Giancarlo Bettuzzi

Considerazioni sulla teoria della concordanza del Gini nel contesto delle distribuzioni statistiche multiple: gli indici di omofilia parziale

Serie Ricerche 1997, n.1

Statistica
Q 30
BIBL. DIP. DI SCIENZE STATISTICHE
UNIVERSITA' DEGLI STUDI DI BOLOGNA
14987

rtimento di Scienze Statistiche “Paolo Fortunati”
Università degli studi di Bologna

Finito di stampare nel mese di Febbraio 1997
presso le Officine Grafiche TECNOPRINT S.N.C.
Via del Legatore 3, Bologna.

1. Introduzione

In una serie di memorie, pubblicate tra il 1914 e il 1916, Corrado Gini pose le basi di una metodologia per lo studio delle relazioni statistiche relative a coppie di caratteri che si associano nelle unità di un collettivo. Com'è noto, egli impostò la problematica in una forma articolata, che implica una gradualità nell'investigazione, distinguendo lo studio della connessione da quello della concordanza; mentre la prima tende ad accertare la presenza di una relazione dai contenuti più ampi e generali possibili, il secondo momento investigativo si propone di stabilire il senso, concordante o discordante, della relazione. E proprio al riguardo di quest'ultimo aspetto relazionale l'autore delineò una teoria degli indici di concordanza fondata sulla nozione di rassomiglianza che gli consentì anche di reinterpretare alcuni indici, quali il coefficiente di correlazione di Bravais-Pearson e l'indice di cograduazione di Spearman, che già erano stati proposti nell'ambito di differenti impostazioni metodologiche. Nella teoria della concordanza di Gini sussiste un intreccio logico tra la nozione di rassomiglianza, imperniata sulle differenze (denominate "discordanze") tra le quantità corrispondenti che afferiscono alla stessa unità di classificazione, e la nozione di cograduazione. Anticipando concetti su cui torneremo nel prosieguo del lavoro, si può dire che esiste rassomiglianza perfetta quando le differenze tra le quantità corrispondenti sono nulle, situazione che implica la perfetta cograduazione delle quantità medesime. In genere, però, nelle distribuzioni statistiche doppie, che formalizzano il nesso relazionale ipotizzato tra i caratteri, mai si presenta rassomiglianza perfetta e, di conseguenza, emerge l'esigenza di determinarne il grado qualificandola in termini di concordanza o di discordanza. In prima approssimazione si può affermare che c'è concordanza quando le quantità corrispondenti tendono ad essere cograduate, mentre una loro propensione alla contrograduazione pone in risalto una situazione di discordanza. Più propriamente Gini stabilì criteri rigorosi per distinguere, nelle varie configurazioni formali, la nozione di concordanza da quella di discordanza (Gini, 1916a) e, a questo proposito, definì un primo criterio di concordanza, denominato criterio α, in base al quale le quantità corrispondenti dei due caratteri si dicono concordanti, o discordanti, a seconda che la somma delle discordanze sia minore, oppure maggiore, della somma che si determinerebbe con riferimento all'ipotesi d'indipendenza. È risaputo che il criterio α è sempre idoneo a distinguere la concordanza dalla discordanza quando gli indici sono quadratici, cioè costruiti assumendo i quadrati delle discordanze. Non altrettanto può dirsi per gli indici semplici la cui costruzione si fonda sulle discordanze prese in valore assoluto; per tali indici furono illustrati gli inconvenienti che potevano manifestarsi e furono enunciati criteri di concordanza di valenza più generale atti ad eliminarli. Nell'economia di questo nostro scritto rivolto alla considerazione di indici quadratici di concordanza tra scostamenti e tra variazioni, riteniamo conveniente non indugiare ulteriormente sull'argomento per attenerci esclusivamente all'enunciato del criterio α. È infatti il caso di rilevare che nella impostazione giniana il termine finora usato di "quantità" inerente ai caratteri che si associano è intenzionalmente impiegato per sottintendere un possibile riferimento sia alle intensità dei caratteri, come pure agli scostamenti dalla media aritmetica o, infine, agli scostamenti standardizzati (ossia rapportati allo scostamento quadratico medio) denominati, anche, variazioni. Con ciò Gini prefigurò la possibilità di definire, sul piano formale, una pluralità di indici ricollegabili alle mutevoli e molteplici istanze conoscitive che possono presentarsi nel quadro della ricerca sostanziale. In questo lavoro ci proponiamo di riconsiderare i più significativi aspetti del contributo giniano con l'intento di estenderli al più ampio contesto delle distribuzioni statistiche multiple. Alla parte propositiva converrà, allora, fare precedere una breve esposizione delle linee fondamentali della teoria di Gini proprio perché alcuni dei risultati richiamati verranno direttamente coinvolti nella costruzione dei nuovi indici.

