Unità Didattica N° 20

La concentrazione

01) Considerazioni generali

02) La curva di Lorentz

03) Il rapporto di concentrazione

Unità didattica N° 20 : LA Concentrazione

Concentrazione

Considerazioni generali

L'ultimo aspetto delle distribuzioni statistiche che si considera è la **concentrazione** . La **concentrazione** ha significato pratico solamente nel caso di *distribuzioni di caratteri trasferibili* . I caratteri trasferibili sono quelli per cui l'ammontare totale del carattere può essere posseduto da una sola unità statistica , mentre tutte le altre ne sono sprovviste , oppure può essere equidistribuito fra tutte le unità in uguale misura . Sono **caratteri trasferibili** : **1)** il reddito delle persone **2)** il numero di addetti delle imprese **3)** la superficie coltivata dalle aziende agricole 4) il numero di voti ottenuti dai partiti politici alle elezioni .

Sono invece *caratteri non trasferibili* quelli che sono propri di ogni singola unità , come ad esempio l 'età , la statura , il peso delle persone . Nel caso di caratteri trasferibili si può valutare se tali caratteri sono equamente distribuiti tra tutte le unità statistiche mediante le **misure di concentrazione** . La *concentrazione* risulta tanto maggiore quanto più l'ammontare totale del carattere è concentrato in poche unità ; se il carattere è **equidistribuito** la concentrazione è nulla .

Si dice che un *carattere è equidistribuito* se per tutte le unità statistiche esso assume lo stesso valore . In tal caso a qualsiasi frazione delle unità statistiche compete una pari frazione dell'ammontare complessivo del carattere . Ad esempio l'indennità di contingenza è equidistribuita in quanto essa è percepita nella stessa misura da tutti i lavoratori dipendenti .

Se un carattere non è equidistribuito si dice che esso è **concentrato** . Ad esempio il reddito è concentrato in quanto i diversi soggetti lo percepiscono in misura diversa . Ovviamente la concentrazione si può manifestare a vari livelli fino al **massimo** in cui tutto l'ammontare globale del carattere compete ad una sola unità statistica mentre le altre sono caratterizzate da valori nulli .

La concentrazione intesa nel senso dianzi indicato non è che un aspetto della variabilità relativa.

Sempre dal punto di vista intuitivo si può pensare al caso di concentrazione massima quando questo ammontare appartiene ad una sola unità , mentre tutte le altre ne sono completamente sprovviste , per cui costruiremo un **indice di concentrazione** che assuma valore zero nel caso di equiripartizione , valore uno nel caso di massima concentrazione e di crescere al crescere della diversità tra i termini , ossia al crescere della sperequazione tra i redditi .Da tutto ciò emerge che concentrazione e variabilità sono due aspetti di uno stesso concetto ; si preferisce parlare di **concentrazione** se il carattere è trasferibile , mentre se esso non è trasferibile è più opportuno parlare di **variabilità** .

Di indici di concentrazione se ne possono costruire una infinità , ma noi utilizzeremo soltanto il più usato . Si tratta del rapporto **R** di concentrazione introdotto dal Gini nel 1912 .

Lo studio della concentrazione relativa ad una distribuzione (ad esempio) di redditi passa attraverso tre fasi : **1)** PRIMA FASE : elaborazione della distribuzione del reddito mediante la costruzione di un tabella contenente la colonna delle percentuali cumulate sia dei redditieri che dei redditi **2)** SECONDA FASE : rappresentazione della distribuzione del reddito mediante la costruzione della **curva di Lorenz** o curva di concentrazione o spezzata di concentrazione .

3) TERZA FASE : stima del grado di concentrazione mediante il calcolo del rapporto di concentrazione .

Consideriamo una distribuzione di un carattere (*trasferibile*) X avente n modalità quantitative (intensità) $x_1, x_2, \cdots\cdots, x_n$ e costruiamo la **curva di Lorenz** (curva di concentrazione , spezzata di concentrazione) . Per fare ciò bisogna riferire il piano ad un sistema ortogonale di assi cartesiani e riportare in ascisse la frazione $p_i = \frac{i}{n}$ (frazione relativa dei redditieri) dei possessori delle i intensità più basse del carattere in esame X , ed in ordinate la frazione $q_i = \frac{\sum_{h=1}^{i} x_h}{\sum_{i=1}^{n} x_i}$ (frazione relativa del reddito) del carattere posseduta dagli i individui . Indi si uniscono i punti così ottenuti con una spezzata e la curva di Lorenz è costruita .

