Università degli studi di Milano Bicocca

SCUOLA DI DOTTORATO IN STATISTICA METODOLOGICA: XXVIII ciclo

SCOMPOSIZIONE PER SOTTOPOPOLAZIONI DELL'INDICE DI BONFERRONI

Relatore:
Prof. Michele Zenga

Dottorando:
Dott. Igor Valli

Coordinatore Dottorato:
Prof.ssa Fulvia Mecatti

Anno Accademico
2014-2015

a Liliana e Salvatore

Indice

Capitolo 1

Introduzione

I primi contributi significativi allo studio della concentrazione della ricchezza e dei redditi risalgono alla fine del XIX secolo. Nel 1895 Vilfredo Pareto, nella memoria ”La Legge della domanda”, afferma che il numero dei contribuenti con reddito Y compreso nell'intervallo $(y, y + dy]$ è dato da $Hy^{-(\alpha+1)}dy$ con H e α costanti da determinarsi in ciascun caso [13]. In particolare, Pareto determinò per diverse distribuzioni il valore di α e dopo applicazioni su dati reali, nella stessa memoria, affermò che una diminuzione di α indica una tendenza ad una minore ineguaglianza delle entrate. Nel 1897 Pareto pubblica un secondo contributo in cui propone $A\alpha(y+a)^{-(\alpha+1)}$ quale legge per determinare il numero di contribuenti con reddito nell'intervallo $(y, y+dy]$, con A, a, α costanti da determinarsi caso per caso. Anche in questa memoria, Pareto conferma di ritenere che il parametro α misurasse la disuguaglianza dei redditi e, precisamente, che al diminuire di α diminuisse la diseguaglianza dei redditi. Indicato con y_0 il reddito più basso e con N_y il numero delle persone con reddito maggiore di y, nel secondo modello paretiano si ha:

$$N_y = A(y + a)^{-\alpha}; \qquad (N_{y_0} - N_y) = A\left[(y_0 + a)^{-\alpha} - (y + a)^{-\alpha}\right].$$

In questi termini, per Pareto, quando il numero di persone con reddito inferiore a y diminuisce in rapporto al numero di persone con reddito superiore a y, la disuguaglianza diminuisce, e pertanto, al diminuire del rapporto

$$\frac{N_{y_0} - N_y}{N_{y_0}} = 1 - \frac{N_y}{N_{y_0}}, \tag{1.1}$$

ovvero all'aumentare del rapporto N_y/N_{y_0}, la disuguaglianza diminuisce. Essendo

$$\frac{N_y}{N_{y_0}} = \left(\frac{y_0 + a}{y + a}\right)^{\alpha} \tag{1.2}$$

Pareto dimostrò la relazione (diretta) tra parametro α e l'ancora ”embrionale” concetto di disuguaglianza. I contributi di Pareto ebbero una vasta risonanza tra gli studiosi dell'epoca; tuttavia, sia il concetto fornito da Pareto di diseguaglianza, che la natura della relazione tra diseguaglianza e parametro α non furono pienamente accolti e diversi studiosi, tra cui Benini (1897), Bresciani-Turroni (1905), Mortara (1911) e Gini, mostrarono

le loro perplessità.
Nel 1909, nella memoria "Il diverso accrescimento delle classi sociali e la concentrazione della ricchezza", Corrado Gini propone per la prima volta, quale misura di diseguaglianza, l'indice δ [13].

Nel 1914 l'indice δ viene rianalizzato nell'articolo "Sulla misura della concentrazione e della variabilità dei caratteri" [7]. A tal riguardo, siano

$$0 \leq y_{(1)} \leq ... \leq y_{(i)} \leq ... \leq y_{(n)} > \ 0 \tag{1.3}$$

gli n valori graduati in senso non decrescente di una variabile Y a valori non negativi e si considerino le seguenti quantità:

$$Q_i(Y) = \sum_{t=1}^{i} y_{(t)} \qquad (i = 1, ..., n) \tag{1.4}$$

$$T(Y) = Q_n(Y)$$

$$q_i = \frac{Q_i(Y)}{T(Y)} \qquad (i = 1, ..., n) \tag{1.5}$$

$$p_i = \frac{i}{n} \qquad (i = 1, ..., n) \tag{1.6}$$

$$\overset{-}{M}_i(Y) = \frac{Q_i(Y)}{i} \qquad (i = 1, ..., n) \tag{1.7}$$

$$\overset{+}{M}_i(Y) = \begin{cases} \dfrac{\sum_{t=i+1}^{n} y_{(t)}}{n-1} = \dfrac{T(Y) - Q_i(Y)}{n-i} & \text{per } i = 1, ..., n-1 \\ y_{(n)} & \text{per } i = n \end{cases} \tag{1.8}$$

$$M(Y) = \frac{T(Y)}{n} = \overset{-}{M}_n(Y) \tag{1.9}$$

dove:

- $Q_i(Y)$ indica i valori cumulati della variabile Y fino all'unità $i - esima$;

- q_i indica la quota relativa dei valori cumulati della variabile Y fino all'unità $i - esima$;

- p_i indica la frequenza relativa cumulata;

- $\overset{-}{M}_i(Y)$, detta *media inferiore*, indica la media aritmetica del *gruppo inferiore*, ovvero delle unità tali che $Y \leq y_{(i)}$;

- $\overset{+}{M}_i$ indica la media aritmetica del *gruppo superiore*, ovvero delle unità tali che $Y > \ y_{(i)}$;

- $M(Y)$ indica la media calcolata su tutto il campione

Ai fini della trattazione successiva, si dimostra che:

$$p_i \geq q_i \tag{1.10}$$

Dimostrazione. Per $i = n$, $p_i = q_i = 1$. Per $i = 1, ..., n-1$, per la proprietà di internalità della media

$$\overset{-}{M}_i(Y) \leq y_{(i)} \leq y_{(i+1)} \leq \overset{+}{M}_{i+1}(Y) \tag{1.11}$$

e pertanto

$$\overset{-}{M}_i(Y) \leq \overset{+}{M}_{i+1}(Y) \tag{1.12}$$

Per la proprietà di associatività della media

$$M(Y) = \frac{1}{n}\left[i\overset{-}{M}_i(Y) + (n-i)\overset{+}{M}_{i+1}(Y)\right] \tag{1.13}$$

e, sfruttando nuovamente la proprietà di internalità della media, si ricava:

$$\overset{-}{M}_i(Y) \leq M(Y) \leq \overset{+}{M}_{i+1}(Y). \tag{1.14}$$

In virtù della (1.14) $\overset{-}{M}_i(Y) \leq M(Y)$ e pertanto

$$\frac{Q_i(Y)}{i} \leq \frac{T(Y)}{n}. \tag{1.15}$$

Infine, in virtù della (1.15)

$$\frac{Q_i(Y)}{Q_n(Y)} \leq \frac{i}{n} \tag{1.16}$$

□

In virtù della (1.10), nel 1914, Gini afferma che [6] "la *concentrazione* del carattere è tanto più forte quanto più è forte, per gli $n-1$ valori di i, la disuguaglianza

$$p_i > \ q_i. \tag{1.17}$$

In altre parole, diremo che la concentrazione di un carattere è tanto più forte, quanto più è piccola la parte che, sull'ammontare totale del carattere, spetta a quella parte dei casi, in cui l'intensità del carattere non supera un certo limite. La concentrazione del carattere si dirà *perfetta* quando l'intensità del carattere sia nulla in $n-1$ casi e $T(Y)$ in un solo caso [...]. Se in tutti i casi il carattere presenterà la stessa intensità [...] diremo che la concentrazione del carattere è nulla o, in altre parole, vi è *equidistribuzione del carattere*". E aggiunge: "Se con una o più costanti si riesce ad esprimere, tra i due membri della disuguaglianza in (1.17) (o di una disuguaglianza che da essa si deduce), una relazione valida [...] potremmo assumere tale costante, o l'insieme di tali costanti, come *indice di concentrazione del carattere*". In luogo della (1.17) Gini utilizza la disuguaglianza

$1 - p_i < \ 1 - q_i$ e quale misura di concentrazione δ considera la media aritmetica delle costanti δ_i tali che

$$1 - p_i = (1 - q_i)^{\delta_i} \tag{1.18}$$

e pertanto:

$$\delta_i = \frac{\log(1 - p_i)}{\log(1 - q_i)}; \qquad \delta = \frac{1}{n-1} \sum_{i=1}^{n-1} \delta_i.$$

Dato che la disuguaglianza in (1.17) è tanto più forte:

1. in valore assoluto, quando maggiore è la differenza $p_i - q_i$;

2. in valore relativo, quando maggiore è il rapporto:

$$R_i(Y) = \frac{p_i - q_i}{p_i}, \tag{1.19}$$

Gini, nello stesso articolo, introduce la misura di diseguaglianza (puntuale) $R_i(Y)$, da cui ottiene il noto *rapporto di concentrazione* $\tilde{R}(Y)$ facendo la media ponderata delle $n-1$ misure puntuali $R_i(Y)$ con pesi p_i:

$$\tilde{R}(Y) = \frac{\sum_{i=1}^{n-1} \frac{p_i - q_i}{p_i} \cdot p_i}{\sum_{i=1}^{n-1} p_i} = \frac{\sum_{i=1}^{n-1} (p_i - q_i)}{\sum_{i=1}^{n-1} p_i}. \tag{1.20}$$

Inoltre, nello stesso articolo, Gini dimostra che

$$\tilde{R}(Y) = \frac{\Delta(Y)}{2M(Y)}, \tag{1.21}$$

ove $\Delta(Y)$ indica la *differenza media* (senza ripetizione):

$$\Delta(Y) = \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{j=1}^{n} \left| y_{(i)} - y_{(j)} \right|. \tag{1.22}$$

Nel 1930 Carlo Emilio Bonferroni, introducendo gli indici di concentrazione nel libro di testo "Elementi di Statisitica Generale" [3], afferma che, un indice di concentrazione deve soddisfare le seguenti condizioni:

1. *sensibilità ai trasferimenti*: "preso $y_{(i)} \leq y_{(j)}$, $(1 \leq i < \ j \leq n)$, se si toglie ad $y_{(i)}$ la quantità h trasferendola ad $y_{(j)}$, cioè se ad $y_{(i)}$ ed $y_{(j)}$ si sostituiscono $y_{(i)} - h$, $y_{(j)} + h$ (con $h > \ 0$), la *concentrazione* s' intende accresciuta, e quindi un indice di essa deve aumentare";

2. *valore massimo per concentrazione massima e minimo per concentrazione minima*: "la concentrazione s' intende massima quando, su n soggetti esaminati, uno solo possiede il carattere considerato: graduando le misure y_i si ha allora $0, 0, ..., 0, y_{(n)}$.

In questo caso, e solo in questo caso, l'indice deve assumere il suo valore massimo, che si può stabilire uguale ad 1. In caso di concentrazione minima, cioè per $y_{(i)} = costante$, l'indice di concentrazione varrà zero";

3. *l'indice deve essere relativo*: "ciò perchè, ovviamente, il giudizio sulla concentrazione non deve dipendere dall'unità di misura delle y_i"

Nel 1930 Bonferroni introduce l'indice di concentrazione $\tilde{V}(Y)$, definito come media aritmetica, calcolata su $n-1$ unità, delle misure puntuali $\tilde{V}_i(Y)$:

$$\tilde{V}_i(Y) = \frac{M(Y) - \bar{M}_i(Y)}{M(Y)} = 1 - \frac{\bar{M}_i(Y)}{M(Y)}, \tag{1.23}$$

$$\tilde{V}(Y) = \frac{1}{n-1}\sum_{i=1}^{n-1}\tilde{V}_i(Y) = \frac{1}{n-1}\sum_{i=1}^{n-1}\frac{M(Y) - \bar{M}_i(Y)}{M(Y)}. \tag{1.24}$$

Dato che $\bar{M}_n(Y) = M(Y)$ segue che $V_n(Y) = 0$ e per tale ragione la sommatoria in (1.24) è applicata a $n-1$ unità.

Nel 1940, nell'articolo "Sul significato di alcuni indici di concentrazione", Mario De Vergottini [14] dimostra che $R_i(Y) = \tilde{V}_i(Y)$. In effetti, dalla (1.19) e dalle (1.5), (1.6), (1.7) (1.9) e (1.23) , per $i = 1, ..., n$:

$$\begin{aligned} R_i(Y) = \frac{p_i - q_i}{p_i} &= \frac{\frac{i}{n} - \frac{\sum_{t=1}^{i} y_t}{n \cdot M(Y)}}{\frac{i}{n}} \\ &= \frac{\frac{i}{n} - \frac{i \cdot \bar{M}_i(Y)}{n \cdot M(Y)}}{\frac{i}{n}} \\ &= 1 - \frac{\bar{M}_i(Y)}{M(Y)} \\ &= \frac{M(Y) - \bar{M}_i(Y)}{M(Y)} = \tilde{V}_i(Y) \end{aligned} \tag{1.25}$$

Usando la (1.25) è quindi possibile esprimere l'indice di concentrazione $\tilde{R}(Y)$ di Gini come media ponderata degli indici puntuali di Bonferroni $\tilde{V}_i(Y)$ con pesi $2 \cdot i/(n-1)$.

Dimostrazione. Ricordando che $\tilde{V}_n(Y) = 0$, in virtù della (1.25) e delle definizioni in

(1.5), (1.6), (1.7) e (1.9):

$$\begin{aligned}\tilde{R}(Y) = \frac{\sum_{i=1}^{n-1} \frac{p_i - q_i}{p_i}\, p_i}{\sum_{i=1}^{n-1} \frac{i}{n}} &= \frac{\sum_{i=1}^{n-1} \tilde{V}_i(Y)\, \frac{i}{n}}{\frac{1}{n}\frac{n(n-1)}{2}} \\ &= \frac{2}{n-1} \sum_{i=1}^{n-1} \tilde{V}_i(Y)\, \frac{i}{n} \\ &= \frac{1}{n} \sum_{i=1}^{n} \tilde{V}_i\, \frac{2\,i}{n-1}\end{aligned} \tag{1.26}$$

$\square$

Nel 2007 Michele Zenga [16] propone un indice di diseguaglianza definito per dati espressi in termini di distribuzione di frequenza, ovvero, per dati definiti dalle coppie:

$$\left\{ (y_h, n_{h.}),\ h = 1, ..., r;\ 0 \le y_1 < \ y_2 < \ ... < \ y_r; \sum_{h=1}^{r} n_{h.} = n \right\}. \tag{1.27}$$

In questi termini, al fine di valutare la concentrazione del carattere Y presente in una popolazione $\mathfrak{P}$, Zenga suddivide $\mathfrak{P}$ in due gruppi incompatibili: un *gruppo inferiore*, definito dalle coppie

$$\{(y_1, n_{1.}), ..., (y_h, n_{h.})\} \tag{1.28}$$

ed un *gruppo superiore*, definito dalle coppie

$$\{(y_{h+1}, n_{h+1.}), ..., (y_r, n_{r.})\}\,, \tag{1.29}$$

dove r indica il numero delle modalità della variabile Y e $n_{h.}$ indica il numero delle unità statistiche tali che $Y = y_h$. Siano inoltre,

$$P_{h.} = \sum_{t=1}^{h} n_{t.} \qquad (h = 1, ..., r) \tag{1.30}$$

$$n = \sum_{h=1}^{r} n_{h.} = P_r \tag{1.31}$$

$$p_{h.} = \frac{P_{h.}}{n} \qquad (h = 1, ..., r) \tag{1.32}$$

$$Q_{h.}(Y) = \sum_{t=1}^{h} y_t \, n_{t.} \qquad (h = 1, ..., r) \tag{1.33}$$

$$T(Y) = \sum_{h=1}^{r} y_h \, n_{h.} = Q_r = \sum_{i=1}^{n} y_{(i)} \qquad (i = 1, ..., n) \tag{1.34}$$

$$q_{h.} = \frac{Q_{h.}}{T(Y)} \qquad (h = 1, ..., r) \tag{1.35}$$

$$\overset{-}{M}_{h.}(Y) = \frac{Q_{h.}(Y)}{P_{h.}} \qquad (h = 1, ..., r) \tag{1.36}$$

$$M(Y) = \frac{T(Y)}{n} = \overset{-}{M}_{r.}(Y) \tag{1.37}$$

$$\overset{+}{M}_{h.}(Y) = \begin{cases} \dfrac{T(Y) - Q_{h.}(Y)}{n - P_{h.}} & h = 1, ..., r-1 \\ y_r & h = r \end{cases} \tag{1.38}$$

dove:

- $P_{h.}$ indica le frequenze cumulate fino alla modalità $h - esima$;

- $p_{h.}$ indica le frequenze relative cumulate fino alla modalità $h - esima$;

- $Q_{h.}(Y)$ indica i valori cumulati della variabile Y fino alla modalità $h - esima$;

- $q_{h.}$ indica la quota relativa cumulata della variabile Y fino alla modalità $h - esima$
- $\overset{-}{M}_{h.}(Y)$ indica la media aritmetica del gruppo inferiore

- $\overset{+}{M}_{h.}(Y)$ indica la media aritmetica del gruppo superiore

In questi termini, Zenga propone [16], quale indice puntuale di ineguaglianza, l'indice $I_h(Y)$,

$$I_h(Y) = \frac{\overset{+}{M}_{h.}(Y) - \overset{-}{M}_{h.}(Y)}{\overset{+}{M}_{h.}(Y)} = 1 - \frac{\overset{-}{M}_{h.}(Y)}{\overset{+}{M}_{h.}(Y)} \tag{1.39}$$

e definisce, tramite media ponderata degli indici puntuali $I_h(Y)$, con pesi dati dalle frequenze relative $n_{h.}/n$, l'indice sintetico $I(Y)$:

$$I(Y) = \sum_{h=1}^{r} I_h(Y)\,\frac{n_{h.}}{n} = \sum_{h=1}^{r} \frac{\overset{+}{M}_{h.}(Y) - \overset{-}{M}_{h.}(Y)}{\overset{+}{M}_{h.}(Y)}\,\frac{n_{h.}}{n}. \tag{1.40}$$

Le caratteristiche e le proprietà che distinguono gli indici $I_h(Y)$ e $I(Y)$, rispetto agli indici finora proposti, saranno presentate nei capitoli successivi.

Capitolo 2

Le principali proprietà degli indici di ineguaglianza

Come accennato nel capitolo precedente, L'approccio utilizzato nei primi e fondamentali contributi allo studio dell'ineguaglianza (concentrazione), è stato quello di costruire delle *misure puntuali* di ineguaglianza, e di sintetizzarle in indici (sintetici) ottenuti come media di dette misure puntuali. A tal proposito [1], indicando con $\mathfrak{I}$ un generico indice di concentrazione, è possibile esprimere alcune misure di ineguaglianza tramite la formula

$$\mathfrak{I} = \frac{\sum_{i=1}^{n-1} \frac{p_i - q_i}{p_i} \cdot \omega_i}{\sum_{i=1}^{n-1} \omega_i}, \tag{2.1}$$

dove, ad esempio, se:

- $\omega_i = p_i$, $\mathfrak{I} = \tilde{R}$ (indice di Gini)
- $\omega_i = 1$, $\mathfrak{I} = \tilde{V}$ (indice di Bonferroni)
- $\omega_i = \dfrac{p_i}{1 - p_i}$, $\mathfrak{I} = V'$ (indice di De Vergottini).

L'unica ragione per cui, in relazione agli indici $\tilde{R}(Y)$ e $\tilde{V}(Y)$, gli Autori considerano $n-1$ misure puntuali (confronti) anzichè n, è che $R_n(Y) = \tilde{V}_n(Y) = 0$, essendo $p_n = q_n = 1$, o analogamente, $\bar{M}_n(Y) = M(Y)$. Tale motivazione appare però non ragionevole. Per tale motivo, ricordando la definizione in (1.20), indichiamo con $G'(Y)$ la media delle n misure puntuali $R_i(Y)$, $i = 1, ..., n$:

$$G'(Y) = \frac{\sum_{i=1}^{n} \frac{p_i - q_i}{p_i}\, p_i}{\sum_{i=1}^{n} p_i} = \frac{1}{n} \sum_{i=1}^{n} R_i(Y)\, \frac{2\,i}{n+1}. \tag{2.2}$$

Ovviamente,

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

tuttavia,

$$\sum_{i=1}^{n-1} p_i = \frac{n-1}{2} < \frac{n+1}{2} = \sum_{i=1}^{n} p_i \tag{2.3}$$

e pertanto, in relazione all'indice di Gini $\tilde{R}(Y)$, la mancata considerazione dell'$n-esima$ misura puntuale non appare giustificata. In virtù della (1.25) e della (2.2) si ha:

$$G'(Y) = \frac{1}{n}\sum_{i=1}^{n} G'_i(Y), \tag{2.4}$$

dove

$$G'_i(Y) = R_i(Y)\,\frac{2\,i}{n+1} \tag{2.5}$$

indica la misura di ineguaglianza puntuale dell'indice sintetico $G(Y)$. In virtù della (2.2) si dimostra [19]:

$$G'(Y) = \frac{n-1}{n+1}\tilde{R}(Y) \tag{2.6}$$

Dimostrazione. Ricordando che $\tilde{V}_n(Y) = 0$, in virtù delle (1.25), (1.26) e (2.2), si ha

$$\tilde{R}(Y) = \frac{\sum_{i=1}^{n-1} \frac{p_i - q_i}{p_i} p_i}{\sum_{i=1}^{n-1} p_i} = \frac{1}{n}\sum_{i=1}^{n-1} \tilde{V}_i(Y)\,\frac{2\,i}{n-1} = \frac{1}{n}\sum_{i=1}^{n} \tilde{V}_i(Y)\,\frac{2\,i}{n-1} \tag{2.7}$$

e pertanto segue la (2.6). □

In modo analogo, indichiamo con $V'(Y)$ la media delle n misure puntuali $\tilde{V}_i(Y)$:

$$V'(Y) = \frac{1}{n}\sum_{i=1}^{n} \tilde{V}_i(Y). \tag{2.8}$$

Ricordando che $\sum_{i=1}^{n-1} \tilde{V}_i(Y) = \sum_{i=1}^{n} \tilde{V}_i(Y)$, risulta:

$$V'(Y) = \frac{n-1}{n}\tilde{V}(Y) \tag{2.9}$$

In virtù delle relazioni in (2.7) e in (2.9), e della definizione in (2.5), nel prosieguo della trattazione saranno presi in considerazione gli indici sintetici $G'(Y)$, $V'(Y)$, $I(Y)$ e i rispettivi indici puntuali $G'_i(Y)$, $\tilde{V}_i(Y)$ e $I_i(Y)$, $i = 1, ..., n$.
Le "anomalie" degli indici puntuali $G'_i(Y)$ e $\tilde{V}_i(Y)$ non riguardano solamente il caso in cui $i = n$. In questa sede, oltre a non sembrare adeguata la spiegazione addotta a considerare solamente le prime $n-1$ misure puntuali, al fine di definire gli indici sintetici $\tilde{R}(Y)$ e $\tilde{V}(Y)$, appare inadeguato assegnare a differenti configurazioni di massima diseguaglianza, medesimo valore di massima diseguaglianza. A tal proposito, si considerino due popolazioni composte da $n_1 = 10$ ed $n_2 = 1000$ unità statistiche rispettivamente ed

in cui, in entrambi i casi, una sola unità statistica possiede il totale $T(Y)$. In entrami i casi, coerentemente alla seconda proprietà introdotta da Bonferroni, si ha

$$\tilde{R}_{n_1}(Y) = \tilde{R}_{n_2}(Y) = \tilde{V}_{n_1}(Y) = \tilde{V}_{n_2}(Y) = 1$$

dove $\tilde{R}_{n_v}(Y)$ e $\tilde{V}_{n_v}(Y)$, $v = 1, 2$, indicano gli indici sintetici $\tilde{R}(Y)$ e $\tilde{V}(Y)$ calcolati nelle due popolazioni di numerosità n_v, rispettivamente. Pertanto, calcolando gli indici di Gini $\tilde{R}(Y)$ e di Bonferoni $\tilde{V}(Y)$ nelle due popolazioni in esame, si giunge alla conclusione che queste possono considerarsi analoghe in termini di ineguaglianza. Tuttavia, nelle applicazioni, in particpolare in quelle di carattere socio-economico, sembra più utile disporre di un indice capace di valutare, in modo più adeguato, il livello di ineguaglianza presente in scenari che, quantomeno per dimensione, risultano diversi. A tal proposito, sembra dunque più opportuno disporre di un indice che, anzichè assumere valore 1 in caso di massima diseguaglianza, indipendentemente dall'ampiezza della popolazione considerata, sia funzione crescente di n e tale che, detta C_n tale funzione:

$$\lim_{n \to \infty} C_n = 1.$$

A tal riguardo, si dimostra che gli indici $G(Y)$, $V'(Y)$ e $I(Y)$ soddisfano tale richiesta. [17]

Dimostrazione. In accordo con la definizione di Bonferroni, in caso di massima diseguaglianza:

$$Y = \left\{0, 0, ..., 0, y_{(n)} = T(Y)\right\}, \qquad \tilde{R}(Y) = \tilde{V}(Y) = 1 \tag{2.10}$$

In questa circostanza, si ha:

1. in virtù della (2.2), della (2.6) e della (2.10),

$$\begin{aligned} G'(Y) &= \frac{n-1}{n+1}\,\tilde{R} \\ &= 1 - \frac{2}{n+1}; \end{aligned} \tag{2.11}$$

2. in virtù della (2.9) e della (2.10),

$$\begin{aligned} V'(Y) &= \frac{n-1}{n}\,\tilde{V}(Y) \\ &= 1 - \frac{1}{n}; \end{aligned} \tag{2.12}$$

3. in virtù della (1.40) e della (2.10),

$$
\begin{aligned}
I(Y) &= \frac{1}{n}\sum_{i=1}^{n} I_i(Y) \\
&= \frac{1}{n}\sum_{i=1}^{n}\left(1-\frac{\overset{-}{M}_i(Y)}{\overset{+}{M}_i(Y)}\right) \\
&= \frac{1}{n}\left[(n-1)+\left(1-\frac{1}{n}\right)\right] \\
&= 1-\frac{1}{n^2}. \qquad (2.13)
\end{aligned}
$$

In virtù delle (2.11), (2.12) e (2.13), si ha

$$
C_n = \begin{cases} 1-\dfrac{2}{n+1} & \text{in relazione all'indice } G'(Y) \\ 1-\dfrac{1}{n} & \text{in relazione all'indice } V'(Y) \\ 1-\dfrac{1}{n^2} & \text{in relazione all'indice } I(Y) \end{cases}
$$

e pertanto,

$$\lim_{n\to\infty} C_n = 1.$$

□

In relazione all'esempio posto di due popolazioni di numerosià $n_1 = 10$ ed $n_2 = 1000$ rispettivamente, si ha $G_{n_1}(Y) = 0,8181$, $V'_{n_1}(Y) = 0,9$, $G_{n_2}(Y) = 0,998$ e $V'_{n_2}(Y) = 0,999$.

