Corso di

STATISTICA ECONOMICA
Modulo 1 - A.A. 2004-2005

Prof. Agostino Tarsitano

Informazioni sul corso e sul docente:

http://www.ecostat.unical.it/Tarsitano/atarsita.htm

Lo studio della concentrazione

Riguarda il modo in cui un fenomeno *trasferibile* si ripartisce tra le unità. In particolare la sua attitudine ad accentrarsi in un numero ridotto di unità.

Frazione che spetta, nel complesso, a coloro che ne possiedono -singolarmente- più della media

Diminuisce ← → **Aumenta**

Minore concentrazione — Maggiore concentrazione

Si parla di disuguaglianza distributiva e si considera la concentrazione come un eccesso di tale fenomeno.

Aree di interesse

- **Concentrazione dei redditi e della ricchezza**
- **Concentrazione industriale**
- **Concentrazione del mercato internazionale**
- **Concentrazioni finanziarie**
- **Condizioni di salute della popolazione**

Trasferibilità

E' trasferibile la variabile la cui intensità globale o una sua parte è attribuibile (anche solo idealmente) ad una sola o a poche unità

Variabili TRASFERIBILI	Variabili NON TRASFERIBILI
Reddito ed altri caratteri numerari	Forma o colori
Diritti di possesso	Peso, altezza, memoria
Popolazioni	
Quote di mercato	

Stato di salute

Le variabili non trasferibili riguardano aspetti intrinseci delle unità e non possono essere trasferiti senza trasferire -in solido- l'unità stessa.

Se i valori della variabile sono livelli raggiungibili da qualsiasi unità ed ha un senso la loro somma o aggregazione allora lo studio di concentrazione è plausibile

Simbologia

Ad ogni modalità riscontrata nella rilevazione corrisponde una quota o un ammontare assoluto di variabile:

$$Ammontare\ \ assoluto:\ \ a_i = X_{(i)} n_i;\ \ i = 1,2,\ldots,k \Rightarrow \sum_{i=1}^{k} a_i = n\mu$$

$$Ammontare\ \ relativo:\ \ g_i = \frac{a_i}{n\mu} = \frac{X_{(i)}}{\mu} f_i;\ \ i = 1,2,\ldots,k \Rightarrow \sum_{i=1}^{k} g_i = 1$$

Se la rilevazione è in classi si useranno le medie parziali μ_i al posto delle $X_{(i)}$.

$$Ammontare\ \ assoluto\ \ cumulato:\ \ A_i = \sum_{j=1}^{i} X_{(j)} n_j;\ \ A_k = n\mu$$

$$Ammontare\ \ relativo\ \ cumulato:\ \ q_i = \sum_{j=1}^{i} g_j;\ \ i = 1,2,\ldots,k;\ \ q_0 = 0$$

Esempio_1

Studenti stranieri per regione:

Regione	X_i	n_i	a_i	A_i	g_i	Q_i	Regione	X_i	n_i	a_i	A_i	g_i	Q_i
Basilicata	0	1	0	0	0.0000	0.0000							
Molise	0	1	0	0	0.0000	0.0000	Sicilia	853	1	853	4334	0.0323	0.1642
Calabria	39	1	39	39	0.0015	0.0015	Campania	1015	1	1015	5349	0.0385	0.2027
Trentino AA	45	1	45	84	0.0017	0.0032	Marche	1029	1	1029	6378	0.0390	0.2417
Sardegna	221	1	221	305	0.0084	0.0116	Toscana	1534	1	1534	7912	0.0581	0.2998
Liguria	406	1	406	711	0.0154	0.0269	Veneto	1799	1	1799	9711	0.0682	0.3680
Abruzzi	537	1	537	1248	0.0203	0.0473	Emilia R.	2067	1	2067	11778	0.0783	0.4463
Friuli VG	714	1	714	1962	0.0271	0.0743	Lombardia	2778	1	2778	14556	0.1053	0.5515
Piemonte	754	1	754	2716	0.0286	0.1029	Lazio	4675	1	4675	19231	0.1771	0.7287
Puglia	765	1	765	3481	0.0290	0.1319	Umbria	7161	1	7161	26392	0.2713	1.0000

Gli <u>ammontari relativi</u> danno conto della quota parte di fenomeno (studenti stranieri) pertinente una singola regione.

Gli <u>ammontari relativi cumulati</u> indicano la quota progressiva spettante alle regioni che singolarmente non ne possiedono più di un dato ammontare.

Esempio_2

Punti vendita per numero di commessi/commesse

Commessi/e		Punti	μ	f	(μi)f	q
1	9	4	3.52	0.1026	0.36	0.0193
10	14	6	11.86	0.1538	1.83	0.1170
15	19	13	16.52	0.3333	5.51	0.4116
20	24	9	21.99	0.2308	5.07	0.6831
25	35	5	29.91	0.1282	3.83	0.8882
36	50	2	42.39	0.0513	2.17	1.0000
		39			18.78	

Le medie di classe evitano il calcolo approssimato con i valori centrali.

La classe 15-19 è quella che assorbe il maggior numero di punti vendita.

La classe 36-50, pur impiegando singolarmente un numero elevato di commessi/e, assorbe solo una quota del 12%

Relazione tra le p_i e le q_i

Gli ammontari relativi cumulati risultano sempre inferiori o uguali alle corrispondenti frequenze relative cumulate di unità.

$$Media\ \ dei\ \ primi\ \text{"}i\text{"}\ Valori:\ \ M_i \le \mu$$

questo perché i valori sono ordinati in senso crescente

$$ciò\ \ implica:\ \ \frac{\sum_{j=1}^{i} X_{(j)} n_j}{\sum_{j=1}^{i} n_j} \le \mu \Rightarrow \sum_{j=1}^{i} X_{(j)} n_j \le \mu \sum_{j=1}^{i} n_j \Rightarrow$$

$$\sum_{j=1}^{i} \frac{X_{(j)}}{\mu} \frac{n_j}{n} \le \sum_{j=1}^{i} \frac{n_j}{n} \Rightarrow q_i \le p_i$$

al 10% delle unità non può spettare più del 10% di variabile perché altrimenti, nel restante 90%, si troverebbero valori più piccoli di quelli inseriti nel primo 10% e questo contraddice l'ordinamento crescente

Concentrazione nulla

La concentrazione è NULLA se tutte le unità possiedono lo stesso ammontare

$$X_{(1)} = X_{(2)} = \dots = X_{(k-1)} = X_{(k)}$$

Questo significa che il primo 15% di unità possiede il 15% di variabile, il primo 45% possiede il 45% e così via.

Non ha in sé alcuna caratteristica ideale:

L'equa distribuzione imporrebbe che tutti gli stabilimenti di un settore avessero lo stesso numero di addetti laddove la teoria economica suggerisce che la distribuzione degli addetti è guidata dalla tendenza all'uguaglianza della produttività del lavoro.

Esempio

L'assegnazione dei diritti di scavo delle miniere in Australia e Brasile avviene ripartendo in maglie uguali i terreni.

In questo caso le quote relative di variabile e di unità coincidono, sia nella forma semplice che in quella aggregata

$$g_i = \frac{X_{(i)}}{\mu} * f_i = \frac{\mu}{\mu} f_i = f_i;\quad i = 1,2,\dots,k \Rightarrow q_i = p_i;\; i = 1,2,\dots,k$$

Concentrazione massima

Una sola unità possiede (oppure è ad essa attribuibile) tutta la variabile

$$X_{(1)} = X_{(2)} = \dots = X_{(k-1)};\quad X_{(k)} = T = totale\ della\ variabile$$

Anche questo è un caso limite a cui non si riconosce nessuna valenza particolare.

Come esempi di questa il latifondo come forma di possesso dei terreni:il caso del faraone egizio

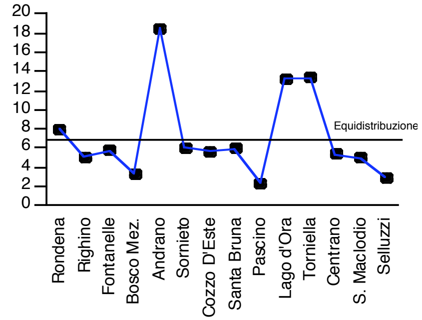
In questo caso le quote relative di variabile sono tutte nulle tranne la k-esima e quelle cumulate sono pure nulle tranne la k-esima che è pari ad uno

$$\begin{cases} g_i = 0;\; i = 1,2,\dots,k-1 \\ g_k = 1 \end{cases};\; \Rightarrow p_i = \begin{cases} 0 \;\; se \;\; i = 1,2,\dots,k-1 \\ 1 \;\; se \;\; i = k \end{cases}$$

Esempio

Popolazione residente nei comuni del comprensorio di Thuria.

Comune	Abitanti	Quota
Rondena	6741	7.92%
Righino	4287	5.04%
Fontanelle	4833	5.68%
Bosco Mez.	2774	3.26%
Andrano	15749	18.50%
Sornieto	5111	6.00%
Cozzo D'Este	4793	5.63%
Santa Bruna	5015	5.89%
Pascino	1971	2.32%
Lago d'Ora	11244	13.21%
Torniella	11366	13.35%
Centrano	4532	5.32%
S. Maclodio	4222	4.96%
Selluzzi	2493	2.93%
	85131	100.00%



Se ogni comune avesse lo stesso numero di abitanti ad ognuno andrebbe una quota del 100/14=7.1%.

Gran parte delle amministrazioni si avvicina a questa soglia, ma la presenza di Andrano con circa 16 mila abitanti porta la distribuzione ad allontanarsi dalla presenza paritaria.

Stima delle medie di classe

Se si ignorano le medie di classe μ_i occorrerà approssimarle

Definizione di media parziale o di classe

$$\mu\left(X \mid L_i \le X \le U_i\right) = \mu_i = \frac{\int_{L_i}^{U_i} x f(x;\theta)dx}{\int_{L_i}^{U_i} f(x;\theta)dx} = \frac{\int_{L_i}^{U_i} x f(x;\theta)dx}{F(U_i;\theta) - F(L_i;\theta)}$$

Se si ipotizza che la densità di frequenza dei dati sia uniforme all'interno della classe allora il valore centrale di questa fornisce la media di classe.

$$\mu_i = c_i = \frac{\left(U_i + L_i\right)}{2}$$

La tecnica di Aigner

Se la distribuzione di frequenza è unimodale è possibile migliorare la stima delle μ_i con le seguenti formule

$$\mu_1 = c_1 + \frac{\left(U_1 - L_1\right)}{6};$$

$$\mu_k = c_k - \frac{\left(U_k - L_k\right)}{6};$$

$$\mu_i = c_i + \left(\frac{f_{i+1} - f_i}{f_i}\right) * \frac{\left(U_i - L_i\right)}{24}; \quad i = 2,3,\ldots,k-1$$

che presuppongono un andamento della funzione di densità o del poligono di frequenza crescente nelle prime classi e decrescente nelle ultime.

Inoltre, i limiti delle classi estreme debbono essere noti e finiti.

