text{ volte}} = x_1 \cdot x_2 \cdot \, ... \, \cdot x_n$$

dovrà essere:

$$\mathrm{G}^n = x_1 \cdot x_2 \cdot \, ... \, \cdot x_n \quad \Rightarrow \quad \mathrm{G} = \sqrt[n]{x_1^{f_1} \cdot x_2^{f_2} \cdot ..... \cdot x_n^{f_n}} \; .$$

Seconda proprietà

La media geometrica G è un valore medio compreso fra il minimo ed il massimo valore della distribuzione data. Infatti, supposto:

$$0 < x_1 \leq x_2 \leq \ \ldots . \leq x_n$$

si ha:

$$\underbrace{x_1 \cdot x_1 \cdot \ \ldots . \cdot x_1}_{n \text{ volte}} \leq x_1 \cdot x_2 \cdot \ \ldots . \cdot x_n \leq \underbrace{x_n \cdot x_n \cdot \ \ldots . \cdot x_n}_{n \text{ volte}}$$

da cui:

$$x_1^n \leq G^n \leq x_n^n \quad \Rightarrow \quad x_1 \leq \mathrm{G} \leq x^n.$$

Terza proprietà

Il reciproco della media geometrica di più numeri è uguale alla media geometrica dei loro reciproci.
Infatti, se ai numeri dati $x_1, x_2, \ldots . , x_n$ si sostituiscono i loro reciproci:

$$\frac{1}{x_1}, \ \frac{1}{x_2}, \ldots\ldots., \frac{1}{x_n}$$

la media geometrica è:

$$\mathrm{G}_1 = \sqrt[n]{\frac{1}{x_1} \cdot \frac{1}{x_2} \cdot \ldots . \cdot \frac{1}{x_n}}$$

che è appunto il reciproco di G.

Quarta proprietà

La media geometrica di più rapporti è uguale al rapporto tra la media geometrica dei numeratori e quella dei denominatori.
Infatti, dati i rapporti:

$$\frac{x_1}{y_1}, \ \frac{x_2}{y_2}, \ldots\ldots\ldots, \frac{x_n}{y_n}$$

la loro media geometrica è:

$$G = \sqrt[n]{\frac{x_1}{y_1} \cdot \frac{x_2}{y_2} \cdot \ldots\ldots \cdot \frac{x_n}{y_n}}$$

che possiamo scrivere nella forma:

$$G = \frac{\sqrt[n]{x_1 \cdot x_2 \cdot \ldots\ldots \cdot x_n}}{\sqrt[n]{y_1 \cdot y_2 \cdot \ldots\ldots \cdot y_n}}.$$

Quinta proprietà

Il logaritmo della media geometrica di n numeri è uguale alla media aritmetica dei logaritmi dei numeri stessi.

Infatti, essendo $G = \sqrt[n]{x_1 \cdot x_2 \cdot \ldots\ldots \cdot x_n}$, passando ai logaritmi si ha:

$$\log G = \frac{1}{n}(\log x_1 + \log x_2 + \cdots\cdots + \log x_n).$$

e nel caso di media geometrica ponderata $G = \sqrt[n]{x_1^{f_1} \cdot x_2^{f_2} \cdot \ldots\ldots \cdot x_n^{f_n}}$ si ha:

$$\log G = \frac{1}{n}(f_1 \cdot \log x_1 + f_2 \cdot \log x_2 + \cdots\cdots + f_n \cdot \log x_n)$$

il cui secondo membro rappresenta la media aritmetica ponderata dei logaritmi dei numeri dati con i rispettivi pesi.

La media ora esaminata si chiama geometrica perché, se calcolata per un numero dispari di termini in progressione geometrica, essa è uguale al termine centrale della progressione stessa.

4. Media quadratica e sue proprietà

Si chiama *media quadratica semplice*, e si indica con M_g, di n numeri $x_1, x_2, \ldots\ldots, x_n$ la radice quadrata della media aritmetica dei quadrati dei numeri dati, cioè:

$$M_q = \sqrt{\frac{x_1^2 + x_2^2 + \cdots + x_n^2}{n}}$$

o, in forma compatta:

$$\mathrm{M}_q = \sqrt{\frac{\sum_{k=1}^{n} x_k^2}{n}}\ .$$

Anche la media quadratica può essere ponderata, ed allora assume la forma:

$$\mathrm{M}_q = \sqrt{\frac{x_1^2 \cdot f_1 + x_2^2 \cdot f_2 + ... + x_n^2 \cdot f_n}{f_1 + f_2 + ... + f_n}}$$

Il valore di M_q, in questo caso, prende il nome di *media quadratica ponderata* dei numeri $x_1, x_2, ..., x_n$ con i pesi $f_1, f_2, ..., f_n$.

<u>*Prima proprietà*</u>
Il numero M_q rappresenta il valore comune che dovrebbe possedere ciascuno degli n numeri $x_1, x_2, ..., x_n$ affinché resti immutata la somma dei loro quadrati.
Infatti, si ha:

$$\underbrace{M_q^2 + M_q^2 + ... + M_q^2}_{n\ volte} = x_1^2 + x_2^2 + ... + x_n^2$$

$$n \cdot M_q^2 = x_1^2 + x_2^2 + ... + x_n^2 \quad \Rightarrow \quad M_q^2 = \frac{x_1^2 + x_2^2 + ... + x_n^2}{n}$$

e quindi:

$$\mathrm{M}_q = \sqrt{\frac{x_1^2 + x_2^2 + \cdots + x_n^2}{n}}\ .$$

<u>*Seconda proprietà*</u>
La media quadratica è un valore medio della distribuzione ed è compresa fra il minimo ed il massimo.
Infatti, se si ha:

$$0 < x_1 \leq x_2 \leq \ .... \leq x_n$$

allora valgono le disuguaglianze:

$$\underbrace{x_1^2 + x_1^2 + \cdots + x_1^2}_{n\ volte} \leq x_1^2 + x_2^2 + \cdots + x_n^2 \leq \underbrace{x_n^2 + x_n^2 + \cdots + x_n^2}_{n\ volte}$$

da cui:

$$n\,x_1^2 \le n\,M_q^2 \le n\,x_n^2$$

Dividendo tutti i termini delle disuguaglianze per n ed estraendo la radice quadrata si ottiene appunti:

$$x_1 \le \mathrm{M}_q \le x_n.$$

Fra le medie considerate la media quadratica è quella che ha valore maggiore e quindi è più influenzata dai valori molto piccoli o molto grandi della distribuzione. Essa è utilizzata per mettere in evidenza l'esistenza di valori che si scostano molto dai valori centrali.

5. Media armonica e sue proprietà

Si chiama *media armonica semplice*, e si indica con H, di n numeri x_1, x_2, ….. , x_n il reciproco della media aritmetica dei reciproci dei numeri dati, cioè:

$$\mathrm{H} = \frac{n}{\frac{1}{x_1}+\frac{1}{x_2}+\cdots+\frac{1}{x_n}}$$

o anche, in forma compatta:

$$\mathrm{H} = \frac{n}{\sum_{k=1}^{n}\frac{1}{x_k}}.$$

In pratica risulta più semplice ricordare l'espressione:

$$\frac{1}{H} = \frac{\frac{1}{x_1}+\frac{1}{x_2}+\cdots+\frac{1}{x_n}}{n} = \frac{1}{n}\sum_{k=1}^{n}\frac{1}{x_k}$$

cioè: *l'inverso della media armonica è la media aritmetica degli inversi dei numeri dati.*

Quando i valori della distribuzione x_1, x_2, ….. , x_n si presentano con le frequenze f_1, f_2, …. , f_n, rispettivamente, la media armonica si scrive sotto la forma:

$$H = \frac{n}{\frac{f_1}{x_1} + \frac{f_2}{x_2} + \cdots + \frac{f_k}{x_k}}$$

e si chiama *media armonica ponderata*.

Prima proprietà

Il numero H rappresenta il valore comune che dovrebbe avere ciascuno degli n numeri $x_1, x_2, \ldots, x_n$ affinché resti immutata la somma dei loro reciproci. Infatti:

$$\underbrace{\frac{1}{H} + \frac{1}{H} + \cdots + \frac{1}{H}}_{n\ volte} = \frac{1}{x_1} + \frac{1}{x_2} + \cdots + \frac{1}{x_n}$$

$$\frac{n}{H} = \frac{1}{x_1} + \frac{1}{x_2} + \cdots + \frac{1}{x_n}$$

da cui, passando ai reciproci:

$$\frac{H}{n} = \frac{1}{\frac{1}{x_1} + \frac{1}{x_2} + \cdots + \frac{1}{x_n}} \quad \Rightarrow \quad H = \frac{n}{\frac{1}{x_1} + \frac{1}{x_2} + \cdots + \frac{1}{x_n}}.$$

Seconda proprietà

La media armonica H è un valore medio e quindi è compreso fra il minimo ed il massimo di quelli dati.

Infatti, supposto che sia $0 < x_1 \le x_2 \le \cdots \le x_n$, si ha che:

$$\underbrace{\frac{1}{x_1} + \frac{1}{x_1} + \cdots + \frac{1}{x_1}}_{n\ volte} \ge \frac{1}{x_1} + \frac{1}{x_2} + \cdots + \frac{1}{x_n} \ge \underbrace{\frac{1}{x_n} + \frac{1}{x_n} + \cdots + \frac{1}{x_n}}_{n\ volte}$$

$$\frac{n}{x_1} \ge \frac{n}{H} \ge \frac{n}{x_n}.$$

Dividendo per n tutti e tre i membri delle disuguaglianze si ottiene infine:

$$\frac{1}{x_1} \geq \frac{1}{H} \geq \frac{1}{x_n} \quad \Rightarrow \quad x_1 \leq \mathrm{H} \leq x_n.$$

la media armonica si usa quando si prende in esame un carattere che risulta essere additivo rispetto ai reciproci dei suoi valori e si vuole ottenere una distribuzione uniforme che però non alteri la somma dei loro reciproci.
Per concludere l'argomento delle medie algebriche osserviamo che le diverse medie, in generale, non sono uguali. Si può dimostrare che se i dati non sono tutti uguali, valgono le relazioni:

$$\mathrm{H} < \mathrm{G} < \mathrm{M} < \mathrm{M}_q.$$

6. Moda

Le medie descritte in precedenza vengono dette *medie algebriche* perché si calcolano mediante operazioni algebriche. In statistica si considerano anche altri valori medi che non provengono da un calcolo algebrico ma dall'esame delle posizioni dei dati nella distribuzione considerata.
Si chiama *moda* (o anche *norma* o *valore modale* o *valore normale* o *valore di massima frequenza*) di un insieme di numeri il valore, che indichiamo con $\mathrm{M_o}$, che si presenta con la frequenza più alta. La moda è quindi un particolare valore medio che indica l'intensità del fenomeno che si verifica con maggiore frequenza.
Ad esempio, per una data popolazione la distribuzione delle famiglie secondo il numero dei loro componenti risulta come segue:

famiglie	con	1	unità	n° 58.324
famiglie	con	2	unità	105.801
famiglie	con	3	unità	108.714
famiglie	con	4	unità	120.312
famiglie	con	5	unità	100.001
famiglie	con	6	unità	40.003
famiglie	con	7	unità	10.321