Analizziamo quanto detto con un esempio . Supponendo di disporre dei seguenti redditi settimanali relativi a 5 operai

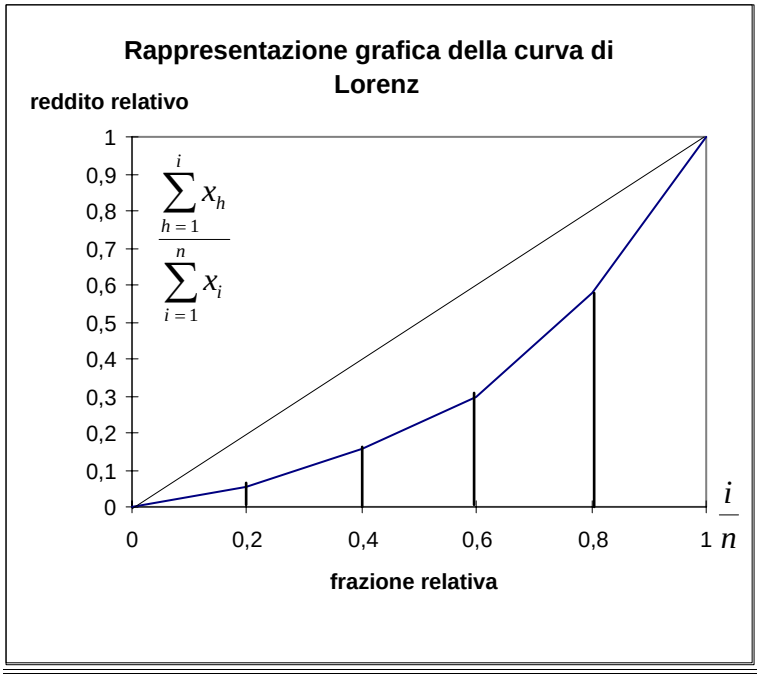
Redditieri	A	B	C	D	E
Redditi (migliaia di lire)	100	70	200	40	300

costruire la **curva di Lorenz** .

La prima operazione da compiere consiste nel disporre tali redditi in ordine non decrescente .

Nella colonna **(1)** elenchiamo i nomi dei **redditieri** (i 5 operai) , nella colonna **(2)** disponiamo i redditi x_i di ciascun operaio , nella colonna **(3)** scriviamo la **cumulata** della colonna (1) , nella colonna **(4)** la frazione relativa $p_i = \frac{i}{n}$ dei redditieri , nella colonna **(5)** il **reddito cumulato** $\sum_{h=1}^{i} x_h$, nella colonna (6) il **reddito relativo** $q_i = \frac{\sum_{h=1}^{i} x_h}{\sum_{i=1}^{n} x_i}$ posseduto dai primi i redditieri .

1	2	3	4	5	6
Redditieri	Redditi x_i	Cumulata della colonna (1) i	Frazione relativa $p_i = \frac{i}{n}$	Reddito cumulato $\sum_{h=1}^{i} x_h$	Reddito relativo $q_i = \frac{\sum_{h=1}^{i} x_h}{\sum_{i=1}^{n} x_i}$
A	40	1	0,2	40	0,056
B	70	2	0,4	110	0,155
C	100	3	0,6	210	0,296
D	200	4	0,8	410	0,577
E	300	5	1	710	1
n=5	$\sum_{i=1}^{n} x_i = 710$				



Adesso generalizziamo la questione e vediamo come è possibile costruire la **curva di Lorenz** per una distribuzione non raggruppata cioè per una distribuzione per unità .

Sia data una distribuzione ordinata in senso non decrescente $x_1, x_2, \cdots, x_i, \cdots, x_n$ con $x_i \leq x_{i+1}$.