Esempio sull'effetto della stima

Redditi		n_i	μ_i	c_i	c'_i
0.5	2	712	1.413	1.250	1.500
2	3	1146	2.536	2.500	2.532
3	4	1586	3.250	3.500	3.514
4	5	1694	4.488	4.500	4.545
5	7.5	3398	6.171	6.250	6.267
7.5	10	2256	8.665	8.750	8.657
10	12.5	1383	11.158	11.250	11.144
12.5	15	852	13.659	13.750	13.703
15	20	999	17.122	17.500	17.466
20	30	688	24.046	25.000	24.669
30	90	452	48.964	60.000	50.000
		15166			

Il miglioramento di Aigner non è costante e nelle classi estreme non ci sono garanzie di plausibilità con nessuna delle due tecniche

Miglioramento paretiano

La distribuzione dei redditi superiori ad un certo livello è, in genere, di tipo paretiano:

$$F(X) = 1 - \left(\frac{a}{X}\right)^b; \quad X \ge a; \quad b > 1$$

Le medie parziali si calcolano in base alla formula:

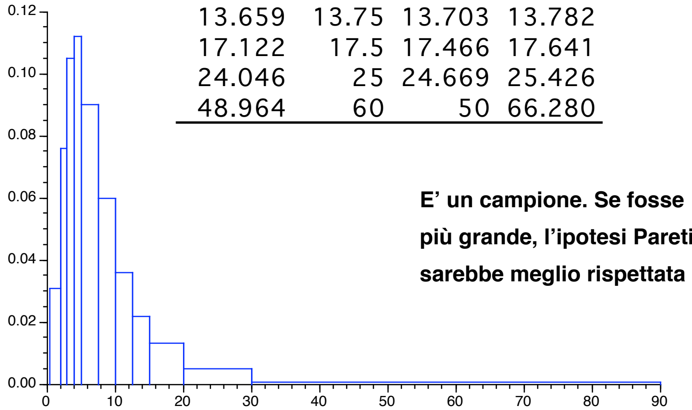
$$\mu_i = \left(\frac{b}{1-b}\right)\left[\frac{U_i^{1-b} - L_i^{1-b}}{L_i^{-b} - U_i^{-b}}\right];$$

$$dove \quad \hat{b} = \frac{Ln\left(1 - p_i\right) - Ln\left(1 - p_{i-1}\right)}{Ln\left(U_i\right) - Ln\left(L_i\right)}$$

Se l'estremo superiore è indeterminato si pone il parametro "b" dell'ultima classe uguale a quello della classe precedente

Verifica

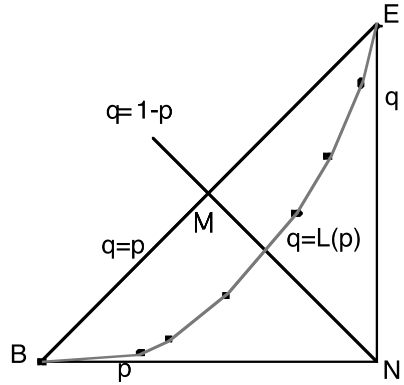
Vere	Ci	Aigner	Pareto
8.665	8.75	8.657	8.777
11.158	11.25	11.144	11.283
13.659	13.75	13.703	13.782
17.122	17.5	17.466	17.641
24.046	25	24.669	25.426
48.964	60	50	66.280



E' un campione. Se fosse più grande, l'ipotesi Paretiana sarebbe meglio rispettata

Il diagramma di Lorenz

Il diagramma è costituito da un triangolo isoscele rettangolo alla cui base sono misurate le frequenze relative cumulate di unità e sull'altezza le quote relative cumulate di variabile



I cateti BN e NE hanno lunghezza uno; l'ipotenusa BE è lunga √2; l'area complessiva del triangolo BNE è 0.5

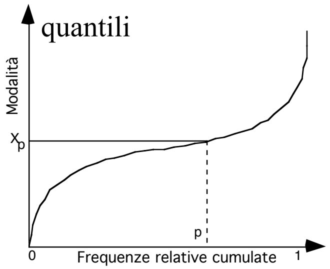
La diagonale incontra la bisettrice nel punto M di coordinate (1/2,1/2)

I punti (p_i,q_i) formano la relazione tra le frequenze relative cumulate di unità p_i=F_i e la corrispondente quota relativa cumulata di variabile q_i.

Funzione di graduazione

Sia X una variabile con dominio (0,∞) con media finita *μ* e funzione di ripartizione F(x)

***(il dominio infinito non deve preoccupare dato che è l'astrazione di un fenomeno in continua espansione*).**



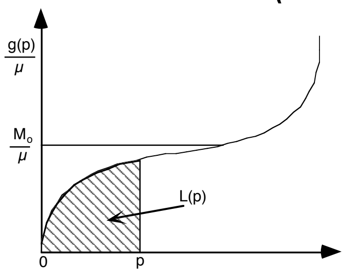
La funzione di graduazione è:

$$F^{-1}(p) = X_p = g(p) = \begin{cases} Min\{x|F(x) \geq p\} & se \;\; 0 < p \leq 1 \\ 0 & se \;\; p = 0 \end{cases}$$

La g(.) è, per costruzione, non decrescente e continua a sinistra per 0<p≤1.

La g(p) è anche nota come funzione quantile.

Funzione di concentrazione
(Curva di Lorenz)



La funzione di concentrazione può essere espressa con l'integrale di Riemann-Stieltjes:

Di solito a=0

$$L(p) = \frac{1}{\mu}\int_0^p [g(t) - a]dt; \quad \mu = \int_0^1 [g(t) - a]dt$$

cioè la L(p) è una funzione di area: ad ogni "p" la funzione associa l'area sottesa alla curva di graduazione (divisa per *μ*) nell'intervallo (0,p).

La definizione -solo apparentemente complessa- è valida sia per variabili continue che discrete

Esempio

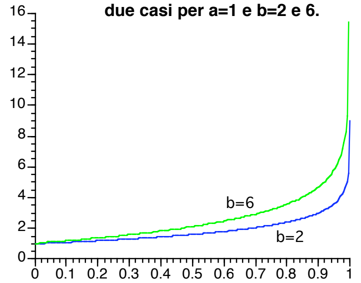
Supponiamo che i dati siano descritti con un modello avente funzione quantile logaritmica

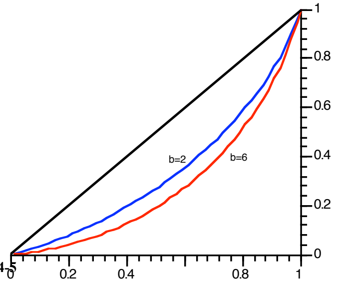
$$g(p) = a - bLn(1-p);$$

$$L(p) = \frac{1}{a+b}\int_0^p \left[a - bLn(1-t)\right]dt$$

$$= p + \left(\frac{b}{a+b}\right)(1-p)Ln(1-p)$$

In figura sono rappresentati due casi per a=1 e b=2 e 6.





Da notare la corrispondenza biunivoca tra funzione quantile e curva di Lorenz

Spezzata di Lorenz

E' il grafico più noto (ma non unico) per lo studio della concentrazione

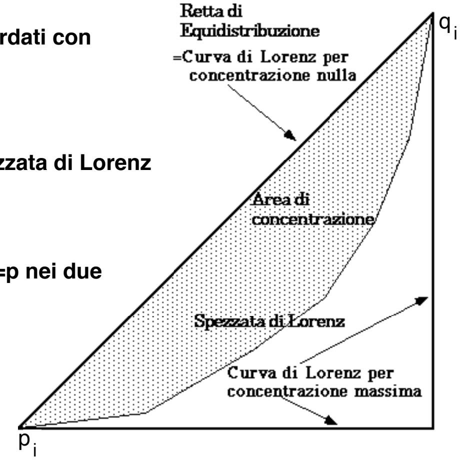
$$L(p) = p_{i-1} + \frac{X_{(i)}}{\mu}(p - p_{i-1}); \quad p \in (p_{i-1}, p_i]; \; i = 1,2,\ldots,n$$

I due vertici (0,0) e (1,1) sono raccordati con segmenti di retta

il grafico che ne risulta è detto spezzata di Lorenz

La spezzata coincide con la retta q=p nei due vertici

Il grafico della spezzata è crescente e rimane sempre al di sotto della retta di equidistribuzione (q ≤ p)



Caratteristiche della spezzata di Lorenz

La spezzata di Lorenz è il grafico di una funzione non decrescente e convessa

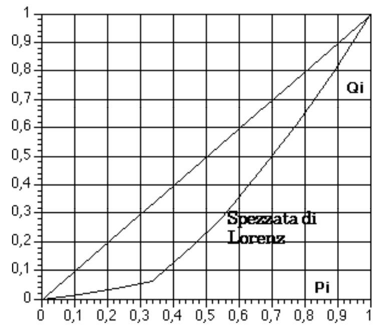
$$L(p) = p_{i-1} + \frac{X_{(i)}}{\mu}(p - p_{i-1}); \quad p \in (p_{i-1}, p_i]; \; i = 1,2,\ldots,n$$

L'inclinazione dei segmenti è positiva (potrebbe essere negativa se qualche modalità avesse valore negativo ad esempio debiti o perdite come reddito negativo).

L'inclinazione dei segmenti è crescente. Infatti le modalità $X_{(i)}$ sono ordinate in senso crescente.

Esempio di spezzata di Lorenz

Gruppi	Matricole	Serie ord.	N_i	f_i	P_i	g_i	Q_i
Scientifico	35475	5115	1	0.1111	0.1111	0.0175	0.0175
Medico	8034	5280	1	0.1111	0.2222	0.0181	0.0356
Ingegneria	48489	8034	1	0.1111	0.3333	0.0275	0.0631
Agrario	5115	32115	1	0.1111	0.4444	0.1099	0.1729
Economico	53610	35475	1	0.1111	0.5556	0.1214	0.2943
Politico-Soc.	32115	45450	1	0.1111	0.6667	0.1555	0.4498
Giuridico	45450	48489	1	0.1111	0.7778	0.1659	0.6157
Letterario	58721	53610	1	0.1111	0.8889	0.1834	0.7991
Diplomi Un.	5280	58721	1	0.1111	1.0000	0.2009	1.0000
	292289		9				



Modalità in classi

Si usa lo stesso schema sostituendo alle modalità ordinate $X_{(i)}$ le medie di classe μ_i qualora siano note oppure una loro stima.

$$L(p) = p_{i-1} + \frac{\mu_i}{\mu}(p - p_{i-1}); \quad p \in (p_{i-1}, p_i]; \; i = 1,2,\ldots,k$$

Le medie sono ordinate in senso crescente. La media globale è μ.

