Statistica, anno LXIV, n. 4, 2004

OSSERVAZIONI SULL'INDICE DI COMPENSAZIONE DISTRIBUTIVA

Walter Maffenini

1. INTRODUZIONE

Negli studi sociali ed economici ci si trova, in alcuni casi, ad analizzare fenomeni scomponibili in più elementi ovvero fenomeni che risultano essere determinati dalla somma di più elementi. Si considerino i consumi delle famiglie: i consumi totali possono essere scomposti in varie categorie di consumi, data questa situazione la composizione dei consumi delle famiglie di una popolazione può variare tra due situazioni estreme:

a) il valore di ciascun tipo di consumo delle famiglie è totalmente determinato dal valore del consumo totale così che le famiglie che hanno il più basso livello di consumo totale hanno anche il più basso livello di consumo per ciascuna delle tipologie di consumi parziali, mentre le famiglie che hanno il più alto livello di consumo totale hanno anche il più alto livello di consumo per ciascuna tipologia di consumi parziali. In questo caso si può dire che non esiste alcun tipo di compensazione dato che nessuno dei consumi parziali sostituisce (compensa) uno qualsiasi degli altri consumi parziali (fissata la posizione del valore del consumo totale),

b) tutte le famiglie hanno lo stesso livello di spesa per i consumi totali ovvero i consumi totali d'ogni famiglia sono uguali in valore. In questo caso si può affermare che c'è massima compensazione dato che i consumi di una categoria sostituiscono (compensano) i consumi di un'altra categoria, così che il consumo totale è il medesimo (quello medio) per ciascuna famiglia della popolazione.

Recentemente è stato proposto un nuovo indice (Zenga, 2003) che permette di misurare l'aspetto di un fenomeno del tipo ora descritto. Questo indice, che è stato indicato con il nome di *indice normalizzato di compensazione,* è basato sulla scomposizione della differenza media della somma Y di k variabili $X_1, ..., X_j, ..., X_k$ ed assume valore 0 nel caso di assenza di compensazione e valore 1 nel caso di massima compensazione.

Si abbia una popolazione finita di N unità, su questa popolazione è stata rilevata la variabile Y composta da k variabili $X_1, ..., X_j, ..., X_k$ così che $Y = X_1 +...+ X_j + ...+ X_k$. Radaelli e Zenga (2002) hanno dimostrato che la differenza media di Y può essere scomposta nel seguente modo:

$$\Delta(Y) = \sum_{j=1}^{k} \Delta(X_j) - \frac{4}{N(N-1)} \sum_{j=1}^{k} \sum_{i=1}^{N} (x_{(ij)} - x_{ij})\, i \tag{1}$$

Essendo state ordinate le N unità della popolazione in senso non decrescente, secondo i valori della variabile Y, si ha che i rappresenta la i-ma posizione di una unità statistica secondo tale ordinamento, x_{ij} indica il valore della variabile X_j presso tale unità statistica, mentre $x_{(ij)}$ indica il valore della variabile X_j che all'interno di tale variabile assume la posizione i-ma nell'ordinamento decrescente.

Pertanto i valori x_{ij} ed i valori $x_{(ij)}$ sono considerati elementi di due matrici, la prima definita matrice reale poiché ordina in senso non decrescente le unità secondo i valori della variabile somma, la seconda definita *matrice di uniforme ordinamento (cograduazione)* poiché ciascuna delle sue colonne, riferite ad una variabile distinta, presentano l'ordinamento non decrescente di tali valori. Pertanto mentre nella matrice reale ogni riga rappresenta una unità statistica, nella matrice di uniforma ordinamento ogni riga può essere riferita a diverse unità statistiche.

La quantità:

$$Q = \frac{4}{N(N-1)} \sum_{j=1}^{k} \sum_{i=1}^{N} (x_{(ij)} - x_{ij})\, i \tag{2}$$

misura l'allontanamento della matrice dei dati reali dalla matrice di uniforme ordinamento.

E' stato, inoltre, dimostrato che $\sum_{j=1}^{k} \sum_{i=1}^{N} (x_{(ij)} - x_{ij})\, i \geq 0$; l'uguaglianza vale se $x_{(ij)} = x_{ij}, \forall (i,j)$ ovvero quando si è in presenza di uniforme ordinamento delle k variabili (si definisce questa situazione come assenza di compensazione), in questo caso Q è pari a zero. Nel caso in cui Q assume il valore zero, $\Delta(Y)$ è uguale a $\sum_{j=1}^{k} \Delta(X_j)$ e raggiunge il suo valore massimo.

Se ciascuna determinazione della variabile somma Y è uguale a $\mu(Y)$, ossia quando si è in presenza di massima compensazione, si ottiene che $\Delta(Y)$ è pari a zero (il suo minimo valore) e in questo caso Q è uguale a $\sum_{j=1}^{k} \Delta(X_j)$.

In conclusione: tenendo fisse le distribuzioni marginali delle variabili X_1, ..., X_j, ..., X_k la quantità $Q = \frac{4}{N(N-1)} \sum_{j=1}^{k} \sum_{i=1}^{N} (x_{(ij)} - x_{ij})\, i$ assume i valori compresi nell'intervallo $[0 \div \sum_{j=1}^{k} \Delta(X_j)]$. Se Q è uguale a zero vi è assenza di compensazione tra le variabili; se Q è uguale a $\sum_{j=1}^{k} \Delta(X_j)$ vi è massima compensazione tra le variabili.

