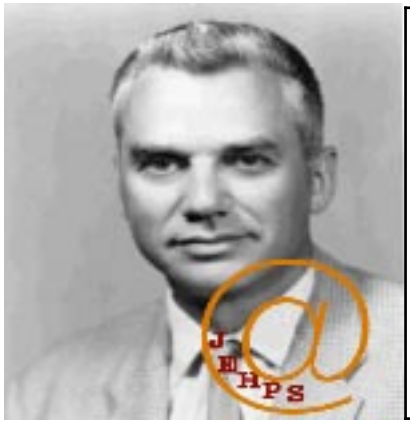

Journ@l Electronique d'Histoire des Probabilités et de la Statistique

Electronic Journ@l for History of Probability and Statistics

Vol 4, n°2; Décembre/December 2008

www.jehps.net

Italian contributions to Data Analysis

Alfredo RIZZI[1]

Introduction

The first public meeting of the International Statistical Institute (Isi) was held in Rome in 1887. A hundred years later the Società Italiana di Statistica (SIS) presented the volume *Italian Contributions to the Methodology of Statistics*[2] at the 46th Isi meeting in Tokyo.

As Leti has noted in the presentation of the volume, it is not uncommon for Italian statisticians to find, in foreign statistical literature, the elucidation of concepts and statistical indexes which originated in Italy, and, for many years, have been an important part of the cultural background of every Italian statistician. The reason for this situation is that Italian statistical work is relatively obscure to most foreign statisticians. In the last forty years the situation has changed quite a bit because English has essentially become the language of scientific communication and thus all important results appear in a limited number of scientific journals.

In the 19th century, the subjects most often studied were those phenomena which display regularity in a large number of observations. Generally the applications were seen in social phenomena.

The 20th century saw the rapid development of a great variety of different approaches to statistics[3]. In research we have always had two differing and conflicting guiding principles: the descriptive and the probabilistic. The first is closer to the historical roots of the discipline.

---

[1] Sapienza Università di Roma. alfredorizzi@fastwebnet.it

[2] Società Italiana di Statistica, *Italian contribution to the methodology of statistics,* edited by Alighiuero Naddeo, Cleup, Padova, 1987

[3] De Cristofaro R., loc cit. in [1]

The normality assumption, although acceptable in the univariate case since we frequently encounter phenomena distributed according to this law in nature, becomes problematic in the multivariate case because the assumption itself carries such restrictive constraints that it is rarely satisfiable in practice.

Italian contributions to descriptive statistics are found throughout all sectors of the discipline.
Until about 1960 a great number of papers concerning the concept of average (or mean or mean value) and the properties of several averages were published [4]. The best known of the works in this field were published by A. Messadaglia (Il calcolo dei valori medi e le sue applicazioni statistiche, 1883, edited posthumousley in 1958, Biblioteca dell'Economista, serie V, vol. XIX) and C. Gini (Le medie, Utet, Torino, 1958).

Since the 1980's it has been impossible to speak of an *Italian School of Statistics* unlike in the past, when the works of Italian statistician were almost entirely concerned with the analysis of the characteristics of complete populations.
In this regard A. Herzel and G. Leti[5] point out : " *Today this no longer holds for Italian statisticians either, for whom inference has passed in a few years from a position of support of the study of complete collectives to a position of the predominance of statistical methodology. This evolution or rather involution, sets the Italians if not at the same level, surely in the wake of most of today's statisticians who have been moulded by the Anglo-Saxon model, and have therefore cancelled those features which had distinguished the Italian School of Statistics.*
While the Italians have adjusted themselves to the prevailing trends, the first attacks on the predominance of inference are now occurring abroad, aiming at re-evaluating the so called descriptive statistics, which is today being re-proposed with new vigour and problems under the label of data analysis".
What the two scientists said certainly gives an exact description of the general trends of research in Italy since 1950. However, this obviously does not mean that the study and analysis of statistical problems of a non-inferential nature have been completely neglected. In this period we have seen many Italian contributions to data analysis.

The return to data analysis did not mean a loss of interest in the study of classical and Bayesian statistical inference, since it is only through the simultaneous examination

---

[4] Frosini, V.B, loc.cit.

[5] Herzel A., Leti G., I contributi degli italiani all'inferenza statistica, in " I fondamenti dell'inferenza statistica", Dipartimento statistico dell'università di Firenze, 1978.
Chiandotto B., Rizzi A., Recent contributions of Italian statisticians to data analysis,
Metron, XXXVIII, n. 1-2, 1980

of the two fields of research (data analysis and inference) and their interrelation that this vital connection which is essential to the progress of knowledge in the field of observation science can be brought about [6]
We must also recall that even abroad the greatest contributions in this field, at least from the point of view of their operational relevance, were only produced in the period 1960-1980, in great part due to the technological progress made in the field of electronic computers by means of which a large amount of observations and data can be processed automatically.