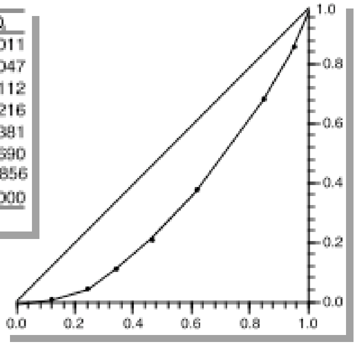
Poiché si ignora il comportamento della variabile nelle classi è necessario fare delle ipotesi per definire la spezzata di Lorenz.

Se la densità all'interno delle classi è simmetrica non si introducono distorsioni nella rappresentazione dei dati

Esempio

Utenze industriali di energia elettrica per ammontare dei consumi

Consumi		Utenti	c_i	f_i	$(X_i)f_i$	p_i	q_i	Q_i
0	5	34	2.50	0.113	0.283	0.113	0.011	0.011
5	10	38	7.50	0.127	0.950	0.240	0.036	0.047
15	20	29	17.50	0.097	1.692	0.337	0.065	0.112
20	25	36	22.50	0.120	2.700	0.457	0.104	0.216
25	30	47	27.50	0.157	4.308	0.613	0.165	0.381
30	40	69	35.00	0.230	8.050	0.843	0.309	0.690
40	50	29	45.00	0.097	4.350	0.940	0.167	0.856
50	75	18	62.50	0.060	3.750	1.000	0.144	1.000
		300		1.000	26.083			



Le quote evidenziano la classe "30-40" come livello di maggiore concentrazione locale (31% dei consumi).

Solo l'1% spetta ai 34 utenti della classe "0-5".

La situazione perciò sembra piuttosto ineguale.

Spezzate alternative

Dalle ipotesi fatte sulle classi discende il comportamento presunto della spezzata nelle stesse classi.

Supponiamo che la distribuzione all'interno di ogni classe si presenti nei due estremi con le frequenze:

Modalità	*Frequenza*
α_i	$\lambda_i f_i$
β_i	$(1-\lambda_i) f_i$

soggette ai vincoli:

$$L_i \le \alpha_i < \beta_i \le U_i$$
$$\alpha_i \lambda_i f_i + \beta_i (1-\lambda_i) f_i = \mu_i$$

Solitamente la curva di Lorenz è costruita ipotizzando l'assenza di variabilità all'interno della classe cioè $\alpha_i=\beta_i=c_i$ (cioè il calor centrale)

Poiché la spezzata è convessa racchiuderà un'area inferiore a quella reale

Spezzate alternative/2

Se aumenta $(\beta_i-\alpha_i)$ e variano le frequenze relative, la spezzata descriverà tutte le posizioni nel triangolo RST.

Minima:

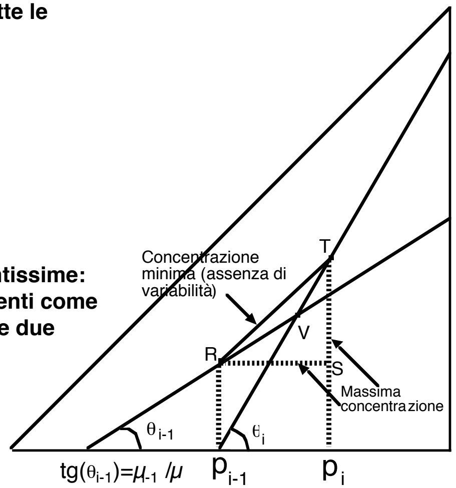
$\beta_i=\alpha_i=c_i$ con frequenza f_i

Massima:

$\alpha_i=L_i$ con frequenza f_i-1/n
$\beta_i=U_i$ con frequenza 1/n

Soluzioni intermedie ne esistono tantissime: ad esempio i due tratti RV e VT aventi come snodo il punto di intersezione delle due tangenti.

Tutte le spezzate sono convesse.



Modelli per la curva di Lorenz/1

Il legame fra la curva di Lorenz e le altre curve è molto stretto per cui lo studio delle curve non è alternativo allo studio dei modelli di distribuzione.

Se infatti è nota la funzione di ripartizione, allora è anche nota la curva di Lorenz:

$$Esponenziale: \quad F(x) = 1 - e^{-\frac{x}{a}}$$

Se si pone F(x)=p s i ottiene $g(p) = -aLn(1-p)$

E quindi si perviene alla seguente curva di Lorenz

$$L(p) = \frac{\int_0^p -aLn(1-t)dt}{\int_0^1 -aLn(1-t)dt} = \frac{p + (1-p)Ln(1-p)}{a}$$

Modelli per la curva di Lorenz/2

L'esame delle curve di Lorenz può essere più fruttuoso e la loro descrizione più facile che non le altre curve

Interpolare la funzione di concentrazione consente il calcolo di qualsiasi quantile cosa che non sempre è possibile interpolando altre funzioni.

Non mancano però riserve a causa dei troncamenti -non sempre plausibili- del campo di variazione per i per i quantili ottenuti a partire dalle curve di Lorenz.

Altri autori giudicano questo approccio privo di una adeguata base teorica ed empirica.

Nello studio degli ordinamenti di Lorenz questa tecnica sta assumendo sempre più importanza.

Un modello per la curva di Lorenz

Cerchiamo una funzione matematica con i seguenti requisiti:

- **Positività** $0 \le L(F;d) \le 1; \; \lim_{F \to 0} L(F;d) = 0; \quad \lim_{F \to 1} L(F;d) = 1;$
- **Nondecrescente** $F_1 < F_2 \Rightarrow L(F_1;d) \le L(F_2;d)$
- **Convessità** $L[aF_1 + (1-a)F_2;b] \le aL(F_1;d) + (1-a)L(F_2;d)$
- **Unicità per l'equidistribuzione:**

$$L(F;d) = F \;\; per \;\; F \in [0,1] \;\; solo \;\; per \;\; un \;\; d = d^*$$

Un modello per la curva di Lorenz/2

Nel caso che la funzione di concentrazione sia espressa da una curva dotata di derivate prime e seconde allora

$$L'(p;d) = \frac{g(p)}{\mu} > 0; \qquad L''(p;d) \ge 0;$$

Che recepiscono l'inclinazione positiva e la convessità

Il fatto che l'ammontare medio della variabile da ripartire sia finito cioè E(x)<∞ comporta:

$$\lim_{p \to 1-} x_p[1-p] = \lim_{p \to 1-} \frac{x_p}{\mu}[1-p] =$$
$$= \lim_{p \to 1-} g(p)[1-p] = \lim_{p \to 1-} L'(p;d)[1-p] = 0$$

Esempi

1.$L(p) = p^b(2-p)^a;\ b \geq a, b+a=1;$ Pietra (1941)

2.$L(p) = p^a e^{-b(1-p)};\ a \geq 1, a+b > \sqrt{a}$ Kakwani e Podder (1973)

3.$L(p) = -abp + (1-a+ab)p^c + a\left[1-(1-p)^b\right];\ a,b,c > 0;\ Maddala\ \ e\ \ Singh\ (1977)$

4.$L(p) = \left[1-(1-p)^a\right]^b;\ Raasche,\ Gaffney,\ Koo,\ Obst\ (1980)$

5.$L(p) = \dfrac{(1-a)^2 p}{\left[(1+a)^2 - 4ap\right]};\quad 0 < a < 1;\quad Aggarwal,\ Singh\ (1984)$

6.$L(p) = pA^{p-1};\ A > 1$ Gupta (1984)

7.$\mathrm{L(p)} = \dfrac{p\left[1+(a-1)p\right]}{1+(a-1)p+b(1-p)}; a,b > 0;\ b-a+1 > 0;\quad Arnold\ (1986).$

8.$L(p) = [ap]^{b(1-p)};\ a,b \geq 0$

Analisi

Controlliamo che la espressione seguente sia una funzione di concentrazione:

$$L(p) = \frac{Ln(1-\beta p)}{Ln(1-\beta)};\quad 0 < \beta < 1$$

Si vede subito che L(0)=0 e L(1)=1. Per la convessità si ha:

$$L'(p) = \left[\frac{-\beta}{Ln(1-\beta)}\right]\frac{1}{(1-\beta p)} > 0;\ L''(p) = \left[\frac{-\beta^2}{Ln(1-\beta)}\right]\frac{\beta}{(1-\beta p)^2} > 0$$

La media della distribuzione è finita

$$\lim_{p\to 1-} L'(p;d)[1-p] = \lim_{p\to 1-}\left[\frac{-\beta}{Ln(1-\beta)}\right]\frac{(1-p)}{(1-\beta p)} = 0$$

Contraddizioni

La derivata prima della funzione di concentrazione è:

$$L'(p) = \frac{F^{-1}(p)}{\mu}$$

Tenuto conto che la curva di Lorenz nei punti estremi tende ad assumere l'inclinazione dei cateti si ha:

$$\lim_{p\to 0+} L'(p) = 0,\qquad \lim_{p\to 1-} L'(p) = \frac{x_{max}}{\mu}$$

Ne consegue che ad esempio

$$L(p) = p^b(2-p)^a \Rightarrow L'(p) = L(p)\left[\frac{b}{p} - \frac{a}{2-p}\right] \Rightarrow 0 \leq x \leq (b-a)$$

per cui tale funzione non è adatta per rappresentare un fenomeno che si espande sistematicamente.

Contraddizioni/2

Nota la curva di Lorenz si può determinare la funzione di ripartizione o di densità

$$L(p) = p^b \Rightarrow L'(p) = bp^{b-1} \Rightarrow bp^{b-1} = \frac{x}{\mu}$$

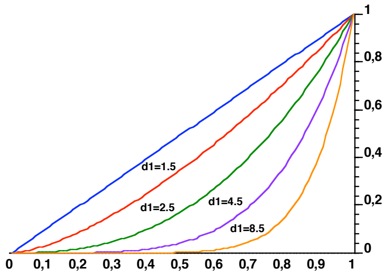
$$\Rightarrow \left[F(x)\right]^{b-1} = \frac{x}{b\mu} \Rightarrow F(x) = \left[\frac{x}{b\mu}\right]^{\frac{1}{b-1}},\ x > 0,\ f(x) = \frac{1}{\mu b(b-1)}\left[\frac{x}{b\mu}\right]^{\frac{2-b}{b-1}}$$

Per b>1 si ottiene una corretta funzione di Lorenz, ma tutte le funzioni di densità sono a forma di L

Tale descrizione è corretta solo per i redditi aldilà di una certa soglia (in genere elevata)

Stima dei parametri

p	q	y=Ln(q)	x=Ln(p)	Regr.LIN(C2:C7;D2:D7;0;1)
0.224	0.035	-3.3524072	-1.4961092	**2.530317**
0.421	0.095	-2.3538784	-0.8651224	0.25295906
0.598	0.190	-1.6607312	-0.5141645	**0.763492**
0.720	0.270	-1.3093333	-0.3285041	16.1408909
0.818	0.345	-1.0642109	-0.2008929	3.52424115
0.893	0.475	-0.7444405	-0.1131687	
1.000	1.000			



$$q = p^{d},\ d > 1$$

$$Ln(q) = dLn(p) \Rightarrow y = mx$$

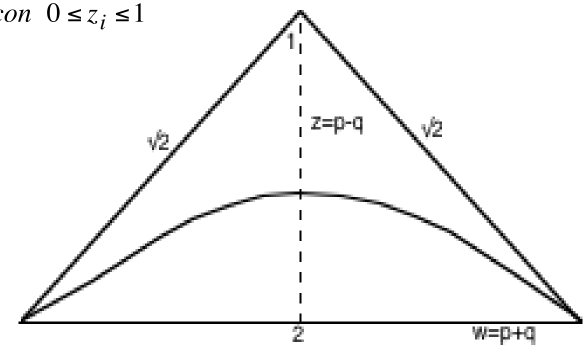
Fatturato		Imprese	f	g	p	q
0	4	48	0.224	0.035	0.224	0.035
4	10	42	0.196	0.060	0.421	0.095
10	20	38	0.178	0.095	0.598	0.190
20	30	26	0.121	0.080	0.720	0.270
30	50	21	0.098	0.075	0.818	0.345
50	80	16	0.075	0.130	0.893	0.475
80	120	23	0.107	0.525	1.000	1.000
		214	1.000			

Le coordinate di Gini

Lo studio della funzione di concentrazione può risultare in certi casi più semplice se è espressa nelle coordinate proposte da Gini nel 1932

$$w_i = p_i + q_i \quad con \quad 0 \le w_i \le 2;$$

$$z_i = p_i - q_i \quad con \quad 0 \le z_i \le 1$$



La funzione di concentrazione z=L(w) somiglia ad un poligono di frequenza unimodale definito in [0,2].