Abbiamo :

$p_i = \frac{i}{n}$ = frazione delle unità meno dotate = frazione sul totale n delle unità che presentano una intensità non superiore ad x_i = frazione relativa

$p_i \cdot 100 = \frac{i}{n} \cdot 100$ = frazione relativa percentuale

$q_i = \frac{x_1 + x_2 + \cdots + x_i}{x_1 + x_2 + \cdots + x_n} = \frac{\sum_{h=1}^{i} x_h}{\sum_{i=1}^{n} x_i}$ = frazione dell'ammontare totale del carattere corrispondente alla

frazione p_i delle unità meno dotate = reddito relativo

$q_i \cdot 100$ = reddito relativo percentuale

p_i rappresenta la parte delle unità meno dotate a cui corrisponde la percentuale di reddito q_i .

Elaborazione teorica della distribuzione nel caso che questa sia una **serie** (carattere qualitativo) .

Data la tabella iniziale :

Redditieri (1)	Redditi x_i
A_1	x_1
A_2	x_2
⋮	⋮
A_n	x_n

si costruisce la seguente tabella completa :

Redditieri A_i	Redditi x_i	Cumulata della colonna (1)	Frazione relativa $p_i = \frac{i}{n}$	Reddito cumulato $\sum_{h=1}^{i} x_h$	Reddito relativo $q_i = \frac{\sum_{h=1}^{i} x_h}{\sum_{i=1}^{n} x_i}$	Frazione Relativa percentuale $p_i \cdot 100$	Reddito relativo percentuale $q_i \cdot 100$
A_1	x_1	1					
A_2	x_2	2					
⋮	⋮	⋮					
A_n	x_n	n					

Una volta effettuata l'elaborazione mediante la compilazione della tabella generale , possiamo rappresentare graficamente la distribuzione di reddito considerata . Tale rappresentazione viene eseguita come segue .

Riferito il piano ad un sistema ortonormale di assi cartesiani **riportiamo** sull'asse delle ascisse la frazione relativa $p_i = \frac{i}{n}$ e sull'asse delle ordinate il reddito relativo $q_i = \frac{\sum_{h=1}^{i} x_h}{\sum_{i=1}^{n} x_i}$. Si individuano dei punti del piano , ciascuno dei quali evidenzia una data percentuale di redditieri che possiede una data percentuale del reddito complessivo . Unendo mediante segmenti i punti predetti otteniamo una spezzata che prende il nome di **curva di Lorenz** o curva di concentrazione o *spezzata di concentrazione* .

Nel caso di una **distribuzione per classi** la frazione delle unità meno dotate , ossia delle unità che presentano una intensità (cioè un reddito) minore od uguale ad x_i è :

$$p_i = \frac{n_1 + n_2 + \cdots\cdots + n_i}{n_1 + n_2 + \cdots\cdots + n_n} = \frac{\sum_{h=1}^{i} n_h}{\sum_{i=1}^{n} n_i} = \frac{\sum_{h=1}^{i} n_h}{N}$$

La corrispondente frazione dell'ammontare del carattere (reddito) è :

$$q_i = \frac{n_1 x_1 + n_2 x_2 + \cdots\cdots + n_h x_h}{n_1 x_1 + n_2 x_2 + \cdots\cdots + n_n x_n} = \frac{\sum_{h=1}^{i} n_h x_h}{\sum_{i=1}^{n} n_i x_i}$$

dove x_i sono i valori centrali delle classi .

Elaborazione teorica della distribuzione nel caso che questa sia una **seriazione** (carattere quantitativo) . Data la tabella iniziale :

classi dei redditi	frequenze n_i (redditieri)
a —— b	n_1
c —— d	n_2
⋮	⋮
e —— f	n_n

Si costruisce la seguente tabella completa

A	B	C	D	E	F	G	H	L	M	N
i	classi	x_i	n_i	$n_i x_i$	$\sum_{h=1}^{i} n_h$	$\sum_{h=1}^{i} n_h x_h$	$p_i = \frac{\sum_{h=1}^{i} n_h}{\sum_{i=1}^{n} n_i}$	$q_i = \frac{\sum_{h=1}^{i} n_h x_h}{\sum_{i=1}^{n} n_i x_i}$	$p_i \cdot 100$	$q_i \cdot 100$
1										
2										
⋮										
n										

A) numero di classi **B)** classi del carattere quantitativo **C)** valori centrali delle singole classi

D) frequenze delle singole classi (redditieri)

