LA CONCENTRAZIONE

Fissiamo l'attenzione sui caratteri quantitativi trasferibili.
Ricordo che un carattere è trasferibile se possiamo immaginare che un'unità possa cedere parte del carattere che possiede ad un'altra unità.
Sono esempi di carattere trasferibile: reddito, fatturato, numero addetti, audience televisiva, clienti.
Sono esempi di carattere non trasferibile: altezza e peso.

Supponiamo per esempio di rilevare il reddito delle famiglie italiane. Ci interessa sapere se il reddito complessivo è equidistribuito tra le famiglie oppure se una grossa parte dell'ammontare complessivo del reddito è posseduto da un numero esiguo di famiglie.

- Nel caso in cui tutte le famiglie detengano lo stesso ammontare di reddito, si parla di **equidistribuzione**;
- nel caso in cui tutto il reddito sia posseduto da una sola famiglia mentre tutte le altre hanno zero reddito, si parla di **massima concentrazione**.

Nella realtà ci troviamo sempre in situazioni intermedie e vogliamo misurare il grado di concentrazione del carattere nella nostra popolazione.

L'importanza di un'analisi di questo tipo è soprattutto riferita allo studio della povertà e quindi continua a caratterizzarsi per la sua attualità visto che sembra allargarsi il divario tra i "molto ricchi" e i "poveri".
L'analisi della concentrazione è importante anche in studi di tipo demografico. Posso, ad esempio, analizzare la distribuzione degli italiani negli oltre 8000 comuni della nazione. In questo modo si può determinare il grado di concentrazione della popolazione nelle grandi città (ovvero studiare il livello di urbanizzazione spesso legato al tipo di attività lavorativa svolta dalle persone).

L'analisi della concentrazione è anche importante internamente all'azienda. Per un'azienda è importante determinare il livello di rischio della propria attività. Un tipo di rischio da tenere sotto controllo è legato all'eventuale concentrazione del suo fatturato per prodotto o per cliente. L'azienda potrebbe, infatti, rendersi conto che il suo fatturato è fortemente legato alla vendita di un solo prodotto: questa cosa la renderebbe a rischio.
Supponiamo che l'azienda operi nel settore tecnologico. Il superamento tecnologico del prodotto di punta porterebbe ad una serie crisi dell'azienda! Per un'azienda è anche rischioso avere una situazione in cui il suo fatturato è legato a pochi grandi clienti. La perdita di uno di questi ridurrebbe, infatti, di molto il fatturato.
Per evitare situazioni di questo tipo, occorre analizzare se il fatturato è dovuto in egual misura a tutti i prodotti (cioè se si ha equidistribuzione) oppure se il suo fatturato deriva in gran misura dalla vendita di pochi prodotti (caso di concentrazione).

Analisi statistica della concentrazione

In generale abbiamo che un carattere è tanto più concentrato quanto maggiore è la frazione dell'ammontare complessivo del carattere che spetta alla frazione di unità più ricche.

Ci aspettiamo che un indice che misura la concentrazione
- sia nullo quando il carattere è equidistribuito tra tutte le unità;
- sia massimo quando una sola unità detiene tutto l'ammontare del carattere.

Consideriamo una distribuzione per unità

$a_1, a_2, ..., a_n$

e supponiamo che le a_i siano già ordinate in modo non decrescente, cioè tale che

$a_1 \leq a_2 \leq ... \leq a_{n-1} \leq a_n$

- Indichiamo con

$$A = a_1 + a_2 + ... + a_n = \sum_{i=1}^{n} a_i$$

l'**ammontare complessivo di carattere nel collettivo.**

- Indichiamo con

$$A_i = a_1 + a_2 + ... + a_i = \sum_{j=1}^{i} a_j$$

l'**ammontare di carattere posseduto dalle i unità più povere,** ovvero dalle prime i unità statistiche (ricordate che le unità sono ordinate dalla più povera alla più ricca).

- Indichiamo con

$$Q_i = \frac{A_i}{A}$$

la frazione di ammontare del carattere, sull'ammontare complessivo, posseduto dalle i unità più povere.

Quindi

$$Q_n = \frac{A_n}{A} = \frac{a_1 + \cdots + a_n}{a_1 + \cdots + a_n} = 1$$

→ Indichiamo con

$$P_i = \frac{i}{n}$$

la frazione, sul totale delle unità, delle i unità più povere.