Le coordinate di Gini/2

Le caratteristiche della curva di Lorenz nel nuovo sistema diventano:

$$L(0) = L(2) = 0;$$

$$L'(w) = \frac{\mu - x}{\mu + x}; \qquad L''(w) < 0;$$

$$\underset{w \to 0+}{Lim}\ L'(w) = 1; \quad \underset{w \to 2-}{Lim}\ L'(w) = -1$$

dove "x" è la variabile di cui si esamina la concentrazione e μ la sua media (ipotizzata finita e positiva).

Le condizioni sulle derivate prevengono debordamenti dal triangolo di massima concentrazione.

In questo sistema nel sistema le curve di massima e minima concentrazione sono:

$$Minima.\ L(w) = 0; \quad Massima.\ L(w) = \begin{cases} w & se \quad 0 \le w \le 1 \\ 2 - w & se \quad 1 \le w \le 2 \end{cases}$$

Limitazioni

Nonostante l'alto patrocinio dello stesso Gini, il nuovo sistema non ha avuto grande fortuna.

L'unica funzione qui definita a godere di qualche notorietà è la curva di Kakwani e Podder (1976):

$$L(w) = aw^{\alpha}(2 - w)^{\beta}; \quad 0 \le a, \alpha, \beta \le 1;$$

che ha però un difetto: le derivate prime non soddisfano le condizioni richieste perché tendono all'infinito.

$$L'(w) = \alpha\frac{L(w)}{w} - \beta\frac{L(w)}{(2 - w)}; \Rightarrow \underset{w \to 0+}{Lim}\ L'(w) \to \infty; \quad \underset{w \to 2-}{Lim}\ L'(w) \to \infty$$

Applicazione

Verifichiamo se siamo di fronte ad una curva di Lorenz

$$z = \frac{a}{\pi}\operatorname{sen}\left(\pi\frac{w}{2}\right); \quad 0 \le w \le 2$$

1) L(0)=0, L(2)=0. Se a=2 la funzione descrive una curva di Lorenz

$$L'(w) = \frac{a}{2}\cos\left(\frac{\pi}{2}w\right), \quad L''(w) = -\frac{a\pi}{4}\operatorname{sen}\left(\frac{\pi}{2}w\right)$$

$$\underset{w\to 0+}{Lim}\ L'(w) = 1 \Rightarrow \underset{w\to 0+}{Lim}\ \frac{a}{2}\cos\left(\frac{\pi}{2}w\right) = \frac{a}{2};$$

$$\underset{w\to 2-}{Lim}\ L'(w) = -1 \Rightarrow \underset{w\to 2-}{Lim}\ \frac{a}{2}\cos\left(\frac{\pi}{2}w\right) = -\frac{a}{2}$$

Misura della concentrazione

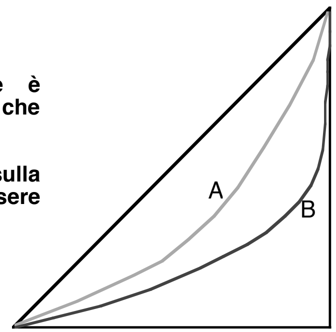
Sin dal suo primo apparire (1905) la curva di Lorenz si è dimostrata uno strumento utile e docile per lo studio della disuguaglianza

"Ci può essere molta diversità di opinioni su cosa si intenda per distribuzione della ricchezza ineguale, ma tutti concordano sull'importanza di sapere se l'attuale distribuzione stia diventando più o meno ineguale".

Nonostante i buoni propositi del suo inventore la curva di Lorenz non può sempre essere utilizzata per valutare la minore o maggiore concentrazione.

Solo se la curva di una distribuzione è interamente contenuta in un'altra si può dire che la contenente è più concentrata

Se le due spezzate si intersecano il giudizio sulla maggiore o minore concentrazione deve essere sospeso



Ordinamenti di Lorenz (Lorenz orderings)

Fonti dei dati sui redditi

Banca d'Italia
Indagine sui bilanci delle famiglie italiane

http://www.bancaditalia.it/statistiche

ISTAT
Indagine sui consumi delle famiglie 1985-2002 e seguenti

http://www.istat.it/

Also useful

http://www.istat.it/binariodie/Stat_per_esempi/index.htm

Requisiti degli indici di concentrazione

L'ideale sarebbe un indice C(X) che aumenti per situazioni di ineguaglianza crescente

Inoltre, deve assumere un valore diverso per ogni diversa distribuzione della variabile.

Questo è impossibile perché gli indici hanno natura sintetica e le inevitabili compensazioni impediscono la corretta diversificazione.

Alcuni requisiti possono però aiutare a scegliere gli indici da usare

- **NORMALIZZAZIONE**
- **INVARIANZA PER TRASFORMAZIONI MOLTIPLICATIVE**
- **DIMINUZIONE PER TRASFORMAZIONI ADDITIVE**
- **SENSIBILITA' AI TRASFERIMENTI**

Normalizzazione

Per comodità di riferimento è almeno necessario che l'indice C(X):

UNIVOCITA' AGLI ESTREMI

a) Sia C(x)=0 se e solo se la distribuzione ha concentrazione nulla
b) Sia C(x)=1 se e solo se la distribuzione ha concentrazione massima

Nulla cambierebbe se il massimo fosse 100 o 1000.

c) Sia C(x) crescente all'aumentare della concentrazione

La normalizzazione non è un requisito essenziale, ma è molto comoda se si confronta la concentrazione di *data set* di diversa numerosità.

Non tutti gli autori sono concordi su tale requisito: alcuni sostengono che è ben diversa la situazione in cui due imprese si bipartiscono il mercato dal caso in cui 1000 imprese controllano ciascuna un millesimo

Esempi

Consideriamo il seguente indice sviluppato in ambito ecologico

$$C(x_1,x_2,\ldots,x_n) = \frac{1}{\sum_{i=1}^{n} g_i e^{-g_i}}; \quad g_i = \frac{x_i f_i}{\mu}, \; i = 1,2,\ldots,k$$

L'indice ha valori nell'intervallo $[e^{-1/n}, e]$ per cui deve essere escluso come misura di concentrazione.

Lo è $(C - e^{-1/n})/(e - e^{-1/n})$

L'indice seguente

$$C(x_1,x_2,\ldots,x_n) = \sum_{i=1}^{k}(f_i - g_i)^2$$

Assume valori nell'intervallo 0 (assenza di concentrazione) e 1-1/n (concentrazione massima). All'aumentare di n, il limite superiore tende all'unità.

Invarianza per trasformazioni molt.

Se si alterano proporzionalmente tutte le modalità, l'indice deve rimanere invariato

$$C(ax_1,ax_2,\ldots,ax_n) = C(x_1,x_2,\ldots,x_n) \quad per \;\; a > 0$$

il requisito consente il confronto della concentrazione di variabili espresse in unità di misura diverse;

Ad esempio la concentrazione dei redditi deve risultare la stessa sia che i redditi siano in lire sia che in euro

C'è da obiettare che chi nulla aveva nella ripartizione X con nulla rimane nella ripartizione aX, ma è chiaro che la sua posizione relativa è peggiorata se a>1 ed è migliorata se a<1.

Esempi

La statistica di Eberhardt è richiamata per verificare l'accentramento spaziale di attività su di un territorio

$$S = \frac{n\sum_{i=1}^{n} x_i^2}{\left(\sum_{i=1}^{n} x_i\right)^2} = \frac{\sum_{i=1}^{k} g_i^2}{\left(\sum_{i=1}^{k} g_i\right)^2}$$

è normalizzata perché varia tra 1/n (assenza di concentrazione) ed uno (massima concentrazione).

E' anche standardizzata dato che lo sono le g_i

$$g_i = \frac{x_i}{\mu} f_i = \frac{a x_i}{a\mu} f$$

Diminuzione rispetto a traslazioni

Se tutte le modalità aumentano di una quantità positiva l'indice deve diminuire

$$C(y)=C(x+a) \quad < \quad C(x) \qquad \text{se } a>0$$

$$g_i = \frac{Y_i}{n\mu_y} = \frac{X_i + a}{n(\mu_x + a)}$$

all'aumentare di "a" le differenze tra modalità tenderanno a sparire (le g_i si avvicinano alle f_i) e la distribuzione tenderà sempre di più alla concentrazione nulla

Tale requisito è utile per le variabili misurate con scale prive di uno zero naturale perché altrimenti si potrebbe ridurre la concentrazione facendo partire la scala dalla costante più conveniente

Sensibilità ai trasferimenti

E' la proprietà più importante e qualificante nello studio della concentrazione

$x_1, x_2, \dots, x_n$

$y_i=x_i$ per $i=1,2, \dots, n;\ i\neq r$ e $i\neq s$;
$y_r=x_r-h$; per $i=r$
$y_s=x_s+h$ per $i=s$

$y_1, y_2, \dots, y_n$

$C(y) < C(x)$ ***La media rimane invariata***

<u>Principio di Pigou-Dalton</u>

Un trasferimento neutrale (order preserving) rispetto alla graduatoria da una unità più "ricca" ad una unità più "povera" deve ridurre l'indice di concentrazione

Sensibilità ai trasferimenti/2

Supponiamo che alla variabile possa applicarsi il principio della utilità marginale decrescente ben noto nel corso di microeconomia.