Utilizzando la (1) si ottiene $\sum_{j=1}^{k} \Delta(X_j)$ secondo la seguente espressione:

$$\sum_{j=1}^{k} \Delta(X_j) = \Delta(Y) + \frac{4}{N(N-1)} \sum_{j=1}^{k} \sum_{i=1}^{N} (x_{(ij)} - x_{ij})\, i \tag{3}$$

da cui si ricava l'indice di compensazione normalizzato:

$$C = \frac{\dfrac{4}{N(N-1)} \sum_{j=1}^{k} \sum_{i=1}^{N} (x_{(ij)} - x_{ij})\, i}{\sum_{j=1}^{k} \Delta(X_j)} = 1 - \frac{\Delta(Y)}{\sum_{j=1}^{k} \Delta(X_j)} \tag{4}$$

L'indice C assume i valori compresi nell'intervallo *(0 ÷ 1),* in particolare:

1) quando C è uguale a 0 ci si trova nel caso di assenza di compensazione tra le k variabili, questo accade quando esiste uniforme ordinamento tra le variabili ovvero le k variabili hanno lo stesso ordinamento: $\sum_{j=1}^{k} \Delta(X_j) = \Delta(Y)$.

2) quando C è uguale a 1, fra le k variabili vi è la massima compensazione, questo accade perché $\Delta(Y) = 0$ cioè la Y è costantemente uguale a $\mu(Y)$.

Nelle situazioni non estreme l'indice assume un valore nell'intervallo *(0 ÷ 1).*

Le finalità di questo lavoro sono le seguenti:

i) proporre una scomposizione dell'indice di compensazione,

ii) evidenziare il ruolo che può assumere il valore dell'indice nel caso di indipendenza distributiva delle k variabili, confrontando il valore dell'indice nel caso di dati reali ed il valore dello stesso indice nell'ipotesi dell'indipendenza distributiva,

iii) applicare l'indice a due situazioni reali: a) la distribuzione congiunta del reddito totale delle famiglie italiane scomposto in: reddito da lavoro dipendente, reddito da lavoro autonomo, reddito da trasferimenti, reddito da capitale; b) la distribuzione congiunta dei voti ottenuti da 653 laureati della facoltà di Economia, dell'Università di Milano-Bicocca, agli esami di quattro insegnamenti tipici di questo tipo di facoltà: Economia aziendale, Economia politica, Matematica generale, Istituzioni di diritto privato.

2. LA SCOMPOSIZIONE DELL'INDICE DI COMPENSAZIONE DISTRIBUTIVA

Si consideri l'indice normalizzato di compensazione distributiva:

$$C = \frac{\dfrac{4}{N(N-1)} \sum_{j=1}^{k} \sum_{i=1}^{N} (x_{(ij)} - x_{ij}) i}{\sum_{j=1}^{k} \Delta(X_j)} . \tag{5}$$

Moltiplicando e dividendo ciascun addendo del numeratore per $\Delta(X_j)$ si ottiene:

$$C = \frac{\sum_{j=1}^{k} \left[\dfrac{4}{N(N-1)\Delta(X_j)} \sum_{i=1}^{N} (x_{(ij)} - x_{ij})\, i \right] \Delta(X_j)}{\sum_{j=1}^{k} \Delta(X_j)} . \tag{6}$$

La quantità:

$$C_j = \frac{4}{N(N-1)\Delta(X_j)} \sum_{i=1}^{N} (x_{(ij)} - x_{ij})\, i \tag{7}$$

può essere interpretata come l'apporto fornito dalla variabile X_j alla compensazione, per $j = 1,2,...,k$. L'espressione di C può allora essere scritta anche nel modo seguente:

$$C = \frac{\sum_{j=1}^{k} C_j \Delta(X_j)}{\sum_{j=1}^{k} \Delta(X_j)} \tag{8}$$

così che l'indice di compensazione risulta essere la media aritmetica ponderata degli apporti della generica variabile X_j alla compensazione, con pesi pari alla differenza media di Gini della stessa variabile X_j. Il contributo della variabile X_j alla compensazione dipende così dal suo grado di concordanza o graduazione con la variabile somma Y: tanto più è bassa la cograduazione di una variabile X_j con la variabile somma Y, tanto più elevato è il suo contributo all'indice normalizzato di compensazione.

Si può, inoltre, calcolare la percentuale di apporto della generica variabile X_j all'indice di compensazione. Dato che:

$$C = \frac{\sum_{j=1}^{k} C_j \Delta(X_j)}{\sum_{j=1}^{k} \Delta(X_j)} =$$

$$= \frac{C_1 \Delta(X_1)}{\sum_{j=1}^{k} \Delta(X_j)} + \frac{C_2 \Delta(X_2)}{\sum_{j=1}^{k} \Delta(X_j)} + \ldots + \frac{C_j \Delta(X_j)}{\sum_{j=1}^{k} \Delta(X_j)} + \ldots + \frac{C_k \Delta(X_k)}{\sum_{j=1}^{k} \Delta(X_j)} \tag{9}$$

la proporzione di apporto, della generica variabile X_j, indicata con A_j, risulta:

$$A_j = \frac{\dfrac{C_j \Delta(X_j)}{\sum_{j=1}^{k} \Delta(X_j)}}{C} = \frac{\dfrac{C_j \Delta(X_j)}{\sum_{j=1}^{k} \Delta(X_j)}}{\dfrac{\sum_{j=1}^{k} C_j \Delta(X_j)}{\sum_{j=1}^{k} \Delta(X_j)}} = \frac{C_j \Delta(X_j)}{\sum_{j=1}^{k} C_j \Delta(X_j)}. \tag{10}$$