E) $n_i x_i$ reddito complessivo di ciascuna classe

F) $\sum_{h=1}^{i} n_h$ **frequenze cumulate**

G) $\sum_{h=1}^{i} n_h x_h$ **Reddito complessivo cumulato**

H) $p_i = \frac{\sum_{h=1}^{i} n_h}{\sum_{i=1}^{n} n_i} = \frac{\sum_{h=1}^{i} n_h}{N}$ = **frequenze relative** = frazione delle unità meno dotate

L) $q_i = \frac{\sum_{h=1}^{i} n_h x_h}{\sum_{i=1}^{n} n_i x_i}$ = **reddito relativo** = frazione dell'ammontare del carattere (reddito)

M) $p_i \cdot 100$ = **frequenza cumulata percentuale**

N) $q_i \cdot 100$ = **reddito complessivo cumulato percentuale**

Una volta effettuata l'elaborazione mediante la compilazione della tabella generale , possiamo rappresentare graficamente la distribuzione di reddito considerata . Tale rappresentazione viene eseguita come segue . Riferito il piano ad un sistema ortonormale di assi cartesiani **riportiamo** sull'asse delle ascisse la frazione relativa $p_i = \frac{n_1 + n_2 + \cdots\cdots + n_i}{n_1 + n_2 + \cdots\cdots + n_n} = \frac{\sum_{h=1}^{i} n_h}{\sum_{i=1}^{n} n_i} = \frac{\sum_{h=1}^{i} n_h}{N}$ e sull'asse

delle ordinate il reddito relativo $q_i = \frac{n_1x_1 + n_2x_2 + \cdots\cdots + n_hx_h}{n_1x_1 + n_2x_2 + \cdots\cdots + n_nx_n} = \frac{\sum_{h=1}^{i} n_hx_h}{\sum_{i=1}^{n} n_ix_i}$. Si individuano dei punti del piano , ciascuno dei quali evidenzia una data percentuale di redditieri che possiede una data percentuale del reddito complessivo . Unendo mediante segmenti i punti predetti otteniamo una spezzata che prende il nome di **curva di Lorenz** o **curva di concentrazione** o *spezzata di concentrazione* .

Nel caso di **equidistribuzione dei redditi** , tutti i punti si trovano sul segmento OB giacente sulla bisettrice fondamentale degli assi cartesiani . La **spezzata di concentrazione** si riduce al segmento **OB** e l'area di concentrazione è nulla . .

La *spezzata di concentrazione* assumerà la forma limite OAB (cioè la spezzata costituita dai segmenti consecutivi OA e AB) nel caso di massima concentrazione , quando cioè tutto il reddito è posseduto da un solo individuo . In questo caso l'*area di concentrazione* diventa massima e coincide con l'area del triangolo rettangolo **OAB** . Si deduce che << **quanto più alta è la concentrazione tanto più la curva di concentrazione si discosta dalla retta di equidistribuzione e tende a sovrapporsi alla spezzata OAB** >>

L'area racchiusa fra la retta di equidistribuzione e la curva di concentrazione prende il nome di **area di concentrazione** .

Consideriamo la distribuzione del reddito mostrata nella seguente tabella: essa mostra come si distribuisce fra i 584 elementi di una collettività il **reddito complessivo** prodotto in un certo periodo .

classi di reddito (in milioni)	frequenze (redditieri) n_i
0——5	104
5——10	160
10——15	132
15——20	88
20——25	62
25——35	26
35——45	12
totale	584
	$\sum_{i=1}^{n} n_i = 584$

Poi elaboriamo la seguente tabella completa .

A	B	C	D	E	F	G	H	L	M	N
i	classi	x_i	n_i	$n_i x_i$	$\sum_{h=1}^{i} n_h$	$\sum_{h=1}^{i} n_h x_h$	$p_i = \frac{\sum_{h=1}^{i} n_h}{\sum_{i=1}^{n} n_i}$	$q_i = \frac{\sum_{h=1}^{i} n_h x_h}{\sum_{i=1}^{n} n_i x_i}$	$p_i \cdot 100$	$q_i \cdot 100$
1	0—5	2,5	104	260	104	260	0,178	0,035	17,8	35
2	5—10	7,5	160	1200	264	1460	0,452	0,199	45,2	19,9
3	10—15	12,5	132	1650	396	3110	0,678	0,424	67,8	42,4
4	15—20	17,5	88	1540	484	4650	0,829	0,634	82,9	63,4
5	20—25	22,5	62	1395	546	6045	0,935	0,824	93,5	82,4
6	25—35	30	26	780	572	6825	0,979	0,93	97,9	93
7	35—45	42,5	12	510	584	7335	1	1	100	100

Da questa tabella risulta che il 17,8% , che sono quelli meno abbienti , percepisce il 3,5 % del reddito complessivo ; il 45,2 % percepisce il 19,9 % , il 67,8 % percepisce il 42,4 % , l ‘82,4 % percepisce il 93 % .