Si rileva subito che la moda è rappresentata dal dato numero 4 cui corrisponde la frequenza più alta (120312). Si può allora dire che la *composizione normale* delle famiglie di quella data popolazione è di 4 persone essendo le famiglie di quel tipo le più numerose.
Si può dire che la moda, sotto certi aspetti, risulta essere un valore più significativo rispetto agli altri valori medi perché, a differenza della

media aritmetica o di altre medie che forniscono valori astratti e possono non coincidere con alcuno dei dati empirici, questa riveste particolare importanza nei problemi statistici in cui occorre mettere in risalto la misura dei fenomeni che ha la maggiore probabilità di verificarsi.
Così, ad esempio, il salario modale dei dipendenti di una azienda è più espressivo rispetto al salario medio aritmetico perché quest'ultimo può essere influenzato da retribuzioni molto alte di un piccolo numero di dirigenti dell'azienda.
Nel caso di una distribuzione di frequenze con modalità raggruppate in classi, dapprima si definisce la *classe modale*, cioè quella classe a cui compete la densità di frequenza più alta e successivamente si assume come moda il valore centrale di tale classe. Ad esempio, considerata la seguente distribuzione di frequenze:

Peso	*N.ro individui*	*Densità di frequenza*
50 – 53	6	2
53 – 59	12	2
59 – 60	3	3

Si nota che la classe modale è la (59 – 60) e moda è $\frac{59+60}{2}=59,5$.
Nel caso che ci sia una sola moda, si dice che la distribuzione è *unimodale*. Qualora ci siano più mode si dice che la distribuzione è *plurimodale*. Nella realtà è difficile che una distribuzione sia pluroimodale; accade piuttosto che la distribuzione presenti dei picchi.

7. Mediana

Si chiama *mediana* (o *valore mediano* o *valore centrale*) di un insieme di numeri disposti in ordine non decrescente o non crescente, il valore M_c che occupa il posto centrale se questi sono in numero dispari oppure la semisomma dei due numeri centrali se questi sono in numero pari.
Così, per esempio, in una successione ordinata di 11 termini la mediana sarà rappresentata dal sesto termine mentre in una successione ordinata di 20 termini la mediana risulterà dalla semisomma del decimo e dell'undicesimo termine.
La mediana, quindi, rappresenta il valore centrale di una distribuzione di intensità di un carattere quantitativo (seriazione) perché essa lascia il 50% dei dati alla sua sinistra ed il rimanente 50% alla sua destra.

Oltre alla mediana, talora, in statistica si usano altri valori medi di posizione che ripartiscono il numero dei dati della distribuzione in un determinato numero di parti uguali. Questi valori medi, a seconda che ripartiscono i dati della seriazione in 4, 5, 10 o 100 parti uguali, prendono il nome, rispettivamente, di *quartili*, *quintili*, *decili* e *centili*. Ci occuperemo qui solo di quartili e percentili.

Quartili

In una distribuzione ordinata in modo crescente o decrescente, ci sono 3 quartili che la dividono in quattro parti uguali. Il primo quartile (che indichiamo con Q_1) ha prima di sé il 25% dei casi; il secondo quartile (Q_2) coincide con la mediana ed ha prima di sé la metà dei casi; il terzo quartile (Q_3) ha prima di sé il 75% dei casi.

Esempio

Consideriamo la seguente distribuzione:

3; 7; 12; 18; 20; 21; 23; 24; 26; 30

I termini di questa distribuzione sono 10, quindi in numero pari. Calcoliamo innanzitutto la mediana Q_2 che è data dalla media aritmetica dei due termini centrali che è 20,5. Per calcolare gli altri due quartili procediamo nel modo seguente.
Il primo quartile Q_1 è compreso tra i termini inferiori alla mediana Q_2, cioè compreso fra 3, 7, 12, 18, 20. Q_1 allora si ottiene calcolando la mediana di questi termini. Poiché il numero dei termini è 5 (dispari) la mediana è data dal termine centrale che è 12. Quindi $Q_1 = 12$. Il terzo quartile è compreso tra i termini superiori alla mediana, quindi fra 21, 23,24, 26, 30. Anche questi sono in numero dispari e la mediana è 24, per cui il terzo quartile è $Q_3 = 24$.

Percentili

I percentili sono quelli che dividono la distribuzione in 100 parti uguali e quindi sono molto simili ai quartili. Per dare un'idea di come si collocano i percentili nella distribuzione, osserviamo che il primo percentile (P_1) supera 1/100 dei casi ed è superato dal rimanente 99/100; il secondo percentile (P_2) supera i 2/100 dei casi ed è a sua volta superato dai restanti 98/100. Così procedendo notiamo che il 25° percentile (P_{25}) corrisponde al primo quartile (Q_1), il 50° percentile (P_{50}) corrisponde al secondo quartile (Q_2) ed alla mediana, il 75° (P_{75}) corrisponde al terzo quartile Q_3.