3. L'INDICE DI COMPENSAZIONE E L'INDIPENDENZA DISTRIBUTIVA

Tradizionalmente nello studio delle relazioni statistiche fra caratteri riveste un ruolo importante l'analisi della condizione di indipendenza distributiva. In questo paragrafo si mostra come calcolare l'indice normalizzato di compensazione in presenza di indipendenza distributiva. In questo modo si potrà apprezzare meglio il valore reale dell'indice di compensazione perché l'interpretazione di quest'ultimo si avvarrà del confronto non solo con i valori estremi ma anche del confronto con il valore riferito a questa ipotesi.

Nel caso in cui N sia elevato è possibile rappresentare gli N valori con una distribuzione di frequenza multivariata. Si supponga[1] che k sia pari a 4 e che le variabili $X_1,\ldots,X_4$ siano discrete con numero di modalità, rispettivamente, pari ad A, B, C, D. I valori che possono essere assunti dalle variabili sono: $x_{11},\ldots,x_{1a},\ldots,x_{1A}$, per X_1; $x_{21},\ldots,x_{2b},\ldots,x_{2B}$, per X_2; $x_{31},\ldots,x_{3c},\ldots,x_{3C}$, per X_3 e $x_{41},\ldots,x_{4d},\ldots,$ x_{4D} per X_4. Con n_{abcd} si indica la frequenza congiunta di: $X_1 = x_{1a}$; $X_2 = x_{2b}$; $X_3 = x_{3c}$; $X_4 = x_{4d}$.

Con $n(x_{1a})$ si indica la frequenza marginale di X_1 in corrispondenza di $X_1 = x_{1a}$:

[1] Pur essendo consapevoli che la notazione matriciale consentirebbe una trattazione generale, si è preferito utilizzare una simbologia semplice e quindi limitata a quattro variabili, ritenendo che il nuovo indice proposto debba essere divulgato, dato il profilo applicativo, in contesti dove non necessariamente siano noti alcuni aspetti matematici.

$$n(x_{1a}) = \sum_{b=1}^{B}\sum_{c=1}^{C}\sum_{d=1}^{D} n_{abcd} \tag{11}$$

Analogo significato avranno i simboli: $n(x_{2b})$, $n(x_{3c})$, $n(x_{4d})$.

Con $\hat{n}_{abcd}$ si indica la frequenza congiunta fra le quattro variabili nell'ipotesi di indipendenza distributiva, quindi:

$$\hat{n}_{abcd} = \frac{n(x_{1a})}{N}\ \frac{n(x_{2b})}{N}\ \frac{n(x_{3c})}{N}\ \frac{n(x_{4d})}{N}\ N =$$

$$= \frac{n(x_{1a})\ n(x_{2b})\ n(x_{3c})\ n(x_{4d})}{N^3} \tag{12}$$

E' possibile associare ad ogni configurazione $(x_{1a}, x_{2b}, x_{3c}, x_{4d})$:

I) la somma delle quattro variabili: $y_{abcd} = x_{1a} + x_{2b} + x_{3c} + x_{4d}$,

II) la frequenza congiunta reale: n_{abcd},

III) la frequenza congiunta nell'ipotesi di indipendenza distributiva: $\hat{n}_{abcd}$.

E' noto che le frequenze teoriche $\hat{n}_{abcd}$ non modificano le frequenze marginali, si avrà, così, che:

$$\hat{n}(x_{1a}) = \sum_{b=1}^{B}\sum_{c=1}^{C}\sum_{d=1}^{D} \hat{n}_{abcd} = n(x_{1a})$$

ed anche che:

$$\hat{n}(x_{2b}) = n(x_{2b});\ \hat{n}(x_{3c}) = n(x_{3c})\ \text{ e }\ \hat{n}(x_{4d}) = n(x_{4d}).$$

Si possono calcolare, per le distribuzioni marginali dei quattro caratteri, le differenze medie: $\Delta(X_1)$, $\Delta(X_2)$, $\Delta(X_3)$, $\Delta(X_4)$ e si può calcolare anche $\Delta(Y)$ sulla distribuzione reale: $\{y_{abcd}; n_{abcd}\}$. Si può infine calcolare la differenza media $\hat{\Delta}(Y)$ della distribuzione di Y nell'ipotesi di indipendenza distributiva: $\{y_{abcd}; \hat{n}_{abcd}\}$.

L'indice normalizzato di compensazione, nel caso di indipendenza distributiva, è:

$$\hat{C} = 1 - \frac{\hat{\Delta}(Y)}{\sum_{j=1}^{k} \Delta(X_j)}. \tag{13}$$

Se si ottiene che $\Delta(Y) < \hat{\Delta}(Y)$ fra i quattro caratteri vi è maggior compensazione di quella che si avrebbe nell'ipotesi di indipendenza distributiva, mentre se $\Delta(Y) > \hat{\Delta}(Y)$ fra i quattro caratteri vi è minor compensazione di quella che si avrebbe nell'ipotesi di indipendenza distributiva.