In definitiva , l’elaborazione di una **distribuzione per classi di reddito** avviene come segue :

Partendo da una distribuzione per classi , la quale è costituita dalla colonna in cui sono indicate le **classi di reddito** e dalla colonna in cui è indicato il **numero di redditieri** (frequenze) corrispondente a ciascuna classe , si aggiungono altre 8 colonne le quali indicano rispettivamente :

1) i **valori centrali** (x_i) delle classi

2) l’ammontare del reddito per ogni singola classe ($n_i x_i$)

3) le **frequenze cumulate** ($\sum_{h=1}^{i} n_h$)

4) il **reddito complessivo cumulato** ($\sum_{h=1}^{i} n_h x_h$)

5) la **frequenza cumulata relativa** ($p_i = \frac{\sum_{h=1}^{i} n_h}{\sum_{i=1}^{n} n_i}$)

6) il reddito cumulato relativo ($q_i = \frac{\sum_{h=1}^{i} n_h x_h}{\sum_{i=1}^{n} n_i x_i}$)

7) la frequenza cumulata percentuale ($p_i \cdot 100$)

8) il reddito cumulato percentuale ($q_i \cdot 100$)

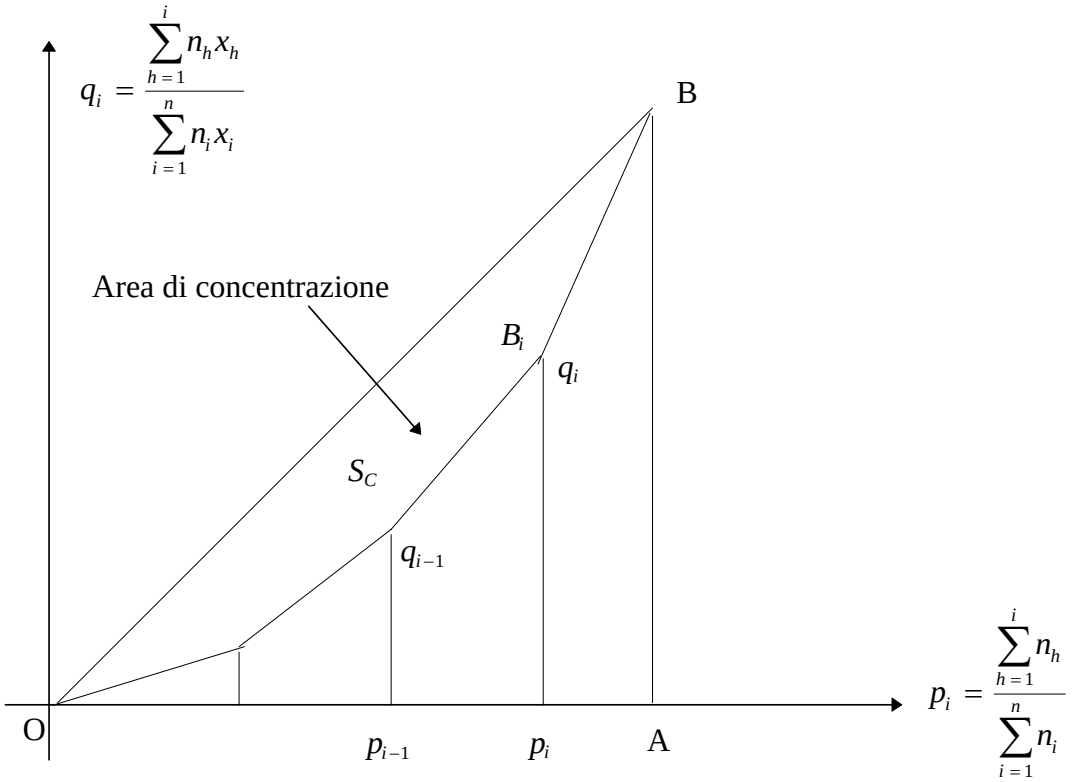
Tra i vari indici proposti per la misura della concentrazione il più usato è il **rapporto di concentrazione R** definito come il rapporto tra l ' *area di concentrazione* e l ' **area di massima concentrazione** , cioè : $R_T = \frac{\text{area di concentrazione}}{\text{area del triangolo OAB}}$