8. La variabilità

Le medie, come già detto, servono a sintetizzare in un solo numero una raccolta di dati a carattere quantitativo e consentono confronti fra le misure di uno stesso fenomeno in momenti e luoghi diversi o fra misure di fenomeni diversi.
Ad esempio può essere utile confrontare il reddito medio degli italiani con quello medio degli abitanti di altri Paesi, oppure fare il confronto con i redditi medi in epoche passate dello stesso nostro Paese. Le medie però non forniscono alcuna indicazione circa la variabilità dei dati. Ad esempio non mettono in evidenza che fra i diversi cittadini vi sono redditi tra loro molto differenti.
Se osserviamo le due successioni di dati:

a) 8, 9, 10, 15, 18
b) 1, 7, 10, 12, 30

si vede che le loro medie aritmetiche:

$$M_a = \frac{8+9+10+15+18}{5} = 12 \qquad \text{e} \qquad M_b = \frac{1+710+12+30}{5} = 12$$

sono uguali come sono uguali pure le mediane, ma le due successioni sono molto diverse per la diversa *variabilità*: nella prima la distribuzione varia da 8 a 18, nella seconda da 1 a 30. Occorre quindi misurare questa variabilità; diamo allora la seguente definizione:

Si chiama campo di variabilità di un insieme di n valori la differenza fra il valore massimo ed il valore minimo.

Se indichiamo la variabilità con R (dall'inglese *range*) e con x_{min} e x_{max} rispettivamente il valore minimo ed il valore massimo, si ha:

$$R = x_{max} - x_{min}.$$

Allora le distribuzioni di valori *a*) e *b*) presentano, rispettivamente, i campi di variabilità:

$$R_a = 18 - 8 = 10 \qquad \text{e} \qquad R_b = 30 - 1 = 29$$

9. Scarto semplice medio dalla media aritmetica

Dato un insieme di n valori di un carattere quantitativo $x_1, x_2, \ldots, x_n$ e detta M_a la media aritmetica ed indicati con:

$$|x_1 - M_a|, \quad |x_2 - M_a|, \ \ldots\ldots, \ |x_n - M_a|$$

i valori assoluti degli scarti, si chiama *scarto semplice medio dalla media aritmetica* degli n numeri dati, al media aritmetica dei valori assoluti degli scarti dalla loro media aritmetica. In altre parole, detto S_{M_a} lo scarto medio semplice, in base alla definizione si ha:

$$S_{M_a} = \frac{|x_1 - M_a| + |x_2 - M_a| + \cdots + |x_n - M_a|}{n}$$

o, in forma compatta:

$$S_{M_a} = \frac{\sum_{k=1}^{n} |x_k - M_a|}{n}.$$

Nel caso che i dati statistici siano distribuiti con le frequenze $f_1, f_2, \ldots, f_n$ lo scarto medio assume la forma:

$$S_{Mp} = \frac{|x_1 - M_p| + |x_2 - M_p| + \cdots + |x_n - M_p|}{f_1 + f_2 + \cdots + f_n}$$

dove M_p rappresenta la media aritmetica ponderata. In forma compatta possiamo scrivere:

$$S_{Mp} = \frac{\sum_{k=1}^{n} |x_k - M_p| \cdot f_k}{n}.$$

Vediamo alcune proprietà dello scarto semplice medio.

<u>*Prima proprietà*</u>

Mentre la somma degli scarti relativi dalla media aritmetica è nulla, la somma dei valori assoluti degli scarti è nulla se e solo se i dati sono tutti uguali tra loro.

Seconda proprietà

Tanto più piccolo è lo scarto semplice medio tanto più i valori si addensano attorno alla media aritmetica.

Terza proprietà

Si definisce anche lo scarto medio dalla mediana ponendo:

$$S_{Me} = \frac{|x_1 - M_e| + |x_2 - M_e| + \cdots + |x_n - M_e|}{n}$$

se i dati sono semplici; se sono ponderati si avrà:

$$S_{Me} = \frac{|x_1 - M_e| \cdot f_1 + |x_2 - M_e| \cdot f_2 + \cdots + |x_n - M_e| \cdot f_n}{n}.$$

Lo scarto semplice medio dalla mediana gode dell'importante proprietà di essere il più piccolo fra tutti gli scarti medi della variabile da qualsiasi valore medio.

Quarta proprietà

Dividendo lo scarto medio per il valore medio rispetto al quale è stato calcolato, si ottiene un corrispondente indice relativo che viene detto *scarto semplice relativo*. Lo scarto semplice relativo dalla media aritmetica è $\frac{S_{M_a}}{M_a}$ mentre quello relativo alla mediana è $\frac{S_{M_e}}{M_e}$.

Esempio

Consideriamo la distribuzione di dati:

3; 4; 5; 6; 11; 13; 14.

Si ha che:

$$M_a = \frac{3+4+5+6+11+13+14}{7} = 8$$

mentre è $M_e = 6$ (essendo 6 il valore del termine centrale). Calcoliamo gli scarti dalla media aritmetica:

$|x_1 - M_a| = |3 - 8| = 5$; $|x_2 - M_a| = |4 - 8| = 4$; $|x_3 - M_a| = |5 - 8| = 3$;

$|x_4 - M_a| = |6 - 8| = 2$; $|x_5 - M_a| = |11 - 8| = 3$; $|x_6 - M_a| = |13 - 8| = 5$

$$|x_7 - M_a| = |14 - 8| = 6.$$

$$S_{M_a} = \frac{5+4+3+2+3+5+6}{7} = 4\,.$$

Lo scarto semplice relativo alla media aritmetica è quindi:

$$\frac{S_{M_a}}{M_a} = \frac{4}{8} = \frac{1}{2} = 0{,}5\,.$$

Calcoliamo ora gli scarti dalla mediana:

$|x_1 - M_e| = |3 - 6| = 3$; $|x_2 - M_e| = |4 - 6| = 2$; $|x_3 - M_e| = |5 - 6| = 1$;

$|x_4 - M_e| = |6 - 6| = 0$; $|x_5 - M_e| = |11 - 6| = 5$; $|x_6 - M_e| = |13 - 6| = 7$;

$$|x_7 - M_e| = |14 - 6| = 8.$$

$$S_{M_e} = \frac{3+2+1+0+5+7+8}{7} = \frac{26}{7} = 3.71.....$$

Lo scarto semplice relativo alla mediana è quindi:

$$\frac{S_{Me}}{M_e} = \frac{3.71...}{86} = 0{,}619....$$

10. Varianza e scarto quadratico medio

Dato un insieme di n valori di un carattere quantitativo $x_1, x_2, \dots, x_n$ si chiama *varianza* di tali numeri la media aritmetica dei quadrati degli scarti dei numeri stessi dalla media aritmetica.