4. LA COMPENSAZIONE TRA LE COMPONENTI DEL REDDITO TOTALE

Si illustrerà ora il significato di compensazione e del relativo indice, ricorrendo ad un esempio numerico con dati artificiali. Si è deciso di seguire questa via, pur presentando nella seconda parte del paragrafo un utilizzo dell'indice di compensazione per lo studio della composizione del reddito reale delle famiglie italiane nell'anno 2000, in un'ottica divulgativa. Si ritiene infatti che per la sua novità l'interpretazione e l'analisi del concetto di compensazione possa essere facilitata dal ricorso ad un esempio numerico che simuli la situazione di una popolazione composta da poche unità.

Si abbiano sei famiglie: F_1, F_2, F_3, F_4, F_5 *e* F_6 e quattro tipi di reddito: X_1 reddito da lavoro dipendente, X_2 reddito da lavoro autonomo, X_3 reddito da trasferimenti, X_4 reddito da capitale e Y_F reddito disponibile familiare (in migliaia di euro) dove $Y_F = X_1 + X_2 + X_3 + X_4$.

TAVOLA 1

Redditi annuali di sei famiglie (F), in migliaia di euro, scomposti secondo le quattro componenti (X)

	X_1	X_2	X_3	X_4	Y_F
F_1	15	13	3	20	51
F_2	9	1	1	10	21
F_3	6	3	7	15	31
F_4	10	0	0	0	10
F_5	3	10	1	2	16
F_6	2	6	2	11	21
T	45	33	14	58	150

Nella tavola 1 sono riportati i redditi annuali delle sei famiglie, in migliaia di euro, scomposti secondo questi quattro tipi di reddito.

TAVOLA 2

Matrice dei dati della tavola 1 ordinati secondo Y_F

	X_1	X_2	X_3	X_4	Y_F
$F_{(1)}$	10	0	0	0	10
$F_{(2)}$	3	10	1	2	16
$F_{(3)}$	9	1	1	10	21
$F_{(4)}$	2	6	2	11	21
$F_{(5)}$	6	3	7	15	31
$F_{(6)}$	15	13	3	20	51
T	45	33	14	58	150

Si è, successivamente, provveduto a permutare le righe della tavola 1 in modo che i valori delle *Y* risultino ordinate in senso non decrescente, ottenendo la tavola 2.

Se i valori della tavola 2 vengono ordinati, per ciascuna delle 4 colonne, in senso non decrescente in modo che:

$$x_{(1j)} \leq x_{(2j)} \leq \ldots \leq x_{(ij)} \leq \ldots \leq x_{(Nj)} \qquad j = 1,2,\ldots,k.$$

si ricava la tavola 3 ovvero la *matrice di uniforme ordinamento*.

TAVOLA 3

Matrice di uniforme ordinamento

	X_1	X_2	X_3	X_4	Y_F
$F^*_{(1)}$	2	0	0	0	2
$F^*_{(2)}$	3	1	2	2	7
$F^*_{(3)}$	6	1	10	10	20
$F^*_{(4)}$	9	2	11	11	28
$F^*_{(5)}$	10	3	15	15	38
$F^*_{(6)}$	15	7	20	20	55
T	45	33	14	58	150

Si noti che aver effettuato l'ordinamento dei valori delle sei famiglie non comporta, necessariamente, che i valori presenti in una riga della matrice appartengano alla stessa unità statistica ovvero alla stessa famiglia.

Sommando per riga i valori della matrice di uniforme ordinamento si ottengono gli ipotetici valori:

$$y_i^* = x_{(i1)} + x_{(i2)} + ... + x_{(ij)} + ... + x_{(ik)} \qquad i = 1,2,..., N$$

i valori $y_{(i)}^*$ che si sono ottenuti sono anch'essi ordinati in senso non decrescente.

La tabella di valori che si è ricavata (tabella di uniforme ordinamento) presenta una situazione di assenza di compensazione. Si può, infatti, notare che l'ipotetica famiglia $F_{(1)}^*$ che occupa il rango più basso per il reddito totale occupa il rango più basso per ciascuno dei quattro tipi di reddito che compongono il reddito totale, mentre l'ipotetica a famiglia $F_{(6)}^*$ che occupa il rango più alto per il reddito totale occupa il rango più alto anche per ciascuno dei quattro tipi di reddito che compongono il reddito totale. L'assenza di compensazione è quindi facilmente interpretabile dato che il rango di ciascun tipo di reddito è determinato dal rango del reddito totale ovvero l'opportunità di formazione del reddito delle famiglie si manifesta, per una determinata famiglia, con la stessa intensità quale che sia il tipo di reddito.

Una situazione, tra le molte, di massima compensazione è, invece, quella descritta nella tavola 4, dove i redditi parziali si compensano tra loro così che ogni famiglia ha lo stesso reddito totale che corrisponde al reddito medio totale delle sei famiglie.

Massima compensazione significa, ad esempio, che una famiglia con reddito basso da lavoro dipendente, riesce ad ottenere altri tipi di reddito che "compensano" tale valore così da avere il reddito totale uguale al reddito medio dell'insieme delle famiglie considerate.