In accordo con le proprietà introdotte da Bonferroni, si procede ora a dimostrare che gli indici $G(Y)$, $V'(Y)$ e $I(Y)$ soddisfano le seguenti proprietà [17]:

1. *valore minimo in caso di concentrazione minima*;

2. *relatività dell'indice (invarianza a trasformazioni di scala)*

3. *sensibilità ai trasferimenti*

Dimostrazione. In relazione alla proprietà di *valore minimo in caso di concentrazione minima*, in caso di equidistribuzione del carattere Y,

$$y_i = M(Y) = \overset{-}{M}_i(Y) = \overset{+}{M}_i(Y), \qquad (i = 1, ..., n); \qquad (2.14)$$

$$\tilde{R}(Y) = \tilde{V}(Y) = 0. \qquad (2.15)$$

Pertanto:

- in virtù della (2.6) e della (2.15),

$$G'(Y) = \frac{n-1}{n+1}\,\tilde{R} = 0; \tag{2.16}$$

- in virtù della (2.9) e della (2.15),

$$V'(Y) = \frac{n}{n-1}\tilde{V}(Y) = 0; \tag{2.17}$$

- in virtù della (1.40),

$$\begin{aligned} I(Y) &= \frac{1}{n}\sum_{i=1}^{n} I_i(Y) \\ &= \frac{1}{n}\sum_{i=1}^{n}\left(1 - \frac{\overset{-}{M}_i(Y)}{\overset{+}{M}_i(Y)}\right) \\ &= \frac{1}{n}\sum_{i=1}^{n}(1-1) = 0. \end{aligned} \tag{2.18}$$

In relazione alla proprietà di *invarianza a trasformazioni di scala*, in virtù delle proprietà della media e posto $X = a\,Y$, $a \in \mathbb{R}^+$:

- in virtù della (2.2),

$$\begin{aligned} G'(X) &= \frac{1}{n}\sum_{i=1}^{n} G'_i(X) \\ &= \frac{1}{n}\sum_{i=1}^{n}\left(1 - \frac{a\,\overset{-}{M}_i(Y)}{a\,M(Y)}\right)\frac{2\,i}{n+1} \\ &= G'(Y) \end{aligned} \tag{2.19}$$

- in virtù della (2.8)

$$\begin{aligned} V'(X) &= \frac{1}{n}\sum_{i=1}^{n} \tilde{V}_i(X) \\ &= \frac{1}{n}\sum_{i=1}^{n}\left(1 - \frac{a\,\overset{-}{M}_i(Y)}{a\,M(Y)}\right) \\ &= V'(Y) \end{aligned} \tag{2.20}$$

- in virtù della (1.40),

$$\begin{aligned} I(X) &= \frac{1}{n}\sum_{i=1}^{n} I_i(X) \\ &= \frac{1}{n}\sum_{i=1}^{n}\left(1-\frac{a\,\overset{-}{M}_i(Y)}{a\,\overset{+}{M}_i(Y)}\right) \\ &= I(Y) \end{aligned} \tag{2.21}$$

Infine, in relazione alla proprietà di *sensibilità ai trasferimenti*, si consideri $y_{(i)} < y_{(i+1)}$ e sia $0 < h < \frac{1}{2}(y_{(i+1)} - y_{(i)})$. Senza perdita in generalità, sia Y' la distribuzione ottenuta trasferendo la quantità h da $y_{(i)}$ a $y_{(i+1)}$, $i = 1, ..., n-1$. In questi termini si ha, in virtù della (1.9),

$$M(Y') = M(Y) = \overset{-}{M}_n(Y') = \overset{-}{M}_n(Y) \tag{2.22}$$

e, per ogni $t \neq i$:

$$\overset{-}{M}_t(Y') = \overset{-}{M}_t(Y); \qquad \overset{+}{M}_t(Y') = \overset{+}{M}_t(Y). \tag{2.23}$$

Pertanto, virtù della (2.22) e della (2.23), per ogni $t \neq i$:

$$G_t'(Y') = G_t'(Y), \qquad \tilde{V}_t(Y') = \tilde{V}_t(Y), \qquad I_t(Y') = I_t(Y). \tag{2.24}$$

Per $t = i$:

$$\overset{-}{M}_t(Y') = \frac{Q_i(Y) - h}{i} = \overset{-}{M}_i(Y) - \frac{h}{i}, \qquad \text{(per } t = i = 1, ..., n-1) \tag{2.25}$$

$$\overset{+}{M}_t(Y') = \begin{cases} \dfrac{T(Y) - (Q_i(Y) - h)}{n-i} = \overset{+}{M}_i(Y) + \dfrac{h}{n-i} & \text{(per } t = i = 1, ..., n-1) \\ y_{(n)} + h & \text{(per } t = n) \end{cases} \tag{2.26}$$

Pertanto, in virtù delle (1.25), (2.8), (2.2), (1.40), (2.25) e (2.26):

- per $t = i = 1, ..., n-1$:

$$\tilde{V}_i(Y') = 1 - \left(\frac{\overset{-}{M}_i(Y) - \frac{h}{i}}{M(Y)}\right) = \left(1 - \frac{\overset{-}{M}_i(Y)}{M(Y)}\right) + \frac{h}{i\,M(Y)} > \tilde{V}_i(Y), \tag{2.27}$$

$$G_i'(Y') = \tilde{V}_i(Y')\,\frac{2\,i}{n+1} > \tilde{V}_i(Y)\,\frac{2\,i}{n+1} = G_i'(Y), \tag{2.28}$$

$$I_i(Y') = 1 - \left(\frac{\overset{-}{M}_i(Y) - \frac{h}{i}}{\overset{+}{M}_i(Y) + \frac{h}{n-i}}\right) > 1 - \left(\frac{\overset{-}{M}_i(Y)}{\overset{+}{M}_i(Y)}\right) = I_i(Y'); \tag{2.29}$$

- per $t = n$, ovvero qualora il trasferimento avvenisse da $y_{(n-1)}$ a $y_{(n)}$,

$$\tilde{V}_n(Y') = \tilde{V}_n(Y) = 0, \tag{2.30}$$
$$G'_n(Y') = G'_n(Y) = 0, \tag{2.31}$$
$$I_n(Y') = 1 - \left(\frac{M(Y)}{y_{(n)} + h}\right) > 1 - \left(\frac{M(Y)}{y_{(n)}}\right) = I_n(Y). \tag{2.32}$$

Pertanto, essendo $G'(Y)$, $V'(Y)$ e $I(Y)$ medie delle rispettive misure puntuali, in virtù delle (2.24), (2.27), (2.28), (2.29) e (2.32) si ha:

$$G'(Y') > G'(Y); \qquad V'(Y') > V'(Y); \qquad I(Y') > I(Y)$$

□

Nel sopra menzionato testo del 1930, Bonferroni studia anche la seguente *sensibilità alle equiaggiunte*. Detto $\mathfrak{I}(\cdot)$ un generico indice di concentrazione, si dice che $\mathfrak{I}(\cdot)$ soddisfa la proprietà di *sensibilità alle equiaggiunte* se, posto $X = Y+h$, $h > 0$, si ha $\mathfrak{I}(X) < \mathfrak{I}(Y)$. Si dimostra ora che gli indici $G(Y)$, $V(Y)$ e $I(Y)$ sono sensibili alle equiaggiunte.

Dimostrazione. Se $X = Y + h$, $h > 0$, si ha

$$M(X) = h + M(Y); \qquad \overset{-}{M}_i(X) = h + \overset{-}{M}_i(Y); \qquad \overset{+}{M}_i(X) = h + \overset{+}{M}_i(Y).$$

Pertanto:

$$G'(X) = \frac{1}{n}\sum_{i=1}^{n} \frac{M(Y) - \overset{-}{M}_i(Y)}{h + M(Y)}\,\frac{2\,i}{n+1} < \frac{1}{n}\sum_{i=1}^{n} \frac{M(Y) - \overset{-}{M}_i(Y)}{M(Y)}\,\frac{2\,i}{n+1} = G'(Y)$$
$$V'(X) = \frac{1}{n}\sum_{i=1}^{n} \frac{M(Y) - \overset{-}{M}_i(Y)}{h + M(Y)} < \frac{1}{n}\sum_{i=1}^{n} \frac{M(Y) - \overset{-}{M}_i(Y)}{M(Y)} = V'(Y)$$
$$I(X) = \frac{1}{n}\sum_{i=1}^{n} \frac{\overset{+}{M}_i(Y) - \overset{-}{M}_i(Y)}{h + \overset{+}{M}_i(Y)} < \frac{1}{n}\sum_{i=1}^{n} \frac{\overset{+}{M}_i(Y) - \overset{-}{M}_i(Y)}{\overset{+}{M}_i(Y)} = I(Y)$$

□

Come fatto notare da Bonferroni [3] in relazione agli indici $\tilde{R}(\cdot)$ e $\tilde{V}(\cdot)$, per $h \to \infty$, si può osservare che gli indici considerati tendono a zero.

A conclusione di questo capitolo si dimostra infine che tra gli indici $G'(Y)$, $V'(Y)$ e $I(Y)$ vale la seguente relazione [17]:

$$G'(Y) \leq V'(Y) \leq I(Y) \tag{2.33}$$

Dimostrazione. In virtù della (1.9), $\dfrac{\overset{-}{M}_i(Y)}{M(Y)} \geq \dfrac{\overset{-}{M}_i(Y)}{\overset{+}{M}_i(Y)}$, $\forall\, i = 1, ..., n$. Pertanto:

$$\begin{aligned}
\sum_{i=1}^{n} \frac{\overset{-}{M}_i(Y)}{M(Y)} &\geq \sum_{i=1}^{n} \frac{\overset{-}{M}_i(Y)}{\overset{+}{M}_i(Y)}, \\
n - \sum_{i=1}^{n} \frac{\overset{-}{M}_i(Y)}{M(Y)} &\leq n - \sum_{i=1}^{n} \frac{\overset{-}{M}_i(Y)}{\overset{+}{M}_i(Y)}, \\
\sum_{i=1}^{n} \left(1 - \frac{\overset{-}{M}_i(Y)}{M(Y)}\right) &\leq \sum_{i=1}^{n} \left(1 - \frac{\overset{-}{M}_i(Y)}{\overset{+}{M}_i(Y)}\right), \\
\frac{1}{n}\sum_{i=1}^{n} \left(1 - \frac{\overset{-}{M}_i(Y)}{M(Y)}\right) &\leq \frac{1}{n}\sum_{i=1}^{n} \left(1 - \frac{\overset{-}{M}_i(Y)}{\overset{+}{M}_i(Y)}\right). \qquad (2.34)
\end{aligned}$$

In virtù delle (2.8), (1.40) e (2.34), segue $V'(Y) \leq I(Y)$.
Per dimostrare la relazione $G'(Y) \leq V'(Y)$ occorre innanzitutto osservare che:

1. $\dfrac{2\,i}{n+1}$ è funzione crescente di i;

2. $\tilde{V}_i(Y) = \dfrac{M(Y) - \overset{-}{M}_i(Y)}{M(Y)}$ è funzione decrescente di i.

Pertanto,

$$\mathrm{Cov}\left(\tilde{V}_i(Y), \frac{2\,i}{n+1}\right) \leq 0 \qquad (2.35)$$

Essendo,

$$
\begin{aligned}
\operatorname{Cov}\left(\tilde{V}_i(Y), \frac{2\,i}{n+1}\right) &= \mathbb{E}\left\{\left[\tilde{V}_i(Y) - \mathbb{E}(\tilde{V}_i(Y))\right]\left[\frac{2\,i}{n+1} - \mathbb{E}\left(\frac{2\,i}{n+1}\right)\right]\right\} \\
&= \mathbb{E}\left\{(\tilde{V}_i(Y) - V'(Y))\left[\frac{2\,i}{n+1} - \frac{2}{n+1}\,\mathbb{E}(i)\right]\right\} \\
&= \mathbb{E}\left[\tilde{V}_i(Y) - V'(Y))\left(\frac{2\,i}{n+1} - \frac{2}{n+1}\frac{1}{n}\sum_{i=1}^{n} i\right)\right] \\
&= \mathbb{E}\left[(\tilde{V}_i(Y) - V'(Y))\left(\frac{2\,i}{n+1} - 1\right)\right] \\
&= \mathbb{E}\left(\tilde{V}_i(Y)\,\frac{2\,i}{n+1} - \tilde{V}_i(Y) - V'(Y)\,\frac{2\,i}{n+1} + V'(Y)\right) \\
&= \mathbb{E}\left(\tilde{V}_i(Y)\,\frac{2\,i}{n+1}\right) - \mathbb{E}\left(\tilde{V}_i(Y)\right) - V'(Y)\,\mathbb{E}\left(\frac{2\,i}{n+1}\right) + \\
&+ \mathbb{E}\left(V'(Y)\right) \\
&= G'(Y) - V'(Y),
\end{aligned}
\tag{2.36}
$$

in virtù della (2.35) e della (2.36) si ottiene $G(Y) \leq V'(Y)$. In conclusione,

$$G'(Y) \leq V'(Y) \leq I(Y),$$

dove l'uguaglianza vale solo in caso di equidistribuzione del carattere Y $\square$

In virtù della (2.5),

$$G'_i(Y) = \tilde{V}_i(Y)\,\frac{2\,i}{n+1}$$

ed essendo $2i/(n+1)$ funzione crescente di i, si può infine osservare che la struttura dei pesi tende a dare, in modo arbitrario, maggiore importanza alle variazioni relative $\tilde{V}_i(Y)$ più piccole.

Capitolo 3

Gli indici di concentrazione nel contesto delle distribuzioni di frequenza

Gli indici $G'(Y)$ e $V'(Y)$ saranno ora ridefiniti per dati espressi in termini di distribuzione di frequenza, ovvero per dati espressi come in (1.27). I vantaggi che una tale procedura comporta saranno più chiari nel capitolo successivo quando si procederà alla scomposizione per sottopopolazioni. In questo capitolo, si analizzerà il *principio della somiglianza* e si dimostrerà che l'indice $V'(Y)$, ridefinito in termini di distribuzioni di frequenza, è, come l'indice $I(Y)$, invariante a replicazioni nella popolazione.

3.1 Principio della somiglianza (o di invarianza a replicazioni nella popolazione)

Si considerino le distribuzioni $Y = y_{(1)}, ..., y_{(i)}, ..., y_{(n)}$ e $Y' = y'_{(1)}, ..., y'_{(j)}, ..., y'_{(k)}$ ed un numero γ intero e maggiore di uno. La distribuzione Y' si dice *somigliante* alla distribuzione Y se è ottenuta replicando γ-volte ogni elemento della distribuzione Y. Ne consegue che, posto n il numero delle osservazioni rilevate per la variabile Y, $k = \gamma \cdot n$ è il numero delle osservazioni rilevate per la variabile Y'. Sia inoltre $\mathfrak{I}(\cdot)$ un generico indice di ineguaglianza. Un indice $\mathfrak{I}(\cdot)$ è detto *invariante a replicazioni nella popolazione* se il suo valore non cambia quando si passa da una distribuzione a una ad essa *somigliante* [15]. Nello studio dell'ineguaglianza, tale proprietà riveste una certa importanza. A tal riguardo, si consideri la distribuzione in Tabella 3.1, relativa alla variabile Y che descrive il reddito di $n = 7$ famiglie, e la distribuzione in Tabella 3.2, somigliante, di parametro $\gamma = 2$, alla distribuzione in Tabella 3.1, relativa alla variabile Y'.

Tabella 3.1: Distribuzione della variabile Y e calcolo degli indici $\tilde{V}_i(Y)$, $V'(Y)$, $G'_i(Y)$ e $G'(Y)$

i	$y_{(i)}$	$Q_i(Y)$	$\bar{M}_i(Y)$	$\tilde{V}_i(Y)$	$\frac{2i}{N+1}$	$G'_i(Y)$
1	2	2	2,00	0,8056	0,25	0,2014
2	4	6	3,00	0,7083	0,50	0,3542
3	4	10	3,33	0,6759	0,75	0,5069
4	10	20	5,00	0,5139	1,00	0,5139
5	13	33	6,60	0,3583	1,25	0,4479
6	13	46	7,67	0,2546	1,50	0,3819
7	26	72	10,28	0,0000	1,75	0,0000
				0,4738=$V'(Y)$		0,3437=$G'(Y)$

Tabella 3.2: Distribuzione della variabile Y' e calcolo degli indici $\tilde{V}_i(Y')$, $V'(Y')$, $G'_i(Y')$ e $G'(Y')$

i	$y_{(i)}$	$Q_i(Y)$	$\bar{M}_i(Y)$	$\tilde{V}_i(Y)$	$\frac{2i}{N+1}$	$G'_i(Y)$
1	2	2	2,00	0,8055	0,13	0,1074
2	2	4	2,00	0,8055	0,27	0,2148
3	4	8	2,67	0,7407	0,40	0,2963
4	4	12	3,00	0,7083	0,53	0,3778
5	4	16	3,20	0,6889	0,67	0,459
6	4	20	3,33	0,6759	0,80	0,5407
7	10	30	4,29	0,5833	0,93	0,5444
8	10	40	5,00	0,5139	1,07	0,5481
9	13	53	5,89	0,4275	1,20	0,5129
10	13	66	6,60	0,3583	1,33	0,4778
11	13	79	7,18	0,3018	1,47	0,4426
12	13	92	7,67	0,2546	1,60	0,4074
13	26	118	9,0769	0,1175	1,73	0,2037
14	26	144	10,28	0,0000	1,87	0,0000
				0,4987= $V'(Y)$		0,3667=$G'(Y)$

Osservando i dati in Tabella 3.1 e in Tabella 3.2 si nota innanzitutto che gli indici $G'(Y)$ e $V'(Y)$, definiti in (2.2) e in (2.8), non sono invarianti a replicazioni nella popolazione. Occorre inoltre osservare che, a unità statistiche con medesimo valore della variabile oggetto di studio, corrispondono in generale indici di ineguaglianza puntuale differenti. Dato che non appare giustificato assegnare misure di ineguaglianza puntuali differenti ad unità che presentano medesino valore della variabile Y oggetto di studio e al fine di poter soddisfare la proprietà di invarianza a replicazioni nella popolazione, si procede a ridefinire ed approssimare l'indice $V'(Y)$ per dati espressi termini di distribuzione di frequenza. A tal proposito, in virtù della (2.8), segue:

$$
\begin{aligned}
V'(Y) &= \frac{1}{n}\sum_{i=1}^{n}\tilde{V}_i(Y)\cdot 1 \\
&= \frac{1}{n}\sum_{h=1}^{r}\left(\sum_{i=1+P_{h-1.}}^{P_{h.}}\tilde{V}_i(Y)\cdot 1\right).
\end{aligned} \tag{3.1}
$$

Al fine di attribuire medesima misura di ineguaglianza puntuale ad unità che presentano uguale valore della variabile oggetto di studio, si porrà:

$$
\tilde{V}_i(Y) = \tilde{V}_{P_{h.}}(Y) = V_h(Y) \qquad \text{per } 1 + P_{h-1.} \leq i \leq P_{h.} \tag{3.2}
$$

Pertanto, in virtù della (3.2) e della relazione $\sum_{i=1+P_{h-1.}}^{P_{h.}} 1 = n_{h.}$, ponendo $P_{0.} = 0$, si ottiene:

$$
\begin{aligned}
V'(Y) &= \frac{1}{n}\sum_{h=1}^{r}\left(\sum_{i=1+P_{h-1.}}^{P_{h.}}\tilde{V}_i(Y)\cdot 1\right) \\
&\simeq \frac{1}{n}\sum_{h=1}^{r}\left(\sum_{i=1+P_{h-1.}}^{P_{h.}}\tilde{V}_{P_{h.}}(Y)\cdot 1\right) \\
&= \frac{1}{n}\sum_{h=1}^{r}V_h(Y)\sum_{i=1+P_{h-1.}}^{P_{h.}} 1 \\
&= \frac{1}{n}\sum_{h=1}^{r}V_h(Y)\cdot n_{h.} = V(Y)
\end{aligned} \tag{3.3}
$$

In virtù della (3.3), si definisce l'indice $V(Y)$

$$
V(Y) = \sum_{h=1}^{r} V_h(Y)\cdot\frac{n_{h.}}{n}, \tag{3.4}
$$

come media ponderata, con pesi $n_{h.}/n$, delle misure puntuali

$$
V_h(Y) = \frac{M(Y) - \bar{M}_{P_{h.}}(Y)}{M(Y)} = \frac{M(Y) - \bar{M}_{h.}(Y)}{M(Y)}. \tag{3.5}
$$

Ovviamente, essendo $\tilde{V}_{P_{h.}}(Y) \leq \tilde{V}_i(Y)$ per $1 + P_{h-1.} \leq i \leq P_{h.}$, segue:

$$
V(Y) \leq V'(Y).
$$

Al fine di chiarire detta procedura, si considerino le distribuzioni di frequenza in Tabella 3.3 e in Tabella 3.4 relative ai dati in Tabella 3.1 e in Tabella 3.2 rispettivamente.