Coerentemente ai concetti richiamati, nei confronti della distribuzione statistica doppia (x_i, y_i), con $i = 1, 2, \ldots, N$, si considerino i quadrati delle discordanze, ossia le quantità $({}_xq_i - {}_yq_i)^2$, ai fini della costruzione degli indici quadratici di concordanza in senso lato, cioè attinenti sia al caso della concordanza che a quello della discordanza; come si è accennato, i simboli ${}_xq_i$ e ${}_yq_i$ possono rappresentare intensità, oppure scostamenti, oppure, ancora, variazioni. Com'è noto, l'espressione

$$^2\mathcal{M} = \Sigma_i({}_xq_i - {}_yq_i)^2 \qquad [1]$$

rappresenta la somma dei quadrati delle discordanze tra le quantità corrispondenti; analogamente, i simboli $^2\mathcal{M}_1$, $^2\mathcal{M}_2$ e $^2\mathcal{M}_0$ indicano la somma dei quadrati delle discordanze che si avrebbero, rispettivamente, nell'ipotesi di cograduazione delle quantità dei due caratteri, in quella di contrograduazione e, infine, nell'ipotesi d'indipendenza. Sulla base di tali notazioni Gini propose la coppia di espressioni generali di indici quadratici

$$^2O = \frac{^2\mathcal{M} - {}^2\mathcal{M}_0}{^2\mathcal{M}_1 - {}^2\mathcal{M}_0}, \qquad \text{se } ^2\mathcal{M} < {}^2\mathcal{M}_0 \qquad [2]$$

$$^2\varepsilon = \frac{^2\mathcal{M} - {}^2\mathcal{M}_0}{^2\mathcal{M}_0 - {}^2\mathcal{M}_2}, \qquad \text{se } ^2\mathcal{M} > {}^2\mathcal{M}_0 \qquad [3]$$

il primo dei quali può assumere valori nell'intervallo [0;1] ed è idoneo a misurare la concordanza, mentre il secondo realizza valori nell'intervallo [–1;0] e misura la discordanza. A differenza degli indici semplici che assumono configurazioni formali diverse a seconda che si considerino le discordanze tra le intensità, tra gli scostamenti o tra le variazioni, gli indici quadratici risultano uguali nei tre casi e, in particolare, l'indice [2] si specifica nell'espressione

$$^2O = \frac{\sigma(X,Y)}{\sigma^{(1)}(X,Y)} \qquad [4]$$

in cui al numeratore figura la covarianza determinata rispetto ai valori osservati e al denominatore la covarianza definita in conformità all'ipotesi di cograduazione delle quantità dei due caratteri; l'indice [3] assume la seguente configurazione