TAVOLA 4

Condizione di massima compensazione dei redditi

	X_1	X_2	X_3	X_4	Y_F
F'_1	10	13	2	0	25
F'_2	15	1	7	2	25
F'_3	3	10	1	11	25
F'_4	2	0	3	20	25
F'_5	9	6	0	10	25
F'_6	6	3	1	15	25
T	45	33	14	58	150

Il valore dell'indice di compensazione per i redditi delle sei famiglie, riportati nella tavola 1, calcolato utilizzando la (4) è il seguente:

$$C = 1 - \frac{\Delta(Y)}{\sum_{j=1}^{4} \Delta(X_j)} = 1 - \frac{16,\overline{6}}{24,4} = 0{,}3169.$$

La compensazione dei redditi è pari al 31,69% del massimo; esiste quindi una certa compensazione tra i vari tipi di reddito come si può notare considerando i dati della tavola 2. La famiglia $F_{(3)}$ che ha basso reddito da lavoro autonomo e da trasferimenti, ha redditi, relativi, più elevati da lavoro dipendente e da capitale; mentre la famiglia $F_{(4)}$ che ha basso reddito da lavoro dipendente e da trasferimenti, ha redditi, relativi, più elevati da capitale e da lavoro autonomo.

Si desidera infine conoscere il contributo delle singole componenti del reddito all'indice di compensazione.

L'utilizzo della (9) per i dati dell'esempio in questione porta al seguente risultato:

$$C = \frac{C_1\Delta(X_1)}{\sum_{j=1}^{4}\Delta(X_j)} + \frac{C_2\Delta(X_2)}{\sum_{j=1}^{4}\Delta(X_j)} + \frac{C_3\Delta(X_3)}{\sum_{j=1}^{4}\Delta(X_j)} + \frac{C_4\Delta(X_4)}{\sum_{j=1}^{4}\Delta(X_j)} =$$

$$= \frac{0{,}6966 - 5{,}9\overline{3}}{24{,}4} + \frac{0{,}4842 - 6.\overline{3}}{24{,}4} + \frac{0{,}1905 - 2{,}8}{24{,}4} + 0 =$$

$$= 0{,}1694 + 0{,}1257 + 0{,}0219 + 0 = 0{,}317.$$

Si ricava, infine, utilizzando la (10), il contributo, in termini percentuali, che ciascuna delle quattro variabili, che compongono il reddito totale ha nella compensazione: X_1, reddito da lavoro dipendente, 53,44%; X_2, reddito da lavoro autonomo, 39,65%; X_3, reddito da trasferimenti, 6,91%; X_4, reddito da capitale, 0%. Si è detto che il contributo di una variabile alla compensazione dipende dal suo livello di cograduazione con la variabile somma che in questo caso è il reddito totale, più è bassa la cograduazione di una componente del reddito con il reddito totale più elevato è il contributo all'indice normalizzato di compensazione *C*. In particolare il reddito da capitale non contribuisce alla compensazione del reddito totale dato che, come si può vedere dalla tavola 2 che riporta i dati dei redditi delle famiglie ordinati per reddito totale, questo tipo di reddito è perfettamente cograduato con il reddito totale.

Si considera ora l'ipotesi di indipendenza distributiva, dei quattro tipi di reddito, per verificare l'effetto della compensazione in questo particolare tipo di distribuzione.

Si sono calcolate (tavola 5) tutte le possibili combinazioni, delle quattro variabili, per ricavare la distribuzione di probabilità del reddito totale, sotto l'ipotesi di indipendenza.

Avendo a disposizione questa distribuzione si ricava, infine, $\hat{\Delta}(Y)$, la differenza media della distribuzione Y nell'ipotesi di indipendenza distributiva.

Si ottiene $\hat{\Delta}(Y) = 13{,}3381$ che risulta minore di $\Delta(Y)$ il che significa che fra le quattro componenti del reddito, dell'esempio proposto, vi è minor compensazione di quella che vi sarebbe nell'ipotesi di indipendenza distributiva.

TAVOLA 5

Distribuzione di probabilità della variabile Y, somma dei quattro tipi di redditi, nell'ipotesi di indipendenza distributiva

Y	p(Y)	Y	p(Y)	Y	p(Y)
2	0.0007716	20	0.0339506	38	0.0177469
3	0.0030864	21	0.0354938	39	0.0162037
4	0.0054012	22	0.0354938	40	0.0131172
5	0.0077160	23	0.0378086	41	0.0108024
6	0.0077160	24	0.0385802	42	0.0100308
7	0.0077160	25	0.0378086	43	0.0100308
8	0.0077160	26	0.0385802	44	0.0054012
9	0.0123456	27	0.0401234	45	0.0054012
10	0.0138888	28	0.0378086	46	0.0054012
11	0.0146604	29	0.0331790	47	0.0023148
12	0.0185185	30	0.0347222	48	0.0023148
13	0.0216049	31	0.0331790	49	0.0023148
14	0.0231481	32	0.0324074	50	0.0023148
15	0.0262345	33	0.0316358	51	0.0007716
16	0.0285493	34	0.0262345	52	0.0007716
17	0.0285493	35	0.0200617	55	0.0007716
18	0.0316358	36	0.0231481	Σ	1.0000000
19	0.0347222	37	0.0231481		

Ne consegue che l'indice normalizzato di compensazione, calcolato in questo caso, diventa:

$$\hat{C} = 1 - \frac{13{,}3381}{24{,}4} = 0{,}4534$$

che risulta, infatti, maggiore dell'indice di compensazione calcolato con i dati dell'esempio proposto.

