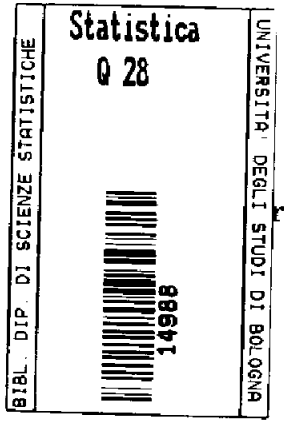
Giancarlo Bettuzzi

Correlazione parziale e teoria della concordanza di Gini

Serie Ricerche 1996, n.3

rtimento di Scienze Statistiche "Paolo Fortunati"
Università degli studi di Bologna



Finito di stampare nel mese di Settembre 1996
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 delineò i fondamenti metodologici di una teoria della concordanza, basata sulla nozione di rassomiglianza, inerente a due caratteri. In questo nostro scritto 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 del Gini proprio perché alcuni dei risultati richiamati verranno direttamente utilizzati per la costruzione dei nuovi indici.

In quella teoria, com'è noto, sussiste un intreccio logico tra nozione di rassomiglianza e quella di cograduazione, ma ciò non comporta l'esaurirsi della prima nella seconda proprio in quanto la rassomiglianza implica il confronto per differenza tra le quantità corrispondenti, che afferiscono alla stessa unità di classificazione; tali differenze sono denominate "discordanze". 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, conseguentemente, 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 il Gini stabilì criteri rigorosi idonei a 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. Com'è noto, 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 di

volta in volta potevano manifestarsi e furono enunciati criteri di concordanza di valenza più generale idonei 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 usato 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 il Gini prefigurò la possibilità di definire, sul piano formale una pluralità di indici ricollegabili alle mutevoli e molteplici istanze conoscitive che possono presentarsi nei contesti della ricerca sostanziale.

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 massi-

mo 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 sia perseguibile quel risultato, 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 determinò le seguenti somme dei quadrati delle discordanze tra gli scostamenti dalle medie dei due caratteri:

$$ {}^2\mathcal{M}_1 = 0; \qquad {}^2\mathcal{M}_2 = 2N[\sigma^2(X) + \sigma^2(Y)]; \qquad {}^2\mathcal{M}_0 = N[\sigma^2(X) + \sigma^2(Y)] \qquad [6] $$

in cui $\sigma^2(X)$ e $\sigma^2(Y)$ indicano, rispettivamente, le varianze delle distribuzioni del carattere X e del carattere Y; in modo analogo, determinò le corrispondenti espressioni inerenti alle discordanze tra le variazioni:

$$ {}^2m_1 = 0; \qquad {}^2m_2 = 4N; \qquad {}^2m_0 = 2N. \qquad [7] $$

Si può allora notare come l'introduzione del massimo assoluto di concordanza in senso lato consenta la costruzione di un'unica struttura formale valida sia per la misura della concordanza che per quella della discordanza, a differenza di quanto accade per gli indici di omofilia. Infatti, osservando le espressioni inerenti agli scostamenti che figurano nella [6], risulta che ${}^2\mathcal{M}_2 = 2\,{}^2\mathcal{M}_0$, e tale relazione, congiuntamente alla nullità di ${}^2\mathcal{M}_1$, permette di ridefinire, rispetto al massimo assoluto di concordanza, la coppia di indici di cui alla [2] e alla [3] nella comune configurazione

$$ \frac{{}^2\mathcal{M}_0 - {}^2\mathcal{M}}{{}^2\mathcal{M}_0}. \qquad [8] $$

Anche per le espressioni che compaiono nella [7] risulta ${}^2m_2 = 2\,{}^2m_0$ e

questo risultato, a sua volta, consente di ricondurre la [2] e la [3] all'unica espressione

$$ \frac{{}^2m_0 - {}^2m}{{}^2m_0} \qquad [8 \text{ bis}] $$

idonea a misurare sia la concordanza che la discordanza tra le variazioni.