Risulta : $0 \leq R_T \leq 1$ $R_T = 1$ nel caso di **massima concentrazione** $R_T = 0$ nel caso di **equidistribuzione** .

area di concentrazione = porzione di piano racchiusa tra la retta di equidistribuzione e la spezzata di concentrazione

Il calcolo di R_T può effettuarsi con un metodo geometrico o con un metodo analitico .

a) **Metodo geometrico**



Vediamo ora come possiamo calcolare l'area di concentrazione è l'area di massima concentrazione

$$R_T = \frac{S_C}{S(OAB)} = \frac{S(OAB) - S(OABB_i \cdots O)}{S(OAB)} = 1 - \frac{S(OABB_i \cdots O)}{S(OAB)} =$$

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

L'area del trapezio avente come altezza $p_i - p_{i-1}$, come base minore q_{i-1} e come base maggiore q_i ci viene data da : $(p_i - p_{i-1})(q_i + q_{i-1})$ e quindi risulta pure :

$$S(OABB_i \cdots O) = \frac{1}{2}\sum_{i=1}^{n}(p_i - p_{i-1})(q_i + q_{i-1})$$

R_T è detto **rapporto di concentrazione calcolato con la formula dei trapezi** .

b) **Metodo analitico**

Il rapporto di concentrazione è uguale al rapporto fra la differenza media semplice Δ ed il doppio della media aritmetica , cioè :

$$R_T = \frac{\Delta}{2 \cdot \bar{x}} = \frac{S_m}{2 \cdot \bar{x} \cdot n(n-1)}$$

essendo $\Delta = \frac{S_m}{n(n-1)}$ S_m = scostamento medio semplice dalla media aritmetica

Si può dimostrare che , per ogni distribuzioni per classi e per ogni distribuzione per modalità , si ha l'uguaglianza tra il rapporto di concentrazione calcolato con la formula dei trapezi e quello calcolato col metodo analitico .

Nel caso di una distribuzione per unità , come indice sintetico di concentrazione possiamo calcolare R_T oppure il seguente :

$$R = \frac{\sum_{i=1}^{n-1}(p_i - q_i)}{\sum_{i=1}^{n-1} p_i}$$

Si noti che la sommatoria vada 1 ad n-1 in quanto per $i = n$ abbiamo $p_n = q_n = 1$ e quindi $p_{n=o} - q_n = 1 - 1 = 0$.

Determinare l'**indice di concentrazione** R sui dati relativi alla piovosità (espressa in **mm**) in Italia dal mese di dicembre 1961 al mese di novembre 1962 e tracciare la curva di Lorenz .

Mese	Dicembre	Gennaio	Febbraio	Marzo	Aprile	Maggio	Giugno	Luglio	Agosto	Settembr	Ottobre	Novembr
	81,0	66,4	53,1	120,1	78,9	60,3	44,6	29,3	10,4	44,6	103,6	188,3