$$^2\varepsilon = -\frac{\sigma(X,Y)}{\sigma^{(2)}(X,Y)} \qquad [5]$$

dove $\sigma^{(2)}(X, Y)$ rappresenta la covarianza calcolata coerentemente all'ipotesi di contrograduazione. Il segno –, che figura nel secondo membro, consegue dalla circostanza che prevede il numeratore negativo per ipotesi e il denominatore negativo per costruzione. Tali indici furono denominati indici di omofilia e la loro caratteristica è quella di misurare la concordanza, o la discordanza, rispetto al massimo relativo che, per definizione, è condizionato dalla specifica configurazione delle distribuzioni statistiche semplici dei due caratteri. Sotto questo profilo, le distribuzioni ora menzionate si propongono come un dato del problema e costituiscono un vincolo per l'operazione di cograduazione, o di contrograduazione, essenziale per la determinazione del denominatore degli indici. In questo contesto è del tutto evidente che la tendenza ad associarsi delle quantità corrispondenti, per quanto forte possa essere, non può oltrepassare il massimo vincolato dall'invarianza delle distribuzioni effettive dei due caratteri. In circostanze diverse, queste distribuzioni possono invece configurarsi come il risultato della tendenza associativa e, pertanto, non costituire un vincolo alle modalità concrete dell'associazione. In tal caso è concettualmente possibile, ed opportuno, modificare le distribuzioni statistiche semplici, componenti la distribuzione statistica doppia assegnata, affinché possa verificarsi sia il massimo assoluto di concordanza che il massimo assoluto di discordanza. A questo proposito Gini introdusse la nozione di carattere "contrario" (Gini, 1915b) e indicò le condizioni che debbono sussistere perché siano perseguibili i massimi assoluti di concordanza e di discordanza, identificandole nella uguaglianza di quattro distribuzioni: le distribuzioni statistiche semplici dei due caratteri e quelle dei loro caratteri contrari. Nel contempo suggerì la soluzione formale del problema ravvisabile nella costruzione di una distribuzione media delle quattro distribuzioni statistiche precedentemente indicate, che risulterà simmetrica (Gini, 1916b). Il riferimento a distribuzioni statistiche semplici, componenti della distribuzione statistica doppia, coincidenti con la distribuzione statistica media sopra specificata consentirà di costruire un modello associativo che, sul piano logico-formale, prevede la realizzazione dei massimi teorici in questione. Relativamente a questi, Gini pervenne all'espressione

$$^{2}\rho = \frac{\sigma(X,Y)}{\frac{1}{2}[\sigma^2(X)+\sigma^2(Y)]} \qquad [6]$$

dell'indice quadratico di correlazione tra scostamenti, che può assumere valori nell'intervallo [−1;1]; nei confronti delle variazioni, invece, determinò l'indice quadratico di correlazione

$$^{2}r = \frac{\sigma(X,Y)}{\sigma(X)\,\sigma(Y)} \qquad [7]$$

coincidente con la già nota espressione del coefficiente di correlazione di Bravais-Pearson.

È appena il caso di rilevare che il termine "correlazione" nel testo giniano è introdotto per significare che gli indici di concordanza sono costruiti con riferimento alla nozione di massimo assoluto di concordanza.

2. Estensione della teoria giniana della concordanza alle distribuzioni statistiche multiple

Si tratta ora di accertare a quali condizioni la teoria della concordanza proposta dal Gini nel caso di associazione di due caratteri possa essere generalizzata ed estesa ad un contesto classificatorio che prevede l'associazione di più di due caratteri. Nel presente lavoro verrà trattato il caso delle distribuzioni statistiche costruite mediante l'associazione di tre caratteri quantitativi X_1, X_2, X_3 nelle N unità di un collettivo. In un tale contesto, com'è noto, la concordanza tra una coppia di variabili, ad esempio fra X_1 e X_2, può essere in parte influenzata dalla terza variabile X_3; va da sé che tale influenza deve essere neutralizzata ogni qual volta si presenta la necessità di misurare la relazione tra X_1 e X_2 svincolata dall'effetto prodotto da X_3. Il problema di eliminazione di cause che ora si prospetta è suscettibile di molteplici soluzioni (A. Gili-G. Bettuzzi, 1984 e 1986; G. Bettuzzi, 1986); quella che adotteremo in questo scritto è quella stessa indicata da Yule, che fu il primo ad occuparsi della questione (Yule, 1897). Pertanto, se indichiamo con x_1, x_2 e x_3 le distribuzioni degli scostamenti rispettivamente dalle medie di X_1, X_2 e X_3, si tratta di detrarre da x_1 e da x_2 quella loro parte dovuta a x_3, che corrisponde, rispettivamente, alla regressione lineare di x_1 su x_3 e di x_2 su x_3. I residui così ottenuti vengono rappresentati mediante la seguente coppia di espressioni:

$${}_{i}x_{1.3} = {}_{i}x_1 - b_{13}\,{}_{i}x_3 \qquad [8]$$

$${}_{i}x_{2.3} = {}_{i}x_2 - b_{23}\,{}_{i}x_3 \qquad [9]$$

in cui b_{13} e b_{23} indicano i coefficienti di regressione lineare. Allora, proprio per quanto è stato argomentato, la questione della misura della concordanza tra i caratteri X_1 e X_2, al netto dell'influenza di X_3, può specificarsi nella sua determinazione rispetto alle distribuzioni poste con la [8] e la [9]. Ma prima ancora di affrontare questo problema conviene richiamare alcuni valori caratteristici delle distribuzioni sopra menzionate, valori che torneranno utili quando dovremo rendere esplicite le strutture degli indici di concordanza. A tal fine, iniziamo col rammentare che la varianza della distribuzione [8] corrisponde al valore

$$\sigma^2_{1.3} = \sigma^2_1(1 - r^2_{13}) \qquad [10]$$

e, analogamente, per la distribuzione [9] risulta

$$\sigma_{2.3}^2 = \sigma_2^2(1 - r_{23}^2) \qquad [11]$$

avendo indicato con σ_1^2 e σ_2^2 le varianze delle variabili X_1 e X_2, e con r_{13} e r_{23} i coefficienti di correlazione lineare tra X_1 e X_3 e tra X_2 e X_3.

Per quanto concerne la covarianza, determinata nei confronti della distribuzione doppia degli scostamenti residui,

$$\sigma_{1.3,2.3} = \frac{\Sigma_i\ {}_ix_{1.3}\ {}_ix_{2.3}}{N} = \frac{\Sigma_i({}_ix_1 - b_{13}\ {}_ix_3)({}_ix_2 - b_{23}\ {}_ix_3)}{N}$$

è elementare pervenire al seguente risultato:

$$\sigma_{1.3,2.3} = \sigma_1\ \sigma_2(r_{12} - r_{13}\ r_{23}). \qquad [12]$$

Infine, possiamo definire le distribuzioni delle variazioni (scostamenti standardizzati):

$${}_iv_{1.3} = \frac{{}_ix_{1.3}}{\sigma_{1.3}} = \frac{{}_ix_1 - b_{13}\ {}_ix_3}{\sigma_1\sqrt{1 - r_{13}^2}} \qquad [13]$$

$${}_iv_{2.3} = \frac{{}_ix_{2.3}}{\sigma_{2.3}} = \frac{{}_ix_2 - b_{23}\ {}_ix_3}{\sigma_2\sqrt{1 - r_{23}^2}} \qquad [14]$$

a loro volta suscettibili di considerazione al fine della costruzione degli indici di concordanza.

3. Indici di concordanza costruiti rispetto al massimo relativo

In un precedente lavoro (G. Bettuzzi, 1996) abbiamo avuto l'opportunità di definire indici quadratici di correlazione parziale coerentemente alla nozione di massimo assoluto di concordanza. Più propriamente, con riferimento alle distribuzioni [8] e [9] si è costruito l'indice quadratico di correlazione parziale tra scostamenti dalla media aritmetica

$${}^2\rho_{12,3} = \frac{\sigma_{1.3,2.3}}{\frac{1}{2}(\sigma_{1.3}^2 + \sigma_{2.3}^2)} = \frac{r_{12} - r_{13}\ r_{23}}{\frac{1}{2}\left[\frac{\sigma_1}{\sigma_2}(1 - r_{13}^2) + \frac{\sigma_2}{\sigma_1}(1 - r_{23}^2)\right]} \qquad [15]$$

mentre l'indice quadratico di correlazione parziale tra variazioni è stato ottenuto intervenendo nei confronti delle distribuzioni [13] e [14]; la sua espressione