Indicata con σ^2 la varianza e con M_a la media aritmetica dei numeri dati, in base alla definizione si ha:

$$\sigma^2 = \frac{(x_1 - M_a)^2 + (x_2 - M_a)^2 + \cdots + (x_n - M_a)^2}{n}\,.$$

Se i dati anziché essere semplici hanno le frequenze, rispettivamente f_1, $f_2, \ldots, f_n$, la varianza è data dalla:

$$\sigma^2 = \frac{(x_1 - M_a)^2 \cdot f_1 \ + \ (x_2 - M_a)^2 \cdot f_2 \ + \ \cdots \ + \ (x_n - M_a)^2 \cdot f_n}{f_1 \ + \ f_2 \ + \ \cdots \ + \ f_n}$$

dove questa volta M_a rappresenta la media aritmetica ponderata dei valori dati. Indicando con ${}_2M_a$ la media aritmetica dei quadrati dei dati, si può dimostrare che:

$$\sigma^2 = {}_2M_a - (\ M_a)^2$$

cioè la varianza è uguale alla differenza fra la media aritmetica dei quadrati dei dati ed il quadrato della media aritmetica dei dati stessi.
Si chiama *scarto quadratico medio di n* valori $x_1, x_2, \ldots, x_n$ la radice quadrata aritmetica della varianza di tali valori. Quindi, per definizione, detto σ lo scarto quadratico medio, si ha:

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

se i dati sono semplici, se sono ponderati si ha:

$$\sigma = \sqrt{\frac{(x_1 - M_a)^2 \cdot f_1 + (x_2 - M_a)^2 \cdot f_2 + \ \cdots \ (x_n \ \ M_a)^2 \cdot f_n}{f_1 \ + \ \cdots \ + \ f_n}} .$$

<u>Esempio</u>
Consideriamo la distribuzione:

1; 3; 6; 7; 13.

Si ha che:

$$M_a = \frac{1+3+6+7+13}{5} = 6$$

e gli scarti sono:

$x_1 - M_a = 1 - 6 = -5$; $x_2 - M_a = 3 - 6 = -3$; $x_3 - M_a = 6 - 6 = 0$;
$x_4 - M_a = 7 - 6 = 1$; $x_4 - M_a = 113 - 6 = 7$

per cui è:

$$\sigma^2 = \frac{25+9+0+1+49}{5} = 16{,}8 .$$

Ma è anche:

$${}_2\text{M}_a = \frac{1^2+3^2+6^2+7^2+13^2}{5} = 52{,}8$$

di conseguenza

$$\sigma^2 = 52{,}8 - 6^2 = 16{,}8$$

11. Differenze medie

La variabilità si può misurare anche utilizzando le differenze di ciascun dato da tutti gli altri. In questo caso si parla di *differenze medie* che si definiscono nel modo seguente:

Si chiama differenza media di una distribuzione di dati una media calcolata sulle differenze fra ciascun dato e tutti gli altri.

Dati n valori $x_1, x_2, \ldots, x_n$ calcoliamo tutte le differenze fra ogni termine e ciascuno degli n termini (quindi compresa la differenza con se stesso) e riportiamoli in una matrice nel modo seguente:

$$\left\| \begin{matrix} x_1 - x_1 & x_2 - x_1 \ldots\ldots\ldots & x_n - x_1 \\ x_1 - x_2 & x_2 - x_2 \ldots\ldots\ldots & x_n - x_2 \\ \ldots\ldots\ldots & \ldots\ldots\ldots \quad \ldots\ldots & \ldots\ldots\ldots \\ x_1 - x_n & x_2 - x_n \ldots\ldots\ldots & x_n - x_n \end{matrix} \right\|$$

Si tratta di una matrice $n \times n$ i cui termini situati sulla diagonale principale sono tutti nulli. Il numero totale delle differenze è n^2; poiché n termini (quelli della diagonale principale) sono nulli, le differenze non nulle sono in numero di $n^2 - n = n(n-1)$.

<u>Esempio</u>
Dati i valori 1, 3, 5, 8, calcolare tutte le differenze possibili e costruire la matrice delle differenze.
Si ha:

$$\begin{Vmatrix} 0 & 2 & 4 & 7 \\ -2 & 0 & 2 & 5 \\ -4 & -2 & 0 & 3 \\ -7 & -5 & -3 & 0 \end{Vmatrix}.$$

Come si può facilmente osservare la matrice è scomponibile in:

a) la diagonale principale contenente tutti elementi nulli;
b) due triangoli simmetrici rispetto a questa diagonale e contenenti differenze uguali in valore assoluto ma opposte nel segno.