4.1. *La compensazione per il reddito delle famiglie italiane*

Si illustrerà l'applicazione, ai redditi delle famiglie italiane per l'anno 2000, dell'indice normalizzato di compensazione C (Pampinella, 2003). I dati sono ricavati dall'indagine campionaria della Banca d'Italia sui bilanci delle famiglie italiane[2]. L'indagine, rivolta a 8001 famiglie italiane, è stata svolta con il metodo CAPI (*Computer Assisted Personal Interviewing*).

I risultati dell'indagine forniscono il reddito disponibile familiare[3] scomposto

[2] *I bilanci delle famiglie italiane nell'anno 2000,* Supplementi al Bollettino Statistico, Anno XII, numero 6, Banca d'Italia, 18 gennaio 2002.

[3] L'unità di misura per questa indagine sono ancora le lire (in migliaia).

secondo quattro componenti: *reddito da lavoro dipendente (X_1), reddito da lavoro autonomo (X_2) reddito da trasferimenti*[4] *(X_3), reddito da capitale*[5] *(X_4).*

L'indice di compensazione, calcolato con i dati desunti dall'indagine, è il seguente:

$$C = 1 - \frac{33156{,}10}{64369{,}95} = 0{,}4849$$

Il valore ottenuto evidenzia una buona compensazione tra le quattro componenti del reddito delle famiglie italiane. Questo significa che le famiglie italiane hanno una tendenza a compensare la presenza di basse entrate, di un certo tipo di reddito, cercando di ottenere entrate più elevate in altri tipi di reddito.

La scomposizione dell'indice di compensazione C, fatta per conoscere il contributo di ciascun tipo di reddito alla compensazione, fornisce il seguente risultato:

$$C = 0{,}1583_{x_1} + 0{,}0801_{x_2} + 0{,}1991_{x_3} + 0{,}0474_{x_4} = 0{,}4849$$

che completa le informazioni sulle caratteristiche della compensazione nel caso dell'analisi della formazione del reddito delle famiglie italiane. In termini percentuali, il contributo più elevato alla compensazione viene dai redditi da trasferimenti (valore pari al 41,06%), quello del reddito da lavoro dipendente ha un valore del 32,65%; il reddito da lavoro autonomo contribuisce per il 16,52% ed infine il reddito da capitale per il 9,77%.

E' stata inoltre studiata la compensazione, fra le quattro componenti del reddito familiare, nell'ipotesi di indipendenza distributiva fra le quattro componenti stesse. Per calcolare l'indice normalizzato di compensazione *C* si deve determinare la distribuzione del reddito totale, cioè la somma delle quattro variabili, nell'ipotesi di indipendenza. Non conoscendo la distribuzione di densità delle variabili la soluzione passa attraverso il calcolo delle somme delle modalità ed il calcolo dei prodotti delle frequenze corrispondenti a tutte le possibili combinazioni di modalità delle quattro variabili.

Per non dover calcolare 8.001^4 somme ed altrettanti prodotti, i calcoli sono stati fatti sui dati raggruppati in quindici classi. Ottenuta la distribuzione di frequenza del reddito totale, nell'ipotesi di indipendenza distributiva, è stata calcolata la differenza media $\hat{\Delta}(Y)$ della distribuzione *Y*, somma delle quattro componenti del reddito.

Per confrontare $\hat{\Delta}(Y)$ con $\Delta(Y)$, differenza media della somma delle quattro componenti del reddito dei dati reali delle famiglie italiane, si è dovuto calcolare quest'ultimo indice utilizzando i dati raggruppati in quindici classi per avere informazioni omogenee.

Si è così ottenuto il seguente risultato:

---

[4] Il reddito da trasferimenti è composto da: pensioni, assistenza economica, borse di studio, assegni alimenti e regali.

[5] Il reddito da capitale è composto da: reddito da fabbricati e reddito da capitale finanziario.

$$\Delta(Y) = 36088{,}43 < \hat{\Delta}(Y) = 40441{,}49$$

che evidenzia che, nel caso di indipendenza distributiva, la compensazione tra le quattro componenti del reddito è minore di quella che si è rilevata nei redditi reali, delle famiglie italiane, dell'anno 2000.

Ne consegue che anche l'indice di compensazione, calcolato nell'ipotesi di indipendenza distributiva:

$$\hat{C} = 1 - \frac{40441{,}49}{65838{,}17} = 0{,}3857$$

risulta minore di quello calcolato sui dati reali raggruppati in classi:

$$C' = 1 - \frac{36088{,}43}{65838{,}17} = 0{,}4519.$$

Le informazioni che si deducono dal confronto tra questi due indici di compensazione mostrano che la tendenza a compensare i redditi allontana l'indice di compensazione dal valore che si ottiene con l'ipotesi di indipendenza spostandolo, in modo significativo, verso la massima compensazione.

5. LA COMPENSAZIONE TRA I VOTI DI QUATTRO INSEGNAMENTI UNIVERSITARI

L'indice di compensazione è stato, inoltre, utilizzato per analizzare le distribuzioni dei voti, ottenuti in quattro insegnamenti del primo anno, dai laureati quadriennali della Facoltà di Economia dell'Università di Milano-Bicocca.

I laureati considerati sono 653 e corrispondono ai laureati quadriennali, della Facoltà di Economia, dalla prima sessione di tesi che si svolse nel mese di marzo del 1997 fino alla sessione di dicembre del 2002.