Quindi $P_n = \frac{n}{n} = 1$

Caso di equidistribuzione

Se il carattere è equidistribuito, allora tutte le unità hanno lo stesso ammontare di carattere, cioè

$a_1 = a_2 = ... = a_{n-1} = a_n = a$

e quindi

$$Q_i = \frac{A_i}{A} = \frac{a_1 + \cdots + a_i}{a_1 + \cdots + a_n} = \frac{\overbrace{a + \cdots + a}^{\text{i volte}}}{\underbrace{a + \cdots + a}_{\text{n volte}}} = \frac{i \cdot a}{n \cdot a} = \frac{i}{n} = P_i$$

Pertanto

$Q_i = \frac{i}{n} = P_i$, $i = 1, ..., n$

Caso di massima concentrazione

Se il carattere è massimamente concentrato, cioè se le prime n-1 unità non possiedono nulla mentre l'n-esima unità possiede tutto il carattere,

$a_1 = 0, a_2 = 0, ..., a_{n-1} = 0, a_n = A$

si ha

$Q_i = \frac{0}{A} = 0$, $i = 1, ..., n-1$

$Q_n = \frac{A}{A} = 1$

Ricapitolando, fino ad ora abbiamo trovato che

- $P_i = Q_i$ $i = 1, ..., n$ se $a_1 = a_2 = ... = a_{n-1} = a_n$
- $P_n = Q_n$ sempre

In tutti gli altri casi, cioè in assenza di equidistribuzione e quando $i \neq n$, si ha

$$Q_i < P_i$$

Dimostrazione

Consideriamo la nostra distribuzione

$a_1, a_2, ..., a_n$

Sappiamo che è ordinata. Notiamo che

1) $\frac{A_i}{i} = \frac{a_1 + a_2 + \cdots + a_i}{i} = \mu_i =$ media dei primi i termini della distribuzione. In altre parole μ_i è l'ammontare medio di carattere posseduto dalle i unità più povere.

2) $\frac{A_n}{n} = \frac{a_1 + a_2 + \cdots + a_n}{n} = \mu_n =$ media di tutti i termini della distribuzione. In altre parole μ_n è l'ammontare medio di carattere posseduto dalle unità.

Dal momento che la distribuzione è ordinata si ha (per la proprietà di monotonia della media aritmetica)

$$\mu_i \leq \mu_n$$

dove il segno di uguaglianza vale solo se $a_1 = a_2 = ... = a_{n-1} = a_n$.

La nostra dimostrazione riguarda il caso in cui non c'è equidistribuzione e quindi

$$\mu_i < \mu_n$$

che può essere scritta

$$\frac{A_i}{i} < \frac{A_n}{n} = \frac{A}{n}.$$

Attraverso una semplice operazione algebrica troviamo

$$\underbrace{\boxed{\frac{A_i}{A_n}}}_{Q_i} < \underbrace{\boxed{\frac{i}{n}}}_{P_i}$$

E quindi $Q_i < P_i$

* * * * * * *

Dal momento che $Q_i = P_i$, $i=1,...n$, solo se c'è equidistribuzione e che altrimenti $Q_i < P_i$, $i=1,…,n-1$, possiamo pensare di misurare la concentrazione confrontando P_i e Q_i. In particolare effettuiamo questo confronto calcolando le differenze $P_i – Q_i$

Ovviamente il confronto ha senso per ogni i tranne $i = n$ perché abbiamo visto che $Q_n = P_n = 1$ e quindi $P_n – Q_n = 0$

Sintetizziamo le differenze $P_i – Q_i$, $i = 1, ..., n-1$, calcolando la loro somma

$\sum_{i=1}^{n-1}(P_i - Q_i)$ che indichiamo con C.

→ $C \geq 0$ perché somma di quantità positive o nulle

→ Se c'è equidistribuzione allora

$Q_i = P_i$ per ogni i e quindi $C = 0$

→ Se c'è massima concentrazione allora

$Q_i = 0$, $i = 1, \ldots, n-1$ e quindi $C = \sum_{i=1}^{n-1} P_i$

→ Nei casi intermedi C assume valori compresi tra 0 e $\sum_{i=1}^{n-1} P_i$, cioè

$$0 < C < \sum_{i=1}^{n-1} P_i$$

Un indice così calcolato non è di immediata interpretazione, visto che ne dobbiamo calcolare il massimo valore assumibile nel caso della nostra distribuzione.

Allora calcoliamo l'indice relativo.

In sostanza dividiamo C per il suo massimo che è $\sum_{i=1}^{n-1} P_i$.