Tabella 3.3: Distribuzione di frequenza della variabile Y in Tabella 3.1 e calcolo degli indici $V_h(Y)$, $V(Y)$, $G_h(Y)$ e $G(Y)$

h	y_h	$n_{h.}$	$P_{h.}$	$Q_{h.}(Y)$	$\bar{M}_h(Y)$	$V_h(Y)$	$\frac{1+2P_{h.}-n_{h.}}{n+1}$	$G_h(Y)$
1	2	1	1	2	2,00	0,8055	0,250	0,2014
2	4	2	3	10	3,33	0,6759	0,625	0,4224
3	10	1	4	20	5,00	0,5139	1,00	0,5139
4	13	2	6	46	7,67	0,2546	1,375	0,3501
5	26	1	7	72	10,28	0,00	1,750	0,00
						0,4544=$V(Y)$		0,2125=$G(Y)$

Tabella 3.4: Distribuzione di frequenza della variabile Y' in Tabella 3.2 e calcolo degli indici $V_h(Y')$, $V(Y')$, $G_h(Y')$ e $G(Y')$

h	y_h	$n_{h.}$	$P_{h.}$	$Q_{h.}(Y')$	$\bar{M}_{h.}(Y')$	$V_h(Y')$	$\frac{1+2P_{h.}-n_{h.}}{n+1}$	$G_h(Y')$
1	2	2	2	4	2,00	0,8055	0,20	0,1611
2	4	4	6	20	3,33	0,6759	0,60	0,4055
3	10	2	8	40	5,00	0,5139	1,00	0,5139
4	13	4	12	92	7,67	0,2546	1,40	0,3565
5	26	2	14	144	10,28	0,00	1,80	0,00
						0,4544=$V(Y')$		0,3141=$G(Y')$

Dato che nella Tabella 3.3 e nella Tabella 3.4 le distribuzioni di Y e di Y' sono espresse in termini di distribuzione di frequenza, le unità aventi medesimo valore della variabile oggetto di studio vengono aggregate in un unico "gruppo". Procedendo in questi termini, a dette unità viene assegnata una media inferiore che, anzichè dipendere dall'ordinamento (arbitrario) con cui vengono graduati inizialmente i dati, dipende dal livello di reddito che esse presentano. Ne consegue che, in virtù della (3.3), a unità con medesimo valore della variabile oggetto di studio corrisponde medesimo scostamento relativo della media inferiore rispetto la media generale e, pertanto, medesima misura di ineguaglianza puntuale $V_h(Y)$.

Tale approccio, in virtù della (1.25), può essere esteso all'indice $G(Y)$ ed assegnare medesimo scostamento relativo della media inferiore dalla media generale $V_h(Y)$ ad unità che presentano stesso valore della variabile considerata. In effetti, adottando l'approccio in (3.2), ponendo $P_{0.} = 0$, in virtù della (1.25) e della relazione

$$\sum_{i=1+P_{h-1.}}^{P_{h.}} i = \frac{(1 + 2P_{h.} - n_{h.}) \cdot n_{h.}}{2},$$

si ricava:

$$
\begin{aligned}
G'(Y) &= \frac{1}{n}\sum_{i=1}^{n} \tilde{V}_i(Y)\cdot\frac{2\,i}{n+1} \\
&= \frac{1}{n}\sum_{h=1}^{r}\left(\sum_{i=1+P_{h-1.}}^{P_{h.}} \tilde{V}_i(Y)\cdot\frac{2\,i}{n+1}\right) \\
&\simeq \frac{1}{n}\sum_{h=1}^{r}\left(\sum_{i=1+P_{h-1.}}^{P_{h.}} \tilde{V}_{P_{h.}}(Y)\cdot\frac{2\,i}{n+1}\right) \\
&= \frac{1}{n}\sum_{h=1}^{r} V_h(Y)\cdot\frac{2}{n+1}\sum_{i=1+P_{h-1.}}^{P_{h.}} i \\
&= \sum_{h=1}^{r} V_h(Y)\cdot\frac{1+2P_{h.}-n_{h.}}{n+1}\cdot\frac{n_{h.}}{n} = G(Y)
\end{aligned}
\tag{3.6}
$$

In virtù della (3.6), ponendo

$$
G_h(Y) = V_h(Y)\cdot\frac{1+2P_{h.}-n_{h.}}{n+1}, \tag{3.7}
$$

è possibile esprimere l'indice $G(Y)$ come media ponderata delle misure $G_h(Y)$ con pesi dati dalle frequenze relative $n_{h.}/n$:

$$
G(Y) = \sum_{h=1}^{r} G_h(Y)\cdot\frac{n_{h.}}{n}. \tag{3.8}
$$

A conclusione del capitolo, si dimostra che l'indice $V(Y)$, definito in (3.3), e l'indice $I(Y)$, definito in (1.40), soddisfano la proprietà di invarianza a replicazioni nella popolazione.

Dimostrazione. Sia Y una variabile distribuita come in (1.3) che assume r valori distinti e sia X una variabile con distribuzione somigliante a quella di Y, con parametro γ intero e maggiore di uno. In virtù della definizione di distribuzione somigliante, segue che anche la variabile X presenta i medesimi r distinti valori di Y, ovvero $x_h = y_h$, $h = 1,..,r$, e che:

$$\begin{aligned}
\overset{-}{M}_{h.}(X) &= \frac{\sum_{t=1}^{h} x_t \cdot n_{t.}(X)}{P_{h.}(X)} \\
&= \frac{\sum_{t=1}^{h} y_t \cdot \gamma \cdot n_{t.}(Y)}{\gamma \cdot P_{h.}(Y)} = \overset{-}{M}_{h.}(Y) \qquad (h = 1, ..., r) \qquad (3.9) \\
\overset{+}{M}_{h.}(X) &= \frac{T(X) - Q_{h.}(X)}{n(X) - P_{h.}(X)} \\
&= \frac{\gamma \cdot T(Y) - \gamma \cdot Q_{h.}(Y)}{\gamma \cdot n(Y) - \gamma \cdot P_{h.}(Y)} = \overset{+}{M}_{h.}(Y) \qquad (h = 1, ..., r-1) \qquad (3.10) \\
\overset{+}{M}_{r.}(X) &= x_r = y_r = \overset{+}{M}_{r.}(Y), \qquad (3.11)
\end{aligned}$$

dove $n_{h.}(X)$ e $n_{h.}(Y)$ indicano le frequenze assolute delle variabili X e Y, rispettivamente, e $P_{h.}(X)$ e $P_{h.}(Y)$ le corrispondenti frequenze assolute cumulate. Pertanto, $\forall\, h = 1, ..., r$ risulta:

$$\overset{-}{M}_h(Y) = \overset{-}{M}_h(X); \qquad \overset{+}{M}_h(X) = \overset{+}{M}_h(X). \qquad (3.12)$$

In virtù della (1.37) segue $M(X) = M(Y)$, mentre, in virtù della (3.5) e della (3.12) segue:

$$V_h(X) = \frac{M(X) - \overset{-}{M}_{h.}(X)}{M(X)} = \frac{M(Y) - \overset{-}{M}_{h.}(Y)}{M(Y)} = V_h(Y). \qquad (3.13)$$

Pertanto, in virtù della (3.3), $V(X) = V(Y)$.
In modo analogo, in virtù della (1.39) e della (3.12):

$$I_h(X) = \frac{\overset{+}{M}_{h.}(X) - \overset{-}{M}_{h.}(X)}{\overset{+}{M}_{h.}(X)} = \frac{\overset{+}{M}_{h.}(Y) - \overset{-}{M}_{h.}(Y)}{\overset{+}{M}_{h.}(Y)} = I_h(Y) \qquad (3.14)$$

e quindi, in virtù della (1.40), $I(X) = I(Y)$. □

Capitolo 4

Scomposizione degli indici di ineguaglianza

La scomposizione di un indice di ineguaglianza ha come obiettivo quello di rispondere alle seguenti domande.

1. Data una variabile $Y = \sum_{j=1}^{c} X_j$ a valori non negativi, somma di c variabili non negative X_j, $j = 1, ..., c$, in che misura la variabile (fonte) X_j contribuisce alla ineguaglianza valutata nella distribuzione di Y?

2. Data una variabile Y a valori non negativi rilevata in una popolazione divisa in k sottopopolazioni disgiunte, qual'è il contributo della sottopopolazione g, $g = 1, ..., k$, all'ineguaglianza valutata nell'intera popolazione?

In risposta a tali quesiti, a partire dal 1960 vengono proposti metodi per scomporre un indice di concentrazione per *fonti* (in risposta alla domanda 1) e per *sottopopolazioni* (in risposta alla domanda 2).
Nella sezione che segue, saranno ricordati i principali contributi di scomposizione per sottopopolazioni degli indici di ineguaglianza di Gini, Bonferroni e Zenga, mentre, nella sezione successiva, sarà proposta una nuova procedura di scomposizione per sottopopolazioni dell'indice di Bonferroni $V(\cdot)$.

4.1 I principali contributi alla scomposizione per sottopopolazioni dell' indice di ineguaglianza di Gini

Il primo tentativo [8] di scomposizione dell'indice di Gini è dovuto a Soltow che, in virtù della (1.21), nel 1960 propone una scomposizione per sottopopolazioni a partire dalla differenza media di Gini con ripetizione ${}_R\Delta$:

$$ {}_R\Delta = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\left|y_{(i)} - y_{(j)}\right| = \frac{n-1}{n}\cdot\Delta \qquad (i,j = 1,...,n). \tag{4.1} $$

A tal riguardo, si consideri una popolazione di ampiezza n divisa in k sottopopolazioni disgiunte di numerosità $n_{.g}$: $\sum_{g=1}^{k} n_{.g} = n$, $g = 1,...,k$ e si ponga:

$$ p_g = \frac{n_{.g}}{n} $$

$$ M_g(Y) = \frac{1}{n_{.g}}\cdot\sum_{t=1}^{n_{.g}} y_{tg} \tag{4.2} $$

$$ {}_R\Delta_{gg} = \frac{1}{n_{.g}^2}\sum_{t=1}^{n_{.g}}\sum_{v=1}^{n_{.g}}\left|y_{(t)} - y_{(v)}\right| \tag{4.3} $$

$$ {}_R\Delta_{g\bar{g}} = \frac{1}{n_{.g}\cdot n_{.\bar{g}}}\sum_{t=1}^{n_{.g}}\sum_{j=1}^{n-n_{.g}}\left|y_{(t)} - y_{(i)}\right|, \tag{4.4} $$

In questi termini: p_g indica la quota di popolazione della sottopopolazione g, $M_g(Y)$ la media della variabile Y nella sottopopolazione g, ${}_R\Delta_{gg}$ la differenza media di Gini con ripetizione calcolata tra le unità statistiche della sottopopolazione g e ${}_R\Delta_{g\bar{g}}$ la differenza media di Gini con ripetizione calcolata tra le $n_{.g}$ unità statistiche della sottopopolazione g e le rimanenti $n - n_{.g}$ unità statistiche considerate come un unico sottogruppo. In virù delle deifinizioni da (4.1) a (**??**) è possibile porre:

$$ \begin{aligned} {}_R\Delta &= \frac{1}{n^2}\sum_{g=1}^{k} {}_R\Delta_{gg}\cdot n_{.g}^2 + \frac{1}{n^2}\sum_{g=1}^{k}\Delta_{g\bar{g}}\cdot n_{.g}\cdot(n - n_{.g}) \\ &= \sum_{g=1}^{k} {}_R\Delta_{gg}\cdot p_g^2 + \sum_{g=1}^{k}\Delta_{g\bar{g}}\cdot p_g\cdot(1-p_g). \end{aligned} \tag{4.5} $$

In virtù della (4.5) e delle relazioni (4.1) e (1.21) , Soltow scompone l'indice di Gini in una componente *within* ed in una componente *between*:

$$ \frac{{}_R\Delta}{2\cdot M(Y)} = \frac{1}{2\cdot M(Y)}\sum_{g=1}^{k} {}_R\Delta_{gg}\cdot p_g^2 + \frac{1}{2\cdot M(Y)}\sum_{g=1}^{k}\Delta_{g\bar{g}}\cdot p_g\cdot(1-p_g) $$

$$ = \sum_{g=1}^{k} {}_R\Delta_{gg}\cdot p_g\cdot\frac{2\cdot n_{.g}\cdot M_g(Y)}{2\cdot n\cdot M(Y)} + \tag{4.6} $$

$$ + \frac{\sum_{g=1}^{k}\Delta_{g\bar{g}}}{M_g(Y) + M_{\bar{g}}(Y)}\cdot p_g\cdot(1-p_g)\cdot\frac{M_g(Y) + M_{\bar{g}}(Y)}{2\cdot M(Y)}, \tag{4.7} $$

dove la componente *within* è indicata nella (4.6) in quanto si considerano differenze medie con ripetizione calcolate all'interno di ogni sottopopolazione mentre la componente *between* è indicata nella (4.7) in quanto si considerano differenze medie con ripetizione calcolate tra le unità statistiche di una sottopopolazione g e le rimanenti $n - n_{.g}$ unità statistiche appartenenti alle rimanenti $k-1$ sottopopolazioni.

Nel 1967, Bhattacharya e Mahalanobis [2] propongono una scomposizione per sottopopolazioni dell'indice di Gini basata anch'essa sulla scomposizione della differenza media di Gini con ripetizione ${}_R\Delta$ in (4.1).

Sia ${}_R\Delta_{g\ell}$ la differenza media di Gini con ripetizione calcolata tra le unità statistiche della sottopopolazione g e ℓ, $(g, \ell = 1, ..., k;\ g \neq \ell)$:

$${}_R\Delta_{g\ell} = \frac{1}{n_{.g} \cdot n_{.\ell}} \sum_{t=1}^{n_{.g}} \sum_{j=1}^{n_{.\ell}} \left| y_{(t)} - y_{(j)} \right| (t = 1, ..., n_{.g};\ j = 1, ..., n_{.\ell}). \tag{4.8}$$

Dato che,

$${}_R\Delta = \sum_{g=1}^{k} {}_R\Delta_{gg} \cdot p_g^2 + \sum_{g=1}^{k} \sum_{\ell \neq g}^{k} {}_R\Delta_{g\ell} \cdot p_g \cdot p_\ell, \tag{4.9}$$

gli Autori procedono definendo la quantità ${}_R\Delta_B$,

$${}_R\Delta_B = \sum_{g=1}^{k} \sum_{\ell \neq g}^{k} |M_g(Y) - M_\ell(Y)| \cdot p_g \cdot p_\ell, \tag{4.10}$$

come *componente "between"* dell'indice ${}_R\Delta$ e derivano, per differenza, la quantità ${}_R\Delta_W$,

$$\begin{aligned} {}_R\Delta_W &= {}_R\Delta - {}_R\Delta_B \\ &= \sum_{g=1}^{k} {}_R\Delta_{gg} \cdot p_g^2 + \sum_{g=1}^{k} \sum_{\ell \neq g}^{k} \left({}_R\Delta_{g\ell} - |M_g(Y) - M_\ell(Y)| \right) \cdot p_g \cdot p_\ell, \end{aligned} \tag{4.11}$$

definendola come *componente "within"* dell'indice ${}_R\Delta$.

Sia S_Y l'intervallo in $\mathbb{R}^+ \cup \{0\}$ che contiene i valori assunti dalla variabile Y rilevati nella popolazione di riferimento e sia S_{Y_g} l'intervallo in $\mathbb{R}^+ \cup \{0\}$ che contiene i valori assunti dalla variabile Y_g, ovvero che contiene i valori della variabile Y rilevati nella sottopopolazione g. Nel solo caso in cui le sottopopolazioni oltre ad essere disgiunte, sono anche *non sovrapposte*, ovvero se $S_{Y_g} \cap S_{Y_\ell} = \emptyset$, per $g \neq \ell$ $g, \ell = 1, ..., k$, si verifica,

$${}_R\Delta_{g\ell} = |M_g(Y) - M_\ell(Y)|, \tag{4.12}$$

e pertanto, essendo nullo il secondo addendo in (4.11), si ha:

$$ {}_R\Delta_W = \sum_{g=1}^{k} {}_R\Delta_{gg} \cdot p_g^2. \tag{4.13} $$

In virtù delle (4.11), (4.12) e (4.13), qualora, si ribadisce, i gruppi oltre ad essere disgiunti, sono anche non sovrapposti, gli Autori pervengono alla seguente scomposizione:

$$ \begin{aligned} {}_R\Delta &= \sum_{g=1}^{k} {}_R\Delta_{gg} \cdot p_g^2 + \sum_{g=1}^{k} \sum_{\ell \neq g}^{k} ({}_R\Delta_{g\ell} - |M_g(Y) - M_\ell(Y)|) \cdot p_g \cdot p_\ell + \\ &\quad + \sum_{g=1}^{k} \sum_{\ell \neq g}^{k} |M_g(Y) - M_\ell(Y)| \cdot p_g \cdot p_\ell \\ &= \sum_{g=1}^{k} {}_R\Delta_{gg} \cdot p_g^2 + \sum_{g=1}^{k} \sum_{\ell \neq g}^{k} |M_g(Y) - M_\ell(Y)| \cdot p_g \cdot p_\ell \\ &= {}_R\Delta_W + {}_R\Delta_B \end{aligned} \tag{4.14} $$

Qualora le sottopopolazioni fossero invece sovrapposte, ovvero $S_{Y_g} \cap S_{Y_\ell} \neq \emptyset$,

$$ {}_R\Delta_W = \sum_{g=1}^{k} {}_R\Delta_{gg} \cdot p_g^2 + \sum_{g=1}^{k} \sum_{\ell \neq g}^{k} {}_R\Delta_{g\ell} \cdot p_g \cdot p_\ell - \sum_{g=1}^{k} \sum_{\ell \neq g}^{k} |M_g(Y) - M_\ell(Y)| \cdot p_g \cdot p_\ell, \tag{4.15} $$

e pertanto, la scomposizione proposta, anzichè fornire informazioni sull'ineguaglianza derivata dal confronto tra unità nella stessa sottopopolazione e tra unità in sottopopolazioni differenti, consta di un terzo termine detrattivo che, come ricordato da Radaelli [8], pur essendo parte della definizione della componente within, dipende da quantità riferite a sottopopolazioni differenti. Nello stesso articolo, Radaelli [8] fa inoltre notare che la procedura proposta da Bhattacharya e Mahalanobis si basa su una definizione *a priori* di componente "between" [8]. Infine, dato che in tale metodologia si procede scomponendo ${}_R\Delta$, ovvero l'indice sintetico di Gini $\tilde{R}(Y)$ in virtù delle relazioni in (1.21) e (4.1), la proposta di Bhattacharya e Mahalanobis non è in grado di valutare in modo diretto l'ineguaglianza presente in, o escludendo, parti di popolazione, ovvero, non è in grado di rilevare l'ineguaglianza nei diversi quantili della distribuzione.

Nel 1969, nell'articolo "Decompositions of Concentration Ratio", Rao [9] propone due tipi di scomposizioni dell'indice sintetico di Gini: una scomposizione per sottopopolazioni e una scomposizione per fonti. Impiegando la "*regola dei trapezi*" [1] per il calcolo dell'indice sintetico di Gini ed impiegando dati espressi come in (1.27), ovvero in termini di distribuzione di frequenza, Rao definisce il *rapporto di concentrazione* c : [2]

[1] Vedere Appendice A
[2] Vedere Appendice A

$$
\begin{aligned}
c &= \sum_{h=1}^{r-1} \left(p_{h.} \cdot q_{(h+1).} - p_{(h+1).} \cdot q_{h.}\right) \\
&= \frac{{}_R\Delta}{2 \cdot M(Y)},
\end{aligned}
\tag{4.16}
$$

da cui segue, in virtù della (1.22), $\tilde{R}(Y) = \dfrac{n}{n-1} \cdot c.$

Detto n_{hg} il numero delle unità nella sottopopolazione g tali che $Y = y_h$ e posto:

$$P_{hg} = \sum_{t=1}^{h} n_{tg} \tag{4.17}$$

$$p_{hg} = \frac{P_{hg}}{n} \tag{4.18}$$

$$Q_{hg}(Y) = \sum_{t=1}^{h} y_t \cdot n_{tg} \tag{4.19}$$

$$q_{hg}(Y) = q_{hg} = \frac{Q_{hg}(Y)}{T(Y)} \tag{4.20}$$

$$q_g = \frac{\sum_{h=1}^{r} y_h \cdot n_{hg}}{\sum_{h=1}^{r} y_h \cdot n_{h.}} = \frac{Q_{rg}(Y)}{T(Y)} = \frac{T_g(Y)}{T(Y)} \tag{4.21}$$

$$p_{hg}^{(n)} = \frac{p_{hg}}{p_g} \tag{4.22}$$

$$q_{hg}^{(n)} = \frac{q_{hg}}{q_g} \tag{4.23}$$

$$c_{g\ell} = \sum_{h=1}^{r} \left(p_{hg}^{(n)} \cdot q_{(h+1)\ell}^{(n)} - p_{(h+1)g}^{(n)} \cdot p_{h\ell}^{(n)}\right) \qquad (g, \ell = 1, ..., k) \tag{4.24}$$

$$a_h = \frac{q_{h.}}{p_{h.}} \tag{4.25}$$

$$b_g = \frac{q_g}{p_g} \tag{4.26}$$

segue:

$$p_{h.} = \sum_{g=1}^{k} p_{hg}; \qquad q_{h.} = \sum_{g=1}^{k} q_{hg}. \tag{4.27}$$

In virtù della relazione in (4.22) e (4.23), essendo,

$$p_{hg} = p_p \cdot p_{hg}^{(n)}, \qquad q_{hg} = q_g \cdot q_{hg}^{(n)}$$

Rao ottiene [3] la seguente definizione dell'indice c:

$$\begin{aligned} c &= \sum_{g=1}^{k}\sum_{\ell=1}^{k} p_g \cdot q_\ell \cdot c_{g\ell} \\ &= \underset{(1\times k)}{\mathbf{p'_g}} \underset{(k\times k)}{\mathbf{c_{g\ell}}} \underset{(k\times 1)}{\mathbf{q_g}} \end{aligned} \tag{4.28}$$

dove, $\mathbf{p'_g}$ indica il vettore trasposto di lunghezza k i cui elementi p_g indicano la quota di popolazione presente nel gruppo g; $\mathbf{c_{g\ell}}$ indica la matrice di dimensione $(k \times k)$ i cui elementi $c_{g\ell}$, indicano i rapporti di concentrazione calcolati considerando le unità statistiche nelle sottopopolazioni g e ℓ, $(g, \ell = 1, ..., k)$ e $\mathbf{q_g}$ indica il vettore di lunghezza k i cui elementi q_g indicano la quota della variabile Y nella sottopopolazione g. Infine, posto $\mathbf{c_{gg}}$ vettore colonna di lunghezza k di elementi c_{gg} e definiti,

$$\underset{(k\times k)}{\mathbf{d_{g\ell}}} = \mathbf{c_{g\ell}} - \mathbf{c_{gg}} \tag{4.29}$$

$$\underset{(k\times k)}{\mathbf{w_{g\ell}}} = \mathbf{c_{g\ell}} - \mathbf{d_{g\ell}} \tag{4.30}$$

Rao ottiene la seguente scomposizione [4]:

$$c = \underset{(1\times k)}{\mathbf{p}'_g} \underset{(k\times 1)}{\mathbf{c}_{gg}} + \underset{(1\times k)}{\mathbf{p}'_g} \underset{(k\times k)}{(\mathbf{b}_\ell \mathbf{d}_{g\ell})} \underset{(k\times 1)}{\mathbf{p}_g} . \tag{4.31}$$

$$= \sum_{g=1}^{k} p_g \cdot c_{gg} + \sum_{g=1}^{k}\sum_{\ell=1}^{k} p_g \cdot p_\ell \cdot (b_g \cdot d_{\ell g} + b_\ell \cdot d_{g\ell}) \tag{4.32}$$

$$= \sum_{g=1}^{k} p_g \cdot c_{gg} + \sum_{g=1}^{k}\sum_{\ell=1}^{k} p_g \cdot p_\ell \cdot \left[\frac{q_g}{p_g} \cdot (c_{\ell g} - c_{\ell\ell}) + \frac{q_\ell}{p_\ell} \cdot (c_{g\ell} - c_{gg})\right], \tag{4.33}$$

con $(\mathbf{b}_g \mathbf{p}_g)$ vettore di lunghezza k di elementi $b_g \cdot p_g = q_g$ e $(\mathbf{b}_\ell \mathbf{d}_{g\ell})$ matrice di dimensione $k \times k$ di elementi $b_\ell \cdot d_{g\ell}$.
In virtù della (8.13) e della (8.14) [5], ed osservando che, per $g = \ell$, $p_g \cdot p_g \cdot (b_g \cdot d_{gg} + b_g \cdot d_{gg}) = 0$, essendo $d_{gg} = 0$ per costruzione, Rao perviene alla (4.32), proponendo dunque una scomposizione per sottopopolazioni definita come somma tra:

- una componente *within*, definita da $\mathbf{p'_g c_{gg}}$, media pesata dei rapporti di concentrazione c_{gg} calcolati nelle distribuzioni delle singole sottopopolazioni, con pesi dati dalla quota di popolazione p_g presente nella $g - esima$ sottopopolazione;

- una componente *between*, definita da $\mathbf{p}'_g(\mathbf{b}_\ell \mathbf{d}_{g\ell})\mathbf{p}_g$, media pesata delle quantità $b_g \cdot d_{\ell g} + b_g \cdot d_{g\ell}$, con pesi dati dai rispettivi "pesi demografici" p_g e p_ℓ.