Un trasferimento da una unità "ricca" ad una "povera" dovrebbe diminuire la concentrazione più di quanto non faccia un trasferimento tra due unità "ricche" di cui una leggermente meno ricca (principio di Kolm)

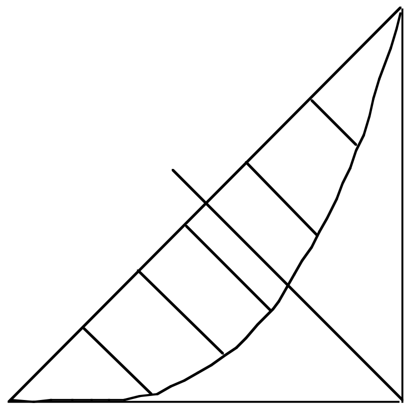
L'effetto è massimo per un trasferimento tra la prima e l'ultima in graduatoria.

Indici derivati da scarti tra quote

Vari indici di concentrazione possono essere derivati dalla distanza (p-q) tra la retta di equidistribuzione e la curva di Lorenz.

In effetti ognuna delle distanze (p-q), rapportata al suo massimo, può essere utilizzata come indice di concentrazione. Fra quelli più noti ci sono:

- **INDICE DI PIETRA-RICCI**
- **QUOTA DIVISORIA**
- **QUOTA MEDIANA**
- **MAGGIORANZA MINIMA**



Concetto

Corrispondono ad una idea di disuguaglianza come deviazione media degli ammontari rispetto allo standard X_m

X_m è una modalità che divide le unità in due sotto popolazioni: quelle con ammontare ≤ X_m e quelle che lo superano.

$$Se \quad y_i = \left(\frac{x_m}{\mu}\right)x_i \Rightarrow \mu_y = \sum_{i=1}^{k} y_i f_i = \sum_{i=1}^{k}\left(\frac{x_m}{\mu}\right)x_i f_i = x_m$$

$$\frac{\sum_{i=1}^{k}\left|y_{(i)} - x_m\right|f_i}{2x_m} = \frac{1}{2x_m}\left[\sum_{y_{(i)}>x_m}\left(y_{(i)} - x_m\right)f_i + \sum_{y_{(i)}\le x_m}\left(x_m - y_{(i)}\right)f_i\right]$$

$$= \frac{1}{2x_m}\left[\left(\sum_{y_{(i)}>x_m} y_{(i)} f_i\right) - x_m\left(1-p_m\right) - \left(\sum_{y_{(i)}\le x_m} y_{(i)} f_i\right) + x_m p_m\right]$$

$$= \frac{1}{2}\left[\left(\sum_{y_{(i)}>x_m}\frac{y_{(i)}}{x_m}f_i\right) - \left(1-p_m\right) - \left(\sum_{y_{(i)}\le x_m}\frac{y_{(i)}}{x_m}f_i\right) + p_m\right]$$

$$= \frac{1}{2}\left[\left(1-q_m\right) - \left(1-p_m\right) - q_m + p_m\right] = p_m - q_m$$

Significato dei D_m

Gli indici sono nulli se e solo se la curva di Lorenz coincide con la retta di equidistribuzione; aumentano se si allontana fino a raggiungere il valore massimo allorché la curva si identifica con i due cateti.

La concentrazione diminuisce se gli scarti relativi dalla modalità X_m, diventano, in media, più piccoli.

$D_m=p_m- q_m$ è la frazione di variabile che, nel complesso, deve essere trasferita dalle unità che ne possiedono più di X_m a quelle che ne possiedono meno affinché si abbia l'equidistribuzione.

Una grave carenze dei D_m, condivisa da tutti gli altri indici sintetici di concentrazione, è la mancanza di univocità: due curve possono avere uno stesso D_m ed essere per il resto diversissime.

Caratteristiche dei D_m

I D_m non risentono di modifiche proporzionali delle modalità poiché ne sono immuni i termini della sua definizione p_m e q_m

Rispetto a variazioni additive si ha:

$$q\left(x_m + a\right) = \sum_{x\le x_m}\frac{x+a}{\mu+a}f_x = \frac{\mu}{\mu+a}\sum_{x\le x_m}\frac{x}{\mu}f_x + \frac{a}{\mu+a}\sum_{x\le x_m}f_x = \frac{\mu q_m + a p_m}{\mu+a}$$

Ne consegue

$$p_m - q\left(x_m + a\right) = p_m - \frac{\mu q_m + a p_m}{\mu+a} = \frac{\mu}{\mu+a}\left[p_m - q_m\right] = \frac{\mu}{\mu+a}D_m$$

per cui ogni D_m è decrescente se α>0.

Reazione ai trasferimenti

Sia $\{X_i$, i=1,2,…, n$\}$ e passiamo a $\{Y_i$, i=1,2,…,n$\}$ distinta per il solo fatto che $Y_r=X_r-\lambda$; $Y_s=X_s+\lambda$, $\lambda>0$ e con r > s; inoltre, supponiamo che tutte le posizioni rimangano inalterate.

L'incremento dell'indice D_m associato alla nuova distribuzione è:

$$D_m(y) - D_m(x) = \left[\sum_{y_{(j)}\le x_m}\frac{y_{(j)}}{\mu}\right] - p_m - \left[\sum_{x_{(j)}\le x_m}\frac{x_{(j)}}{\mu}\right] + p_m$$

$$= \frac{y_{(r)}}{\mu} + \frac{y_{(s)}}{\mu} - \frac{x_{(r)}}{\mu} - \frac{x_{(s)}}{\mu}$$

$$= \frac{x_{(r)}-\lambda}{\mu} + \frac{x_{(s)}+\lambda}{\mu} - \frac{x_{(r)}}{\mu} - \frac{x_{(s)}}{\mu} = -\frac{\lambda}{\mu} + \frac{\lambda}{\mu} = 0$$

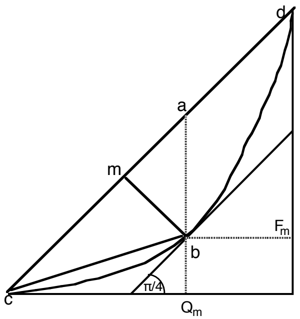
perché risponda ad un trasferimento è necessario che le due modalità si collochino ognuna in un lato diverso rispetto alla soglia X_m

Poiché la sensibilità ai trasferimenti è fondamentale, l'uso di questi indici è inappropriato

Indice di Pietra-Ricci

E' la distanza massima, parallela alla diagonale q=1-p, tra la curva di Lorenz e la retta di equidistribuzione, cioè MB.

Per la stessa funzione si può usare il segmento AB che è proporzionale a MB e che ha il vantaggio di variare tra zero ed uno:



Il segmento “ab” è inoltre pari alla metà dell’area del triangolo “cbd” ovvero del triangolo di area massima iscrivibile all’interno della curva di Lorenz.

$$area\ "cbd" = \frac{cd * mb}{2} = \frac{\sqrt{2}\,\frac{ab}{\sqrt{2}}}{2} = \frac{ab}{2}$$

Il segmento AB è pari alla differenza $D_2 = p_\mu - q_\mu$ di quote cumulate di unità e di variabile associate alla media aritmetica.

Significato

L'indice ha un parallelo con una misura di variabilità relativa

$$D_2 = \left(F_\mu - Q_\mu\right) = \frac{1}{2}\left[1 - Q_m - \left(1 - F_m\right) - Q_m + F_m\right] = \frac{1}{2}\left[\sum_{i=m+1}^{k} \frac{X_i}{\mu} f_i - \sum_{i=m+1}^{k} f_i - \sum_{i=1}^{m} \frac{X_i}{\mu} f_i + \sum_{i=1}^{m} f_i\right]$$

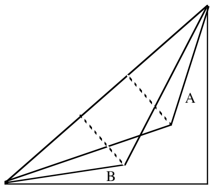
$$= \frac{1}{2\mu}\left[\sum_{i=m+1}^{k} X_i f_i - \mu \sum_{i=m+1}^{k} f_i - \sum_{i=1}^{m} X_i f_i + \mu \sum_{i=1}^{m} f_i\right] = \frac{1}{2\mu}\left[\sum_{i=m+1}^{k} X_i f_i - \mu \sum_{i=m+1}^{k} f_i - \sum_{i=1}^{m} X_i f_i + \mu \sum_{i=1}^{m} f_i\right]$$

$$= \frac{1}{2\mu}\left[\sum_{i=m+1}^{k} \left(X_i - \mu\right) f_i - \sum_{i=1}^{m} \left(X_i - \mu\right) f_i\right] = \frac{\sum_{i=1}^{k} \left|X_i - \mu\right| f_i}{2\mu}$$

una distribuzione è meno concentrata di un'altra se gli scarti relativi da "*μ*" sono, in media, più piccoli.

L'indice di Pietra-Ricci è pari alla frazione di variabile che deve essere trasferita dal gruppo più "ricco" (> μ) al gruppo più "povero" (< μ) per azzerare la concentrazione

Proprietà



Manca di univocità

L'indice è normalizzato (varia tra zero ed uno) $D_2 = p_m - q_m$

L'indice è invariante rispetto a trasformazioni di scala

Diminuisce se aumentano tutti aumenta se diminuiscono tutti

$$Q(X + a) = Q(X) + \frac{a}{a + \mu}\left[F(X) - Q(X)\right] \Rightarrow D_2(X + a) = \frac{\mu}{\mu + a} D_2(X)$$

Si avvede del trasferimento se avviene tra unità su lati diversi di *μ* altrimenti si realizza una compensazione che lo lascia invariato.

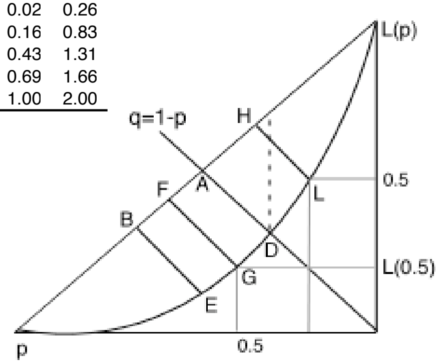
$$X_r < \mu \ \ e \ \ X_s > \mu \Rightarrow \left|X_r + d - \mu\right| + \left|X_s - d - \mu\right| = \mu - d - X_r + X_s - d - \mu = \left|X_r - \mu\right| + \left|X_s - \mu\right| - 2d$$

Esempi/1

AD=Quota divisoria BE= Pietra-Ricci

FG= Quota mediana HL=Maggioranza minima

Addetti		Imprese	μ	f	g	p	q	p+q
0	10	18	6.66	0.10	0.00	0.10	0.00	0.11
1	50	23	33.89	0.13	0.02	0.24	0.02	0.26
5	99	71	76.13	0.42	0.14	0.67	0.16	0.83
10	499	36	286.15	0.21	0.26	0.88	0.43	1.31
50	999	14	731.59	0.08	0.26	0.97	0.69	1.66
1000	5000	5	2333.33	0.03	0.30	1.00	1.00	2.00
		167	225.66	1.00				