$${}^2r_{12,3} = \frac{\sigma_{1.3,2.3}}{\sigma_{1.3}\ \sigma_{2.3}} = \frac{r_{12} - r_{13}\ r_{23}}{\sqrt{1 - r_{13}^2}\ \sqrt{1 - r_{23}^2}} \qquad [16]$$

riproduce quella del coefficiente di correlazione parziale di Yule e consente di reinterpretare quest'ultimo come indice quadratico di correlazione parziale tra variazioni. È stato, infine, opportunamente rilevato che i due indici quadratici di correlazione parziale, di cui alla [15] e alla [16], differiscono nei denominatori che rappresentano, rispettivamente, la media aritmetica e la media geometrica delle varianze delle distribuzioni [8] e [9], e da ciò segue la relazione secondo cui è ${}^2r_{12,3} \geq {}^2\rho_{12,3}$.

L'indice quadratico di correlazione parziale tra scostamenti può essere scelto per misurare la concordanza tra caratteri espressi nella stessa unità di misura; quando i caratteri che si associano sono eterogenei, o più in generale quando si vuole eliminare la loro differente variabilità, si può ricorrere al calcolo dell'indice quadratico di correlazione parziale tra variazioni.

La problematica giniana, volta alla costruzione di indici definiti rispetto alla nozione di massimo relativo di concordanza e di discordanza, può essere ripresa per estenderla all'associazione degli scostamenti, oppure delle variazioni, dei caratteri X_1 e X_2 al netto dell'influenza esercitata da X_3, e ciò consente di configurare espressioni di indici quadratici, che ci permettiamo di denominare di "omofilia parziale", quando si assumano quali punti di riferimento le formule [2] e [3] di Gini. Il problema che si presenta ora nei confronti di questi ultimi indici è quello di determinarne l'espressione del denominatore in quanto hanno in comune con gli indici di correlazione, considerati nel precedente lavoro citato, quella del numeratore. Iniziamo col considerare l'indice quadratico tra scostamenti; in piena coerenza all'impostazione giniana, e quindi in stretta analogia alle espressioni [2] e [3] richiamate, possiamo introdurre le seguenti strutture formali valide nei confronti delle distribuzioni [8] e [9]

$$ {}^2_x O_{12,3} = \frac{{}^2\mathcal{M} - {}^2\mathcal{M}_0}{{}^2\mathcal{M}_1 - {}^2\mathcal{M}_0} \qquad \text{se} \qquad {}^2\mathcal{M} < {}^2\mathcal{M}_0 \tag{17} $$

$$ {}^2_x \varepsilon_{12,3} = \frac{{}^2\mathcal{M} - {}^2\mathcal{M}_0}{{}^2\mathcal{M}_0 - {}^2\mathcal{M}_2} \qquad \text{se} \qquad {}^2\mathcal{M} > {}^2\mathcal{M}_0 \tag{18} $$

in cui, come già si è avuto modo di chiarire nel lavoro (G. Bettuzzi, 1996), risulta

$$ {}^2\mathcal{M} = \Sigma_i({}_i x_{1.3} - {}_i x_{2.3})^2 = N\,\sigma^2_{1.3} + N\,\sigma^2_{2.3} - 2N\,\sigma_{1.3,2.3} \tag{19} $$

ed anche

$$ {}^2\mathcal{M}_0 = N\,\sigma^2_{1.3} + N\,\sigma^2_{2.3} \tag{20} $$

stante la nullità della covarianza $\sigma_{1.3,\,2.3}$ nell'ipotesi d'indipendenza.