[3] Vedere Appendice B, sezione B.1
[4] Vedere Appendice B, sezione B.2
[5] Vedere Appendice B, sezione B.2

In relazione alla scomposizione proposta, tuttavia, se da una parte appare ragionevole assumere come componente within la media pesata delle misure di ineguaglianza c_{gg}, perchè misure di ineguaglianza calcolate nelle distribuzioni delle singole sottopopolazioni, in virtù della (4.33) risulta invece poco chiara la definizione di componente between.[6] Infine, come ricordato in relazione alla proposta di Bhattacharya e Mahalanobis, anche questa scomposizione non permette di rilevare l'ineguaglianza presente nei diversi quantili della distrbuzione.

Nel 1989 Silber [11] propone una scomposizione per sottopopolazioni dell'indice di Gini basata sulla scomposizione di una matrice G introdotta, secondo l'autore, per agevolarne il calcolo. Precisamente, il metodo proposto permette di calcolare l'indice c introdotto in (4.16) e di dedurre l'indice di Gini tramite la nota relazione $\tilde{R}(Y) = n/(n-1) \cdot c$.

Si consideri una matrice $\mathbf{G}$ di dimensione $(n \times n)$ di elementi g_{ij}, $i, j = 1, ..., n$ tali che:

$$g_{ij} = \begin{cases} 0 & \text{per } i = j \\ 1 & \text{per } i > j \\ -1 & \text{per } i < j. \end{cases}$$

Siano inoltre:

$$s_{(i)} = \frac{y_{(i)}}{T(Y)} \tag{4.34}$$

$$\underset{(n\times 1)}{\mathbf{e}(n)} = [1/n...1/n] \tag{4.35}$$

$$\underset{(n_{.g}\times 1)}{\mathbf{e}(n_{.g})} = [1/n...1/n] \tag{4.36}$$

$$\underset{(n\times 1)}{\mathbf{s}(n)} = \left[s_{(1)}...s_{(n)}\right] \text{ con } s_{(i)} \geq s_{(j)} \text{ per } i < j, \quad i, j = 1, ..., n \tag{4.37}$$

[6]Quest'ultima rappresenta la media pesata delle differenze tra la misura di ineguaglianza $c_{g\ell}$ nelle distribuzioni g ed ℓ meno la misura di ineguaglianza c_{gg} calcolata nella sottopopolazione g con pesi dati dal prodotto dei rispettivi pesi demografici p_g e p_ℓ e dalla quota q_g/p_g.

$$\underset{(n\times 1)}{\nu(n)} = \left[s_{(tg)}...s_{(t\ell)}\right] \text{ con: } \begin{cases} t = 1, ..., n_{.g}; \\ g, \ell = 1, ..., k; \\ (g,\ell) : M_g(Y) \geq M_\ell(Y) \end{cases} \tag{4.38}$$

$$\underset{(n_{.g}\times 1)}{\nu(n_{.g})} = \left[s_{(1g)}...s_{(n_{.g}g)}\right] \tag{4.39}$$

$$\underset{(n\times 1)}{\mathbf{d}(n)} = \mathbf{s}(n) - \nu(n). \tag{4.40}$$

Più preciasamente, il vettore $\nu(n)$ è il vettore di elementi $s_{(i)}$, ma a differenza del vettore $\mathbf{s}(n)$, i cui elementi $s_{(i)}$ sono ordinati in modo decrescente indipendentemente dall'appartenenza di una unità statistica ad una particolare sottopopolazione, nel vettore $\nu(n)$ vi è un "doppio ordinamento": le unità, e quindi le quote $y_{(tg)}/T(Y)$, sono raggruppate dapprima in base alla sottopopolazione cui appartengono ed ordinate in modo che $M_g(Y) \geq M_\ell(Y)$, $g, \ell = 1, ..., k$; successivamente, all'interno di ogni sottopopolazione, dette unità sono ordinate in modo decrescente. In virtù delle definizioni date, Silber ricava,

$$c = \underset{(1\times n)}{\mathbf{e}'(n)}\,\underset{(n\times n)}{\mathbf{G}(n;n)}\,\underset{(n\times 1)}{\mathbf{s}(n)} = \underset{(1\times n)}{\mathbf{e}'(n)}\,\underset{(n\times n)}{\mathbf{G}(n)}\,\underset{(n\times 1)}{\nu(n)} \tag{4.41}$$

$$= \sum_{g=1}^{k} \underset{(1\times n_{.g})}{\mathbf{e}'(n_{.g})}\,\underset{(n_{.g}\times n_{.g})}{\mathbf{G}(n_{.g};n_{.g})}\,\underset{(n_{.g}\times 1)}{\nu(n_{.g})} + \tag{4.42}$$

$$+ \sum_{g=1}^{k}\sum_{\ell\neq g} \underset{(1\times n_{.g})}{\mathbf{e}'(n_{.g})}\,\underset{(n_{.g}\times n_{.\ell})}{\mathbf{G}(n_{.g};n_{.\ell})}\,\underset{(n_{.\ell}\times 1)}{\nu(n_{.\ell})} + \tag{4.43}$$

$$+ \underset{(1\times n)}{\mathbf{e}'(n)}\,\underset{(n\times n)}{\mathbf{G}(n;n)}\,\underset{(n\times 1)}{\mathbf{d}(n)} \tag{4.44}$$

dove: la (4.42) è indicata dall'Autore come componente *within* in quanto considera quantità relative a una medesima sottopopolazione, la (4.43) è indicata come componente *between* in quanto considera quantità relative a sottopopolazioni differenti, mentre la (4.44) è indicata come *fattore di interazione* e misura l'effetto dell'ordinamento necessario al colcolo dell'indice di Gini. Più precisamente, essa indica "una misura dell'intensità delle permutazioni che si rileva quando anzichè ordinare i valori $s_{(i)}$ in ordine decrescente, l'ordinamento avviene dapprima ordinando le unità per valori decrescenti delle medie $M_g(Y)$ delle sottopopolazioni cui esse appartengono e successivamente per valori decrescenti di $s_{(tg)}$" [11]. Ponendo l'attenzione sulla componente within in (4.42) e sulla componente between in (4.43) e posto il fattore di interazione $\mathbf{e}'(n)\,\mathbf{G}(n;n)\,\mathbf{d}(n) = \Psi$, la scomposizione in (4.41) risulta:

$$c = \sum_{g=1}^{k}\sum_{t=1}^{n_{.g}} \frac{(1-2t+n_{.g})}{n} \cdot q_{tg} + \sum_{g=1}^{k}\sum_{g\neq\ell} (-1)^{\mathbb{1}(g,\ell)} \cdot \frac{n_{.g}}{n} \cdot q_\ell + \Psi \tag{4.45}$$

dove, $\forall\, g, \ell = 1, ..., k,$

$$q_{(tg)} = \frac{y_{(tg)}}{T(Y)}; \quad q_\ell = \frac{\sum_{t=1}^{n_{.\ell}} y_{(t\ell)}}{T(Y)}; \quad \mathbb{1}(g,\ell) = \begin{cases} 1 & \text{se } \ell > g \\ 0 & \text{altrove.} \end{cases} \tag{4.46}$$

In virtù della (4.45) è possibile osservare che il termine within risulta essere la somma pesata delle quote q_{tg} con pesi $(1 - 2t + n_{.g})/n$, mentre la componente between, risulta essere la somma pesata della quota q_ℓ con pesi pari a $n_{.g}/n$, dove il segno è dato dal confronto tra g ed ℓ definito in (4.46). In virtù della (4.45), pertanto, oltre alla presenza di un terzo termine di "interazione" Ψ che rende meno immediata l'interpretazione della scomposizione proposta, anche l'interpretazione delle componenti within e between, non essendo espresse in termini di medie, appare poco immediata. Infine, in analogia alle scomposizioni qui presentate dell'indice di Gini, non è possibile valutare l'ineguaglianza prensente nei quantili della distribuzione essendo, anche in questo caso, una scomposizione dell' indice di ineguaglianza sintetico. Seguendo tale approccio, nel 2012 Silber [11] proporrà anche per l'indice di Bonferroni una scomposizione per fonti e per sottopopolazioni.

4.2 I principali contributi alla scomposizione per sottopopolazioni dell' indice di ineguaglianza di Bonferroni

Nel 1990, Agostino Tarsitano [12] propone una scomposizione per sottopopolazioni dell'indice di Bonferroni. Posto:

$$F_{hg} = \frac{P_{hg}}{n_{.g}} \tag{4.47}$$

$$F_{h.} = \frac{Ph.}{n} \tag{4.48}$$

$$T_g(Y) = \sum_{t=1}^{n_{.g}} y_{tg} \tag{4.49}$$

$$M_g(Y) = \frac{T_g(Y)}{n_{.g}} \tag{4.50}$$

$$h_g = \frac{T_g(Y)}{T(Y)} = \frac{n_{.g} \cdot M_g(Y)}{n \cdot M(Y)} = \frac{M_g(Y) \cdot p_g}{M(Y)} \tag{4.51}$$

$$\tilde{V}_g(Y) = \frac{1}{r-1} \sum_{h=1}^{r} \frac{M_g(Y) - \bar{M}_{hg}(Y)}{M_g(Y)}, \tag{4.52}$$

Tarsitano mostra che,

$$\tilde{V}(Y) = B_W + B_A \tag{4.53}$$

$$= \sum_{g=1}^{k} h_g \cdot \tilde{V}_g(Y) + \sum_{g=1}^{k} \frac{h_g}{M_g(Y)} \cdot \sum_{h=1}^{r-1} \left(1 - \frac{F_{hg}}{F_{h.}}\right) \cdot \bar{M}_{hg}(Y), \tag{4.54}$$

dove,

$$B_W = \sum_{g=1}^{k} h_g \cdot \tilde{V}_g(Y) \tag{4.55}$$

$$B_A = \sum_{g=1}^{k} \frac{h_g}{M_g(Y)} \cdot \sum_{h=1}^{r-1} \left(1 - \frac{F_{hg}}{F_{h.}}\right) \cdot \bar{M}_{hg}(Y). \tag{4.56}$$

In virtù delle (4.55) e (4.56), Tarsitano scompone l'indice di Bonferroni in un termine "within" e un termine "across": il termine within B_W, dato dalla media ponderata degli indici di Bonferroni calcolati nelle k sottopopolazioni con pesi dati dal prodotto tra il peso relativo delle sottopopolazioni p_g e la quota relativa h_g; il termine across, dato dalla media ponderata di quantità differenti. Per comprendere meglio il termine "across", si consideri il caso in cui vi è ineguaglianza nulla all'interno delle sottopopolazioni, ovvero, $y_{tg} = M_g(Y)$, $\forall\, t = 1, ..., n_{.g}$; $g = 1, ..., k$. In questo caso, $B_W = 0$ e,

$$V(Y) = \sum_{g=1}^{k} h_g \cdot \left[\frac{1}{r-1} \sum_{h=1}^{r} \left(1 - \frac{F_{hr}}{F_{h.}}\right)\right] = B_B. \tag{4.57}$$

In virtù della (4.57) e dell'assunto di ineguaglianza nulla all'interno delle singole sottopopolazioni, l'indice di ineguaglianza di Bonferroni risulta definito sulla base dell'ineguaglianza presente tra le sottopopolazioni che compongono la popolazione complessiva e pertanto il termine B_B viene interpretato come componente between dell'indice sintetico $\tilde{V}(Y)$ e ad esso coincide nel caso in cui la componente within è nulla. Si consideri invece, il caso in cui le frequenze soddisfano la condizione di indipendenza statistica, ovvero, il caso in cui

$$n_{hg} = \frac{n_{h.} \cdot n_{.g}}{n}; \qquad P_{hg} = n_{h.} \cdot p_g; \qquad \frac{F_{hg}}{F_{h.}} = 1 \tag{4.58}$$

e le distribuzioni condizionate sono uguali alla distribuzione totale. In questo caso, $\tilde{V}(Y) = B_W$ e di conseguenza, l'indice di bonferroni risulta essere la media aritmetica degli indici B_W. In termini generali, la componente B_A può essere espressa come,

$$B_A = B_B + B_I, \tag{4.59}$$

dove:

$$B_I = \frac{1}{r-1} \sum_{h=1}^{r-1} \sum_{g=1}^{k} h_g \cdot \left(1 - \frac{\bar{M}_{hg}(Y)}{M_g(Y)}\right) \cdot \left(\frac{F_{hg}}{F_{h.}} - 1\right) \tag{4.60}$$

indica un termine di interazione il cui valore dipende dal legame con cui il rango di una unità statistica è "influenzato" dal gruppo di appartenenza. La scomposizione proposta consta dunque di tre termini additivi:

$$\tilde{V}(Y) = B_W + B_B + B_I \tag{4.61}$$

Come sottolinea l'Autore stesso, a causa del fattore $F_{hg}/F_{h.}$ in B_A, la conoscenza delle caratteristiche aggregate $\tilde{V}_g(Y), p_g, M_g(Y)$ non è sufficiente per calcolare l'indice $\tilde{V}(Y)$ e, di conseguenza, l'indice $\tilde{V}(Y)$, non essendo "aggregativo", non è scomponibile per sottopopolazioni.

Nel 2012, in analogia ai risultati ottenuti da Silber nel 1989 [11] in relazione all'indice di Gini, Silber e Martin [4] propongono una scomposizione dell'indice di Bonferroni $V(Y)$ basata sulla scomposizione della matrice $\mathbf{B}$, una matrice di dimensione $(n \times n)$ di elementi b_{ij}, $i, j = 1, ..., n$, tali che:

$$b_{ij} = \begin{cases} 0 & \text{se } i = j \\ -n/i & \text{se } j < i \\ n/j & \text{se } j > i, \end{cases}$$

da cui segue, $b_{ij} = -b_{ji}$.
A tal proposito, siano inoltre:

1. $\mathbf{s}(n)$ il vettore di lunghezza n di elementi graduati $s_{(i)} : s_{(i)} \leq s_{(j)}$ per $i < j$, $i, j = 1, ..., n$ con $s_{(i)} = y_{(i)}/T(Y)$;

2. $\tau(n)$ il vettore di lunghezza n di elementi $s_{(i)}$ ordinati, dapprima, per valori cresenti della media del gruppo di appartenenza $M_g(Y)$ e, successivamente, in modo crescente in base al vero valore $y_{(i)}$ rilevato;

3. $\bar{\tau}(n)$, il vettore di lunghezza n di elementi $\bar{s}_{(g)} = \dfrac{M_g(Y)}{T(Y)}$ con $n_{.1}$ elementi $\bar{s}_{(1)}$, $n_{.2}$ elementi $\bar{s}_{(2)}$ e così via, tale che $\bar{s}_{(1)} \leq ... \leq \bar{s}_{(g)} \leq ... \leq \bar{s}_{(\ell)} \leq ... \leq \bar{s}_{(k)}$, per $M_g(Y) \leq M_\ell(Y)$, $g, \ell = 1, ..., k$

In virù di dette definizioni, Silber e Martin scompongono l'indice di Bonferroni nel modo seguente:

$$V(Y) = \underset{(1\times n)}{\mathbf{e}'(n)}\,\underset{(n\times n)}{\mathbf{B}(n;n)}\,\underset{(n\times 1)}{\mathbf{s}(n)} \tag{4.62}$$

$$= \underset{(1\times n)}{\mathbf{e}'(n)}\,\underset{(n\times n)}{\mathbf{B}(n;n)}\,\underset{(n\times 1)}{\tau(n)} + \underset{(1\times n)}{\mathbf{e}'(n)}\,\underset{(n\times n)}{\mathbf{B}(n;n)}\,\underset{(n\times 1)}{\mathbf{s}(n)} - \underset{(1\times n)}{\mathbf{e}'(n)}\,\underset{(n\times n)}{\mathbf{B}(n;n)}\,\underset{(n\times 1)}{\tau(n)} \tag{4.63}$$

$$= \left[\underset{(1\times n)}{\mathbf{e}'(n)}\,\underset{(n\times n)}{\mathbf{B}(n;n)}\,\underset{(n\times 1)}{\tau(n)} - \underset{(1\times n)}{\mathbf{e}'(n)}\,\underset{(n\times n)}{\mathbf{B}(n;n)}\,\underset{(n\times 1)}{\overline{\tau}(n)}\right] + \tag{4.64}$$

$$+ \left[\underset{(1\times n)}{\mathbf{e}'(n)}\,\underset{(n\times n)}{\mathbf{B}(n;n)}\,\underset{(n\times 1)}{\overline{\tau}(n)}\right] + \tag{4.65}$$

$$+ \left[\underset{(1\times n)}{\mathbf{e}'(n)}\,\underset{(n\times n)}{\mathbf{B}(n;n)}\,\underset{(n\times 1)}{\mathbf{s}(n)} - \underset{(1\times n)}{\mathbf{e}'(n)}\,\underset{(n\times n)}{\mathbf{B}(n;n)}\,\underset{(n\times 1)}{\tau(n)}\right] \tag{4.66}$$

dove il termine $\mathbf{e}'(n)\,\mathbf{B}(n;n)\,\overline{\tau}(n)$, in (4.64) è definito dagli Autori componente *between* dell'indice $V(Y)$. In analogia con quanto fatto in relazione alle proposte precedenti, può essere utile esprimere la (4.62) in termini di scalari. A tal proposito, la scomposizione della (4.62) risulta:

$$V(Y) = \mathbf{e}'(n)\,\mathbf{B}(n;n)\,\mathbf{s}(n) \tag{4.67}$$

$$= \left\{\sum_{g=1}^{k}\left[\sum_{t=1}^{n_{.g}} s_{(tg)}\left(\sum_{i=1}^{n}\sum_{\substack{j=1\\ j\neq i}}^{n}(-1)^{\mathbb{1}(i,j)}\frac{1}{j}\right)\right] + \right.$$

$$\left. -\sum_{g=1}^{k} M_g(Y)\sum_{i=1+N_{g-1}}^{N_g}\left((-1)^{\mathbb{1}(i,j)}\frac{1}{j}\right)\right\} + \tag{4.68}$$

$$+ \left\{\sum_{g=1}^{k} M_g(Y)\sum_{i=1+N_{g-1}}^{N_g}\left((-1)^{\mathbb{1}(i,j)}\frac{1}{j}\right)\right\} + \tag{4.69}$$

$$+ \left\{\sum_{i=1}^{n}\left(s_{(i)}\sum_{\substack{j=1\\ j\neq i}}^{n}(-1)^{\mathbb{1}(i,j)}\frac{1}{j}\right) - \sum_{g=1}^{k} M_g(Y)\sum_{g=1+N_{g-1}}^{N_g}\left((-1)^{\mathbb{1}(i,j)}\frac{1}{j}\right)\right\}, \tag{4.70}$$

dove,

$$\mathbb{1}(i,j) = \begin{cases} 1 & \text{se } j < i \\ 0 & \text{se } j > i \end{cases} \tag{4.71}$$

I termini in (4.68), (4.69) e (4.70) coincidono ovviamente con le componenti in (4.64), (4.65) e (4.66), rispettivamente e, come nella procedura di scomposizione dell'indice di Gini del 1989 proposta da Silber [11], tale approccio, oltre a non fornire una interpretazione immediata dei diversi termini non permette nè di valutare l'ineguaglianza puntuale, nè di valutare il contributo delle sottopopolazioni all'indice di ineguaglianza complessivo.

4.3 Scomposizione per sottopopolazioni dell' indice di ineguaglianza di Zenga

Nel 2015 Zenga propone una scomposizione per sottopopolazioni dell'indice di ineguaglianza puntuale $I_h(Y)$ e sintetico $I(Y)$ da lui introdotti. Sfruttando la proprietà associativa della media e un approccio, definito dall'Autore, "a due fasi", che prevede, nella prima fase, la scomoposizione della misura puntuale $I_h(Y)$ e, nella seconda, la scomposizione dell'indice sintetico $I(Y)$ aggregando opportunamente le quantità definite nella prima fase, Zenga perviene ad una scomposizione iniziale di $k \times k$ componenti. A tal proposito, riportando per completezza le quantità definite in (4.17) e (4.19) si definiscono le seguenti.

$$P_{hg} = \sum_{t=1}^{h} n_{tg} \qquad (h = 1, ..., r;\ g = 1, ..., k) \tag{4.17}$$

$$P_{h.} = \sum_{g=1}^{k} P_{hg} = \sum_{t=1}^{h} n_{t.} \tag{4.72}$$

$$n = \sum_{h=1}^{r} \sum_{g=1}^{k} n_{hg} = \sum_{g=1}^{k} n_{.g} = \sum_{h=1}^{r} n_{h.} \tag{4.73}$$

$$Q_{hg}(Y) = \sum_{t=1}^{h} y_{tg}\, n_{tg} \qquad (h = 1, ..., r) \tag{4.19}$$

$$Q_{h.}(Y) = \sum_{g=1}^{k} Q_{hg}(Y) = \sum_{t=1}^{h} y_t \cdot n_{t.} \tag{4.74}$$

$$T(Y) = \sum_{h=1}^{r} y_h\, n_{h.} = \sum_{g=1}^{k} T_g(Y) = \sum_{g=1}^{k} Q_{rg}(Y), \tag{4.75}$$

$$o(g) = \min h : n_{hg} > 0 \tag{4.76}$$

$$u(g) = \max h : n_{hg} > 0 \tag{4.77}$$

$$\overset{-}{M}_{hg}(Y) = \begin{cases} y_{o(g)} & \text{se } h < o(g) \\ \dfrac{Q_{hg}(Y)}{P_{hg}} & \text{se } h \geq o(g) \end{cases} \tag{4.78}$$

$$\overset{+}{M}_{hg}(Y) = \begin{cases} \dfrac{T_g(Y) - Q_{hg}(Y)}{n - P_{h.}} & \text{se } h < u(g) \\ y_{u(g)} & \text{se } h \geq u(g) \end{cases} \tag{4.79}$$

$$M(Y) = \frac{T(Y)}{n} = \overset{-}{M}_{r.}(Y) = \sum_{g=1}^{k} \overset{-}{M}_{rg}(Y) \cdot \frac{n_{.g}}{n} \tag{4.80}$$

$$p(g|h) = P_{hg}/P_{h.} \tag{4.81}$$

$$a(g|h) = \begin{cases} \dfrac{n_{.g} - P_{hg}}{n - P_{h.}} & \text{se } h = 1, ..., r-1 \\ \dfrac{n_{rg}}{n_{h.}} & \text{se } h = r \end{cases} \tag{4.82}$$

La procedura proposta da Zenga consiste nello scomporre dapprima il numeratore dell'indice puntuale $I_h(Y)$,

$$I_h(Y) = \frac{\overset{+}{M}_{h.}(Y) - \overset{-}{M}_{h.}(Y)}{\overset{+}{M}_{h.}(Y)},$$

sfruttando la proprietà associativa della media aritmetica in modo da ottenere i primi $k \times k$ contributi di base ("*basic decomposition*" [20]) $B_{hlg}(Y)$. Osservando che con

$$\overset{+}{M}_{h.}(Y) - \overset{-}{M}_{h.}(Y),$$

si confrontano due "popolazioni" differenti:

1. una "popolazione" relativa alla distribuzione del gruppo inferiore

   $$\{(y_{1.}, n_{1.}), ..., (y_{h.}, n_{h.}),\ h = 1, ..., r\}$$

   delle unità tali che $Y \leq y_h$, composta dalle k sottopopolazioni con distribuzione

   $$\{(y_{1g}, n_{1g}), ..., (y_{hg}, n_{hg}),\ g = 1, ..., k\}$$

2. una "popolazione" relativa alla distribuzione del gruppo superiore

   $$\{(y_{h+1.}, n_{h+1.}), ..., (y_{r.}, n_{r.}),\ h = 1, ..., r\}$$

   delle unità tali che $Y > y_h$, composta dalle k sottopopolazioni con distribuzione

   $$\{(y_{h+1g}, n_{h+1g}), ..., (y_{rg}, n_{rg}),\ g = 1, ..., k\}$$

ed essendo,

$$\overset{-}{M}_{h.}(Y) = \sum_{g=1}^{k} \overset{-}{M}_{hg} \cdot p(g|h) \tag{4.83}$$

$$\overset{+}{M}_{h.}(Y) = \sum_{g=1}^{k} \overset{+}{M}_{hg} \cdot a(g|h), \tag{4.84}$$