In questo modo si ottiene il **rapporto di concentrazione di Gini**.

$$g = \frac{\sum_{i=1}^{n-1}(P_i - Q_i)}{\sum_{i=1}^{n-1} P_i} = \frac{\sum_{i=1}^{n-1} P_i - \sum_{i=1}^{n-1} Q_i}{\sum_{i=1}^{n-1} P_i} =$$

$$= 1 - \frac{\sum_{i=1}^{n-1} Q_i}{\sum_{i=1}^{n-1} P_i} \qquad (1)$$

Osserviamo che $\sum_{i=1}^{n-1} P_i = \sum_{i=1}^{n-1} \frac{i}{n} = \frac{1}{n}\sum_{i=1}^{n-1} i = \frac{(n-1)n}{2n} = \frac{n-1}{2}$

Perché sappiamo che $\sum_{i=1}^{n} i = \frac{n(n+1)}{2} \Rightarrow \sum_{i=1}^{n-1} i = \frac{(n-1)n}{2}$

Pertanto il apporto di concentrazione di Gini si può anche scrivere

$$g = 1 - \frac{2}{n-1}\sum_{i=1}^{n-1} Q_i \qquad (2)$$

Esempio: consideriamo le sei principali reti televisive italiane. Analizziamo se nella fascia oraria 20,30 – 22, 30 queste hanno avuto lo stesso numero di telespettatori oppure se solo una o due reti hanno registrato un forte ascolto. In altre parole cerchiamo di scoprire se il carattere "numero di telespettatori" si è equidistribuito tra le reti oppure è concentrato in una o due reti soltanto.

I dati sono i seguenti:

RAI 1	14 milioni
RAI 2	3 milioni
RAI 3	4 milioni
Canale5	9 milioni
Italia1	2 milioni
Rete4	3 milioni

Il numero di unità è pari a 6, $n = 6$.

Ordiniamo le osservazioni rispetto al carattere rispetto al carattere e calcoliamo i P_i e Q_i.

Rete TV	a_i	P_i	A_i	Q_i
Italia1	2	1/6 = 0.17	2	2/35 = 0.06
Rete4	3	2/6 = 0.33	5	5/35 = 0.14
RAI 2	3	3/6 = 0.5	8	8/35 = 0.23
RAI 3	4	4/6 = 0.67	12	12/35 = 0.34
Canale5	9	5/6 = 0.83	21	21/35 = 0.6
RAI 1	14	6/6 = 1	35	35/35 = 1
	35			

↑

A = ammontare totale (= n° totale di spettatori)

$$\sum_{i=1}^{5} Q_i = 0.06 + 0.14 + 0.23 + 0.34 + 0.6 = 1.37$$

$$\Rightarrow g = 1 - \frac{2}{6-1} 1.37 = 1 - 0.548 = 0.452$$

ciò ci porta a dire che c'è una discreta concentrazione degli spettatori in alcune reti TV.

Osservazione: l'indice di concentrazione che abbiamo visto è calcolabile a partire dai dati sotto forma di distribuzione per unità .

Passiamo ad illustrare un altro modo di misurare la concentrazione. Questa metodologia può essere usata sia per i dati in forma di distribuzione unitaria sia per i dati in forma di distribuzione di frequenze.

LA CURVA DI LORENZ

Fino ad ora abbiamo visto che, una volta ordinate in modo non decrescente le nostre osservazioni, associamo ad ogni unità due numeri: P_i e Q_i.

Rappresentiamo graficamente le coppie di punti (P_i, Q_i).

A tal fine consideriamo il piano cartesiano e poniamo:

- P_i sull'asse delle ascisse
- Q_i sull'asse delle ordinate

Abbiamo visto che

1. $0 \leq P_i \leq 1$ per ogni i
2. $0 \leq Q_i \leq 1$ per ogni i

Disegnamo la curva di Lorenz (o spezzata di concentrazione) unendo i punti di coordinate (P_0, Q_0), (P_1, Q_1), ..., (P_n, Q_n)

Poiché $P_0 = Q_0 = 0$ e $P_n = Q_n = 1$

abbiamo che

- il punto di coordinate (P_0, Q_0) è (0, 0)
- il punto di coordinate (P_n, Q_n) è (1, 1)