Ora si tratta di determinare le quantità ${}^2\mathcal{M}_1$ e ${}^2\mathcal{M}_2$ che consentono di definire i denominatori dei nuovi indici; in conformità alla fondamentale espressione [1], e se si usa il simbolo (i) per indicare scostamenti cograduati, risulta:

$$ \begin{aligned} {}^2\mathcal{M}_1 &= \Sigma_{(i)}({}_{(i)}x_{1.3} - {}_{(i)}x_{2.3})^2 = \\ &= \Sigma_{(i)}\,{}_{(i)}x^2_{1.3} + \Sigma_{(i)}\,{}_{(i)}x^2_{2.3} - 2\Sigma_{(i)}\,{}_{(i)}x_{1.3}\,{}_{(i)}x_{2.3} = \\ &= N\sigma^2_{1.3} + N\sigma^2_{2.3} - 2N\sigma^{(1)}_{1.3,\,2.3} \end{aligned} \tag{21} $$

dove $\sigma^{(1)}_{1.3,\,2.3}$ è la covarianza nell'ipotesi di cograduazione degli scostamenti residui [8] e [9]. In modo del tutto analogo risulta che

$$ \begin{aligned} {}^2\mathcal{M}_2 &= \Sigma_{(i)}({}_{(i)}x_{1.3} - {}_{N-(i)+1}x_{2.3})^2 = \\ &= \Sigma_{(i)}\,{}_{(i)}x^2_{1.3} + \Sigma_{(i)}\,{}_{N-(i)+1}x^2_{2.3} - 2\Sigma_{(i)}\,{}_{(i)}x_{1.3}\,{}_{N-(i)+1}x_{2.3} = \\ &= N\sigma^2_{1.3} + N\sigma^2_{2.3} - 2N\sigma^{(2)}_{1.3,\,2.3} \end{aligned} \tag{22} $$

avendo indicato con $\sigma^{(2)}_{1.3,\,2.3}$ la covarianza nell'ipotesi di contrograduazione. Siamo ora in grado di dare espressione agli indici quadratici di omofilia in questione. Per quanto attiene alla [17], sulla base delle formule [19], [20] e [21], scriveremo

$$ {}^2_x O_{12,3} = \frac{\sigma_{1.3,2.3}}{\sigma^{(1)}_{1.3,2.3}} \tag{23} $$

mentre la [18], tenuto conto della [19] della [20] e della [22], si specifica nel seguente modo

$$ {}^2_x \varepsilon_{12,3} = -\frac{\sigma_{1.3,2.3}}{\sigma^{(2)}_{1.3,2.3}} \tag{24} $$

È del tutto immediato ravvisare la stretta analogia formale che accomuna questi ultimi due indici con quelli del Gini espressi nella [4] e nella [5]. Mentre l'indice [23] misura la concordanza e, per costruzione, può assumere valori compresi nell'intervallo [0;1], l'indice [24] misura la discordanza variando nell'intervallo [– 1;0].

Se adesso consideriamo l'associazione in termini di variazioni, possiamo sempre riferirci alla [2] e alla [3] del Gini, per porre le strutture degli indici quadratici di omofilia parziale

$$ {}^2_v O_{12,3} = \frac{{}^2m - {}^2m_0}{{}^2m_1 - {}^2m_0} \qquad \text{se} \qquad {}^2m < {}^2m_0 \tag{25} $$

$$ {}^2_v \varepsilon_{12,3} = \frac{{}^2m - {}^2m_0}{{}^2m_0 - {}^2m_2} \qquad \text{se} \qquad {}^2m > {}^2m_0 \tag{26} $$

in cui le quantità che figurano al numeratore, come si è avuto modo di specificare nel precedente lavoro (G. Bettuzzi, 1996), sono determinate nella forma seguente:

$$ {}^2m = \Sigma_i({}_i v_{1.3} - {}_i v_{2.3})^2 = 2N - 2N\,\frac{\sigma_{1.3,2.3}}{\sigma_{1.3}\,\sigma_{2.3}} \tag{27} $$

mentre si perviene al risultato

$$ {}^2m_0 = 2N \tag{28} $$

in considerazione del valore nullo della covarianza $\sigma_{1.3,\,2.3}$ nell'ipotesi d'indipendenza a cui si richiama la quantità ${}^2m_0$.