Esempi/2

$$L(p)=p^2;\ \ L(0)=0,\ \ L(1)=1,\ \ L'(p)=2p>0,\ \ L''(p)=2>0$$

$$Quota\ \ divisoria:\ L(p)+p=1\Rightarrow p^2+p-1=0$$

$$\Rightarrow p_D=\frac{\sqrt{5}-1}{2}=0.618, q_D=0.382;\ \ D_1=0.236$$

$$Pietra-Ricci:\ \ max\{p-L(p)\};1-L'(p)=0$$

$$p=0.5;\ \ L''(p)=-2<0\Rightarrow D_2=0.5-(0.5)^2=0.25$$

$$Quota\ \ mediana:\ 0.5-L(0.5)\Rightarrow D_3=0.5-(0.5)^2=0.25$$

$$Maggioranza\ \ minima:\ 0.5=L(p_{0.5})\Rightarrow p_{0.5}=\sqrt{0.5}\Rightarrow$$

$$D_4=0.707-0.500=0.207$$

Indici lineari di concentrazione

Rientrano in questo gruppo indici espressi come medie ponderate delle differenze relative tra quote cumulate di unità e di variabile:

$$C_n=\frac{\sum_{i=1}^{k}J_n\left(\frac{i}{n+1}\right)X_{(i)}}{n\mu};\quad \sum_{i=1}^{k}J_n\left(\frac{i}{n+1}\right)=0$$

- **Gini**
- **Bonferroni**
- **Mehran**
- **De Vergottini**
- **Amato**
- **Piesch**

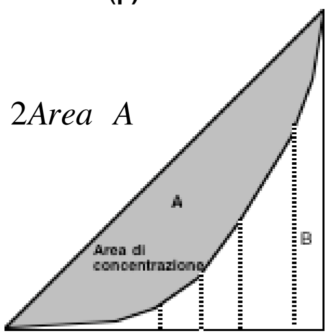
Tali indici hanno diverse caratteristiche in comune (cambia solo il modo di definire i pesi)

Rapporto di concentrazione

E' l'indice più noto e più discusso di concentrazione

Si basa sull'area compresa tra la retta q=p e la curva L(p)

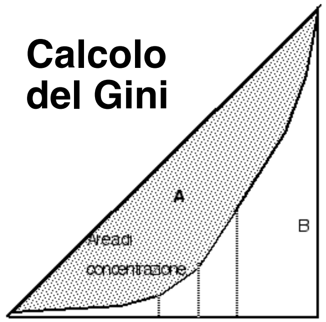
$$R=\frac{Area\ \ A}{Area\ \ A\ +Area\ \ B}=\frac{Area\ \ A}{1/2}=2Area\ \ A$$



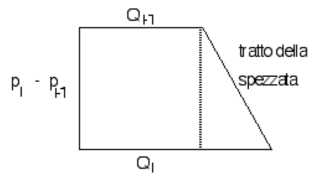
L'area è nulla se la curva di Lorenz si sovrappone alla retta q=p ed è pari a 0.5 quando c'è massima concentrazione.

Ne consegue che R è un indice normalizzato (tra zero ed uno)

Calcolo del Gini



Per il calcolo della area sono disponibili diversi metodi. In particolare si può utilizzare la regola dei trapezi applicata però all'area B e sfruttando la relazione: area "A" = 0.5 - area "B".



$$Area=Altezza*\left(\frac{Base\ \text{minore}+Base\ \ maggiore}{2}\right)\Rightarrow Area_i=\frac{(p_i-p_{i-1})(q_i+q_{i-1})}{2};$$

$$R=1-\left[\sum_{i=1}^{n}(p_i-p_{i-1})(q_i+q_{i-1})\right]=1-\left[\sum_{i=1}^{n}f_i(q_i+q_{i-1})\right]$$

Poiché la curva di Lorenz è convessa, l'uso della formula dei trapezi porta ad approssimare per eccesso l'area B (e quindi per difetto l'area A e quindi R è sottostimato

Formula alternativa

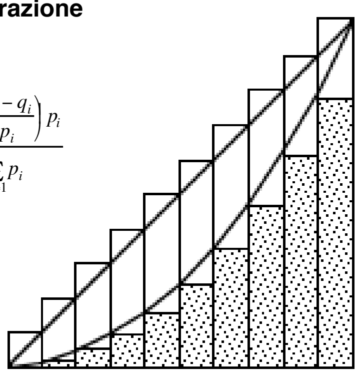
$$A = \frac{\sum_{i=1}^{n} p_i}{n}; \qquad B = \frac{\sum_{i=1}^{n} q_i}{n}$$

"A" è una approssimazione dell'area di massima concentrazione (rettangoli bianchi)

"B" approssima l'area complementare (rettangoli punteggiati).

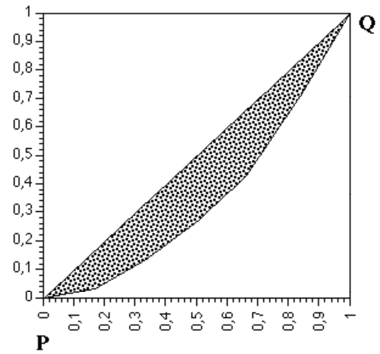
La loro differenza relativa dà una misura approssimata dell'area di concentrazione

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



Calcolo del rapporto di concentrazione

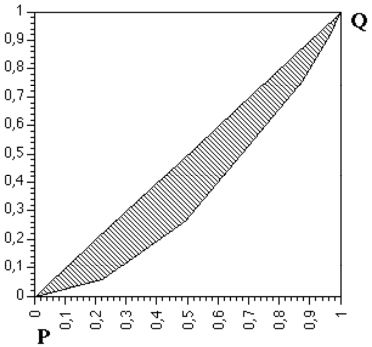
Ministeri	Progetti Legge 64	Serie ord.	n_i	f_i	P_i	g_i	Q_i	$Q_{i-1}+Q_i$	$f_i*(Q_{i-1}+Q_i)$
Agricoltura	3	1	1	0.1667	0.1667	0.0333	0.0333	0.0333	0.0056
Beni cult.	8	3	1	0.1667	0.3333	0.1000	0.1333	0.1667	0.0278
Bilancio	5	4	1	0.1667	0.5000	0.1333	0.2667	0.4000	0.0667
Difesa	1	5	1	0.1667	0.6667	0.1667	0.4333	0.7000	0.1167
Mezzogi.	4	8	1	0.1667	0.8333	0.2667	0.7000	1.1333	0.1889
Parte. Stat.	9	9	1	0.1667	1.0000	0.3000	1.0000	1.7000	0.2833
	30		6	1.0000		1.0000			0.6889



R = 1- 0.6889 = 0.3111

Dati raggruppati

Classi di reddito		n_i	C_i	f_i	P_i	$C_i n_i$	g_i	Q_i	$Q_{i-1}+Q_i$	$f_i*(Q_{i-1}+Q_i)$
0	10	22	5	0.2200	0.2200	110	0.0564	0.0564	0.0564	0.0124
10	20	28	15	0.2800	0.5000	420	0.2154	0.2718	0.3282	0.0919
20	30	37	25	0.3700	0.8700	925	0.4744	0.7462	1.0179	0.3766
30	40	9	35	0.0900	0.9600	315	0.1615	0.9077	1.6538	0.1488
40	50	4	45	0.0400	1.0000	180	0.0923	1.0000	1.9077	0.0763
		100		1.0000		1950	1.0000			0.7061



R = 1- 0.7061 = 0.2939

Qual è l'effetto di approssimare le medie di classe con i loro valori centrali?

Proprietà di R

- L'indice passa da zero (assenza di concentrazione) ad uno (massima concentrazione) aumentando con l'aumentare della disuguaglianza nella distribuzione.
- Esprime la percentuale di variabile che deve essere trasferita da ciascuna unità all'altra che la precede nella graduatoria per ottenere la distribuzione uniforme
- Il rapporto di concentrazione varia tra zero ed uno.

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

Il numeratore è sempre non negativo (perché le $p_i \geq q_i$) ed è sempre inferiore o uguale al denominatore (perché le q_i sono non negative)

Limiti per il Gini

Per una distribuzione in classi esistono infinite spezzate di Lorenz che passano per tutti gli "n" punti e quindi si possono calcolare infiniti rapporti di concentrazione.

Gastwirth (1972) considera l'esito dei trapezi il limite inferiore di R laddove il limite superiore è dato da:

$$R_u = R_l + \sum_{i=1}^{k} \gamma_i; \qquad R_l = 1 - \sum_{i=1}^{k} (p_i - p_{i-1})(q_i + q_{i-1});$$

$$\gamma_i = \frac{1}{\mu}(p_i - p_{i-1})^2 \left[\frac{(\mu_i - L_i)}{(U_i - L_i)}\right]\left[\frac{(U_i - \mu_i)}{(U_i - L_i)}\right]$$

i "γ_i" riducono l'effetto del raggruppamento in classi. Per il loro calcolo è necessario conoscere le medie di classe.

Le correzioni dipendono dalle ampiezza, ma anche dalla scelta dei loro estremi.

Esempio

Redditi USA - 1968

L_i	U_i	f_i	μ_i	p_i	q_i	RL	γ
0	1	0.0482	0.54140	0.0482	0.00323	0.0002	0.0000
1	2	0.0825	1.46363	0.1308	0.01815	0.0018	0.0001
2	3	0.0722	2.44572	0.2029	0.03994	0.0042	0.0001
3	4	0.0690	3.43890	0.2719	0.06925	0.0075	0.0002
4	5	0.0662	4.43732	0.3381	0.10550	0.0116	0.0003
5	6	0.0760	5.40118	0.4141	0.15618	0.0199	0.0006
6	7	0.0785	6.39292	0.4925	0.21813	0.0294	0.0009
7	10	0.2140	8.30454	0.7066	0.43763	0.1404	0.0118
10	15	0.1911	11.90433	0.8977	0.71857	0.2210	0.0186
15		0.1024	22.26150	1.0001	1.00011	0.1760	0.0762
		1.0001	μ=8.1			0.6118	0.1089
						0.3882	0.0134
							0.4016

I calcoli portano all'intervallo [0.3883, 0.4016]. Il risultato è già ottimo (il valore vero è intorno a 0.4014).

Sono possibili limiti ancora ancora più stringenti, ma legati a particolari ipotesi sulla funzione di distribuzione che si nasconde sotto la rilevazione.

Reazioni ai trasferimenti

Analizziamo la dazione di un ammontare "d" dalla X_i alla X_j con $X_i > X_j$ nell'ipotesi che il trasferimento sia neutrale.

L'effetto sull'indice è

$$\Delta R = -\frac{2d}{\mu}\left(\frac{i-j}{n}\right)$$

Fissata la quota da trasferire ($d/n\mu$) e fissato pure "n", l'effetto del trasferimento dipende solo dal numero di posizioni tra la X_j ed X_i.

I trasferimenti tra modalità intorno ad una moda avranno più effetto che non quelli tra unità lontane dalle mode perché qui sarà minore (j-i).

Se in una classe di reddito 10'000-11'000 euro ci sono 20 unità e nella classe 50'000-51'000 ce ne sono 10, l'effetto di spostare 100 euro dall'estremo superiore all'estremo inferiore avrà maggiore impatto sul Gini nel primo caso.