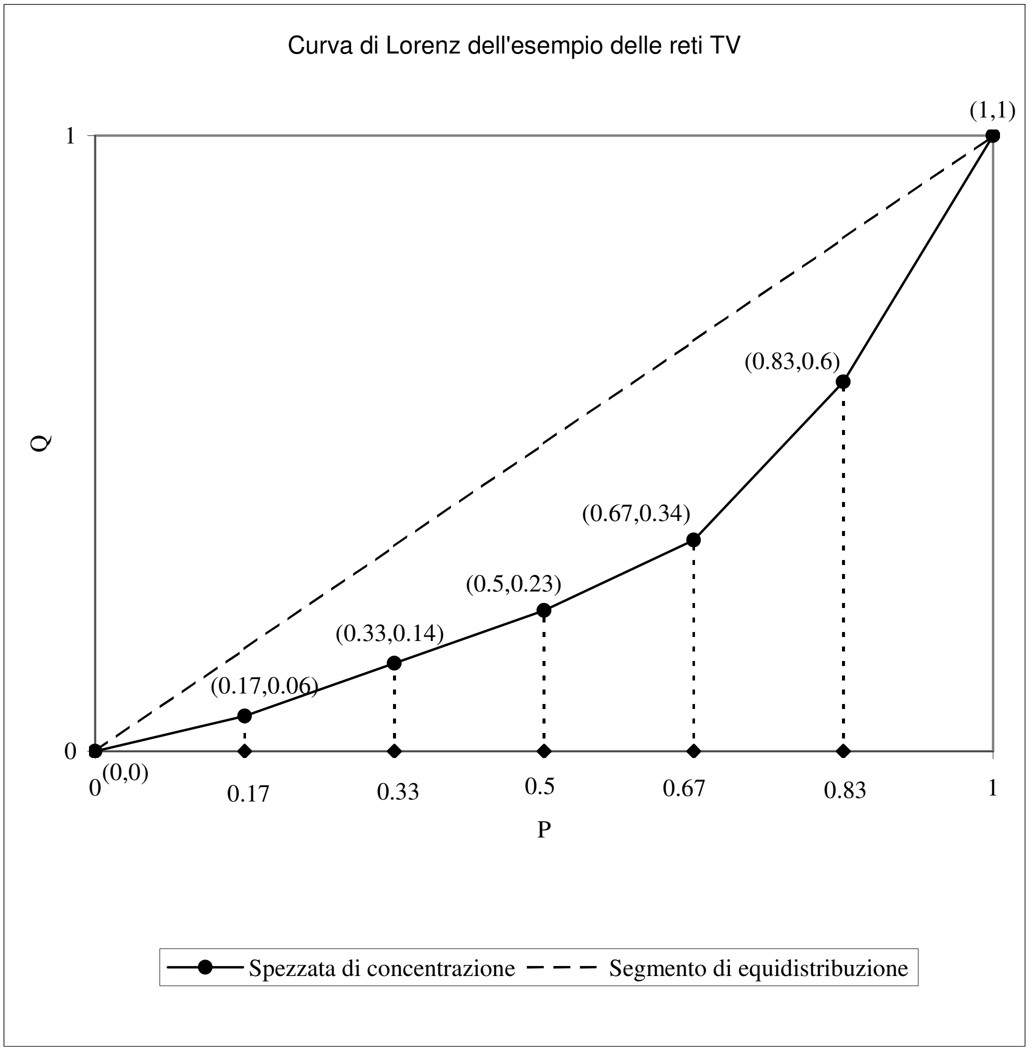


Figura 1

Alcune proprietà:

i) la curva di Lorenz è interamente contenuta nel triangolo di estremi (0,0), (1,0) e (1,1). Ciò è dovuto al fatto che $Q_i \leq P_i$.

ii) La curva di Lorenz è non decrescente perché $Q_i - Q_{i-1} \geq 0$.

Vediamo la dimostrazione:

$$Q_i - Q_{i-1} = \frac{a_1 + a_2 + \cdots + a_{i-1} + a_i}{A} - \frac{a_1 + a_2 + \cdots + a_{i-1}}{A} = \frac{a_i}{A} \geq 0$$

Il segno di uguaglianza vale solo se le prime i osservazioni sono nulle (cioè se le prime i unità non possiedono nulla del carattere).

iii) La curva di Lorenz è convessa (cioè ha incrementi non decrescenti).

Vediamo la dimostrazione:

Dobbiamo dimostrare che $(Q_{i+1} - Q_i) - (Q_i - Q_{i-1}) \geq 0$.

Seguendo la procedura vista sopra troviamo

$$(Q_{i+1} - Q_i) + (Q_i - Q_{i-1}) = \frac{a_{i+1}}{A} - \frac{a_i}{A} = \frac{a_{i+1} - a_i}{A} \geq 0$$

perché le osservazioni sono ordinate in modo non decrescente rispetto al carattere e quindi $a_{i+1} - a_i \geq 0$. L'uguaglianza vale solo se l'unità i e l'unità $i+1$ possiedono lo stesso ammontare di carattere.