Per quanto attiene alle restanti quantità che figurano al denominatore della [25] e della [26], scriveremo

$$
\begin{aligned}
{}^2m_1 &= \Sigma_{(i)}({}_{(i)}v_{1.3} - {}_{(i)}v_{2.3})^2 = \\
&= \Sigma_{(i)}\,{}_{(i)}v^2_{1.3} + \Sigma_{(i)}\,{}_{(i)}v^2_{2.3} - 2\Sigma_{(i)}\,{}_{(i)}v_{1.3}\,{}_{(i)}v_{2.3} = \\
&= 2N - 2N\frac{\sigma^{(1)}_{1.3,2.3}}{\sigma_{1.3}\,\sigma_{2.3}}
\end{aligned}
\qquad [29]
$$

e, similmente

$$
\begin{aligned}
{}^2m_2 &= \Sigma_{(i)}({}_{(i)}v_{1.3} - {}_{N-(i)+1}v_{2.3})^2 = \\
&= \Sigma_{(i)}\,{}_{(i)}v^2_{1.3} + \Sigma_{(i)}\,{}_{N-(i)+1}v^2_{2.3} - 2\Sigma_{(i)}\,{}_{(i)}v_{1.3}\,{}_{N-(i)+1}v_{2.3} = \\
&= 2N - 2N\frac{\sigma^{(2)}_{1.3,2.3}}{\sigma_{1.3}\,\sigma_{2.3}}\,.
\end{aligned}
\qquad [30]
$$

Sulla scorta dei valori trovati si può pervenire agevolmente alle seguenti espressioni

$$
{}^2_vO_{12,3} = \frac{\sigma_{1.3,2.3}}{\sigma^{(1)}_{1.3,2.3}} \qquad [31]
$$

$$
{}^2_v\varepsilon_{12,3} = -\frac{\sigma_{1.3,2.3}}{\sigma^{(2)}_{1.3,2.3}}\,. \qquad [32]
$$

Si può notare che anche gli indici quadratici di omofilia parziale, al pari di quelli definiti da Gini, presentano la stessa espressione indipendentemente dal fatto che si considerino scostamenti oppure variazioni, infatti la [23] e la [31] coincidono e così pure la [24] con la [32].

Concludiamo queste brevi note ponendo in evidenza la relazione tra indice quadratico di omofilia parziale e indice quadratico di correlazione parziale tra variazioni. La [31], infatti, è suscettibile della seguente trasformazione

$$
{}^2_vO_{12,3} = \frac{\dfrac{\sigma_{1.3,2.3}}{\sigma_{1.3}\,\sigma_{2.3}}}{\dfrac{\sigma^{(1)}_{1.3,2.3}}{\sigma_{1.3}\,\sigma_{2.3}}} = \frac{{}^2r_{12.3}}{{}^2r^{(1)}_{12,3}} \qquad [33]
$$

che consente di caratterizzare l'indice quadratico di omofilia come rapporto tra il valore dell'indice quadratico di correlazione determinato nei confronti della distribuzione statistica doppia delle variazioni dei caratteri X_1 e X_2 al netto dell'influenza del carattere X_3 e, al denominatore, il valore che lo stesso indice realizza nei confronti delle variazioni residue cograduate, coerentemente all'ipotesi di massimo relativo di concordanza. Analogamente, si può intervenire nei riguardi della [32]; risulta

$$
{}^2_v\varepsilon_{12,3} = -\frac{\dfrac{\sigma_{1.3,2.3}}{\sigma_{1.3}\,\sigma_{2.3}}}{\dfrac{\sigma^{(2)}_{1.3,2.3}}{\sigma_{1.3}\,\sigma_{2.3}}} = -\frac{{}^2r_{12.3}}{{}^2r^{(2)}_{12,3}} \qquad [34]
$$

dove ${}^2r^{(2)}_{12.3}$, a sua volta, esprime il valore dell'indice quadratico di correlazione calcolato rispetto alle variazioni residue contrograduate, che scontano l'ipotesi del massimo relativo di discordanza.