Da questa matrice si deducono anche alcuni indici di variabilità chiamati *differenze medie* che possono essere così definiti:

1) Differenza media assoluta
Si chiama differenza media assoluta la media aritmetica dei valori assoluti delle differenze.
Tale differenza si dice *con ripetizione* se si considerano tutte le n^2 differenze (cioè comprese le nulle); *senza ripetizione* se si considerano solo le $n(n-1)$ differenze ottenute escludendo i termini nulli della diagonale principale.
Indicando con Δ_r e con Δ rispettivamente le differenze medie con ripetizione e senza, si ha:

$$\Delta_r = \frac{|x_1 - x_1| + |x_2 - x_1| + \cdots + |x_{n-1} - x_n| - |x_n - x_n|}{n^2}$$

$$\Delta = \frac{|x_2 - x_1| + |x_3 - x_1| + \cdots + |x_{n-2} - x_n| - |x_{n-1} - x_n|}{n(n-1)}.$$

Esempio
Consideriamo la distribuzione:

8; 12; 15; 19; 24.

Le differenze con ripetizione sono date da:

$D'_{n,k} = n^k$ che nel nostro caso fornisce il valore $D'_{5,2} = 5^2 = 25$

mentre quelle senza ripetizione sono:

$D_{n,k} = n(n-1)$ che nel nostro caso fornisce il valore $D_{5,2} = 5(5 - 1) = 20$.

La matrice delle differenze è:

$$\begin{Vmatrix} 0 & 4 & 7 & 11 & 16 \\ 4 & 0 & 3 & 7 & 12 \\ 7 & 3 & 0 & 4 & 9 \\ 11 & 7 & 4 & 0 & 5 \\ 16 & 12 & 9 & 5 & 0 \end{Vmatrix}$$

Sommando i termini di ciascuna colonna si ottengono i numeri:

38; 26; 23; 27; 42

e la somma di tutte queste differenze è:

38 + 26 + 23 + 27 + 42 = 156

per cui si ha:

$$\Delta_r = \frac{156}{25} = 6{,}24$$

$$\Delta = \frac{156}{20} = 7{,}80\,.$$

2) <u>*Differenza media quadratica*</u>

Si chiama *differenza media quadratica con ripetizione*, la media quadratica di tutte le differenze (comprese quelle nulle).

Si chiama *differenza quadratica media senza ripetizione* la media quadratica delle differenze non nulle.

Se indichiamo con ${}_2\Delta_r$ e con ${}_2\Delta$ rispettivamente la differenza media quadratica con ripetizione e senza ripetizione, in base alla definizione si ha:

$${}_2\Delta_r = \sqrt{\frac{(x_1 - x_1)^2 + (x_2 - x_1)^2 + \cdots + (x_n - x_n)^2}{n^2}}$$

$$_2\Delta = \sqrt{\frac{(x_2-x_1)^2+(x_3-x_1)^2+\cdots+(x_{n-1}-x_n)^2}{n(n-1)}}$$

e si può dimostrare che:

$$_2\Delta_r = \sigma\sqrt{2}$$

$$_2\Delta = \sigma\sqrt{\frac{2n}{n-1}}\,.$$

12. La concentrazione

Un altro aspetto importante di una distribuzione statistica è la *concentrazione*.

Consideriamo la distribuzione x_1, x_2, …. , x_n di n redditi relativi ad n individui di una popolazione ed indichiamo con:

$$S = x_1 + x_2 + \ldots + x_n$$

Il reddito complessivo di tali individui. Può capitare che per un certo x_k, con $k \in [1,2,\ldots,n]$ valga la:

$$x_k = \frac{99}{100}S\,.$$

allora vuol dire che la ricchezza è concentrata quasi totalmente in un individuo e quindi gli altri $(n-1)$ individui posseggono redditi trascurabili rispetto ad x_k. In questa situazione si dice che la distribuzione è *fortemente concentrata.*

Può capitare anche che la ricchezza totale sia equamente distribuita, ossia che i redditi sono ripartiti in modo tale che ciascun individuo ha un reddito pari ad $\frac{S}{n} = M_a$, cioè è tale che $x_1 = x_2 = \ldots = x_n = M_a$. In questo caso si dirà che la distribuzione è *uniforme* (o anche che la *concentrazione è nulla*).

Sostanzialmente possiamo allora dire che un fenomeno ha un alto grado di concentrazione se molta parte della sua intensità complessiva è attribuita a pochi casi in cui il fenomeno si manifesta, ed essa è tanto minore quanto più la sua intensità complessiva è equamente distribuita. Il

carattere della concentrazione viene indagato soprattutto nei confronti di fenomeni economici quali, ad esempio, i patrimoni, i redditi, le imposte, i salari, ecc.
Per misurare la concentrazione faremo riferimento ad una distribuzione di ricchezza posseduta da una popolazione composta da n individui.
Supponiamo, dapprima, che la ricchezza di quella popolazione sia equidistribuita tra gli n individui ognuno dei quali, quindi, possiede la quota $\frac{S}{n}$. Per rappresentare graficamente la concentrazione fissiamo un sistema di assi cartesiani ortogonali (in ascisse riportiamo il numero degli individui ed in ordinate la ricchezza) ed in esso riportiamo i punti:

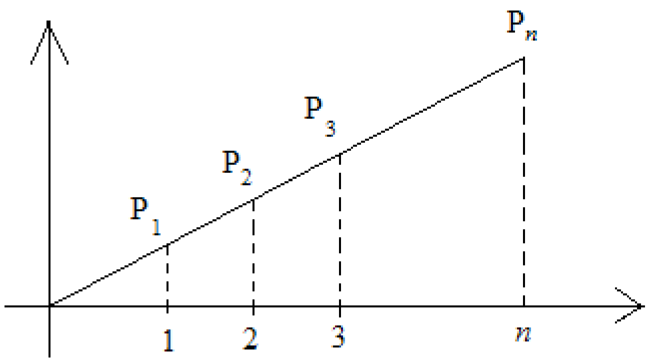
$$P_1\left(1, \frac{S}{n}\right), \quad P_2\left(2,\ 2\frac{S}{n}\right), \ \ldots\ldots , \ P_n\left(n,\ n\frac{S}{n}\right).$$