Proprio utilizzando i risultati illustrati nella [6] Gini, com'è noto, pervenne all'espressione

$$ {}^2\rho = \frac{\sigma(X,Y)}{\frac{1}{2}[\sigma^2(X) + \sigma^2(Y)]} \qquad [9] $$

dell'indice quadratico di correlazione tra scostamenti, che può assumere valori nell'intervallo [–1;1]; avendo invece presenti i risultati espressi nella [7] determinò l'indice quadratico di correlazione tra variazioni

$$ {}^2r = \frac{\sigma(X,Y)}{\sigma(X)\,\sigma(Y)} \qquad [10] $$

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:

$$ {}_i x_{1.3} = {}_i x_1 - b_{13}\, {}_i x_3 \qquad [11] $$

$$ {}_i x_{2.3} = {}_i x_2 - b_{23}\, {}_i x_3 \qquad [12] $$

e la questione della misura della concordanza tra i caratteri X_1 e X_2, al netto dell'influenza di X_3, può specificarsi, per quanto è stato argomentato, nella sua determinazione rispetto alle distribuzioni poste con la [11] e la [12]. 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 [11] corrisponde al valore

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

e, analogamente, per la distribuzione [12] risulta

$$ \sigma^2_{2.3} = \sigma^2_2(1 - r^2_{23}) \qquad [14] $$

avendo indicato con σ^2_1 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

$$ \sigma_{1.3,2.3} = \frac{\Sigma_i\, {}_i x_{1.3}\, {}_i x_{2.3}}{N} = \frac{\Sigma_i ({}_i x_1 - b_{13}\, {}_i x_3)\,({}_i x_2 - b_{23}\, {}_i x_3)}{N} $$

è elementare pervenire al seguente risultato:

$$ \sigma_{1.3,\,2.3} = \sigma_1\, \sigma_2(r_{12} - r_{13}\, r_{23}). \qquad [15] $$

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

$$ {}_i v_{1.3} = \frac{{}_i x_{1.3}}{\sigma_{1.3}} = \frac{{}_i x_1 - b_{13}\, {}_i x_3}{\sigma_1 \sqrt{1 - r^2_{13}}} \qquad [16] $$

$$ {}_i v_{2.3} = \frac{{}_i x_{2.3}}{\sigma_{2.3}} = \frac{{}_i x_2 - b_{23}\, {}_i x_3}{\sigma_2 \sqrt{1 - r^2_{23}}} \qquad [17] $$

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

3. La correlazione parziale

Torniamo a considerare le indicazioni del Gini in materia di indici di correlazione, costruiti rispetto alla nozione di massimo assoluto di concordanza. Per ottenere l'espressione di un indice quadratico di correlazione parziale tra variazioni si tratta di determinare le grandezze che figurano nella [8 bis] con riferimento alle distribuzioni [16] e [17] di variazioni; rammentando la fondamentale espressione [1], risulta

$$ {}^2m = \Sigma_i ({}_i v_{1.3} - {}_i v_{2.3})^2 = \Sigma_i\, {}_i v^2_{1.3} + \Sigma_i\, {}_i v^2_{2.3} - 2\,\Sigma_i\, {}_i v_{1.3}\, {}_i v_{2.3} = $$

$$ = \Sigma_i \left(\frac{{}_i x_{1.3}}{\sigma_{1.3}} \right)^2 + \Sigma_i \left(\frac{{}_i x_{2.3}}{\sigma_{2.3}} \right)^2 - 2\,\Sigma_i\, \frac{{}_i x_{1.3}}{\sigma_{1.3}} \cdot \frac{{}_i x_{2.3}}{\sigma_{2.3}} = $$

$$ = \frac{1}{\sigma^2_{1.3}} N\, \sigma^2_{1.3} + \frac{1}{\sigma^2_{2.3}} N\, \sigma^2_{2.3} - \frac{2}{\sigma_{1.3}\, \sigma_{2.3}} N\, \sigma_{1.3,2.3} $$

vale a dire

$$ {}^2m = 2N - 2N \frac{\sigma_{1.3,2.3}}{\sigma_{1.3}\, \sigma_{2.3}}\,; \qquad [18] $$

va da sé che è lecito pervenire al risultato seguente

$$^2m_0 = 2N \qquad [19]$$

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

Se ora si pone mente ai risultati espressi nella [18] e nella [19], è possibile dare la seguente connotazione della [8 bis]

$$^2r_{12,3} = \frac{^2m_0 - {}^2m}{^2m_0} = \frac{\sigma_{1.3,2.3}}{\sigma_{1.3}\,\sigma_{2.3}}. \qquad [20]$$