Mentre i numeratori che compaiono al terzo membro della [33] e della [34] possono essere espressi in funzione dei coefficienti di correlazione lineare di ordine zero, com'è già stato fatto per gli indici quadratici di correlazione parziale, non altrettanto può dirsi per i denominatori poiché la cograduazione e la contrograduazione delle distribuzioni delle variazioni residue [13] e [14], che stanno alla base della determinazione dei denominatori medesimi, in generale non implicano la cograduazione e la controgradazione delle variazioni inerenti alle distribuzioni statistiche doppie componenti della distribuzione statistica tripla assegnata. Pertanto, il calcolo dell'indice quadratico di omofilia parziale, a differenza degli indici quadratici di correlazione parziale, passa attraverso l'effettiva determinazione delle distribuzioni degli scostamenti residui [8] e [9], o delle variazioni residue [13] e [14].

Dipartimento di Scienze Statistiche
"Paolo Fortunati"
Università degli Studi di Bologna

Giancarlo Bettuzzi

Riferimenti bibliografici

G. Bettuzzi (1986), Sulla definizione degli indici quadratici di correlazione tra variazioni, *Statistica*, vol. XLVI, n. 3.

G. Bettuzzi (1996), Correlazione parziale e teoria della concordanza di Gini, *Dipartimento di Scienze Statistiche "Paolo Fortunati" dell'Università degli Studi di Bologna, Serie Ricerche 1996 n. 3.*

P. Fortunati (1967), Alcune considerazioni sulla impostazione giniana delle misure di concordanza, *Atti della XXV Riunione Scientifica della Società Italiana di Statistica*, Bologna 1967.

A. Gili (1978), Indici quadratici di concordanza, *Istituto di Statistica dell'Università di Bologna*, CLUEB, Bologna, 1978.

A. Gili-G. Bettuzzi (1984), In tema di indici quadratici di concordanza tra scostamenti: struttura generale, *Statistica*, vol. XLIV, n. 4.

A. Gili-G. Bettuzzi (1986), In tema di indici quadratici di concordanza tra scostamenti: indici di correlazione, *Statistica*, vol. XLVI, n. 1.

C. Gini (1914), Di una misura della dissomiglianza tra due gruppi di quantità e delle sue applicazioni allo studio delle relazioni statistiche, *Atti del Reale Istituto Veneto di Scienze, Lettere ed Arti*, Tomo LXXIV – Parte seconda, Venezia 1914.

C. Gini (1915a), Indici di omofilia e di rassomiglianza e loro relazioni col coefficiente di correlazione e con gli indici di attrazione, *Atti del Reale Istituto Veneto di Scienze, Lettere ed Arti*, Tomo LXXIV – Parte seconda, Venezia 1915.

C. Gini (1915b), Nuovi contributi alla teoria delle relazioni statistiche, *Atti del Reale Istituto Veneto di Scienze, Lettere ed Arti*, Tomo LXXIV – Parte seconda, Venezia 1915.

C. Gini (1916a), Sul criterio di concordanza tra due caratteri, *Atti del Reale Istituto Veneto di Scienze, Lettere ed Arti*, Tomo LXXV – Parte seconda, Venezia 1916.

C. Gini (1916b), Indici di concordanza, *Atti del Reale Istituto Veneto di Scienze, Lettere ed Arti*, Tomo LXXV – Parte seconda, Venezia 1916.

A. Predetti (1972), *Operatori statistici su aggregati di osservazioni di due o più caratteri*, A. Giuffrè Ed., Milano 1972.

T. Salvemini (1939), Sugli indici di omofilia, *Atti della I Riunione Scientifica della Società Italiana di Statistica*, Pisa 1939.

T. Salvemini (1945-46), *Lezioni di statistica metodologica. Parte III: Le relazioni statistiche*, Facoltà di Scienze Statistiche, Demografiche ed Attuariali, Università di Roma, 1945-46.

L. Vajani (1978), *Statistica descrittiva*, Etas Libri, Milano 1978.

G.U. Yule (1897), On the theory of correlation, *Journal of the Royal Statistical Society*, vol. LX, 1897.

G.U. Yule (1932), *An introduction to the theory of statistics*, C. Griffin and Co., London 1932.

G.U. Yule-M.G. Kendall (1965), *An introduction to the theory of statistics*, C. Griffin and Co., London 1965.