Vediamo come è fatta la curva di Lorenz nei casi estremi

→ **Equidistribuzione.**

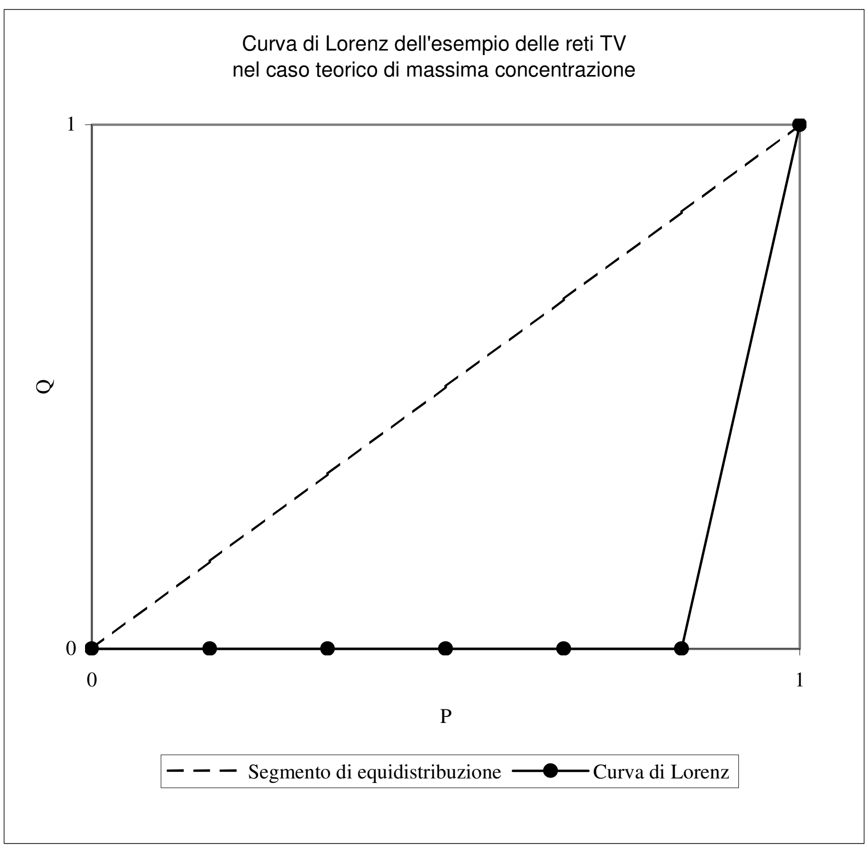
Nel caso di equidistribuzione abbiamo visto che $Q_i = P_i$ per ogni i.

Quindi tutti i punti si trovano esattamente sulla bisettrice del primo quadrante. Pertanto la situazione di equidistribuzione si rappresenta con il **segmento di equidistribuzione** che unisce il punto (0,0) con il punto (1,1). (si veda la linea tratteggiata nella figura sopra)

→ **Massima concentrazione.**

Ricordiamo che c'è massima concentrazione quando le prime $n-1$ unità del collettivo non possiedono nulla e tutto l'ammontare del carattere è posseduto dall'n-esima unità.

Nell'esempio delle reti televisive, supponendo che tutte le reti (tranne RAI 1) non abbiano nessun telespettatore, si ha



Vediamo che la spezzata rimane sempre pari a 0 e inizia a crescere solo a partire dalla 5° unità (cioè, in generale, la $(n-1)$-esima unità) perché le prime $n-1$ unità non possiedono nulla e possiede tutto l'ultima unità.

Quindi nel caso di massima concentrazione la curva di Lorenz unisce i punti di coordinate (0,0), $(P_1,0)$, ..., $(P_{n-1}, 0)$ e (1,1), cioè coincide con l'asse delle ascisse fino all'unità $n-1$ e poi raggiunge il punto (1,1).

Come leggere il grafico della curva di Lorenz?

Da quanto visto sopra si evince che la spezzata di concentrazione è:

☺ Tanto più vicina al segmento di equidistribuzione quanto minore è la concentrazione

☺ Tanto più lontana dal segmento di equidistribuzione (e quindi più vicina all'asse delle ascisse) quanto maggiore è la concentrazione.

Osservazione: dalla Figura1 vediamo che i punti di coordinate

$C_1 = (0,17, 0.06)$, $C_2 = (0.33, 0.14)$ e $C_3 = (0.5, 0.23)$

sono allineati. In particolare C_2 e C_3 si riferiscono alle unità RAI 2 e Rete4 che hanno lo stesso numero di telespettatori (cioè presentano la stessa modalità).