Zenga ricava,

$$
\begin{aligned}
I_h(Y) &= \frac{\overset{+}{M}_{h.}(Y) - \overset{-}{M}_{h.}(Y)}{\overset{+}{M}_{h.}(Y)} \\
&= \sum_{g=1}^{k} \sum_{\ell=1}^{k} \frac{\overset{+}{M}_{hg}(Y) - \overset{-}{M}_{h\ell}(Y)}{\overset{+}{M}_{h.}(Y)} \cdot p(\ell|h) \cdot a(g|h) \qquad (4.85) \\
&= \sum_{g=1}^{k} \sum_{\ell=1}^{k} B_{hlg}(Y), \qquad (4.86)
\end{aligned}
$$

dove,

$$
B_{h\ell g}(Y) = \frac{\overset{+}{M}_{hg}(Y) - \overset{-}{M}_{h\ell}(Y)}{\overset{+}{M}_{h.}(Y)} \cdot p(\ell|h) \cdot a(g|h) \qquad (4.87)
$$

indica il contributo all'indice puntuale $I_h(Y)$ che deriva dai confronti tra la media inferiore $\overset{-}{M}_{h\ell}(Y)$ e la media superiore $\overset{+}{M}_{hg}(Y)$, ponderati per la frequenza relativa del gruppo inferiore e del gruppo superiore, rispettivamente, delle diverse k sottopopolazioni. In virtù della (4.86) è possibile osservare che la scomposizione dell'indice $I_h(Y)$ ha la struttura della differenza media di Gini, a meno del fattore moltiplicativo $1/\overset{+}{M}_{h.}(Y)$, in quanto, per definizione $\overset{+}{M}_{hg}(Y) \geq \overset{-}{M}_{hg}(Y)$, $\forall\, h = 1, ..., r$ e $\forall\, \ell, g = 1, ..., k$. Come nel caso degli indici precedentemente esposti, anche in relazione all'indice di Zenga è possibile individuare un contributo *within* ed un contributo *between*. Seguendo l'approccio a due fasi, tali quantità vengono calcolate dapprima in relazione alla misura di ineguaglianza puntuale $I_h(Y)$ e successivamente, in relazione all'indice sintetico $I(Y)$. A tar riguardo, osservando che per $g = \ell$,

$$
B_{h\ell\ell}(Y) = \frac{\overset{+}{M}_{h\ell}(Y) - \overset{-}{M}_{h\ell}(Y)}{\overset{+}{M}_{h.}(Y)} \cdot p(\ell|h) \cdot a(\ell|h)
$$

indica il contributo a $I_h(Y)$ derivante dal confronto tra media superiore $\overset{+}{M}_{h\ell}(Y)$ e media inferiore $\overset{-}{M}_{h\ell}(Y)$ della medesima sottopopolazione; mentre, per $g \neq \ell$,

$$
B_{h\ell g}(Y) = \frac{\overset{+}{M}_{hg}(Y) - \overset{-}{M}_{h\ell}(Y)}{\overset{+}{M}_{h.}(Y)} \cdot p(\ell|h) \cdot a(g|h)
$$

indica il contributo a $I_h(Y)$ derivante dal confronto tra media superiore $\overset{+}{M}_{hg}(Y)$ e media inferiore $\overset{-}{M}_{h\ell}(Y)$ di sottopopolazioni differenti, si ha:

$$\begin{aligned}
I_h(Y) &= \sum_{\ell=1}^{k}\sum_{g=1}^{k} \frac{\overset{+}{M}_{hg}(Y) - \overset{-}{M}_{h\ell}(Y)}{\overset{+}{M}_{h.}(Y)} \cdot p(\ell|h) \cdot a(g|h) \\
&= \sum_{\ell=1}^{k} \frac{\overset{+}{M}_{h\ell}(Y) - \overset{-}{M}_{h\ell}(Y)}{\overset{+}{M}_{h.}(Y)} \cdot p(\ell|h) \cdot a(\ell|h) + && (4.88)\\
&\qquad + \sum_{\ell=1}^{k}\sum_{g\neq\ell} \frac{\overset{+}{M}_{hg}(Y) - \overset{-}{M}_{h\ell}(Y)}{\overset{+}{M}_{h.}(Y)} \cdot p(\ell|h) \cdot a(g|h) && (4.89)\\
&= \sum_{\ell=1}^{k} B_{hlW}(Y) + \sum_{\ell=1}^{k}\sum_{g\neq\ell} B_{h\ell g}(Y) && (4.90)\\
&= \sum_{\ell=1}^{k} B_{hlW}(Y) + \sum_{\ell=1}^{k} B_{hlB}(Y) && (4.91)\\
&= B_{h.W}(Y) + B_{h.B}(Y) && (4.92)
\end{aligned}$$

dove,

$$\begin{aligned}
B_{h\ell W}(Y) &= \frac{\overset{+}{M}_{h\ell}(Y) - \overset{-}{M}_{h\ell}(Y)}{\overset{+}{M}_{h.}(Y)} \cdot p(\ell|h) \cdot a(\ell|h) = B_{h\ell\ell}(Y) && (4.93)\\
B_{h\ell B}(Y) &= \sum_{g\neq\ell} B_{hlg}(Y) && (4.94)\\
B_{h.W}(Y) &= \sum_{\ell=1}^{k} B_{h\ell W}(Y) && (4.95)\\
B_{h.B}(Y) &= \sum_{\ell=1}^{k} B_{h\ell B}(Y) && (4.96)\\
& && (4.97)
\end{aligned}$$

In questi termini, si osserva che $B_{h.W}$, somma dei confronti $B_{h\ell W}(Y)$ tra medie superiori ed inferiori all'interno della medesima sottopopolazione, si configura in via naturale come componente *within* dell'indice puntuale $I_h(Y)$, mentre $B_{h.B}(Y)$, somma dei confronti tra medie superiori ed inferiori di sottopopolazioni differenti, si configura in via naturale come componente between dell'indice puntuale $I_h(Y)$. La metodologia proposta da Zenga permette altresì di definire un contributo finora trascurato, ovvero, permette di definire il contributo delle singole sottopopolazioni all'indice di ineguaglianza (puntuale e sintetico).

In questi termini, in virtù della (4.84), l'Autore osserva che:

$$\begin{aligned}
I_h(Y) &= \sum_{\ell=1}^{k}\left[\sum_{g=1}^{k} B_{h\ell g}(Y)\right] \\
&= \frac{1}{\overset{+}{M}_{h.}(Y)}\cdot\sum_{\ell=1}^{k}\left\{\sum_{g=1}^{k}\left[\overset{+}{M}_{hg}(Y)-\overset{-}{M}_{h\ell}(Y)\right]\cdot p(\ell|h)\cdot a(g|h)\right\} \\
&= \frac{1}{\overset{+}{M}_{h.}(Y)}\cdot\sum_{\ell=1}^{k}\left\{p(\ell|h)\sum_{g=1}^{k}\overset{+}{M}_{hg}(Y)\cdot a(g|h)-\overset{-}{M}_{h\ell}\cdot p(\ell|h)\cdot\sum_{g=1}^{k}a(g|h)\right\} \\
&= \sum_{\ell=1}^{k}\frac{\overset{+}{M}_{h.}(Y)-\overset{-}{M}_{h\ell}(Y)}{\overset{+}{M}_{h.}(Y)}\cdot p(\ell|h) \\
&= \sum_{\ell=1}^{k} B_{hl.}(Y) \qquad (4.98)
\end{aligned}$$

dove,

$$B_{hl.}(Y)=\frac{\overset{+}{M}_{h.}(Y)-\overset{-}{M}_{h\ell}(Y)}{\overset{+}{M}_{h.}(Y)}\cdot p(\ell|h). \qquad (4.99)$$

In virtù della (4.99), Zenga definisce la quantità $B_{hl.}(Y)$ come il prodotto tra la variazione relativa della media inferiore $\overset{-}{M}_{h\ell}(Y)$ nella sottopopolazione ℓ rispetto alla media superiore dell'intera popolazione $\overset{+}{M}_{h.}(Y)$ e la frequenza relativa del gruppo inferiore $p(\ell|h)$. Ne deriva che, anche in questo caso, la quantità $B_{h\ell.}(Y)$ si configura in via naturale come *contributo della sottopopolazione* ℓ all'indice di ineguaglianza puntuale $I_h(Y)$. In effetti, il contributo $B_{h\ell.}(Y)$ può anche essere espresso come somma tra il contributo within $B_{h\ell W}$ e la somma dei contributi $B_{h\ell g}(Y)$ che coinvolgono confronti tra la media superiore del gruppo g e la media inferiore del gruppo ℓ, per $\ell \neq g$. Formalmente, in virtù delle (4.93) e (4.94) il contributo $B_{hl.}(Y)$ può quindi essere scomposto nel modo seguente:

$$\begin{aligned}
B_{h\ell.}(Y) &= \sum_{g=1}^{k} B_{h\ell g}(Y) \qquad (4.100)\\
&= B_{h\ell\ell}(Y)+\sum_{\substack{g=1\\ g\neq\ell}}^{k} B_{h\ell g}(Y) \qquad (4.101)\\
&= B_{h\ell W}+B_{h\ell B} \qquad (4.102)
\end{aligned}$$

In virtù dei risultati in (4.86), (4.98), (4.101) e (4.102), l'indice di ineguaglianza puntuale $I_h(Y)$ oltre a poter essere scomposto in un contributo within ed in un contributo between, può essere espresso come somma dei contributi delle singole sottopopolazioni:

$$\begin{aligned}
I_h(Y) &= \sum_{\ell=1}^{k} \sum_{g=1}^{k} B_{h\ell g}(Y) && (4.103)\\
&= \sum_{\ell=1}^{k} B_{h\ell .}(Y) && (4.104)\\
&= \sum_{\ell=1}^{k} \left[B_{h\ell W}(Y) + B_{h\ell B}(Y)\right] && (4.105)\\
&= B_{h.W}(Y) + B_{h.B}(Y) && (4.106)
\end{aligned}$$

Ottenute le scomposizioni da (4.103) a (4.106), relative all'indice di ineguaglianza puntuale $I_h(Y)$, è possibile passare alla "seconda fase" e ricavare scomposizioni analoghe per l'indice di ineguaglianza sintetico $I(Y)$. In virtù della (1.40), è possibile ottenere i $k \times k$ contributi $B_{.\ell g}(Y)$:

$$\begin{aligned}
I(Y) &= \sum_{h=1}^{r} I_h(Y) \cdot \frac{n_{h.}}{n}\\
&= \sum_{h=1}^{r} \left[\sum_{g=1}^{k} \sum_{\ell=1}^{k} B_{h\ell g}(Y)\right] \cdot \frac{n_{h.}}{n}\\
&= \sum_{g=1}^{k} \sum_{\ell=1}^{k} \left[\sum_{h=1}^{r} B_{h\ell g}(Y) \cdot \frac{n_{h.}}{n}\right]\\
&= \sum_{g=1}^{k} \sum_{\ell=1}^{k} B_{.\ell g}(Y) && (4.107)\\
&= \sum_{\ell=1}^{k} \left[B_{.\ell\ell}(Y) + \sum_{g \neq \ell} B_{.\ell g}(Y)\right] && (4.108)\\
&= \sum_{\ell=1}^{k} B_{.\ell .}(Y) && (4.109)\\
&= \sum_{\ell=1}^{k} \left[B_{.\ell W}(Y) + B_{.\ell B}(Y)\right] && (4.110)\\
&= B_{..W}(Y) + B_{..B}(Y) && (4.111)
\end{aligned}$$

dove:

$$B_{.\ell g}(Y) = \sum_{h=1}^{r} B_{h\ell g}(Y) \cdot \frac{n_{h.}}{n}$$

è la media ponderata degli indici $B_{h\ell g}(Y)$, con pesi $n_{h.}/n$ ed indica i $k \times k$ contributi all'indice sintetico $I(Y)$;

$$B_{.\ell.}(Y) = B_{.\ell\ell}(Y) + \sum_{g \neq \ell} B_{.\ell g}(Y)$$

indica il contibuto della sottopopolazione ℓ all'indice sintetico $I(Y)$

$$B_{.\ell W}(Y) = B_{.\ell\ell}$$

e

$$B_{.\ell B} = \sum_{g \neq \ell} B_{.\ell g}(Y)$$

indicano rispettivamente i contributi *within* e *between* della sottopopolazione ℓ all'indice sintetico $I(Y)$

$$B_{..W}(Y) = \sum_{\ell=1}^{k} B_{.\ell W}(Y)$$

e

$$B_{..B}(Y) = \sum_{\ell=1}^{k} B_{.\ell B}(Y)$$

La misura introdotta da Zenga, così come la metodologia "two-steps" proposta per la scomposizione per sottopopolazioni degli indici di ineguaglianza $I_h(Y)$ e $I(Y)$, permette quindi di definire sia una componente *within* ed una componente *between* degli indici $I_h(Y)$ e $I(Y)$; sia il *contributo delle singole sottopopolazioni* agli indici di ineguaglianza $I_h(Y)$ e $I(Y)$, a loro volta esprimibili in termini di una componente *within* e di una componente *between* che, opportunamente aggregate, permettono di ottenere le componenti *within* $B_{h.W}$ e $B_{..W}$ e *between* $B_{h.B}$ e $B_{..B}$ relative agli indici $I_h(Y)$ e $I(Y)$, rispettivamente, valutate senza prendere in considerazione il contributo delle singole sottopopolazioni.

In virtù di tali risultati, nel capitolo successivo si procederà estendendo detta metodologia all'indice di ineguaglianza di Bonferroni.

Capitolo 5

Scomposizione per sottopopolazioni dell'indice di Bonferroni

Come ricordato nel Cap.1, l'indice di ineguaglianza di Bonferroni $\tilde{V}(Y)$ in (1.24) è la media aritmetica di $n-1$ variazioni relative $\tilde{V}_i(Y)$, $i = 1, ..., n-1$, della media inferiore $\bar{M}_i(Y)$ rispetto la media generale $M(Y)$. Per i motivi esposti nel Cap. 2 e nel Cap. 3, l'indice di Bonferroni è stato modificato, dapprima, definendo l'indice $V'(Y)$ in (2.8) come media di n misure puntuali $\tilde{V}_i(Y)$; successivamente, attribuendo medesima misura di ineguaglianza puntuale alle unità che presentano medesimo valore della variabile Y. In questo modo, per tutti gli $n_{h.}$ valori y_h è stata attribuita una misura pari a $V_h(Y) = \tilde{V}_{Ph.}(Y)$. Formalmente:

$$\tilde{V}_i(Y) = \tilde{V}_{P_{h.}}(Y) = V_h(Y), \qquad \text{per} \qquad 1 + P_{h.} - n_{h.} \leq i \leq P_{h.},$$

con $h = 1, ..., r$, ove r indica il numero dei distinti valori assunti dalla variabile Y. In tal modo è stato definito l'indice di Bonferroni $V(Y)$ in (3.4) per dati espressi in termini di distribuzione di frequenza che, oltre a conservare le proprietà dell'indice $\tilde{V}(Y)$, è altresì *invariante a replicazioni nella popolazione.* In questo capitolo, estendendo la metodologia a "due fasi" proposta da Zenga [20], si procederà alla scomposizione per sottopopolazioni dell'indice di Bonferroni $V(Y)$: nella "prima fase" si scomporrà la misura puntuale $V_h(Y)$ definita in (3.5) e qui riportata

$$V_h(Y) = \frac{M(Y) - \bar{M}_{h.}(Y)}{M(Y)}; \tag{3.5}$$

nella "seconda fase", aggregando opportunamente le quantità definite nella fase precedente, si scomporrà l'indice sintetico $V(Y)$ definito in (3.4) e di seguito riportato:

$$V(Y) = \sum_{h=1}^{r} V_h(Y) \cdot \frac{n_{h.}}{n}. \tag{3.4}$$

Nel capitolo successivo, la procedura proposta sarà infine applicata ai redditi netti disponibili di $n = 8151$ famiglie italiane, suddivise in base alla loro area di residenza (Nord, Centro, Sud e Isole). [1]

5.1 Definizioni e notazioni

Sia Y una variabile a valori non negativi osservata su n unità e si supponga che dette unità possano essere suddivise in k sottoinsiemi disgiunti, che chiameremo *sottopopolazioni*, di numerosità $n_{.g}$, $g = 1, ..., k$.
Sia inoltre $\{0 \leq y_1 < \ ... < \ y_h < \ ... < \ y_r\}$ l'insieme degli r distinti valori ordinati della variabile Y, osservati nelle k sottopopolazioni. I dati così definiti possono essere rappresentati come in Tabella 5.1.

Tabella 5.1: Distribuzione di frequenza dell'intera popolazione suddivisa in k sottopopolazioni.

	Sottopopolazioni					
	1	$\dots$	g	$\dots$	k	Tot.
y_1	n_{11}	$\dots$	n_{1g}	$\dots$	n_{1k}	$n_{1.}$
$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
y_h	n_{h1}	$\dots$	n_{hg}	$\dots$	n_{hk}	$n_{h.}$
$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
y_r	n_{r1}	$\dots$	n_{rg}	$\dots$	n_{rk}	$n_{r.}$
Tot.	$n_{.1}$	$\dots$	$n_{.g}$	$\dots$	$n_{.k}$	n

5.2 Scomposizione degli indici puntuali $V_h(Y)$ e dell'indice sintetico $V(Y)$

In analogia con la metodologia proposta da Zenga [20] per la scomposizione per sottopopolazioni dell'indice puntuale $I_h(Y)$ e sintetico $I(Y)$, si procederà scomponendo dapprima la differenza

$$M(Y) - \bar{M}_{h.}(Y),$$

numeratore dell'indice (puntuale) di Boneferroni $V_h(Y)$ in (3.5).

In virtù delle definizioni in (1.36) e (1.37) si ricava:

$$M(Y) - \bar{M}_{h.}(Y) = \frac{T(Y)}{n} - \frac{Q_{h.}(Y)}{P_{h.}} \tag{5.1}$$

$$= \frac{P_{h.} \cdot T(Y) - n \cdot Q_{h.}(Y)}{n \cdot P_{h.}}. \tag{5.2}$$

[1] Fonte dati: "I bilanci delle famiglie italiane nell'anno 2012", Banca d'Italia, 2014 [5]

In virtù delle (4.72), (4.74) e (4.75) e dalla relazione $n = \sum_{g=1}^{k} n_{.g}$, dal numeratore della (5.2) si ottiene:

$$\begin{aligned}
P_{h.} \cdot T - n \cdot Q_{h.}(Y) &= \sum_{g=1}^{k}\sum_{\ell=1}^{k} P_{h\ell} \cdot T_g(Y) - \sum_{g=1}^{k}\sum_{\ell=1}^{k} n_{.g} \cdot Q_{h\ell}(Y) \\
&= \sum_{g=1}^{k}\sum_{\ell=1}^{k} \frac{T_g(Y)}{n_{.g}} \cdot n_{.g} \cdot P_{h\ell} - \sum_{g=1}^{k}\sum_{\ell=1}^{k} \frac{Q_{h\ell}(Y)}{P_{h\ell}} \cdot P_{h\ell} \cdot n_{.g} \\
&= \sum_{g=1}^{k}\sum_{\ell=1}^{k} M_g(Y) \cdot n_{.g} \cdot P_{h\ell} - \sum_{g=1}^{k}\sum_{\ell=1}^{k} \bar{M}_{h\ell}(Y) \cdot P_{h\ell} \cdot n_{.g} \\
&= \sum_{g=1}^{k}\sum_{\ell=1}^{k} \Big[M_g(Y) - \bar{M}_{h\ell}(Y)\Big] \cdot P_{h\ell} \cdot n_{.g} \qquad (g, \ell = 1, ..., k). \qquad (5.3)
\end{aligned}$$

Dividendo ambo i lati della (5.3) per $n \cdot P_{h.}$ si ottiene nuovamente la (5.2) e si ricava,

$$\begin{aligned}
\frac{P_{h.} \cdot T - n \cdot Q_{h.}(Y)}{n \cdot P_{h.}} &= \frac{\sum_{g=1}^{k}\sum_{\ell=1}^{k} \Big[M_g(Y) - \bar{M}_{h\ell}(Y)\Big] \cdot P_{h\ell} \cdot n_{.g}}{n \cdot P_{h.}} \\
&= \sum_{g=1}^{k}\sum_{\ell=1}^{k} \Big[M_g(Y) - \bar{M}_{h\ell}(Y)\Big] \cdot \frac{P_{h\ell}}{P_{h.}} \cdot \frac{n_{.g}}{n} \qquad (5.4) \\
&= \sum_{g=1}^{k}\sum_{\ell=1}^{k} \Big[M_g(Y) - \bar{M}_{h\ell}(Y)\Big] \cdot p(\ell|h) \cdot \frac{n_{.g}}{n}
\end{aligned}$$

dove,

$$p(\ell|h) = \frac{P_{h\ell}}{P_{h.}} \qquad (5.5)$$

indica la frequenza relativa cumulata della modalità y_h nel gruppo inferiore della sottopopolazione ℓ, $\ell = 1, ..., k$.
Riassumendo i passaggi dalla (5.1) alla (5.4) si ha quindi:

$$M(Y) - \bar{M}_{h.}(Y) = \sum_{\ell=1}^{k}\sum_{g=1}^{k} \Big[M_g(Y) - \bar{M}_{h\ell}(Y)\Big] \cdot p(\ell|h) \cdot \frac{n_{.g}}{n}. \qquad (5.6)$$

Dividendo infine la (5.6) per $M(Y)$ si ottiene la scomposizione per sottopopolazioni dell'indice di ineguaglianza puntuale $V_h(Y)$:

$$
\begin{aligned}
V_h(Y) &= \frac{M(Y) - \bar{M}_{h.}(Y)}{M(Y)} \\
&= \sum_{l=1}^{k} \sum_{g=1}^{k} \left[\frac{M_g(Y) - \bar{M}_{h\ell}(Y)}{M(Y)} \right] \cdot p(\ell|h) \cdot \frac{n_{.g}}{n} \\
&= \sum_{l=1}^{k} \sum_{g=1}^{k} V_{h\ell g}(Y)
\end{aligned}
\tag{5.7}
$$

dove,

$$
V_{h\ell g}(Y) = \frac{M_g(Y) - \bar{M}_{h\ell}(Y)}{M(Y)} \cdot p(\ell|h) \cdot \frac{n_{.g}}{n} \tag{5.8}
$$

rappresenta il contributo all'indice di ineguaglianza puntuale $V_h(Y)$ che deriva dal confronto tra la media del gruppo inferiore $\bar{M}_{h\ell}(Y)$ della sottopopolzione ℓ e la media generale $M_g(Y)$ della sottopopolazione g, $g, \ell = 1, ..., k$.