Gini e differenza media

Il rapporto di concentrazione si collega in modo diretto alla differenza semplice media.

$$\Delta = \frac{2\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}\left|x_i - x_j\right|}{n^2} = \frac{2\sum_{i=1}^{n}\sum_{j=1}^{i}\left[x_{(i)} - x_{(j)}\right]}{n^2}$$

$$= \frac{2}{n^2}\sum_{i=1}^{n}\left[ix_{(i)} - \sum_{j=1}^{i} x_{(j)}\right] = 2\mu\sum_{i=1}^{n}\left[\frac{i}{n}\frac{x_{(i)}}{n\mu} - \frac{1}{n}\left(\sum_{j=1}^{i}\frac{x_{(j)}}{n\mu}\right)\right]$$

$$= 2\mu\sum_{i=1}^{n}\left[p_i(q_i - q_{i-1}) - (p_i - p_{i-1})q_i\right] = 2\mu\left[q_i p_{i-1} - p_i q_{i-1}\right]$$

$$dove \quad p_i = \frac{i}{n}; \quad p_i - p_{i-1} = \frac{1}{n}; \quad q_i - q_{i-1} = \frac{x_{(i)}}{n\mu}$$

e pertanto: R=D/2μ

Il valore di R scaturisce dal confronto a coppie tra tutti i valori riscontrati nella rilevazione

Indice di Gini e curve analitiche

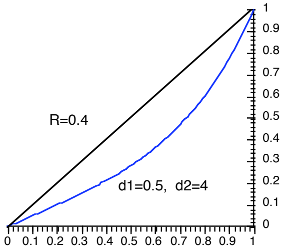
La determinazione di R è agevolata dal calcolo integrale.

$$R = 1 - 2\int_0^1 L(p)\,dp$$

$$L(p;d_1,d_2) = d_1 p + (1-d_1)p^{d_2};\quad 0 \le d_1 \le 1;\ d_2 \ge 1$$

$$R = 2\left(\frac{1}{2} - Area\ \ B\right) = 1 - 2\int_0^1 d_1 p + (1-d_1)p^{d_2}\,dp = 1 - d_1 p^2\Big|_0^1 - 2\frac{1-d_1}{1+d_2}p^{1+d_2}\Big|_0^1 =$$

$$= 1 - d_1 - 2\frac{(1-d_1)}{(1+d_2)} = \frac{(1-d_1)}{(1+d_2)}(1+d_2-2)$$



Una volta determinata la combinazione di parametri che meglio descrive la curva, il calcolo degli indici segue immediatamente.

Esempio

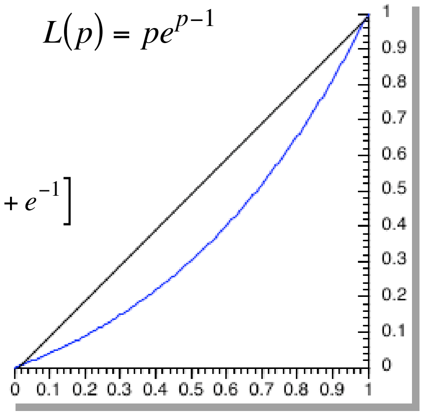
La determinazione di R è agevolata dal calcolo integrale.

$$L(p) = pe^{p-1}$$

$$R = 1 - 2\int_0^1 pe^{p-1}dp$$

$$= 1 - 2\left[(p-1)e^{p-1}\Big|_0^1\right] = 1 - 2\left[0 + e^{-1}\right]$$

$$= 1 - \frac{2}{e} = 0.2642$$



$$\int pe^{p-1}dp = e^{p-1}p - \int e^{p-1}dp = e^{p-1}p - e^{p-1} = e^{p-1}(p-1)$$

Altri Indici lineari di concentrazione

$$C_w = \int_0^1 w(p)g(p)\,dp; \qquad \int_0^1 w(p)\,dp = 0$$

Principio di Pigou-Dalton $w'(p) > 0$

Principio di Kolm $w''(p) < 0$

Gli indici variano tra zero ed uno.

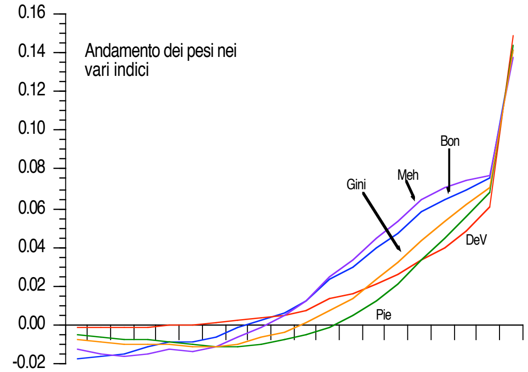
Assumono il valore nullo se la variabile è distribuita equamente tra le unità; il valore uno è raggiunto se una sola modalità è non negativa.

Sono standardizzati

Diminuiscono se si somma ad ogni valore una costante positiva.

Casi particolari

Indice	*Pesi* $W(p)$
Bonferroni	$Log(p) + 1$
De Vergottini	$Log\left(\frac{1}{1-p}\right) - 1$
Piesch	$\frac{3p^2 - 1}{2}$
Mehran	$1 - 3(1-p)^2$
Gini	$2p - 1$



i pesi del Gini e del Piesch sono minori per le modalità centrali;

il De Vergottini assegna pesi crescenti alle modalità più grandi con poca differenza per quelle piccole e medio-piccole;

il Mehran attribuisce pesi maggiori alle modalità medio-grandi e nel Bonferroni le modalità minori hanno maggiore rilevanza.

L'indice di Piesch è quello in cui le modalità intermedie contribuiscono meno a definire il livello di disuguaglianza.

Indice di Bonferroni

Modalità raggruppate $\qquad$ *Modalità singole*

$$\frac{1}{k}\sum_{i=1}^{k-1}\left(\frac{p_i - q_i}{p_i}\right) \qquad \frac{1}{n}\sum_{i=1}^{n}\frac{x_{(i)}}{\mu}\left(1-\sum_{j=i}^{n}\frac{1}{j}\right)$$

Soddisfa il principio dei trasferimenti di Pigou-Dalton

E' di "sinistra" perché la sensibilità aumenta man mano che il trasferimento va verso unità più povere (principio di Kolm).

$$\Delta B = \frac{d}{(n-1)\mu} * \sum_{r=i}^{j-1}\frac{1}{r}$$

Esempio

Dichiarazioni ICIAP. Numero di contribuenti per superficie del negozio

Superficie		Negozi	μ_i	f_i	g	p	q	(p-q)/p	(p-q)/(1-p)	p/(1-p)
0	25	481	16.67	0.497	0.138	0.497	0.138	0.723	0.714	0.988
26	50	244	37.52	0.252	0.157	0.749	0.295	0.606	1.808	2.984
51	100	127	74.04	0.131	0.162	0.880	0.457	0.481	3.533	7.345
101	150	36	124.76	0.037	0.077	0.917	0.534	0.418	4.639	11.100
151	200	23	175.06	0.024	0.069	0.941	0.603	0.359	5.739	15.982
201	300	18	251.19	0.019	0.078	0.960	0.681	0.291	6.921	23.821
301	400	21	348.14	0.022	0.126	0.981	0.807	0.178	9.404	52.778
400	600	9	497.22	0.009	0.077	0.991	0.883	0.108	11.534	106.556
601	800	6	696.35	0.006	0.072	0.997	0.955	0.042	13.431	321.667
801	1000	3	867.33	0.003	0.045	1.000	1.000	0.000		
		968	60.10	1.000	1.000			3.205	57.723	543.219

$$B = \frac{3.205}{10} = 0.321$$

Indici Gastwirth-Kakwani

Questa classe include molte misure interessanti

$$Continue:\ GK_r = \int_a^\infty r(x)h(x)dx; \quad Discrete:\ GK_r = \sum_{i=1}^{k} r(x_i)f_i$$

$$r(x)\ \ è \begin{cases} Convessa \\ Omogenea\ di\ grado\ zero\ :\ r(ax) = r(x) \\ r(\mu) = 0 \end{cases}$$

oppure la media geometrica

- **Lunghezza della curva di Lorenz**
- **Indice quadratico di Bonferroni**
- **Varianza dei logaritmi**
- **Indici di Theil**
- **Indice di Herfindahl**

Sensibilità ai trasferimenti

Se r(x) è convessa allora l'indice verifica il principio di Pigou-Dalton.

La variazione di sensibilità ad un trasferimento infinitesimo positivo "d" da "x" ad un'altra con modalità inferiore è data da:

$$t(x) = r'(x) - r'(x-d)$$

Esempio:

$$r(x) = Ln^2\left(\frac{x}{G}\right),\ \ G =\ media\ geometrica,\quad x \le eG$$

$$t(x) = 2\frac{G}{x}Ln\left(\frac{x}{G}\right) - 2\frac{G}{x-d}Ln\left(\frac{x-d}{G}\right) = 2G\left[Ln\left(\frac{x}{x-d}\right)\right]$$

L'effetto è positivo fintanto che d≥0 cioè la sensibilità cresce man mano che il ricevente ha un valore maggiore (indice "rightist")

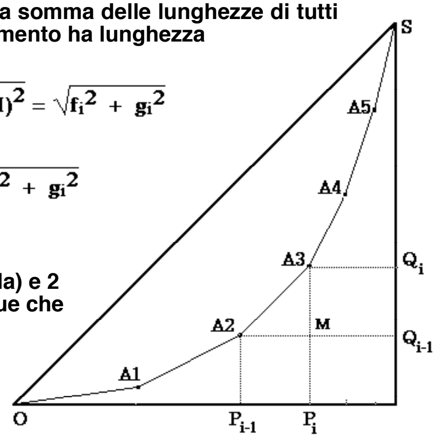
Lunghezza della curva

La concentrazione nulla e la massima si riflettono in posizioni estreme della spezzata di Lorenz. La lunghezza della spezzata può fornire perciò una base per misurare la concentrazione

La lunghezza "L" della spezzata è pari alla somma delle lunghezze di tutti i segmenti componenti. Il generico segmento ha lunghezza

$$A_2A_3 = \sqrt{(A_2M)^2 + (A_3M)^2} = \sqrt{f_i^2 + g_i^2}$$

La lunghezza "L" sarà quindi $L = \sum_{i=1}^{k} \sqrt{f_i^2 + g_i^2}$



l'indice L varia tra √2 (concentrazione nulla) e 2 (concen trazione massima). Ne consegue che L'indice

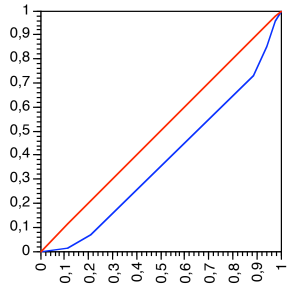
$$\lambda = \frac{L - \sqrt{2}}{2 - \sqrt{2}}$$