L'ordinata di ognuno di questi punti rappresenta la ricchezza posseduta da tanti individui quanti ne indica l'ascissa del punto stesso. Si nota facilmente che le ordinate sono proporzionali alle rispettive ascisse e che i punti dati si trovano tutti su una stessa retta passante per l'origine delle coordinate e la sua equazione è $y = \frac{S}{n}x$. Infatti si ha:

$$m = \frac{y_2 - y_1}{x_2 - x_1} = \frac{2\frac{S}{n} - \frac{S}{n}}{2-1} = \frac{S}{n}$$

e anche:

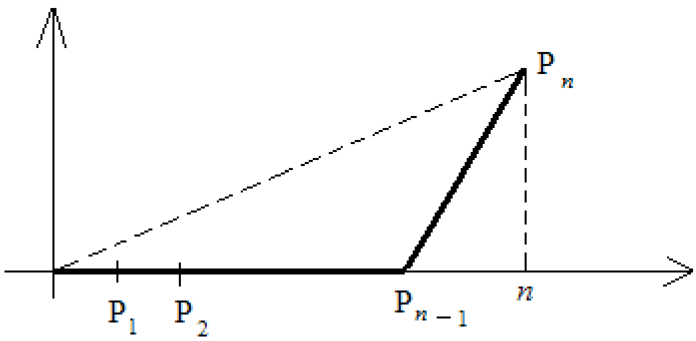
$$m = \frac{y_3 - y_1}{x_3 - x_1} = \frac{3\frac{S}{n} - \frac{S}{n}}{3-1} = \frac{S}{n}.$$



Questa retta prende il nome di *retta di equidistribuzione* perché geometricamente rappresenta la ricchezza della popolazione in esame ripartita in parti uguali fra tutti i suoi componenti.
Supponiamo ora che in quella stessa collettività la ricchezza sia tutta concentrata in un solo individuo, cosicché tutti gli altri non ne posseggono affatto. Disponiamo allora gli n individui in modo che la ricchezza del primo valga zero; la ricchezza posseduta dai primi due insieme valga zero; e così via fino alla ricchezza posseduta dai primi $(n-1)$ individui (insieme) sia nulla. L'ultimo individuo possiede tutta la ricchezza globale S. Si hanno così i punti

$$P_1(1,0), \quad P_2(2,0), \dots, P_{n-1}(n-1, 0), \quad P_n(n, S)$$

che andiamo a rappresentare graficamente:

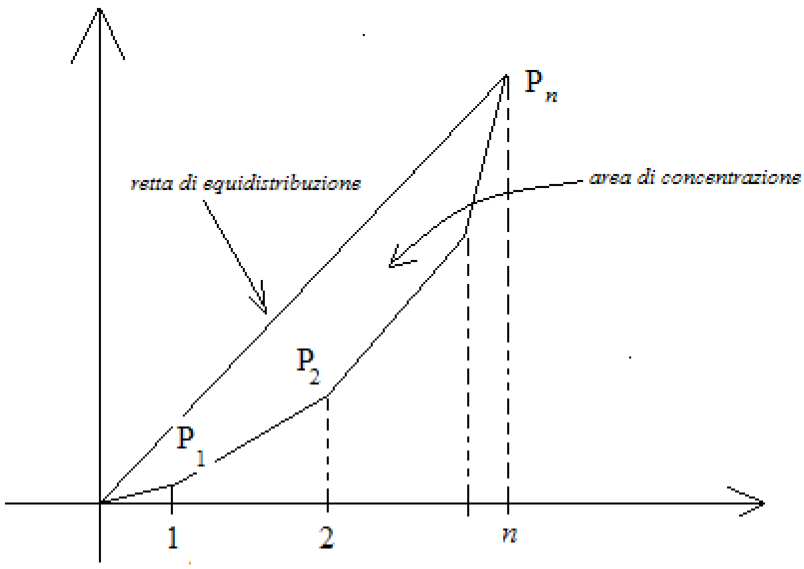


Si osserva facilmente che la linea che rappresenta tale fenomeno coincide con l'asse delle ascisse fino ad $(n-1)$ e si innalza fino a P_n per indicare che per tutti gli n individui essa è diventata S. Se n è sufficientemente grande, $(n-1)$ tende a confondersi con n ed allora l'angolo $O\hat{P}_{n-1}P_n$ tende ad essere retto ed il grafico che rappresenta la concentrazione è costituito dalla linea avente la forma dell'angolo $O\hat{P}_{n-1}P_n$ e si chiama *spezzata di concentrazione massima.*
In realtà la ricchezza di una qualsiasi collettività non è mai equidistribuita né sarà mai concentrata soltanto su un singolo individuo. Consideriamo quindi il caso generale. Indichiamo con $x_1, x_2, \dots, x_n$ le quote individuali di ricchezza posseduta dagli n individui e disponiamo tali quote in ordine non decrescente, cioè consideriamo la successione:

$$0 < x_1 \leq x_2 \leq x_3 \leq \dots \leq x_n$$

e rappresentiamo graficamente i punti:

$P_1(1, x_1)$, $P_2(2, x_1+x_2)$, ... , $P_n\,(n, x_1+x_2+ \cdots + x_n)$



La spezzata $OP_1P_2\ldots P_n$ si chiama *spezzata di concentrazione* relativa alla data distribuzione. (Osserviamo che l'ordinata di ciascun punto P_i rappresenta la ricchezza posseduta da tanti individui quanti ne indica l'ascissa del punto stesso).
Per n sufficientemente grande il diagramma sarà rappresentato, anziché da una spezzata, da una curva crescente avente la concavità rivolta verso l'alto ed è chiamata *curva di concentrazione*. Tra la retta della equidistribuzione e la curva di concentrazione è compresa un'area che viene detta *area di concentrazione.*
È ovvio che tanto più la ricchezza è concentrata, tanto più la curva di concentrazione presenta un'accentuata concavità e tanto più grande risulta l'area di concentrazione. Dopo quanto detto appare chiaro che un'efficace misura della concentrazione può essere espressa dal rapporto tra l'area di concentrazione del suo diagramma ed il massimo che tale area potrebbe assumere. Si dà quindi la seguente definizione:

Se si divide l'area C di concentrazione per l'rea T del triangolo avente per lati la retta di equidistribuzione e la spezzata di concentrazione massima, si ottiene un numero R chiamato rapporto di concentrazione della distribuzione data e quindi si ha:

$$R = \frac{C}{T}.$$

Se R=0 vuol dire che C=0 e quindi il fenomeno è equidistribuito. Se R=1 vuol dire che C=T, ossia la ricchezza appartiene ad un solo individuo e quindi la concentrazione è massima.
Per determinare il valore del rapporto di concentrazione si deve calcolare l'area di concentrazione, cioè l'area di quella parte di piano racchiusa fra la curva di concentrazione e la retta di equidistribuzione e quindi fare il rapporto fra queste e l'area del triangolo avente per lati la retta di equidistribuzione e la spezzata di concentrazione massima.
Questo procedimento però risulta laborioso, ed allora si ricorre ad un altro procedimento in quanto si può dimostrare che *il rapporto di concentrazione R è dato dal rapporto tra la differenza media assoluta senza ripetizione* Δ *ed il doppio della media aritmetica, cioè:*

$$R = \frac{\Delta}{2M_a}$$

<u>Esempio</u>
Studiare la concentrazione della distribuzione

1, 3, 5, 8.

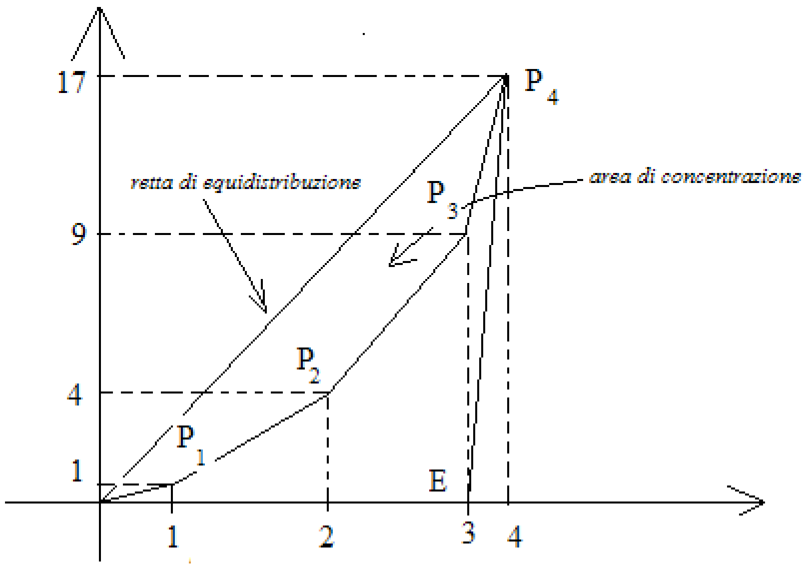
Si ha che:

$$M_a = \frac{1+3+5+8}{4} = \frac{17}{4} = 4{,}25$$

$$\Delta = \frac{\left(|x_2 - x_1| + |x_3 - x_1| + |x_4 - x_1| + |x_3 - x_2| + |x_4 - x_2| + |x_4 - x_3|\right) \cdot 2}{n \cdot (n-1)}$$

$$\Delta = \frac{(2+4+7+2+5+3) \cdot 2}{4 \cdot 3} = 3{,}8333\ldots..$$

Consideriamo i punti (1,1), (2,4), (3,9) e (4,17), che andiamo a rappresentare graficamente:



Si ha che: la retta di equidistribuzione è la OP_4; la spezzata di concentrazione massima è OEP_4; la spezzata della nostra distribuzione è $OP_1P_2P_3P_4$.
L'area T del triangolo OEP_4 di concentrazione massima risulta:

$$T = \frac{3 \cdot 17}{2} = 25,5 .$$

Troviamo ora l'area di concentrazione. Si ha:

$$A_1 = \frac{1 \cdot 1}{2} = \frac{1}{2} = 0,5$$

$$A_2 = \frac{(1+4) \cdot 1}{2} = \frac{5}{2} = 2,5$$

$$A_3 = \frac{(4+9) \cdot 1}{2} = \frac{13}{2} = 6,5$$

$$A_4 = \frac{(9+17) \cdot 1}{2} = 13$$

$$A_1 + A_2 + A_3 + A_4 = 22,5$$

$$A_0 = \frac{4 \cdot 17}{2} = 34$$

e quindi, infine:

$$C = 34 - 22,5 = 11,5.$$

Quindi il rapporto di concentrazione è:

$$R = \frac{C}{T} = \frac{11,5}{25,5} = 0,45...$$

oppure:

$$R = \frac{3,8333}{2 \cdot 4,25} = 0,45....$$