In questi termini, dato che,

$$
\sum_{\ell=1}^{k} p(\ell|h) = 1 \quad \text{e} \quad \sum_{g=1}^{k} \frac{n_{.g}}{n} = 1, \tag{5.9}
$$

segue,

$$
\sum_{\ell=1}^{k} \sum_{g=1}^{k} p(\ell|h) \cdot \frac{n_{.g}}{n} = 1,
$$

e pertanto, l'indice di ineguaglianza puntuale $V_h(Y)$ coincide con la media ponderata delle $k \times k$ differenze relative:

$$
\frac{M_g(Y) - \bar{M}_{h\ell}(Y)}{M(Y)}
$$

con pesi $p(\ell|h) \cdot \frac{n_{.g}}{n}$.[2]

Dalla (5.7), con opportune aggregazioni, è inoltre possibile ottenere ulteriori scomposizioni della misura puntuale $V_h(Y)$. Infatti, dalla (5.9) e dalla relazione,

$$
\sum_{g=1}^{k} M_g(Y) \cdot \frac{n_{.g}}{n} = M(Y),
$$

[2]Tale risultato è coerente con la struttura di pesi (coefficienti) indicata da Shorroks per la classe delle misure scomponibili di ineguaglianza [10]

si ricava:

$$\begin{aligned}
V_h(Y) &= \sum_{g=1}^{k}\sum_{\ell=1}^{k} V_{h\ell g}(Y)\\
&= \sum_{\ell=1}^{k}\sum_{g=1}^{k} \frac{M_g(Y)-\bar{M}_{h\ell}(Y)}{M(Y)}\cdot p(\ell|h)\cdot\frac{n_{.g}}{n}\\
&= \frac{1}{M(Y)}\sum_{\ell=1}^{k}\left[\sum_{g=1}^{k} M_g(Y)\cdot p(\ell|h)\cdot\frac{n_{.g}}{n}-\sum_{g=1}^{k}\bar{M}_{h\ell}(Y)\cdot p(\ell|h)\cdot\frac{n_{.g}}{n}\right]\\
&= \frac{1}{M(Y)}\sum_{\ell=1}^{k}\left[p(\ell|h)\cdot\sum_{g=1}^{k} M_{.g}\cdot\frac{n_{.g}}{n}-\bar{M}_{h\ell}(Y)\cdot p(\ell|h)\cdot\sum_{g=1}^{k}\frac{n_{.g}}{n}\right]\\
&= \frac{1}{M(Y)}\sum_{\ell=1}^{k}\left[p(\ell|h)\cdot M(Y)-\bar{M}_{h\ell}(Y)\cdot p(\ell|h)\right]\\
&= \sum_{\ell=1}^{k}\frac{M(Y)-\bar{M}_{h\ell}(Y)}{M(Y)}\cdot p(\ell|h)\\
&= \sum_{\ell=1}^{k} V_{h\ell.}(Y),
\end{aligned} \tag{5.10}$$

dove,

$$V_{h\ell.}(Y) = \frac{M(Y)-\bar{M}_{h\ell}(Y)}{M(Y)}\cdot p(\ell|h). \tag{5.11}$$

In questi termini, la quantità $V_{h\ell.}(Y)$ in (5.11) rappresenta la variazione relativa della media inferiore $\bar{M}_{h\ell}(Y)$ della sottopopolazione ℓ, rispetto la media generale $M(Y)$, pesata per le frequenze relative $p(\ell|h)$ del gruppo inferiore e pertanto, $V_{h\ell.}(Y)$ si configura in via naturale come contributo della sottopopolazione ℓ alla misura di ineguaglianza puntuale $V_h(Y)$. In virtù della (5.10) si dimostra, quindi, che l'indice puntuale $V_h(Y)$ è la media pesata delle k variazioni relative

$$\frac{M(Y)-\bar{M}_{h\ell}(Y)}{M(Y)},$$

con pesi $p(\ell|h) = \dfrac{P_{h\ell}}{P_{h.}}$.

In analogia alle precedenti proposte di scomposizione, saranno ora definite una componente *within* ed una componente *between* dell'indice di ineguaglianza puntuale e sintetico. A tal proposito, il contributo $V_{h\ell.}(Y)$ sarà ora scomposto in una componente *within*, ricavata dai confronti tra la media inferiore $\bar{M}_{h\ell}(Y)$ e la media $M_\ell(Y)$ della medesima sottopopolazione ($\ell = 1, ..., k$), e da una componente *between*, ricavata dai confronti tra le medie $M_g(Y)$ e $\bar{M}_{h\ell}(Y)$ di sottopopolazioni differenti ($\ell, g = 1, ..., k;\ g \neq \ell$). In questi termini, in virtù della (5.10) si ricava:

$$\begin{aligned} V_{h\ell.}(Y) &= \sum_{g=1}^{k} V_{h\ell g}(Y) = \sum_{g=1}^{k} \frac{M_g(Y) - \bar{M}_{h\ell}(Y)}{M(Y)} \cdot p(\ell|h) \cdot \frac{n_{.g}}{n} \\ &= \frac{M_\ell(Y) - \bar{M}_{h\ell}(Y)}{M(Y)} \cdot p(\ell|h) \cdot \frac{n_{.\ell}}{n} + \sum_{\ell \neq g} \frac{M_g(Y) - \bar{M}_{h\ell}(Y)}{M(Y)} \cdot p(\ell|h) \cdot \frac{n_{.g}}{n} \qquad (5.12) \\ &= V_{h\ell\ell}(Y) + \sum_{\ell \neq g} V_{h\ell g} \end{aligned}$$

Dato che:

$$\sum_{\ell \neq g} \frac{n_{.g}}{n} = 1 - \frac{n_{.\ell}}{n}$$

la (5.11) può essere riscritta nel modo seguente:

$$V_{h\ell.}(Y) = p(\ell|h) \left[\frac{M_\ell(Y) - \bar{M}_{h\ell}(Y)}{M(Y)} \cdot \frac{n_{.\ell}}{n} \right] + \qquad (5.13)$$

$$+ p(\ell|h) \left[\sum_{\ell \neq g} \frac{M_g(Y) - \bar{M}_{h\ell}(Y)}{M(Y)} \cdot \frac{\frac{n_{.g}}{n}}{1 - \frac{n_{.\ell}}{n}} \right] \cdot \left(1 - \frac{n_{.\ell}}{n}\right). \qquad (5.14)$$

Il confronto tra la (5.13) e la (5.14) mostra che la variazione relativa della media inferiore $\bar{M}_{h\ell}(Y)$ rispetto la media generale $M(Y)$ è la media pesata:

- del rapporto $\dfrac{M_\ell(Y) - \bar{M}_{h\ell}(Y)}{M(Y)}$ con pesi $\dfrac{n_{.\ell}}{n}$ e
- della media pesata $\displaystyle\sum_{\ell \neq g} \frac{M_g(Y) - \bar{M}_{h\ell}(Y)}{M(Y)} \cdot \frac{\frac{n_{.g}}{n}}{1 - \frac{n_{.\ell}}{n}}$ con pesi $1 - \dfrac{n_{.\ell}}{n}$

Nella (5.12) quindi,

$$V_{h\ell\ell}(Y) = p(\ell|h) \cdot \left[\frac{M_\ell(Y) - \bar{M}_{h\ell}(Y)}{M(Y)} \right] \cdot \frac{n_{.\ell}}{n} = V_{h\ell W}(Y) \qquad (5.15)$$

può essere interpretato come componente *within* del contributo $V_{h\ell.}$. In modo analogo:

$$\sum_{g \neq \ell} V_{h\ell g}(Y) = p(\ell|h) \cdot \left[\sum_{g \neq \ell} \frac{M_g(Y) - \bar{M}_{h\ell}(Y)}{M(Y)} \right] \cdot \frac{n_{.g}}{n} = V_{h\ell B}(Y) \qquad (5.16)$$

può essere interpretato come componente *between* del contributo $V_{h\ell.}(Y)$. Formalmente:

$$V_{h\ell.}(Y) = V_{h\ell W}(Y) + V_{h\ell B}(Y) \tag{5.17}$$

Dalle relazioni in (5.10) e (5.17) si ricava:

$$V_h(Y) = \sum_{\ell=1}^{k} V_{h\ell.}(Y) = \sum_{\ell=1}^{k} V_{h\ell W}(Y) + \sum_{\ell=1}^{k} V_{h\ell B}(Y) = V_{h.W}(Y) + V_{h.B}(Y), \tag{5.18}$$

dove,

$$V_{h.W}(Y) = \sum_{\ell=1}^{k} V_{h\ell W}(Y) \quad \text{e} \quad V_{h.B}(Y) = \sum_{\ell=1}^{k} V_{h\ell B}(Y), \tag{5.19}$$

possono essere interpretati rispettivamente come contributo *within* complessivo e come contributo *between* complessivo, delle diverse sottopopolazioni, all'indice di ineguaglianza puntuale $V_h(Y)$. È opportuno notare che le componenti *within* e *between* possono essere ricavate direttamente dalla (5.7) nel modo seguente:

$$V_h(Y) = \sum_{\ell=1}^{k} \sum_{g=1}^{k} V_{h\ell g} \tag{5.20}$$

$$= \sum_{\ell=1}^{k} \sum_{g=1}^{k} \frac{M_g(Y) - \bar{M}_{h\ell}(Y)}{M(Y)} \cdot p(\ell|h) \cdot \frac{n_{.g}}{n} \tag{5.21}$$

$$= \sum_{\ell=1}^{k} \frac{M_\ell(Y) - \bar{M}_{h\ell}(Y)}{M(Y)} \cdot p(\ell|h) \cdot \frac{n_{.\ell}}{n} \tag{5.22}$$

Ovvero, sostituendo in (5.21) e (5.22) le definizioni in (5.15), (5.17) e (5.19) si ottiene nuovamente:

$$\begin{aligned} V_h(Y) &= \sum_{\ell=1}^{k} V_{h\ell W} + \sum_{\ell=1}^{k} V_{h\ell B} \\ &= V_{h.W} + V_{h.B} \end{aligned}$$

Ricavate le scomposizioni dell'indice di ineguaglianza puntuale $V_h(Y)$, si può procedere alla "seconda fase" e scomporre l'indice di ineguaglianza sintetico $V(Y)$. In virtù della (3.4) e in analogia ai risultati in (5.11) si ricavano i $k \times k$ contributi $V_{.\ell g}(Y)$ e i k

contributi delle sottopopolazioni $V_{.\ell.}(Y)$ dell'indice sintetico $V(Y)$ nel seguente modo:

$$\begin{aligned} V(Y) &= \sum_{h=1}^{r} V_h(Y) \cdot \frac{n_{h.}}{n} \\ &= \sum_{h=1}^{r}\sum_{\ell=1}^{k}\sum_{g=1}^{k} V_{h\ell g}(Y) \cdot \frac{n_{h.}}{n} = \sum_{\ell=1}^{k}\sum_{g=1}^{k}\sum_{h=1}^{r} V_{h\ell g}(Y) \cdot \frac{n_{h.}}{n} \\ &= \sum_{\ell=1}^{k}\sum_{g=1}^{k} V_{.\ell g}(Y) \\ &= \sum_{\ell=1}^{k} V_{.\ell.}(Y) \end{aligned} \tag{5.23}$$

dove,

$$V_{.\ell g}(Y) = \sum_{h=1}^{r} V_{h\ell g}(Y) \cdot \frac{n_{h.}}{n} = \sum_{h=1}^{r} \left\{ \frac{M_g(Y) - \bar{M}_{h\ell}(Y)}{M(Y)} \cdot p(\ell|h) \cdot \frac{n_{.g}}{n} \right\} \cdot \frac{n_{h.}}{n}. \tag{5.24}$$

In altre parole, $V_{.\ell g}(Y)$ è pertanto la media ponderata dei contributi $V_{h\ell g}(Y)$ con pesi $n_{h.}/n$ e rappresenta uno dei $k \times k$ "*basic contributions*" dell'indice di ineguaglianza sintetico $V(Y)$.

In relazione al contributo $V_{.\ell.}(Y)$, dalla (5.23) ed in virtù della (5.11) segue che:

$$V_{.\ell.}(Y) = \sum_{g=1}^{k} V_{.\ell g}(Y), \tag{5.25}$$

$$= \sum_{h=1}^{r} \left\{ \sum_{g=1}^{k} \frac{M_g(Y) - \bar{M}_{h\ell}(Y)}{M(Y)} \cdot p(\ell|h) \cdot \frac{n_{.g}}{n} \right\} \cdot \frac{n_{h.}}{n} \tag{5.26}$$

$$= \sum_{h=1}^{r} \left\{ \frac{M(Y) - \bar{M}_{h\ell}(Y)}{M(Y)} \cdot p(\ell|h) \right\} \cdot \frac{n_{h.}}{n} \tag{5.27}$$

$$= \sum_{h=1}^{r} V_{h\ell.}(Y) \cdot \frac{n_{h.}}{n} \tag{5.28}$$

Pertanto, $V_{.\ell.}(Y)$ è la media ponderata dei contributi $V_{h\ell.}(Y)$ delle k sottopopolazioni all'indice di ineguaglianza sintetico $V_h(Y)$, e pertanto, $V_{.\ell.}(Y)$ rappresenta, in via naturale, il contributo della sottopopolazione ℓ all'indice sintetico $V(Y)$.

Anche per l'indice di ineguaglianza sintetico $V(Y)$, è possibile ricavare delle componenti *within* e *between*. A tal proposito si procederà definendo: dapprima una componente *within* ed una componente *between* del contributo $V_{.\ell.}(Y)$; successivamente, con opportune aggregazioni, una componente *within* complessiva e una componente *between* complessiva dell'indice sintetico $V(Y)$. Infatti, dalla (5.24) e (5.28) e dalle relazioni da (5.20) a (5.22) è possibile osservare che:

$$\begin{aligned}
V_{.\ell.}(Y) &= \sum_{h=1}^{r}\sum_{g=1}^{k} V_{h\ell.} \cdot \frac{n_{h.}}{n} \\
&= \sum_{h=1}^{r} \left\{ \frac{M_\ell(Y) - \bar{M}_{h\ell}(Y)}{M(Y)} \cdot p(\ell|h) \cdot \frac{n_{.\ell}}{n} + \right. \\
&\qquad \left. + \sum_{g \neq \ell} \frac{M_g(Y) - \bar{M}_{h\ell}(Y)}{M(Y)} \cdot p(\ell|h) \cdot \frac{n_g}{n} \right\} \cdot \frac{n_{h.}}{n} \\
&= \sum_{h=1}^{r} \left\{ \frac{M_\ell(Y) - \bar{M}_{h\ell}(Y)}{M(Y)} \cdot p(\ell|h) \cdot \frac{n_{.\ell}}{n} \right\} \cdot \frac{n_{h.}}{n} + \\
&\qquad + \sum_{h=1}^{r} \left\{ \sum_{\ell \neq g} \frac{M_g(Y) - \bar{M}_{h\ell}(Y)}{M(Y)} \cdot p(\ell|h) \cdot \frac{n_{.g}}{n} \right\} \cdot \frac{n_{h.}}{n} \\
&= \sum_{h=1}^{r} V_{h\ell\ell}(Y) \cdot \frac{n_{h.}}{n} + \sum_{h=1}^{r}\sum_{g \neq \ell} V_{h\ell g}(Y) \cdot \frac{n_{h.}}{n} \\
&= \sum_{h=1}^{r} V_{h\ell\ell}(Y) \cdot \frac{n_{h.}}{n} + \sum_{g \neq \ell}\sum_{h=1}^{r} V_{h\ell g}(Y) \cdot \frac{n_{h.}}{n} \\
&= V_{.\ell\ell} + \sum_{g \neq \ell} V_{.\ell g},
\end{aligned} \tag{5.29}$$

Di conseguenza, in virtù delle relazioni in (5.29), posto:

$$V_{.\ell\ell}(Y) = \sum_{h=1}^{r} \left\{ \frac{M_\ell(Y) - \bar{M}_{h\ell}(Y)}{M(Y)} \cdot p(\ell|h) \cdot \frac{n_{.\ell}}{n} \right\} \cdot \frac{n_{h.}}{n} = V_{.\ell W}(Y) \tag{5.30}$$

e

$$V_{.\ell B}(Y) = \sum_{g \neq \ell} V_{.\ell g}(Y) \tag{5.31}$$

$V_{.\ell W}$ e $V_{.\ell B}$ si configurano in via naturale come i contributi *within* e *between*, rispettivamente, della sottopopolazione ℓ, all'indice sintetico $V(Y)$.

Pertanto, in virtù della (5.23) e delle (5.30) e (5.31):

$$\begin{aligned} V(Y) &= \sum_{\ell=1}^{k} V_{.\ell.}(Y) \\ &= \sum_{\ell=1}^{k} \left\{ V_{.\ell\ell}(Y) + \sum_{g \neq \ell} V_{.\ell g}(Y) \right\} \\ &= \sum_{\ell=1}^{k} V_{.\ell W}(Y) + \sum_{\ell=1}^{k} V_{.\ell B}(Y) \\ &= V_{..W}(Y) + V_{..B}(Y), \end{aligned} \tag{5.32}$$

dove,

$$V_{..W}(Y) = \sum_{\ell=1}^{k} V_{.\ell W}(Y) \quad \text{e} \quad V_{..B}(Y) = \sum_{\ell=1}^{k} V_{.\ell B}(Y) \tag{5.33}$$

indicano rispettivamente il contributo *within* complessivo ed il contributo *between* complessivo dell' indice di ineguaglianza sintetico $V(Y)$.
In particolare, in virtù della (5.33) e delle (5.30) e (5.19) è possibile osservare che,

$$\begin{aligned} V_{..W}(Y) &= \sum_{\ell=1}^{k} V_{.\ell\ell}(Y) \\ &= \sum_{h=1}^{r} \left\{ \sum_{\ell=1}^{k} \frac{M_{\ell}(Y) - \bar{M}_{h\ell(Y)}}{M(Y)} \cdot p(\ell|h) \cdot \frac{n_{.\ell}}{n} \right\} \cdot \frac{n_{h.}}{n} \\ &= \sum_{h=1}^{r} \left\{ \sum_{\ell=1}^{k} V_{h\ell W}(Y) \right\} \frac{n_{h.}}{n} \\ &= \sum_{h=1}^{r} V_{h.W}(Y) \cdot \frac{n_{h.}}{n}, \end{aligned} \tag{5.34}$$

e pertanto, il contributo *within* complessivo $V_{..W}(Y)$ dell'indice sintetico $V(Y)$ è la media ponderata degli r contributi complessivi $V_{h.W}(Y)$, con pesi $n_{h.}/n$, dell'indice di ineguaglianza puntuale $V_h(Y)$. Analogamente, in virtù delle (5.33), (5.31), (5.23), (5.17) e (5.19) essendo,

$$\begin{aligned}
V_{..B}(Y) &= \sum_{\ell=1}^{k} V_{.\ell B}(Y) \\
&= \sum_{\ell=1}^{k} \sum_{g \neq \ell} V_{.\ell g}(Y) \\
&= \sum_{h=1}^{r} \left\{ \sum_{\ell=1}^{k} \sum_{g \neq \ell} \frac{M_g(Y) - \bar{M}_{h\ell(Y)}}{M(Y)} \cdot p(\ell|h) \cdot \frac{n_{.g}}{n} \right\} \cdot \frac{n_{h.}}{n} \qquad (5.35) \\
&= \sum_{h=1}^{r} \left\{ \sum_{\ell=1}^{k} V_{h\ell B}(Y) \right\} \cdot \frac{n_{h.}}{n} \\
&= \sum_{h=1}^{r} V_{h.B}(Y) \cdot \frac{n_{h.}}{n}
\end{aligned}$$

il contributo *between* complessivo $V_{..B}(Y)$ dell'indice sintetico $V(Y)$ è la media ponderata degli r contributi complessivi $V_{h.B}(Y)$, con pesi $n_{h.}/n$, dell'indice di ineguaglianza puntuale $V_h(Y)$.

Capitolo 6

Applicazione

I dati utilizzati in questa applicazione sono forniti dalla survey sui redditi e la ricchezza di $n = 8151$ famiglie, condotta nel 2012 dalla Banca d'Italia [5]. La variabile di riferimento Y è il reddito netto familiare disponibile, ottenuto aggregando: il reddito da lavoro dipendente (X_1), le pensioni e trasferimenti (X_2), il reddito da lavoro autonomo (X_3) e il reddito derivante da proprietà (X_4) [1]. Le $n = 8151$ unità (famiglie) sono state suddivise in base all'area di residenza: "Nord" (1), "Centro" (2), "Sud e Isole" (3). Nelle analisi effettuate sono stati considertati i pesi campionari $w_i > 0$ ($i = 1, ..., n$; $\sum w_i = n = 8151$) forniti dalla Banca d'Italia per ogni famiglia. Nel caso in esame sono stati rilevati $r = 7287$ differenti valori della variabile Y.

6.1 Caratteristiche aggregate delle tre macroregioni italiane

In Tabella 6.1 sono riportate: la media aritmetica $M_\ell(Y)$, la mediana $Med_\ell(Y)$, l'indice di ineguaglianza sintetico $V_{.\ell}(Y)$, la somma dei pesi $n_{.\ell}$ e la frequenza relativa $n_{.\ell}/n$, relativi alla $\ell-$esima sottopopolazione, $\ell = 1, 2, 3$, e le medesime quantità relative alla popolazione intera. L'indice di ineguaglianza sintetico $V_{.\ell}(Y)$ relativo alla $\ell - esima$ sottopopolazione è dato da:

$$V_{.\ell}(Y) = \sum_{h=1}^{r} V_{h\ell}(Y) \cdot \frac{n_{h\ell}}{n_{.\ell}} = \sum_{h=1}^{r} \frac{M_\ell(Y) - \bar{M}_{h\ell}(Y)}{M_\ell(Y)} \cdot \frac{n_{h\ell}}{n_{.\ell}}$$

dove

$$V_{h\ell}(Y) = \frac{M_\ell(Y) - \bar{M}_{h\ell}(Y)}{M_\ell(Y)}$$

rappresenta l'indice puntuale di Bonferroni calcolato nella sottopopolazione ℓ. I dati riportati in Tabella 6.1 mostrano che:

[1] Per maggiori chiarimenti si rimanda a [18]

- il reddito medio delle famiglie del ”Sud” è molto distante rispetto al reddito medio familiare delle altre due aree considerate.

- il ”Nord” registra una misura di ineguaglianza superiore rispetto alle altre aree, mentre il ”Centro” risulta l’area con minore diseguaglianza di redditi; l’indice di ineguaglianza relativo a tutta la popolazione è invece leggermente superiore a quella del ”Nord”.

- La misura sintetica $V(Y) = 0,4795$ indica che, in media, la media inferiore è pari al $[1 - V(Y)] \cdot 100 = 52,06\%$ della media generale, o detto in altri termini, che, in media, il gruppo dei più poveri ha un reddito medio pari al $52,05\%$ del reddito medio totale.

In Figura 6.1 si riporta il grafico dell’indice puntuale $V_{h(p_{h.})}(Y)$ relativo all’intera popolazione; i punti hanno coordinate:

$$\left(p_{h.} = \frac{P_{h.}}{n}, \quad V_{h(p_{h.})}(Y)\right),$$

dove $h(p_{h.}) = h : P(Y \leq y_h) = p_{h.}$.

In figura 6.2 vengono invece riportati i grafici degli indici puntuali $V_{h(p_{h\ell})\ell}(Y)$ relativo alle tre sottopopolazioni; i punti hanno coordinate:

$$\left(p_{h\ell} = \frac{P_{h\ell}}{n_{.\ell}}, \quad V_{h(p_{h\ell})\ell}(Y)\right),$$

dove $h(p_{h\ell}) = h : P(Y \leq y_h) = p_{h\ell}$.

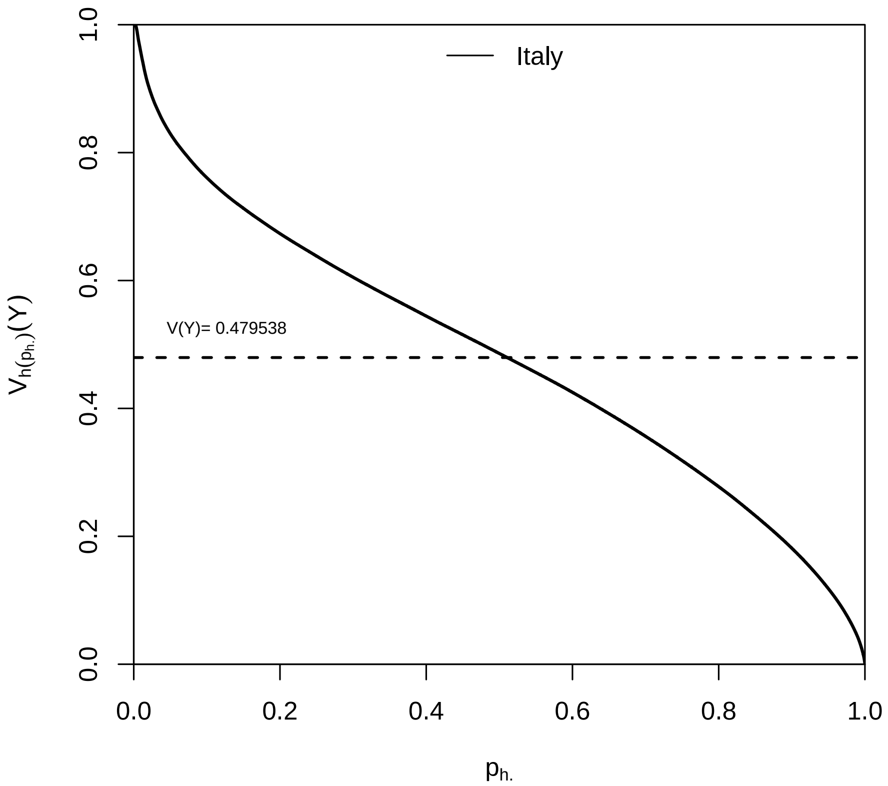


Figura 6.1: Grafico dell' indice di ineguaglianza puntuale di Bonferroni $V_{h(p_{h.})}(Y)$, relativo all'intera popolazione.