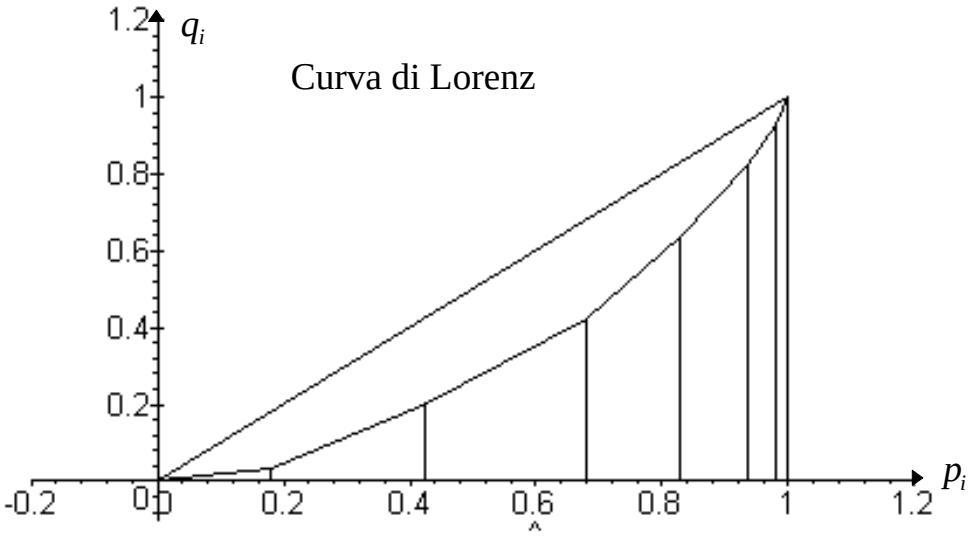
I dati vanno ordinati in senso non decrescente .

Mesi	x_i	i	$p_i = \frac{i}{n}$	$\sum_{h=1}^{i} x_h$	$q_i = \frac{\sum_{h=1}^{i} x_h}{\sum_{i=1}^{n} x_i}$	$p_i - q_i$
Agosto	10,4	1	0,0833	10,4	0,011816	0,071484
Luglio	29,5	2	0,1666	39,9	0,0453357	0,1212643
Giugno	44,6	3	0,25	84,5	0,096	0,154
Settembre	44,6	4	0,33	129,1	0,1466878	0,1833122
Febbraio	53,1	5	0,4166	182,2	0,20702	0,20958
Maggio	60,3	6	0,5	242,5	0,2755368	0,2244632
Gennaio	66,4	7	0,5833	308,9	0,3509828	0,2323172
Aprile	78,2	8	0,6666	387,1	0,4398363	0,2267637
Dicembre	81,0	9	0,75	468,1	0,5318713	0,2181287
Ottobre	103,6	10	0,8333	571,7	0,6495852	0,1837148
Marzo	120,1	11	0,9166	691,8	0,786047	0,130553
Novembre	188,3	12	1	880,1	1	0
	880,1		$\sum_{i=1}^{11} p_i$ =6,4963	3996,3	$\sum_{i=1}^{11}(p_i - q_i)$ = 1,9555811	

$$R = \frac{\sum_{i=1}^{11}(p_i - q_i)}{\sum_{i=1}^{11} p_i} = \frac{1,95558}{6,4963} = 0,3$$

Il rapporto di concentrazione può essere calcolato anche utilizzando la seguente formula :

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

Mesi	x_i	$p_i = \frac{i}{n}$	$p_i - p_{i-1}$	$q_i = \frac{\sum_{h=1}^{i} x_h}{\sum_{i=1}^{n} x_i}$	$q_i + q_{i-1}$	$(p_i - p_{i-1})(q_i + q_{i-1})$
Agosto	10,4	0,0833	0,0833	0,011816	0,011816	0,000984
Luglio	29,5	0,1666	0,0833	0,0453357	0,0571517	0,0047607
Giugno	44,6	0,25	0,0834	0,096	0,1413357	0,0117874
Settembre	44,6	0,33	0,08	0,1466878	0,2426878	0,019415
Febbraio	53,1	0,4166	0,0866	0,20702	0,3537078	0,0306311
Maggio	60,3	0,5	0,0834	0,2755368	0,4825568	0,0402452
Gennaio	66,4	0,5833	0,0833	0,3509828	0,6265196	0,0521891
Aprile	78,2	0,6666	0,0833	0,4398363	0,7908191	0,0658752
Dicembre	81,0	0,75	0,0834	0,5318713	0,9717076	0,0810404
Ottobre	103,6	0,8333	0,0833	0,6495852	1,1814565	0,0984153
Marzo	120,1	0,9166	0,0833	0,786047	1,4356322	0,1195882
Novembre	188,3	1	0,0834	1	1,786047	0,148956
	880,1					0,6738876

$$R_T = 1 - \sum_{i=1}^{12}(p_i - p_{i-1})(q_i + q_{i-1}) = 1 - 0{,}6738876 = 0{,}326$$