Gli insegnamenti analizzati sono i seguenti: *Economia aziendale* (X_1), *Economia politica* (X_2), *Matematica generale* (X_3), *Istituzioni di diritto privato* (X_4). Sono stati scelti questi insegnamenti perché sono quelli comuni a tutti gli indirizzi e a tutti corsi di laurea della facoltà di Economia.

Il primo interrogativo, da porsi, è quale possa essere il significato dell'indice di compensazione in questa condizione che si presenta con caratteristiche alquanto specifiche, rispetto a quelle della scomposizione dei redditi e dei consumi. Si considerino le due situazioni estreme: nel caso in cui non ci sia compensazione (indice C uguale a zero) significa che il laureato che ha la più bassa somma dei voti dei quattro insegnamenti, ha anche i più bassi voti per ciascuno dei quattro insegnamenti e così di seguito, in modo che chi ha la somma dei quattro insegnamenti più alta, ha anche il voto più alto per ciascuno dei quattro insegnamenti ovvero la posizione della serie ordinata di laureati per voto totale si riflette, anche, nella posizione dei voti per ciascun insegnamento. Questa situazione sta ad indicare che il livello del voto, per ciascuno dei quattro insegnamenti, dipende non tanto dalle attitudini quanto dalla capacità intellettuale complessiva dello studente e/o dal suo impegno nello studio.

Nel caso in cui la compensazione sia massima (indice C uguale ad uno) il voto totale per i quattro insegnamenti è il medesimo per ogni laureato. Si è in una situazione in cui gli studenti hanno conseguito voti che si compensano tra loro, così che, ad esempio, chi ha un voto superiore a quello medio in matematica e diritto privato, avrà un voto inferiore a quello medio in economia aziendale ed economia politica. In questa situazione, si può ritenere che ciascuno studente ha delle attitudini particolari che fanno sì che egli ottenga voti migliori in alcuni insegnamenti rispetto ad altri, inoltre queste differenti attitudini si compensano, in modo tale che il voto totale finale sia il medesimo per ogni studente.

L'indice di compensazione calcolato sulla distribuzione dei voti conseguiti dai laureati nei quattro insegnamenti considerati è pari a:

$$C = 1 - \frac{9{,}2065}{14{,}2884} = 0{,}3557.$$

La compensazione non è molto elevata si può allora affermare che le capacità intellettuali complessive degli studenti e il loro impegno nello studio hanno una indubbia influenza nella determinazione del livello dei voti di ciascun insegnamento.

La scomposizione dell'indice di compensazione C, per i voti dei laureati della Facoltà di Economia dell'Università di Milano-Bicocca, secondo il contributo alla compensazione dei quattro insegnamenti presenta il seguente risultato:

$$C = 0{,}0772_{x_1} + 0{,}0778_{x_2} + 0{,}1051_{x_3} + 0{,}0956_{x_4} = 0{,}3557.$$

TAVOLA 6

Distribuzione delle frequenze relative dei quattro insegnamenti secondo il voto

Insegnamento/Voto	X_1	X_2	X_3	X_4
18	0.02143	0.05513	0.25114	0.03369
19	0.00459	0.03215	0.07350	0.03522
20	0.02909	0.03981	0.12710	0.05819
21	0.02756	0.07503	0.10719	0.04747
22	0.02603	0.07810	0.08116	0.05819
23	0.04287	0.09494	0.08269	0.10260
24	0.08882	0.13782	0.11179	0.11638
25	0.13323	0.12098	0.05513	0.11179
26	0.15773	0.08728	0.04134	0.11485
27	0.16998	0.09494	0.02909	0.07810
28	0.16359	0.07810	0.01378	0.10719
29	0.00000	0.03215	0.00153	0.00153
30	0.12098	0.05513	0.02450	0.10107
33	0.01225	0.01837	0.00000	0.03369
Totale	1.00000	1.00000	1.00000	1.0000

In questa situazione il contributo dei voti degli insegnamenti alla compensazione in termini percentuali è del 29,56% per l'insegnamento di Matematica generale, del 26,86% per l'insegnamento di Diritto privato, del 21,87% per l'insegnamento di Economia politica ed infine del 21,71% per l'insegnamento di Economia aziendale.

Si è, infine, considerata la compensazione tra i voti dei quattro insegnamenti nell'ipotesi di indipendenza distributiva.

Per analizzare la compensazione nell'ipotesi di indipendenza distributiva si sono calcolate (tavola 6), per ciascuno dei quattro insegnamenti, le probabilità associate ad ogni voto (da 18 a 30, più 33 per la lode).

Si è, successivamente, ricavata la distribuzione di probabilità della variabile somma Y, sempre nell'ipotesi di indipendenza (tavola 7).

Utilizzando i dati di questa distribuzione si è calcolata la differenza media $\hat{\Delta}(Y)$ della somma dei voti con la quale si è ottenuto l'indice di compensazione $\hat{C}$.