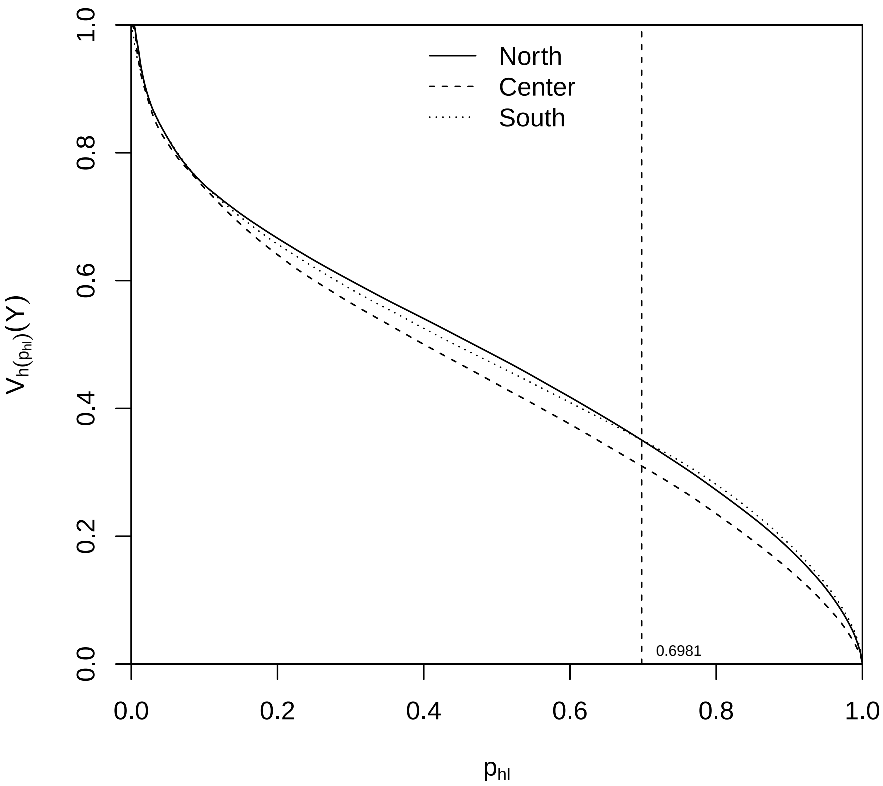


Figura 6.2: Grafici degli indici di ineguaglianza puntuale di Bonferroni $V_{h(p_{h\ell})\ell}(Y)$, $\ell = 1, 2, 3$.

6.2 Scomposizione per area geografica degli indici puntuali $V_h(Y)$ e dell'indice sintetico $V(Y)$

In questa sezione si illustra la scomposizione dell'indice sintetico $V(Y)$ e dell'indice puntuale $V_{h(p_{h.})}$ per tre valori di $p_{h.}$. I valori scelti per $p_{h.}$ sono:

- $p_{h.} = 0.10$ in quanto $V_{h(0{,}10)}(Y) = V_{460}(Y) = 0,7607$ confronta il reddito medio del 10% delle famiglie più povere col reddito medio complessivo. Il valore di $1 - V_{460}(Y) = 0,2393$ indica che il reddito medio delle famiglie più povere è il $23,93\%$ del reddito medio di tutta la popolazione.

- $p_{h.} = 0.50$ in quanto $V_{h(0{,}50)}(Y) = V_{3064}(Y) = 0,4859$ confronta il reddito medio delle famiglie con reddito $Y \leq Med(Y)$ col reddito medio di tutta la popolazione.

Il valore $1 - V_{3064}(Y) = 0,5141$ indica che il reddito delle famiglie con reddito inferiore al reddito mediano è pari al $51,41\%$ del reddito familiare medio.

- $p_{h.} = 0.95$, in quanto $V_{h(0,95)}(Y) = V_{6841}(Y) = 0,1181$ confronta il reddito medio delle famiglie con reddito minore o uguale a $y_{6841} = 68.819,23$ col reddito medio complessivo. Il valore $1 - V_{6841}(Y) = 0,8819$ indica che il reddito delle famiglie con reddito inferiore al 95−esimo percentile è pari all' $88,19\%$ del reddito medio di tutta la popolazione.

Tabella 6.1: Applicazione: Statistiche descrittive

	Sottopopolazione			Italia
	$\ell = 1$	$\ell = 2$	$\ell = 3$	
$n_{.\ell}$	3.971,949	1.537,372	2.641,679	8.151=n
$n_{.\ell}/n$	0,4873	0,1886	0,3241	1,0000
$Med_\ell(Y)$	27.527,57	29.824,24	19.123,67	24.590,10=Med(Y)
$M_\ell(Y)$	33.543,17	34.000,09	23.517,86	30.380,22=$M(Y)$
$V_{.\ell}(Y)$	0.4740	0.4421	0.4695	0,4795=$V(Y)$

Tabella 6.2: Frequenze cumulate e quantili per tre valori di $p_{h.} = P(Y \leq y_h)$ del reddito totale Y di tutta la popolazione

$p_{h.}$	$h(p_{h.})$	$P_{h(p_{h.}).}$	$P_{h(p_{h.}).}/n$	y_h
$p_{h.} = 0,10$	460	815,20	0,10	10.600,00
$p_{h.} = 0,50$	3.064	4.075,65	0,50	24.590,10
$p_{h.} = 0,95$	6.841	7.743,48	0,95	68.819,23
$p_{h.} = 1,00$	7.287	8.151,00	1,00	368.689,70

Tabella 6.3: Quantità necessarie al calcolo dei contibuti $V_{h(0,10)\ell g}(Y) = V_{460\ell g}(Y)$

$p_{h.} = 0,10; h = 460$	Sottopopolazione			$M(Y)$
	Nord	Centro	Sud	
$M_\ell(Y)$	33.543,17	34.000,09	23.517,86	30.380,22

$\bar{M}_{h\ell}(Y)$	$\bar{M}_{h1}(Y)$	$\bar{M}_{h2}(Y)$	$\bar{M}_{h3}(Y)$	$\bar{M}_{h.}(Y)$
	7.091,45	7.554,05	7.310,03	7.270,21

$\frac{n_{.\ell}}{n}$	$\frac{n_{.1}}{n}$	$\frac{n_{.2}}{n}$	$\frac{n_{.3}}{n}$	Tot.
	0,4873	0,1886	0,3241	1,00

$p(\ell\|h) = \frac{P_{h\ell}}{P_{h.}}$	$\frac{P_{h1}}{P_{h.}}$	$\frac{P_{h2}}{P_{h.}}$	$\frac{P_{h3}}{P_{h.}}$	Tot.
	0,3383	0,1399	0,5218	1,00

Tabella 6.4: Scomposizione di $V_{h(0,1)}(Y) = V_{460}(Y) = 0,7607$ nei contributi $V_{460\ell g}(Y)$, $V_{460\ell.}(Y)$, $V_{460\ell W}(Y)$, $V_{460\ell B}(Y)$, $V_{460.W}(Y)$, $V_{460.B}(Y)$

$V_{460\ell g}(Y)$	ℓ			
g	1	2	3	
1	0,1435	0,0583	0,2196	
2	0,0565	0,0230	0,0865	
3	0,0593	0,0238	0,0902	
$V_{460\ell.}(Y)$	0,2593	0,1051	0,3963	0,7607=$V_{460}(Y)$
$V_{460\ell W}(Y)$	0,1435	0,0230	0,0902	0,2567=$V_{460.W}(Y)$
$V_{460\ell B}(Y)$	0,1158	0,0821	0,3060	0,5040=$V_{460.B}(Y)$

La Tabella 6.4 mostra che il contributo $V_{h(0,10)}(Y) = V_{460\ell g}(Y)$ maggiore è fornito da $V_{h(0,1)31}(Y) = 0,2196$:

$$\begin{aligned} V_{h(0,1)31}(Y) &= \frac{M_1(Y) - \bar{M}_{h(0,1)3}(Y)}{M(Y)} \cdot \frac{n_{.1}}{n} \cdot p(3|460) \\ &= \frac{33.543,17 - 7.310,03}{30.380,22} \cdot 0,4873 \cdot 0,5218 \\ &= 0,8635 \cdot 0,4873 \cdot 0,5218 = 0,2196 \end{aligned}$$

Considerando invece il contributo $V_{h(0,1)13}(Y)$ si ha invece che:

$$\begin{aligned}
V_{h(0,1)13}(Y) &= \frac{M_3(Y) - \bar{M}_{h(0,1)1}(Y)}{M(Y)} \cdot \frac{n_{.3}}{n} \cdot p(1|460) \\
&= \frac{23.517,86 - 7.091,45}{30.380,22} \cdot 0,3241 \cdot 0,3383 \\
&= 0,5407 \cdot 0,3241 \cdot 0,3383 = 0,0593
\end{aligned}$$

I risultati sopra riportati, ricavati dalla Tabella 6.4, mostrano che il contributo $V_{h(0,1)31}(Y)$ dipende dalla differenza tra il reddito medio familiare del Nord e il reddito delle famiglie del Sud con reddito inferiore o uguale a $y_{460} = 10.600,00$ Euro e dalle rispettive frequenze relative. Il contributo $V_{h(0,1)13}(Y)$ dipende invece dal confronto tra il reddito medio delle famiglie del Sud ed il reddito delle famiglie del Nord con reddito inferiore o uguale a $y_{460} = 10.600,00$ Euro e dalle rispettive frequenze relative. La differenza tra le quantità qui riportate fornisce quindi la spiegazione della notevole differenza tra i due contributi.
Per quanto riguarda i contributi delle singole sottopopolazioni all'indice di ineguaglianza puntuale $V_{h(0,1)}(Y) = V_{460}(Y)$ si ha:

$$\begin{aligned}
V_{h(0,1)1.}(Y) &= \frac{M(Y) - \bar{M}_{h(0,1)1.}(Y)}{M(Y)} \cdot p(1|460) \\
&= \frac{30.380,22 - 7.091,45}{30.380,22} \cdot 0,3383 \\
&= 0,7666 \cdot 0,3383 = 0,2593
\end{aligned}$$

$$\begin{aligned}
V_{h(0,1)2.}(Y) &= \frac{M(Y) - \bar{M}_{h(0,1)2.}(Y)}{M(Y)} \cdot p(2|460) \\
&= \frac{30.380,22 - 7.554,05}{30.380,22} \cdot 0,1399 \\
&= 0,7513 \cdot 0,1399 = 0,1051
\end{aligned}$$

$$\begin{aligned}
V_{h(0,1)3.}(Y) &= \frac{M(Y) - \bar{M}_{h(0,1)3.}(Y)}{M(Y)} \cdot p(3|460) \\
&= \frac{30.380,22 - 7.310,03}{30.380,22} \cdot 0,5218 \\
&= 0,7594 \cdot 0,5218 = 0,3963
\end{aligned}$$

I valori riportati relativi ai contributi $V_{460\ell.}(Y) = V_{h(0,10)\ell.}(Y)$ mostrano che le variazioni relative delle medie inferiori $\bar{M}_{h(0,1)\ell}(Y)$ delle k sottopopolazioni rispetto alla media generale $M(Y)$ sono simili, mentre risultano differenti i rispettivi pesi $p(\ell|h)$. La considerevole differenza dei contributi delle tre macroaree all'indice di ineguaglianza puntuale $V_{460}(Y) = 0,7607$ è dunque attribuibile in misura maggiore alla differente frequenza relativa $p(\ell|h)$ delle famiglie con reddito minore uguale a $y_{460} = 10.600$ Euro nella sottopopolazione ℓ. In particolare, dato che il 52% del 10% delle famiglie più povere in Italia è rappresentato dalle famiglie del Sud, ciò spiega perchè il Sud contribuisce in misura maggiore all'indice di ineguaglianza puntuale $V_{h(0,10)}(Y) = 0,7607$

Tabella 6.5: Quantità necessarie al calcolo dei contibuti $V_{h(0,5)\ell g}(Y) = V_{3.064\ell g}(Y)$

$p = 0,50;\ h = 3.064$	Sottopopolazione			$M(Y)$
	Nord	Centro	Sud	
$M_\ell(Y)$	33.543,17	34.000,09	23.517,86	30.380,22

$\bar{M}_{h\ell}(Y)$	$\bar{M}_{h1}(Y)$	$\bar{M}_{h2}(Y)$	$\bar{M}_{h3}(Y)$	$\bar{M}_{h.}(Y)$
	16.013,33	16.449,92	14.972,15	15.618,20

$\frac{n_{.g}}{n}$	$\frac{n_{.1}}{n}$	$\frac{n_{.2}}{n}$	$\frac{n_{.3}}{n}$	Tot.
	0,4873	0,1886	0,3241	1,00

$p(\ell\|h) = \frac{P_{h\ell}}{P_{h.}}$	$\frac{P_{h1}}{P_{h.}}$	$\frac{P_{h2}}{P_{h.}}$	$\frac{P_{h3}}{P_{h.}}$	Tot.
	0,4196	0,1416	0,4389	1,00

Tabella 6.6: Scomposizione di $V_{h(0,5)}(Y) = V_{3064}(Y) = 0,4859$ nei contributi $V_{3064\ell g}(Y)$, $V_{3064\ell.}(Y)$, $V_{3064\ell W}(Y)$, $V_{3064\ell B}(Y)$, $V_{3064.W}(Y)$, $V_{3064.B}(Y)$

$V_{3064\ell g}(Y)$	ℓ			
g	1	2	3	
1	0,1180	0,0388	0,1307	
2	0,0469	0,0154	0.,0518	
3	0,0336	0,0107	0,0400	
$V_{3064\ell.}(Y)$	0,1984	0,0649	0,2226	0,4859=$V_{3064}(Y)$
$V_{3064\ell W}(Y)$	0,1180	0,0154	0,0400	0,1734=$V_{3064.W}(Y)$
$V_{3064\ell B}(Y)$	0,0804	0,0495	0,1826	0,3125=$V_{3064.B}(Y)$

Tabella 6.7: Quantità necessarie al calcolo dei contibuti $V_{h(0,95)\ell g}(Y) = V_{6841\ell g}(Y)$

p=0,95; h=6.841	Sottopopolazione			$M(Y)$
	Nord	Centro	Sud	
$M_\ell(Y)$	33.543,17	34.000,09	23.517,86	30.380,22

$\bar{M}_{h\ell}(Y)$	$\bar{M}_{h1}(Y)$	$\bar{M}_{h2}(Y)$	$\bar{M}_{h3}(Y)$	$\bar{M}_{h.}(Y)$
	28.856,94	30401.45	21.819,47	26.791,44

$\frac{n_{.g}}{n}$	$n_{.1}/n$	$n_{.2}/n$	$n_{.3}/n$	Tot.
	0,4873	0,1886	0,3241	1,00

$p(\ell\|h) = \frac{P_{h\ell}}{P_{h.}}$	$P_{h1}/P_{h.}$	$P_{h2}/P_{h.}$	$P_{h3}/P_{h.}$	Tot.
	0,4792	0,1864	0,3344	1,00

Tabella 6.8: Scomposizione di $V_{h(0,95)}(Y) = V_{6841}(Y) = 0,1181$ nei contributi $V_{6841\ell g}(Y)$, $V_{6841\ell.}(Y)$, $V_{6841\ell W}(Y)$, $V_{6841\ell B}(Y)$, $V_{6841.W}(Y)$, $V_{6841.B}(Y)$

$V_{6841\ell g}(Y)$	ℓ			
g	1	2	3	
1	0,0360	0,0094	0,0629	
2	0,0153	0,0042	0,0253	
3	-0,0273	-0,0137	0,0061	
$V_{6841\ell.}(Y)$	0,0240	-0,0001	0,0942	0,1181=$V_{6841}(Y)$
$V_{6841\ell W}(Y)$	0,0360	0,0042	0,0061	0,0462=$V_{6841.W}(Y)$
$V_{6841\ell B}(Y)$	-0,0120	-0,0043	0,0882	0,0719=$V_{6841.B}(Y)$

In analogia a quanto fatto per $p = 0,10$ si ricavano ora i contributi $V_{6841\ell g}(Y) = V_{h(0,95)\ell g}(Y)$ e $V_{6841\ell.}(Y) = V_{h(0,95)\ell.}(Y)$ all'indice di ineguaglianza puntuale $V_{h(0,95)}(Y) = V_{6841}(Y) = 0,1181$. Dalla Tabella 6.8 si nota che il contributo maggiore all'indice $V_{6841}(Y)$ è fornito da $V_{h(0,95)31}(Y) = 0,0629$:

$$\begin{aligned} V_{h(0,95)31}(Y) &= \frac{M_1(Y) - \bar{M}_{h(0,95)3}(Y)}{M(Y)} \cdot \frac{n_{.1}}{n} \cdot p(3|6841) \\ &= \frac{33.543,17 - 21.819,47}{30.380,22} \cdot 0,4873 \cdot 0,3344 \\ &= 0,3859 \cdot 0,4873 \cdot 0,3344 = 0,0629 \end{aligned}$$

Per quanto riguarda invece il contributo $V_{h(0,95)13}(Y) = -0,0273$ si ha:

$$\begin{aligned}
V_{h(0,95)13}(Y) &= \frac{M_3(Y) - \bar{M}_{h(0,95)1}(Y)}{M(Y)} \cdot \frac{n_{.3}}{n} \cdot p(1|6841) \\
&= \frac{23.517,86 - 28.856,94}{30.380,22} \cdot 0,3241 \cdot 0,4792 \\
&= -0,1757 \cdot 0,3241 \cdot 0,4792 = -0,0273
\end{aligned}$$

In relazione ai contributi $V_{h(0,95)\ell.}(Y)$ delle singole sottopopolazioni all'indice $V_{h(0,95)}(Y)$ si ricava:

$$\begin{aligned}
V_{h(0,95)1.}(Y) &= \frac{M(Y) - \bar{M}_{h(0,95)1}(Y)}{M(Y)} \cdot p(1|6841) \\
&= \frac{30.380,22 - 28.856,94}{30.380,22} \cdot 0,4792 \\
&= 0,0501 \cdot 0,4792 = 0,0240
\end{aligned}$$

$$\begin{aligned}
V_{h(0,95)2.}(Y) &= \frac{M(Y) - \bar{M}_{h(0,95)2}(Y)}{M(Y)} \cdot p(2|6841) \\
&= \frac{30.380,22 - 30.401,45}{30.380,22} \cdot 0,1864 \\
&= -0,0007 \cdot 0,1864 = -0,0001
\end{aligned}$$

$$\begin{aligned}
V_{h(0,95)3.}(Y) &= \frac{M(Y) - \bar{M}_{h(0,95)3}(Y)}{M(Y)} \cdot p(3|6841) \\
&= \frac{30.380,22 - 21.819,47}{30.380,22} \cdot 0,3344 \\
&= 0,2818 \cdot 0,3344 = 0,0942
\end{aligned}$$

Da quanto ottenuto si può constatare che:

- il confronto tra il reddito medio del Sud ed il reddito medio delle famiglie del Nord con reddito inferiore o uguale a $y_{6841} = 68.819,23$ Euro comporta una "impercettibile" riduzione dell'indice di ineguaglianza puntuale $V_{6841}(Y)$;

- Il contributo del Centro all'ineguaglianza puntuale $V_{6841}(Y)$ è di fatto nullo.

Tabella 6.9: Scomposizione dell'indice sintetico $V(Y) = 0,4795$ nei contributi $V_{.\ell g}(Y)$, $V_{.\ell.}(Y)$, $V_{.\ell W}(Y)$, $V_{.\ell B}(Y)$, $V_{..W}(Y)$, $V_{..B}(Y)$

$V_{.\ell g}(Y)$	ℓ			
g	1	2	3	
1	0,1114	0,0364	0,1366	
2	0,0443	0,0145	0,0541	
3	0,0289	0,0084	0,0449	
$V_{.\ell.}(Y)$	0,1847	0,0593	0,2355	0,4795=$V(Y)$
$V_{.\ell W}(Y)$	0,1114	0,0145	0,0449	0,1708=$V_{..W}(Y)$
$V_{.\ell B}(Y)$	0,0733	0,0448	0,1907	0,3088=$V_{..B}(Y)$

In relazione alla scomposizione dell'indice sintetico $V(Y)$ è possibile osservare che la macroarea che contribuisce in misura maggiore alla determinazione dell'indice di ineguaglianza sintetico è il Sud in quanto $V_{.3.}(Y) = 0,2355$. Tale risultato è ottenuto ricordando che $V_{.3.}(Y)$ rappresenta la media ponderata dei contributi $V_{h3.}(Y)$ con pesi $n_{h.}/n$.

Al fine di una maggiore interpretazione dei risultati ottenuti, si introdurranno ora i contributi relativi.

6.3 Contributi Relativi

Al fine di una maggiore comprensione del contributo delle sottopopolazioni agli indici di ineguaglianza puntuali e sintetici calcolati su tutta la popolazione, si introducono i seguenti contributi relativi:

$$\nu_{h\ell.}(Y) = \frac{V_{h\ell.}(Y)}{V_h(Y)} = \frac{M(Y) - \bar{M}_{h\ell}(Y)}{M(Y) - \bar{M}_{h.}(Y)} \cdot p(\ell|h) \qquad h = 1, ..., r-1 \tag{6.1}$$

$$\nu_{.\ell.}(Y) = \frac{V_{.\ell.}(Y)}{V(Y)} \tag{6.2}$$

In virtù delle (5.11) e (3.5) si può notare che il contributo relativo $\nu_{h\ell.}(Y)$ in (6.1), ovvero il contributo relativo della sottopopolazione ℓ all'indice di ineguaglianza puntuale $V_h(Y)$, è il rapporto tra la variazione (relativa) della media inferiore $\bar{M}_{h\ell}(Y)$ della sottopopolazione ℓ rispetto la media generale $M(Y)$ e la variazione (relativa) della media inferiore di tutta la popolazione $\bar{M}_{h.}(Y)$ rispetto la media generale $M(Y)$, moltiplicata per $p(\ell|h)$, frequenza relativa delle unità statistiche con reddito $Y \leq y_h$ nella sottopopolazione ℓ.

In relazione al contributo relativo della sottopopolazione ℓ all'indice di ineguaglianza sintetico $V(Y)$, si osserva che $\nu_{.\ell.}(Y)$ è il rapporto tra $V_{.\ell.}(Y)$, media ponderata dei contributi $V_{h\ell}(Y)$ con pesi $n_{h.}/n$ e la media ponderata delle misure di ineguaglianza puntuali

$V_h(Y)$, con pesi $n_{h.}/n$.

In Tabella 6.10 sono presentati i contributi relativi delle sottopopolazioni all'indice puntuale $V_{h(p_{h.})}(Y)$, per $p_{h.} = 0,10,\ 0,50,\ 0,95$ e i contributi relativi delle sottopopolazioni all'indice di ineguaglianza sintetico $V(Y)$.

Tabella 6.10: Contributi relativi $\nu_{460\ell.}(Y), \nu_{3064\ell.}(Y), \nu_{6841\ell.}(Y), \nu_{.\ell.}(Y)$

	ℓ			Tot.
	Nord	Centro	Sud	
$\nu_{460\ell.}(Y)$	0,3409	0,1381	0,5209	1,00
$\nu_{3064\ell.}(Y)$	0,4084	0,1336	0,4581	1,00
$\nu_{6841\ell.}(Y)$	0,2034	-0,0011	0,7977	1,00
$\nu_{.\ell.}(Y)$	0,3851	0,1237	0,4912	1,00
$n_{.\ell}/n$	0,4873	0,1886	0,3241	1,00

In conclusione, dalla Tabella 6.10 si può notare che il Sud contribuisce in misura maggiore all'ineguaglianza complessiva. Il valore dell'indice sintetico $V(Y) = 0,4795$ è infatti dovuto per il $49,12\%$ ai contributi forniti da questa macroregione. In relazione alle componenti *within* e *between* si osserva infine che l'ineguaglianza presente all'interno delle macroregioni ha un'incidenza del $0,1708/0,4795 \cdot 100 = 35,62\%$, mentre l'ineguaglianza derivante dai confronti tra macroregioni differenti ha un'incidenza del $(1-0,3562)\cdot 100 = 64,38\%$.