E' un indice normalizzato

Calcolo di "λ"

La società CEMENTIR SpA presentava la seguente ripartizione del capitale sociale. Valutare la concentrazione

Classi di azioni		azionisti n_i	azioni $\mu_i n_i$	f_i	g_i	$\sqrt{f_i^2 + g_i^2}$
fino a	100	21	935	0.1135	0.0147	0.1145
101	200	18	3380	0.0973	0.0533	0.1109
201	400	125	41880	0.6757	0.6600	0.9446
401	800	10	7963	0.0541	0.1255	0.1366
801	1000	7	6522	0.0378	0.1028	0.1095
1001	e più	4	2770	0.0216	0.0437	0.0487
		185	63450	1.0000	1.0000	1.4649



La concentrazione non è elevata perché gran parte degli azionisti si colloca nella classe 201-400. infatti l'indice normalizzato vale

$$\lambda = \frac{1.4649 - \sqrt{2}}{2 - \sqrt{2}} = 0.0865$$

Proprietà di λ

L'indice rientra nella classe di Gastwirth-Kakwani

$$r(x) = \sqrt{1 + (x/\mu)^2} - \sqrt{2};$$

$$r'(x) = x \Big/ \sqrt{x^2 + \mu^2} \geq 0 \quad se \quad x \geq 0; \quad r''(x) = \mu^2 \Big/ \sqrt{(x^2 + \mu^2)^3} \geq 0$$

$$t'(x) = \left(\frac{\mu}{\sqrt{2}}\right) \left\{ 1 \Big/ \sqrt{(x^2 + \mu^2)^3} - 1 \Big/ \sqrt{\left[(x-d)^2 + \mu^2\right]^3} \right\} \leq 0$$

che è negativa per ogni "x" e quindi l'indice di concentrazione λ reagisce a tutti i trasferimenti, ma con sensibilità crescente al diminuire della "x" ovvero l'effetto è più rilevante quando la "x" è piccola (indice "leftist").

Concentrazione industriale

La concentrazione industriale è collegata alle varie forme di mercato: monopolio, oligopolio, concorrenza perfetta, etc.

Lo studio della concentrazione industriale si esplica nell'esame della ripartizione del carattere "ampiezza" fra le unità di un rilevazione di imprese.

L' ampiezza può essere espressa in termini di dipendenti, di profitti, di vendite, di fatturato, quotazioni azionarie.

Una misura utile di concentrazione industriale dovrebbe:

1) Dipendere dalle dimensioni delle imprese;

2) Dal numero di imprese ("n" compare esplicitamente)

3) Dalla situazione di mercato (sensibilità)

Indice entropico di Theil

E' la media ponderata (con pesi g_i) delle differenze logaritmiche tra quote relative di carattere e frequenze relative di unità

$$T = \sum_{i=1}^{k} g_i Ln\left(\frac{g_i}{f_i}\right) = \sum_{i=1}^{k} g_i\left[Ln(g_i) - Ln(f_i)\right] = \sum_{i=1}^{k} g_i Ln(g_i) - \sum_{i=1}^{k} g_i Ln(f_i)$$

La scala logaritmica ha un effetto telescopico: avvicina i valori grandi ed allontana i valori piccoli

Per calcolare l'indice è necessario che nessuna unità abbia modalità zero o negativa.

Tuttavia si può dimostrare che il termine: $g_i Ln(g_i)$ tende a zero man mano che g_i diventa piccola

Proprietà dell'indice di Theil

Secondo tale indice una distribuzione è più concentrata di un'altra se rispetto a questa ha una minore entropia cioè ha meno disordine

- **L'indice T assume valore minimo zero (massima entropia) quando tutte le unità hanno la stessa modalità.**
- **assume valore massimo Ln(n) (entropia nulla) quando tutte una sola unità è diversa da zero.**
- **E' standardizzato perché coinvolge solo delle frazioni**
- **Diminuisce per variazioni additive**

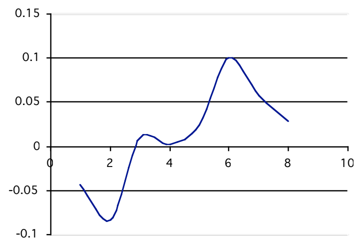
Calcolo del Theil

Superficie di vendita dei supermercati in Liguria

Superficie		Supermercati	μ_i	f_i	g_i	g_i*Ln(μ_i/μ)
0	200	41	133.33	0.118	0.051	-0.043
200	300	159	246.28	0.458	0.363	-0.084
300	350	74	323.17	0.213	0.222	0.009
350	400	38	373.30	0.110	0.132	0.024
400	500	17	446.32	0.049	0.070	0.025
500	750	9	1016.20	0.026	0.085	0.100
750	1000	7	864.58	0.020	0.056	0.057
1000	1500	2	1166.67	0.006	0.022	0.029
		347	310.79	1.000	1.000	0.117

L'indice di Theil, T=0.117 deve essere comparato con il suo massimo in rilevazioni di questa ampiezza: Ln(347)=5.85.

Ne consegue che in Liguria sussiste una situazione di sostanziale equità di ripartizione della superficie di vendita tra i supermercati.



Altre proprietà del Theil

L'indice rientra nella classe di Gastwirth-Kakwani

$$r(x) = \frac{x}{\mu} Ln\left(\frac{x}{\mu}\right) \quad convessa \quad poiché \quad r''(x) = \frac{1}{x} > 0$$

$$t(x) = \frac{1}{\mu} Ln\left(\frac{x}{x-d}\right)$$

che è positiva per ogni "x" e quindi l'indice di Theil reagisce ai trasferimenti con sensibilità decrescente al diminuire della "x" ovvero l'effetto è più rilevante quando la "x" è ai livelli alti

L'indice di Herfindahl

E’ dato dalla media aritmetica dei quadrati delle singole ampiezze rapportate alla media aritmetica:

$$H_1 = \sum_{i=1}^{k}\left(\frac{\mu_i}{\mu}\right)^2 f_i \;=\sum_{i=1}^{k}\frac{g_i^2}{f_i}$$

l’indice vale uno nel caso di un mercato formato da imprese di eguale ampiezza; vale invece “n” se il mercato è un monopolio

E’ legato al coefficiente di variazione:

$$CV^2 = \sum_{i=1}^{k}\left(\frac{\mu_i-\mu}{\mu}\right)^2 f_i = H_1 - 1$$

e quindi rimane invariato per trasformazioni moltiplicative e diminuisce per trasformazioni additive

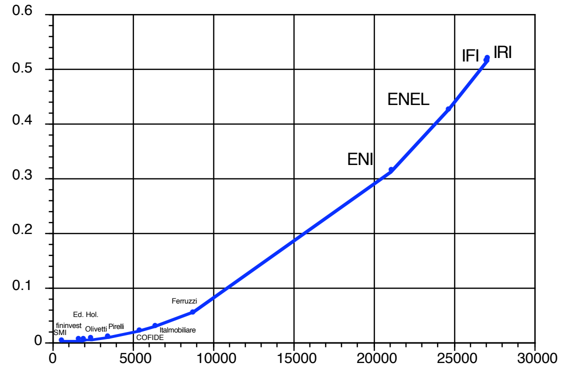
Secondo H una distribuzione è più concentrata di un'altra quando gli scarti relativi al quadrato dalla media aritmetica sono, in media, più piccoli.

Esempio

Le ”top 12” italiane.

Capitale netto (in miliardi di lire) dei principali gruppi italiani.

Gruppo	Capitale netto	$(1/n)(\mu_i/\mu)^2$
IFI	27182	0.5158
IRI	27078	0.5119
ENEL	24697	0.4258
ENI	21115	0.3112
Ferruzzi	8812	0.0542
Italmobiliare	6423	0.0288
COFIDE	5297	0.0196
Pirelli	3483	0.0085
Edizione Holding	2437	0.0041
Olivetti	2052	0.0029
fininvest	1861	0.0024
SMI	672	0.0003
Media	10925.75	1.8856



Minimo: 1
Massimo: 12
H=1.9

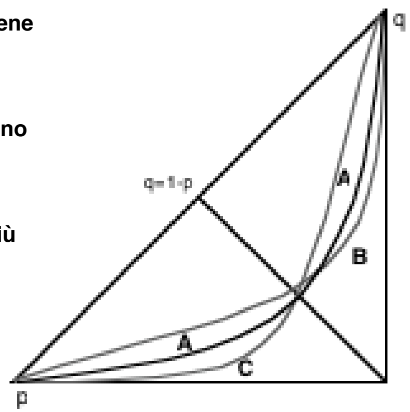
Confronto di curve Lorenz

La natura di “medie” degli indici di concentrazione impedisce di specificare bene la distribuzione

La B e C hanno lo stesso Gini, ma esprimono due situazione distinte:

Nella B vi è un ristretto numero di unità più “abbienti” e numerose unità “povere” che quasi si equidistribuiscono il resto dell’ammontare.

Nella C c’è un folto gruppo di “poveri” cui compete gran parte della variabile e poche unità “ricche” che si ripartiscono in modo uniforme l’ammontare non assegnato alle altre.

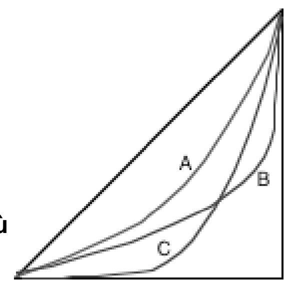


Confronto di curve Lorenz/2

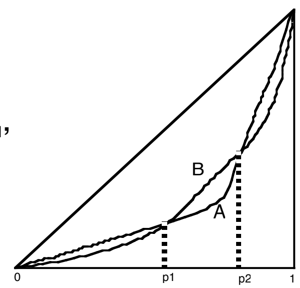
Se una curva è tutta inclusa in un’altra allora una situazione è meno ineguale dell’altra.

Se le curve si intersecano non sarà possibile stabilirlo univocamente:

Se si dà più importanza ai livelli minori allora è la C che è più ineguale, se invece pesano di più i livelli maggiori lo sarà la B.



Nella 2ª figura il confronto dà esito univoco solo ragionando per zone separate: $(0, p_1)$, (p_1, p_2) e $(p_2, 1)$. Nel primo e nel terzo è meno concentrata la “A”; in quella centrale è meno concentrata la “B”.



Classificazione di Fellman

Questo autore dimostra che se due variabili sono legate dalla relazione funzionale y=t(x) allora:

$$L_y(p) = L_x(p) \quad per \quad 0 \le p \le 1 \quad se \quad \frac{t(x)}{x} = costante;$$

$$L_y(p) \le L_x(p) \quad per \quad 0 \le p \le 1 \quad se \quad \frac{t(x)}{x} = monotona \quad crescente$$

$$L_y(p) \ge L_x(p) \quad per \quad 0 \le p \le 1 \quad se \quad \frac{t(x)}{x} = monotona \quad decrescente;$$

Una politica di tassazione progressiva cioè con un rapporto t(x)/x monotono crescente, ottiene minore ineguaglianza a tutti i livelli.