Il fatto che C_1 C_2 e C_3 siano allineati ci indica che per disegnare il tratto della spezzata che va da C_1 a C_3, C_2 non occorre perché giace esattamente su quel tratto di spezzata.

Quindi per disegnare la spezzata possiamo anche considerare la distribuzione di frequenze (invece della distribuzione di unità) che è

N° telespettatori	n_i	P_i (=F_i)	A_i	Q_i
2 (= x_1)	1	0.17	2	0.06
3 (= x_2)	2	0.50	2+3·2=8	0.23
4 (= x_3)	1	0.67	8+4=12	0.34
9 (= x_4)	1	0.83	12+9=21	0.6
14 (= x_5)	1	1	21+14=35	1
	6			

In generale per costruire la curva di Lorenz non è necessario determinare tutti i punti (uno per ogni unità) ma solo gli estremi dei segmenti le cui coordinate si ricavano dalla distribuzione di frequenze. Data una distribuzione di frequenze

x_i	x_1	x_2	…	x_k
n_i	n_1	n_2	…	n_k

I punti necessari per costruire la curva di Lorenz quindi sono:

- (0,0)
- $\left(P_i = \frac{1}{n}\sum_{j=1}^{i} n_j , Q_i = \frac{\sum_{j=1}^{i} n_j x_j}{\sum_{i=1}^{k} n_i x_i} \right) \quad i = 1, ..., k-1$
- (1,1)

A partire dalla curva di Lorenz è possibile determinare un indice di concentrazione.

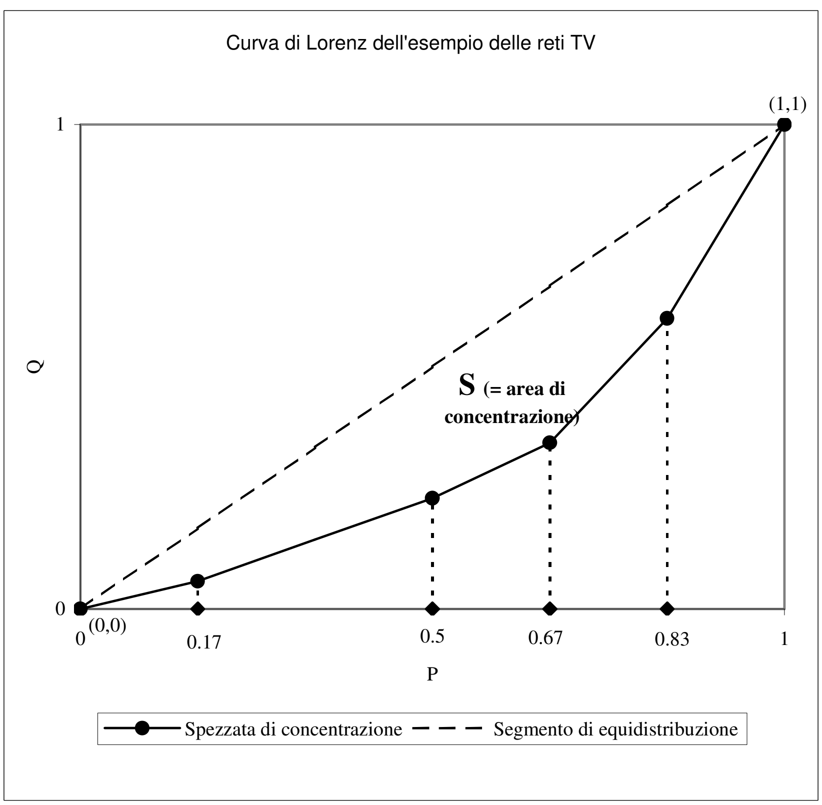
Abbiamo detto che, quanto più la curva di Lorenz è lontana dal segmento di equidistribuzione, tanto più forte è la concentrazione.

Quindi possiamo prendere come misura assoluta della concentrazione l'area compresa tra il segmento di equidistribuzione e la curva di Lorenz. Come linea guida abbiamo che

☺ Quanto minore è l'area tra il segmento di equidistribuzione e la curva di Lorenz, tanto minore è la concentrazione

☺ Quanto maggiore è l'area tra il segmento di equidistribuzione e la curva di Lorenz, tanto maggiore è la concentrazione.