Capitolo 7

Conclusioni

In questa sezione saranno considerati i vantaggi che la metodologia proposta presenta rispetto alle altre procedure di scomposizione per sottopopolazioni dell'indice di Bonferroni.

- Innanzitutto, va ricordato (Cap. 3) che la valutazione dell'indice (sintetico) di Bonferroni $V(Y)$, nel contesto delle distribuzioni di frequenza, oltre a conservare le usuali proprietà che un indice di ineguaglianza deve soddisfare, è altresì *invariante a replicazioni nella popolazione.*

- Inoltre, mentre le scomposizioni per sottopopolazioni finora proposte riguardano unicamente l'indice sintetico, la metodologia presentata in questo lavoro fornisce anche la scomposizione delle misure puntuali di ineguaglianza $V_h(Y)$, agevolando di conseguenza la valutazione dell'ineguaglianza presente in quote $p_{h.}$ di popolazione: $h = 1, ..., r$, ove r indica il numero dei distinti valori che assume la variabile oggetto di studio.

- La scomposizione per sottopopolazioni qui illustrata, oltre a fornire le componenti *within* e *between* degli indici puntuali e sintetici, permette inoltre di valutare il *contributo delle singole sottopopolazioni* all'indice di ineguaglianza (sia a livello puntuale che sintetico).

Un aspetto ribadito più volte nel testo riguarda il valore assunto dall'indice $V(Y)$. Dato che per il calcolo dell'indice $V(Y)$ si considerano tutti gli n valori assunti dalla variabile Y e che, per $1+P_{h.}-n_{h.} \leq i \leq P_{h.}$, si assume $\tilde{V}_i(Y) = V_h(Y)$[1], $i = 1, \ldots, n,\ h = 1, \ldots, r$, segue che $V(Y) \leq \tilde{V}(Y)$. In riferimento all'Applicazione qui presentata è però opportuno ricordare che:

[1]Si ricorda che con tale assunto viene assegnata medesima misura di ineguaglianza puntuale alle unità che presentano medesino valore della variabile Y

- il valore assunto dall'indice di Bonferroni è $\tilde{V}(Y) = 0,4796$, mentre il valore dell'indice calcolato con la metodologia proposta è $V(Y) = 0,4795$

- degli $r = 7287$ distinti valori che assume la variabile Y qui considerata, 6995 presentano valore $n_{h.} = 1$

In relazione all'indice di Bonferroni non si possono tuttavia non evidenziare le "anomalie" legate alla definizione dell'indice. Dato che per $h \to r$, $\bar{M}_{h\ell}(Y) \to M_\ell(Y)$ e, per $h = r$, $\bar{M}_{h\ell}(Y) = M_\ell(Y)$, può accadere che, qualora $M_g(Y) < \bar{M}_{h\ell}(Y)$, alcuni contributi

$$V_{h\ell g}(Y) = \frac{M_g(Y) - \bar{M}_{h\ell}(Y)}{M(Y)} \cdot p(\ell|h) \cdot \frac{n_{.g}}{n} \qquad (g, \ell = 1, \ldots, k;\, g \neq \ell)$$

abbiano segno negativo. [2]

Tale "anomalia", inoltre, interessa anche i termini $V_{h\ell.}(Y)$, contributi delle singole sottopopolazioni all'indice di ineguaglianza puntuale. Infatti, essendo,

$$V_{h\ell.}(Y) = \frac{M(Y) - \bar{M}_{h\ell}(Y)}{M(Y)} \cdot p(\ell|h),$$

quando si verifica $\bar{M}_{h\ell}(Y) > M(Y)$, il contributo della sottopopolazione $\ell - esima$ all'ineguaglianza, risulta negativo. A tal riguardo, in riferimento all'Applicazione qui presentata, sono stati rilevati i seguenti valori:

$$M(Y) = 30.380,22;\ M_1(Y) = 33.543,57;\ M_2(Y) = 34.000,09;\ M_3(Y) = 23.517,86.$$

Ne consegue che il contributo all'ineguaglianza della macro-area "Sud e Isole" ($\ell = 3$) non è mai negativo, mentre, vi sono valori di h tali che i contributi delle macroaree "Nord" ($\ell = 1$) e "Centro" ($\ell = 2$) risultano negativi.

In particolare, per:

1. $h \geq 7055$, $\bar{M}_{h1}(Y) > M(Y)$ e pertanto $V_{h1.}(Y) < 0$;

2. $h \geq 6850$, $\bar{M}_{h2}(Y) > M(Y)$ e pertanto $V_{h2.}(Y) < 0$.

Si puntualizza inoltre che tale "anomalia", interessa anche i contributi $V_{.\ell g}(Y)$ dell'indice sintetico $V(Y)$. A titolo di esempio, si consideri la distribuzione in Tabella 7.1.

[2] A tal riguardo si consideri la Tabella 6.8

Tabella 7.1: Distribuzione di frequenza di $n = 30$ unità statistiche suddivise in $k = 3$ sottopopolazioni.

y_h	n_{h1}	n_{h2}	n_{h3}
0	1	1	1
2	6	1	0
10	1	6	0
13	1	1	5
26	1	1	4
$n_{.\ell}$	10	10	10

In riferimento ai dati in Tabella 7.1, applicando la metodologia proposta, si ottengono le scomposizioni riportate in Tabella 7.2, da cui è possibile osservare i contributi $V_{.21}(Y) = -0,0061$ e $V_{.31}(Y) = -0,0171$.

Tabella 7.2: Contributi $V_{.\ell g}(Y)$, $V_{.\ell .}(Y)$, $V_{.\ell W}(Y)$, $V_{.\ell B}(Y)$, $V_{..W}(Y)$ e $V_{..B}(Y)$ all'indice di ineguaglianza sintetico $V(Y) = 0,5116$

$V_{.\ell g}(Y)$	$\ell = 1$	$\ell = 2$	$\ell = 3$
$g = 1$	0,0447	-0,0061	-0,0171
$g = 2$	0,1004	0,0355	0,0065
$g = 3$	0,1950	0,1061	0,0466

$V_{.\ell .}(Y)$	0,3401	0,1355	0,0360	$V(Y) = 0,5116$
$V_{.\ell W}(Y)$	0,0447	0,0355	0,0466	$V_{..W}(Y) = 0,1268$
$V_{.\ell B}(Y)$	0,2954	0,1	-0,0106	$V_{..B}(Y) = 0,3848$

Si evidenzia infine che i contributi *within* $V_{h\ell W}(Y)$ all'indice puntuale, interessando confronti tra medie della medesima sottopopolazione, non possono mai essere negativi e di conseguenza, non possono essere mai negativi i contributi *within* $V_{.\ell W}(Y)$ e $V_{..W}(Y)$ dell'indice sintetico $V(Y)$. Al contrario, in virtù del contributo $V_{.3B}(Y) = -0,0106$ in Tabella 7.2, è possibile osservare che contributi $V_{.\ell B(Y)}$, e di conseguenza, i contributi $V_{..B}(Y)$, possono essere negativi. In generale, quindi, essendo $V_{.\ell .}(Y) = V_{.\ell W}(Y) + V_{.\ell B}(Y)$, può accadere che anche i termini $V_{.\ell .}(Y)$, contributi delle singole sottopopolazioni all'indice sintetico, risultino negativi.

In relazione agli indici di ineguaglianza ricordati in questo lavoro, l'unico indice a non presentare le "anomalie" ora osservate è l'indice di Zenga [16]. Dato che,

$$I(Y) = \sum_{h=1}^{r} I_h(Y) \cdot \frac{n_{h.}}{n}, \quad \text{dove} \quad I_h(Y) = \frac{\overset{+}{M}_{h.}(Y) - \overset{-}{M}_{h.}(Y)}{\overset{+}{M}_{h.}(Y)},$$

i confronti considerati in dette misure riguardano medie tra i seguenti gruppi disgiunti:

- il *gruppo inferiore*, definito dalle coppie $\{(y_t, n_{t.}), t = 1, \ldots, h\}$;

- il *gruppo superiore*, definito dalle coppie $\{(y_{h+1}, n_{h+1.}), h = 1, \ldots, r-1\}$.

Per costruzione, pertanto, $\overset{+}{M}_{hg}(Y) \geq \overset{-}{M}_{h\ell}(Y)$, dove l'uguaglianza vale unicamente nel caso teorico di ineguaglianza nulla.

Si conclude infine puntualizzando che le sopra menzionate anomalie dell'indice di Bonferroni, si estendono anche all'indice di Gini, ricordando [3] la seguente relazione fra le misure puntuali dei due indici:

$$G_h(Y) = V_h(Y) \cdot \frac{1 + 2 \cdot P_{h.} - n_{h.}}{n+1}.$$

[3] Vedere pagina 21

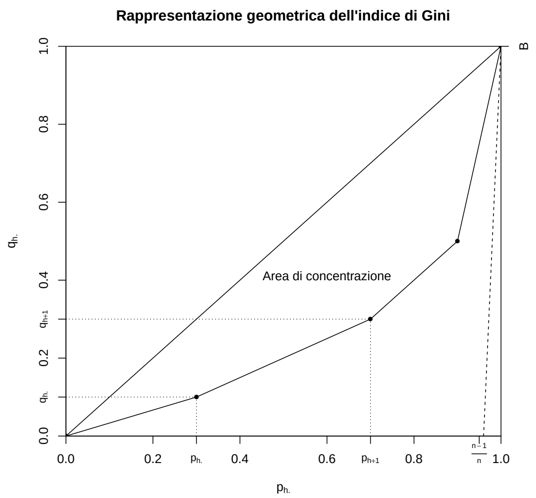
Capitolo 8

Appendice

8.1 Appendice A

La "regola dei trapezi" è un metodo che permette di determinare l'indice di ineguaglianza di Gini $\tilde{R}(Y)$ utilizzando un approccio di tipo geometrico.

Figura 8.1: Interpretazione geometrica dell' indice di Gini $\tilde{R}(Y)$



Si consideri la Figura 8.1. In termini geometrici, l'indice $\tilde{R}(Y)$ rappresenta il rapporto tra l'area di concentrazione, ovvero l'area compresa tra la retta di equiripartizione (la bisettrice) e la spezzata di coordinate $(p_{h.}, q_{h.})$, e l'area di massima concentrazione, ovvero l'area di concentrazione che si genera qualora $n_{1.} = n - 1$ unità statistiche presentano valore $y_1 = 0$ e $n_{2.} = 1$ unità statistica presenta valore $y_2 = T(Y)$. In altri termini:

$$\tilde{R}(Y) = \frac{\text{Area di concentrazione}}{\text{Area di massima concentrazione}}. \tag{8.1}$$

L'area $A(\cdot)$ di massima concentrazione è data dall'area del triangolo di vertici $0\frac{n-1}{n}B$:

$$A(0\frac{n-1}{n}B) = \frac{1}{2} \cdot \frac{n-1}{n} \cdot 1 = \frac{n-1}{2 \cdot n}. \tag{8.2}$$

L'area di concentrazione si può ottenere invece per differenza tra l'area soggiacente la bisettrice, ovvero l'area del triangolo $01B$ e la somma delle aree degli r trapezi t_h, $h = 1, ..., r$, soggiacenti la spezzata di coordinate $(p_{h.}, q_{h.})$. Detta $A(01B)$ l'area soggiacente la bisettrice e $A(t_h)$ l'area dell' h-esimo trapezio, $h, = 1, ..., r$, si ha:

$$\begin{aligned} A(O1B) &= \frac{1}{2} && (8.3)\\ A(t_h) &= \frac{(\text{Base maggiore} + \text{Base minore}) \cdot \text{altezza}}{2} \\ &= \frac{(q_{h.} + q_{h+1.}) \cdot (p_{h+1.} - p_{h.})}{2} \\ &= \frac{q_{h+1.} \cdot p_{h+1.} - q_{h+1.} \cdot p_{h.} + q_{h.} \cdot p_{h+1.} - q_{h.} \cdot p_{h.}}{2}. && (8.4) \end{aligned}$$

Pertanto:

$$\sum_{h=1}^{r-1} A(t_h) = \sum_{h=1}^{r-1} \frac{p_{h+1.} \cdot p_{h+1.} - q_{h+1.} \cdot p_{h.} + q_{h.} \cdot p_{h+1.} - p_{h.} \cdot p_{h.}}{2}. \tag{8.5}$$

Dato che $A(O1B) = 1/2$ e che $\sum_{h=1}^{r-1} p_{h+1.} \cdot q_{h+1.} - \sum_{h=1}^{r-1} p_{h.} \cdot q_{h.} = p_r \cdot q_r = 1$, si ha:

$$\sum_{h=1}^{r-1} A(t_h) = \frac{\sum_{h=1}^{r-1}(q_{h+1.} \cdot q_{h.} - p_{h+1.} \cdot q_{h.})}{2}. \tag{8.6}$$

In virtù della (8.1) e dei risultati in (8.2) e in (8.6):

$$\begin{aligned} \tilde{R}(Y) &= \frac{\text{Area di concentrazione}}{\text{Area di massima concentrazione}} \\ &= \frac{\sum_{h=1}^{r-1}(q_{h+1.} \cdot q_{h.} - p_{h+1.} \cdot q_{h.})}{2} \cdot \frac{2 \cdot n}{n-1} \\ &= \frac{n}{n-1} \cdot \sum_{h=1}^{r-1}(q_{h+1.} \cdot q_{h.} - p_{h+1.} \cdot q_{h.}) && (8.7)\\ &= \frac{\Delta(Y)}{2 \cdot M(Y)} \\ &= \frac{1}{2 \cdot M(Y)} \cdot \frac{\sum_{h=1}^{r} \sum_{s=1}^{r} |y_h - y_s| \cdot n_{h.} \cdot n_{s.}}{n \cdot (n-1)} && (h, s = 1, ..., r)^1. \end{aligned}$$

Se al denominatore della (8.1), si sostituisce l'Area di massima concentrazione $A(0\frac{n-1}{n}B)$ con l'area $A(01B)$ del triangolo $01B$, in virtù della (8.3) e della (8.7) si ha:

$$
\begin{aligned}
\frac{\text{Area di concentrazione}}{\text{A(01B)}} &= 2 \cdot \frac{\sum_{h=1}^{r-1}(q_{h+1.} \cdot p_{h.} - p_{h+1.} \cdot q_{h.})}{2} \\
&= \frac{n-1}{n} \cdot \left[\frac{n}{n-1} \cdot \sum_{h=1}^{r-1}(q_{h+1.} \cdot q_{h.} - p_{h+1.} \cdot q_{h.}) \right] \\
&= \frac{n-1}{n} \cdot \frac{\Delta(Y)}{2 \cdot M(Y)} \\
&= \frac{n-1}{n} \cdot \left[\frac{1}{2 \cdot M(Y)} \cdot \frac{\sum_{h=1}^{r} \sum_{s=1}^{r} |y_h - y_s| \cdot n_{h.} \cdot n_{s.}}{n \cdot (n-1)} \right] \quad (h, s = 1, ..., r) \\
&= \frac{{}_R\Delta(Y)}{2 \cdot M(Y)} \\
&= c. \qquad (8.8)
\end{aligned}
$$

In virtù della (8.8) si dimostra la (4.16).

8.2 Appendice B

8.2.1 Appendice B.1

$$
\begin{aligned}
c &= \sum_{h=1}^{r-1} \left(p_{h.} \cdot q_{(h+1).} - p_{(h+1).} \cdot q_{h.} \right) \\
&= \sum_{h=1}^{r-1} \sum_{g=1}^{k} \sum_{\ell=1}^{k} \left(p_{hg} \cdot q_{(h+1)\ell} - p_{(h+1)g} \cdot q_{h\ell} \right) \\
&= \sum_{h=1}^{r-1} \left[\sum_{g=1}^{k} \sum_{\ell=1}^{k} \left(p_g \cdot p_{hg}^{(n)} \cdot q_\ell \cdot q_{(h+1)\ell}^{(n)} - p_g \cdot p_{(h+1)g}^{(n)} \cdot q_\ell \cdot q_{h\ell}^{(n)} \right) \right] \\
&= \sum_{\ell=1}^{k} \sum_{g=1}^{k} \left[p_g q_\ell \sum_{h=1}^{r-1} \left(p_{hg}^{(n)} q_{(h+1)\ell}^{(n)} - p_{(h+1)g}^{(n)} q_{h\ell}^{(n)} \right) \right] \\
&= \sum_{g=1}^{k} \sum_{\ell=1}^{k} p_g \cdot q_\ell \cdot c_{g\ell} \\
&= \underset{(1 \times k)}{\mathbf{p}'_{\mathbf{g}}} \; \underset{(k \times k)}{\mathbf{c}_{\mathbf{g}\ell}} \; \underset{(k \times 1)}{\mathbf{q}_{\mathbf{g}}} \qquad (8.9)
\end{aligned}
$$

8.2.2 Appendice B.2

In virtù delle ....:

$$\mathbf{c}_{g\ell} = \mathbf{w}_{gl} + \mathbf{d}_{gl}; \quad q_g = b_g \cdot p_g, \quad w_{gl} = c_{gg},\ \forall\, \ell = 1, ..., k, \tag{8.10}$$

Osservando che, fissato g, essendo w_{gl} costante per costruzione, si ha:

$$c_{gg} = \sum_{\ell=1}^{k} w_{gl} \cdot q_\ell = w_{gl} \cdot \sum_{\ell=1}^{k} q_\ell = w_{gl},\ \forall \ell = 1, ..., k \tag{8.11}$$

Pertanto:

$$\begin{aligned} c &= \mathbf{p}_g' \mathbf{c}_{g\ell} \mathbf{q}_g \\ &= \mathbf{p}_g' \left(\mathbf{w}_{g\ell} + \mathbf{d}_{g\ell}\right) \mathbf{q}_g \\ &= \mathbf{p}_g' \mathbf{w}_{g\ell} \mathbf{q}_g + \mathbf{p}_g' \mathbf{d}_{g\ell} \mathbf{q}_g \\ &= \mathbf{p}_g' \mathbf{w}_{g\ell} \left(\mathbf{b}_g \mathbf{p}_g\right) + \mathbf{p}_g' \left(\mathbf{b}_\ell \mathbf{d}_{g\ell}\right) \mathbf{p}_g \\ &= \underset{(1\times k)}{\mathbf{p}_g'} \underset{(k\times 1)}{\mathbf{c}_{gg}} + \underset{(1\times k)}{\mathbf{p}_g'} \underset{(k\times k)}{\left(\mathbf{b}_\ell \mathbf{d}_{g\ell}\right)} \underset{(k\times 1)}{\mathbf{p}_g} \end{aligned} \tag{8.12}$$

Dato che,

$$\mathbf{p}_g' \mathbf{c}_{gg} = \sum_{g=1}^{k} p_g \cdot c_{gg}; \tag{8.13}$$

$$\mathbf{p}_g' (\mathbf{b}_\ell \mathbf{d}_{g\ell}) \mathbf{p}_g = \sum_{g=1}^{k} \sum_{\ell=1}^{k} p_g \cdot p_\ell \cdot (b_g \cdot d_{\ell g} + b_\ell \cdot d_{g\ell}) \qquad (g, \ell = 1, ..., k) \tag{8.14}$$

segue:

$$c = \sum_{g=1}^{k} p_g \cdot c_{gg} + \sum_{g=1}^{k} \sum_{\ell=1}^{k} p_g \cdot p_\ell \cdot (b_g \cdot d_{\ell g} + b_\ell \cdot d_{g\ell}). \tag{8.15}$$

Ringraziamenti

Desidero ora ringraziare le persone che, durante questi anni di dottorato, hanno contribuito alla mia formazione e a rendere meno faticose e più serene le numerose e lunghe giornate di lavoro.

Un ringraziamento particolare va al mio relatore, il Prof. Michele Zenga, che, per l'intera durata del dottorato, non ha mai smesso di contribuire alla mia formazione, accademica e non. Per lui, un ringraziamento davvero sentito.

Grazie ai miei colleghi ed amici del dottorato, che menziono in ordine puramente alfabetico: i dottori Alberto Arcagni, Gaia Bertarelli, Agnese Maria Di Brisco, Nicola Donelli, Marcella Mazzoleni e Veronica Sora. Senza di loro, questo percorso avrebbe potuto avere anche esito diverso.

Grazie alle Prof.sse Mariangela Zenga, Angiola Pollastri e Fulvia Pennoni per il coinvolgimento nelle attività didattiche e la loro disponibilità.

Grazie al Prof. Piero Quatto e ai dottori Lucio De Capitani, Francesco Porro, Leo Pasquazzi, Emanuela Furfaro e Gianmarco Vacca per i loro preziosi e gentili suggerimenti.

Grazie infine, alla coordinatrice della Scuola, la Prof.ssa Fulvia Mecatti, per aver dato fiducia e spazio ad un candidato di formazione non prettamente statistica.

A tutti loro, un sincero ringraziamento.

Bibliografia

[1] Vittorio Amato. Contributo della scuola statistica italiana alla costruzione di classi di indici di concentrazione. In *La Distribuzione Personale del Reddito*. Vita e Pensiero, Milano, 1987. (A cura di Michele Zenga).

[2] N. Battacharya and B. Mahalanobis. Regional disparities in household consumption in India. *American Statistical Association*, 62(317):143–161, 1967.

[3] Carlo Emilio Bonferroni. *Elementi di Statistica Generale*. Litografia Felice Gili, Torino, 1938.

[4] Elena Bàrcena-Martin and Jacques Silber. On the generalization and decomposition of the Bonferroni index. *Social Choice and Welfare*, 41:763–787, 2013.

[5] Romina Gambacorta, Stefano Iezzi, Giuseppe Ilardi, Andrea Neri, and Alfonso Rosolia. I bilanci delle famiglie italiane nell'anno 2014. Technical report, Banca d'Italia, 2014.

[6] Corrado Gini. *Variabilità e Mutabilità*. Tipografia Paolo Ruffini, Bologna, 1912.

[7] Corrado Gini. Sulla misura della concentrazione e della variabilità dei caratteri. *Atti del Reale Istituto Veneto di Scienza, Lettere e Arti*, 73:1203–1248, 1914.

[8] Paolo Radaelli. On the decomposition by subgroups of the Gini index and Zenga's uniformity and inequality indexs. *International Statistical Review*, 78(1):81–101, 2010.

[9] V. M. Rao. Two decompositions of concentration ratio. *Journal of The Royal Statistical Society*, 132(3):418–425, 1969.

[10] A. F. Shorrocks. The class of additively decomposable inequality measures. *Econometrica*, 48(3), 1980.

[11] Jacques Silber. Factor components, population subgroups and the computation of the Gini index of inequality. *The Review of Economics and Statistics*, 71(1):107–115, 1989.

[12] Agostino Tarsitano. The Bonferroni index of inequality. In *Income and Wealth Distribution, Inequality and Poverty*, pages 228–242. Springer-Verlag, Berlin, 1990. (A cura di C. Dagum, M. Zenga).

[13] Autori Vari. *La Distribuzione Personale del Reddito*. Vita e Pensiero, Milano, 1987.

[14] Mario De Vergottini. Sul significato di alcuni indici di concentrazione. *Giornale degli Economisti e Annali di Economia*, (5/6):317–347, 1940.

[15] Michele Zenga. Effetti della normalizzazione sul principio della somiglianza e sulla scomponibilità degli indici di concentrazione. In *La Distribuzione Personale del Reddito*. Vita e Pensiero, Milano, 1987. (A cura di Michele Zenga).

[16] Michele Zenga. Inequality curve and inequality index based on the ratios between lower and upper means. *Statistica & Applicazioni*, 5(1):3–27, 2007.

[17] Michele Zenga. *Lezioni di Statistica Descrittiva*. G. Giappichelli, Torino, 2007.

[18] Michele Zenga. Decomposition by sources of the Gini, Bonferroni and Zenga inequality indexes. *Statistica & Applicazioni*, 11(2):133–161, 2013.

[19] Michele Zenga. *Lezioni di Statistica Descrittiva*. G. Giappichelli, Torino, 2015.

[20] Michele Zenga. On the decomposition by subpopulations of the point and synthetic Zenga (2007) inequality indexes. *Metron*, 2016. DOI: 10.1007/s40300-016-0086-7.