TAVOLA 7

Distribuzione di probabilità della variabile somma Y nell'ipotesi di indipendenza distributiva

Y	p(Y)	Y	p(Y)	Y	p(Y)	Y	p(Y)
72	0.00001	87	0.01939	102	0.04446	117	0.00049
73	0.00002	88	0.02438	103	0.03878	118	0.00031
74	0.00005	89	0.02991	104	0.03303	119	0.00015
75	0.00011	90	0.03578	105	0.02750	120	0.00010
76	0.00021	91	0.04164	106	0.02259	121	0.00005
77	0.00037	92	0.04737	107	0.01772	122	0.00001
78	0.00064	93	0.05246	108	0.01399	123	0.00002
79	0.00108	94	0.05686	109	0.01064	124	0.00001
80	0.00174	95	0.05975	110	0.00777	125	0.000001
81	0.00270	96	0.06176	111	0.00578	126	0.000003
82	0.00405	97	0.06197	112	0.00405	127	0.000000
83	0.00581	98	0.06071	113	0.00275	128	0.000000
84	0.00824	99	0.05827	114	0.00192	129	0.000000
85	0.01209	100	0.05446	115	0.00124	Σ	1.000000
86	0.01496	101	0.04969	116	0.00077		

La differenza media nell'ipotesi di indipendenza distributiva è risultata inferiore alla differenza media dei voti calcolata su dati reali:

$$\hat{\Delta}(Y) = 7{,}2275 < \Delta(Y) = 9{,}2065$$

il che sta ad indicare che la compensazione fra i voti ottenuti dai laureati della facoltà di Economia di Milano-Bicocca, nei quattro insegnamenti considerati, è minore di quella che si ottiene nell'ipotesi di indipendenza.

Risulta così che l'indice di compensazione, nell'ipotesi di indipendenza, è:

$$\hat{C} = 1- \frac{7{,}2275}{14{,}2884} = 0{,}4942.$$

La minor compensazione che si nota, rispetto all'ipotetica situazione d'indipendenza, rafforza la considerazione, proposta in precedenza, che l'effetto impegno nello studio e l'effetto capacità intellettuali spostano l'indice verso la compensazione nulla.

6. CONCLUSIONI

Si è da principio considerata la scomposizione dell'indice normalizzato C di compensazione per conoscere la percentuale d'apporto, alla compensazione, della generica variabile X_j. Si è in seguito riproposto il confronto tra il valore dell'indice di compensazione in caso di indipendenza distributiva ed il valore dell'indice nel caso di dati reali, al fine di poter meglio interpretare il significato del concetto di compensazione.

L'applicazione di questi indicatori della compensazione ad un esempio numerico ed a due casi reali (la distribuzione del reddito totale delle famiglie italiane, rilevata nell'anno 2001 tramite indagine campionaria e i voti in quattro insegnamenti dei laureati della facoltà di Economia dell'Università di Milano-Bicocca) ha confermato la sensibilità dell'indice C, proposto da Zenga, a misurare la compensazione nel caso di fenomeni complessi scomponibili in una pluralità di elementi.

La ricerca dovrà continuare nella verifica delle caratteristiche e delle potenzialità d'analisi di tale indice nello studio della compensazione di altri fenomeni socio-economici.

Dipartimento di Metodi quantitativi
per le Scienze Economiche ed Aziendali,
Facoltà di Economia,
Università degli Studi di Milano-Bicocca

WALTER MAFFENINI

RIFERIMENTI BIBLIOGRAFICI

F. PAMPINELLA, (2003), *Indice normalizzato di compensazione,* Tesi di Dottorato, Dottorato di Ricerca in Statistica Metodologica, XIV ciclo, Università degli Studi di Trento.

P. RADAELLI, M. ZENGA, (2002), *Decomposition of the Mean Difference of the Sum of Variates*, Rapporto di Ricerca del Dipartimento di Metodi Quantitativi per l'Economia, Università degli Studi di Milano-Bicocca.

M. ZENGA, (2003), *Distributive compensation ratio derived from the decomposition of the mean difference of a sum*, "Statistica & Applicazioni", Vol. I, n. 1, pp. 19-27.

RIASSUNTO

Osservazioni sull'indice di compensazione distributiva

Recentemente Zenga ha proposto un indice di compensazione normalizzato basato sulla scomposizione della differenza media della somma Y di k variabili $X_1, ..., X_k$.

In questo articolo si è proposta una scomposizione di questo indice per conoscere la percentuale d'apporto, alla compensazione, della generica variabile X_j. Si è in seguito riproposto il confronto tra il valore dell'indice di compensazione in caso di indipendenza distributiva ed il valore dell'indice nel caso di dati reali, al fine di poter meglio interpretare il significato del concetto di compensazione.

L'applicazione di questi indicatori della compensazione ad un esempio numerico ed a due casi reali (la distribuzione del reddito totale delle famiglie italiane, rilevata nell'anno

2001 tramite indagine campionaria e i voti in quattro insegnamenti dei laureati della facoltà di Economia dell'Università di Milano-Bicocca) ha confermato la sensibilità dell'indice a misurare la compensazione nel caso di fenomeni complessi scomponibili in una pluralità di elementi.

SUMMARY

Some notes about the distributive compensation ratio

Recently Zenga has proposed a normalized compensation index based on a decomposition of the mean difference of the sum Y of the k variates $X_1, ..., X_k$.

This paper proposes a decomposition of the above index to evaluate the percent contribution of a generic variate X_j to compensation. Moreover, the importance of a comparison between the observed value of the index and its value under the hypothesis of distributive independence is emphasized, to understand exhaustively the concept of compensation.

The index's capability of measuring compensation of a complex phenomena, possibly composed by a number of different elements, is shown by a numerical example and by two applications to real data (a sample survey of total income of Italian families in 2001 and a dataset concerning marks of graduate students of Milano-Bicocca University in four exams).