Torniamo al nostro esempio e consideriamo la sua distribuzione di frequenze (vista sopra)



Dobbiamo calcolare l'area S per misurare la concentrazione.
S è detta **superficie di concentrazione** (o area di concentrazione).

S = area (triangolo di vertici (0,0), (1,0), (1,1)) – area sotto la curva di Lorenz

E' immediato osservare che:

$$\text{area (triangolo di vertici (0,0), (1,0), (1,1))} = \frac{1}{2}$$

Quindi

$$S = \frac{1}{2} - \text{area sotto la curva di Lorenz}$$

Nel nostro esempio

Area sotto la curva di Lorenz =

= area (triangolo di vertici (0,0), (0.17, 0), (0.17, 0.6)) +
+ area (trapezio di vertici (0.17, 0), (0.5, 0), (0.5, 0.23), (0.17, 0.6)) +
+ area (trapezio di vertici (0.5, 0), (0.67, 0), (0.67, 0.34), (0.5, 0.23)) +
+ area (trapezio di vertici (0.67, 0), (0.83, 0), (0.83, 0.6), (0.67, 0.34)) +
+ area (trapezio di vertici (0.83, 0), (1,0), (1,1), (0.83, 0.6)) =

$$= \frac{P_1 \cdot Q_1}{2} +$$
$$+ \frac{(Q_2 + Q_1)(P_2 - P_1)}{2} +$$
$$+ \frac{(Q_3 + Q_2)(P_3 - P_2)}{2} +$$
$$+ \frac{(Q_4 + Q_3)(P_4 - P_3)}{2} +$$
$$+ \frac{(Q_5 + Q_4)(P_5 - P_4)}{2} =$$

$$= \frac{1}{2} P_1 \cdot Q_1 + \frac{1}{2} \sum_{i=2}^{5} (Q_i + Q_{i-1})(P_i - P_{i-1})$$

$P_0 = Q_0 = 0$, allora possiamo scrivere che $P_1 \cdot Q_1 = (Q_1 + Q_0)(P_1 - P_0)$. Quindi

$$\text{area sotto la curva di Lorenz} = \frac{1}{2} \sum_{i=1}^{5} (Q_i + Q_{i-1})(P_i - P_{i-1})$$

allora

$$S = \frac{1}{2} - \frac{1}{2} \sum_{i=1}^{5} (Q_i + Q_{i-1})(P_i - P_{i-1})$$

Quanto abbiamo visto con l'esempio (che ha una distribuzione di frequenze con 5 modalità) vale in generale.

Data una distribuzione di frequenze, la superficie di concentrazione S è data da

$$S = \frac{1}{2}\left[1 - \sum_{i=1}^{k}(Q_i + Q_{i-1})(P_i - P_{i-1})\right]$$

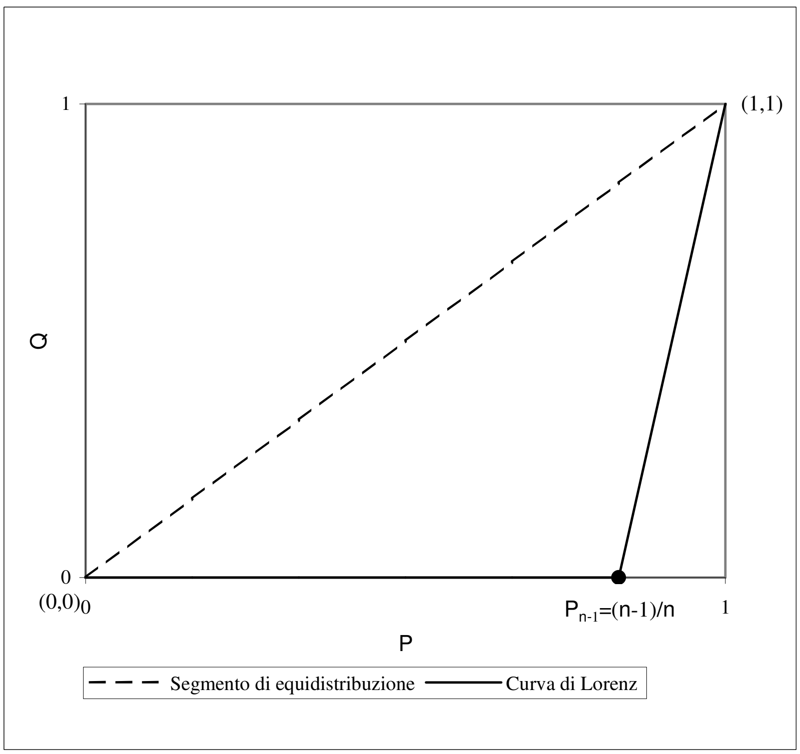
Vediamo i casi estremi

→ Equidistribuzione

S = 0 perché la spezzata coincide con il segmento di equidistribuzione.