La [20], sul piano formale, ha una struttura analoga alla [10] poiché al numeratore figura la covarianza calcolata rispetto alla distribuzione statistica doppia dei residui di variazioni, di cui alla [16] e alla [17], e al denominatore il prodotto dei loro scostamenti quadratici medi. Infine, sulla base delle relazioni [13],[14] e [15] è del tutto immediato dare della [20] la seguente espressione

$$^2r_{12,3} = \frac{r_{12} - r_{13}\,r_{23}}{\sqrt{1-r_{13}^2}\,\sqrt{1-r_{23}^2}} \qquad [21]$$

che riproduce quella del coefficiente di correlazione parziale di Yule e che consente di reinterpretarlo come indice quadratico di correlazione parziale tra variazioni.

Coerentemente ai contenuti metodologici della teoria giniana della concordanza, ci si può porre come ulteriore obiettivo quello di pervenire alla identificazione di un indice quadratico di correlazione parziale tra scostamenti. Il punto d'avvio è, questa volta, costituito dall'espressione [8] che dovrà essere sviluppata coinvolgendo nella operazione le distribuzioni degli scostamenti di cui alla [11] e alla [12]; in analogia all'espressione [1] scriveremo

$$^2\mathcal{M} = \Sigma_i({}_ix_{1.3} - {}_ix_{2.3})^2 = \Sigma_i\,{}_ix_{1.3}^2 + \Sigma_i\,{}_ix_{2.3}^2 - 2\,\Sigma_i\,{}_ix_{1.3}\,{}_ix_{2.3}$$

ossia, risulta

$$^2\mathcal{M} = N\,\sigma_{1.3}^2 + N\,\sigma_{2.3}^2 - 2N\,\sigma_{1.3,2.3} \qquad [22]$$

mentre si perviene all'espressione

$$^2\mathcal{M}_0 = N\,\sigma_{1.3}^2 + N\,\sigma_{2.3}^2 \qquad [23]$$

stante la nullità della covarianza $\sigma_{1.3,2.3}$ nell'ipotesi d'indipendenza. I risultati indicati con la [22] e la [23] permettono di dare alla [8] la seguente configurazione:

$$^2\rho_{12,3} = \frac{^2\mathcal{M}_0 - {}^2\mathcal{M}}{^2\mathcal{M}_0} = \frac{\sigma_{1.3,2.3}}{\frac{1}{2}(\sigma_{1.3}^2 + \sigma_{2.3}^2)}, \qquad [24]$$

che presenta una stretta analogia formale con l'espressione [9] introdotta dal Gini. Ponendo mente alle espressioni [13], [14] e [15] e agevole dare alla [24] la conformazione:

$$^2\rho_{12,3} = \frac{r_{12} - r_{13}\,r_{23}}{\frac{1}{2}\left[\frac{\sigma_1}{\sigma_2}(1-r_{13}^2) + \frac{\sigma_2}{\sigma_1}(1-r_{23}^2)\right]} \qquad [25]$$

che è quella di un indice che, in considerazione del procedimento seguito, può essere denominato indice quadratico di correlazione parziale tra scostamenti. Dal confronto della [20] con la [24] si può notare che il coefficiente di correlazione parziale tra variazioni e il coefficiente di correlazione parziale tra scostamenti differiscono unicamente nei denominatori che sono espressi, rispettivamente, dalla media geometrica e dalla media aritmetica delle varianze delle distribuzioni [11] e [12], e la differente caratterizzazione pertanto giustifica la relazione $^2r_{12,3} \geq {}^2\rho_{12,3}$.

L'indice quadratico di correlazione parziale tra scostamenti può essere scelto quando la misura della concordanza è rivolta nei confronti di 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.

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

Riferimenti bibliografici

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