→ Massima concentrazione

S assume il suo valore massimo. Si veda il grafico seguente relativo al caso di massima concentrazione; la curva di Lorenz è pari a zero fino all'unità n-1 compresa e poi va a congiungersi al punto (1,1).



L'area S è data dall'area del triangolo di vertici (0,0), $\left(\frac{n-1}{n},0\right)$ e (1,1).

Calcoliamo quest'area, come sopra, per differenza.

max(S) = area (triangolo di vertici (0,0), (1,0), (1,1)) – area (triangolo divertici $\left(\frac{n-1}{n},0\right)$, (1,0), (1,1)) =

$$= \frac{1}{2} - \left(1 - \frac{n-1}{n}\right) \cdot \frac{1}{2} = \frac{1}{2} - \frac{n-n+1}{n} \cdot \frac{1}{2} = \frac{1}{2}\left(1 - \frac{1}{n}\right) =$$

$$= \frac{1}{2} \cdot \frac{n-1}{n}$$

Una volta calcolato max(S), possiamo calcolare l'indice relativo della concentrazione che indichiamo con R

$$R = \frac{S}{max(S)}$$

e quindi

$$R = \frac{\frac{1}{2}\left[1 - \sum_{i=1}^{k}(Q_i + Q_{i-1})(P_i - P_{i-1})\right]}{\frac{1}{2}\frac{n-1}{n}} = \frac{n}{n-1}\left[1 - \sum_{i=1}^{k}(Q_i + Q_{i-1})(P_i - P_{i-1})\right]$$

Abbiamo che

☺ $R = 0$ se c'è equidistribuzione

☺ $R = 1$ se c'è massima concentrazione

Nota: quando n è molto grande allora $\frac{n}{n-1} \cong 1$. Quindi

$$R = 1 - \sum_{i=1}^{k}(Q_i + Q_{i-1})(P_i - P_{i-1})$$

Osservazione: si può dimostrare (non in questo corso) che

$$R = g$$

Esempio: calcoliamo R nell'esempio dell'audience.

N° telespettatori	n_i	P_i	Q_i	$P_i - P_{i-1}$	$Q_i + Q_{i-1}$	$(P_i - P_{i-1})(Q_i + Q_{i-1})$
		0	0			
2	1	0.17	0.06	0.17-0=0.17	0.06+0=0.06	0.102
3	2	0.50	0.23	0.5-0.17=0.33	0.23+0.06=0.29	0.0957
4	1	0.67	0.34	0.67-0.5=0.17	0.34+0.23=0.57	0.0969
9	1	0.83	0.6	0.83-0.67=0.16	0.6+0.34=0.94	0.1504
14	1	1	1	1-0.83=0.17	1+0.6=1.6	0.272
	6					0.6252

$$R = \frac{6}{5}(1 - 0.6252) = 0.45$$

Caso delle distribuzioni in classi

Quando abbiamo distribuzioni in classi e per ogni classe si conosce l'ammontare complessivo del carattere, si può facilmente sia costruire la curva di Lorenz sia misurare la concentrazione mediante l'indice R.

Ovviamente in questo caso c'è sottostante l'ipotesi che l'ammontare del carattere di una classe sia equamente distribuito tra le unità che appartengono a quella classe. In altre parole si assume che le unità della classe posseggano lo stesso ammontare.

L'indice R si calcola nello stesso modo in cui si calcola R per le distribuzioni di frequenze.

Esempio (da Di Ciaccio-Borra): distribuzione delle imprese per classi di addetti (dati in migliaia)

Classi di addetti	n_i	Ammontare	P_i	Q_i	$P_i - P_{i-1}$	$Q_i + Q_{i-1}$	$(P_i - P_{i-1})(Q_i + Q_{i-1})$
			0	0			
0 – 2	2043	2718.3	0.72	0.302	0.72	0.302	0.2174
3 – 9	636	2845.6	0.944	0.618	0.224	0.92	0.2061
10 – 19	103	1352	0.98	0.768	0.036	1.386	0.0499
20 – 49	43.4	1281	0.995	0.910	0.015	1.678	0.0252
50 – 99	11.8	808.7	1	1	0.005	1.910	0.0096
	2837.2	9005.6					0.5082

n è molto grande (n = 2837.2) quindi

$R = 1 – 0.5082 = 0.4918$