The important contributions of the French school have been studied and appreciated by many Italian scholars of Statistics. I remember the fundamental papers of J.-P. Benzécri, J.-M. Bouroche, F. Cailliez, E. Diday, Y. Escoufier, J.P Fenelon, B. Fichet, L. Lebart, M. Le Chevalier, A. Morineau, J.-P. Pages, G. Saporta, and of many others.

Gini's work (Variabilità e mutabilità. Contributo allo studio delle distribuzioni e relazioni statistiche, Studi economici giuridici dell'università di Cagliari, 1912, pp 1-159) marked a turning point in the conception of statistical tools, particularly concerning variability and concentration. Gini recalled some studies in the field of Astronomy, the resulting theory of accidental errors and therefore some variability measures as the probable deviation, the mean (absolute) deviation about the mean, and standard deviation. He stressed the conceptual coherence that exists between these tools and the premises concerning the dispersion of the specific phenomena from which they were drawn.

He pointed out the purely formal use of these implementations to measure the variability in various fields, e.g. Biology, Demography Astronomy and Economics. Starting from these considerations Gini introduced the mean difference, the basis of which is to be found in the answer to the question "How much do the observed quantities differ with respect to each other?" in contrast to another question "How much do the observed quantities differ with respect to their arithmetic mean?"

V:Castellano (Il contributo di Gini alla metodologia statistica, vol.1, Istituto di Statistica della facoltà di Scienze statistiche demografiche ed attuariali, Roma, pag 3-27) exhaustively clarified how every type of mean deviation may be expressed without recourse to any mean value. He remarks "*The question of the use of the mean difference as well as of the other indices of dispersion is of a completely conceptual kind, and as such is indisputably solved at the outset by Gini's clear approach".*

---

[6] We are here referring to V. Castellano (17) who affirms that "*... any science is statistics at the stage of observation of the external world and, vice versa, statistics is the empirical moment of observation sciences". But statistics is also the set of all " ... all the procedures ... aiming at testing the validity of an assumption or model (or theory)or of somehow connecting rationally the conditions of a process or development or premises to consequences* ".

Many other Italian scholars worked out specific variability measures following their hypotheses and observations of specific phenomena. We remember Niceforo (1923) for his research in Criminology, Salvioni (1886-1888) and Viola (1933) for their studies about morphological types and Boldrini who, beginning in 1931, constructed the " theory of rational measurement of dispersion ".
These studies always began with a concrete problem, then introduced a particular measure of dispersion and studied the formal properties of the indexes. None of these scholars were mathematicians; they used mathematics after a deep analysis of the logic of the phenomenon.

Gini is well known for the introduction of some basic concepts in the field of concentration including the concentration ratio R [7]. In different papers (1909, 1910, 1912, 1914) he gave a definition of income concentration which was different from the one given by Pareto. Following his new way of considering income concentration and the frequency distribution concentration Gini (1914) (Sulla misura della concentrazione e della variabilità dei caratteri. Atti del Reale Istituto Veneto di Scienze, Lettere ed Arti, aa 1913-1914, tomo LXXIII, parte II, pag. 1203-1248) obtained the δ index and in two successive papers he analysed some other procedures to calculate and interpret δ from real data.

Gini compared δ to Pareto's famous α index and showed that by increasing α the income inequality decreased.
He showed that:

- It is possible to obtain R by utilizing the Lorenz curve.
- R is related to the mean difference Δ and to the arithmetic mean M, by means of the relation: $R= \Delta/(2M)$.
- If the frequency distribution follows the Pareto's law then $R=1/(2 \alpha -1)$

In this paper we will recall Italian contributions to data analysis before the 1980s. In particular we will present the work done in Metrics, quantification theory, cluster analysis and factor analysis.

Quantification Theory

It can easily be concluded that the solution to the problem of quantification- the substitution of some real numbers for the qualitative modalities of an observed character- is of primary relevance within the field of data analysis. It is enough to think, for instance, of the need to homogenize standards and scales when a cluster analysis is to be carried out on units distinguished by qualitative and quantitative modalities of several qualitative data and variables, and of the impossibility of applying many factor analysis techniques to non-numerical data.

[7] Zenga, M., loc. cit. in [1]

In this regard we should note that one of the greatest and most common obstacles encountered by researchers in the social sciences (particularly psychology and sociology which, among the social sciences, have the most recourse to factor analysis) is the problem of data quantification.
The problem of quantification, not to be mistaken with that of the alternative between qualitative and quantitative analysis[8], was studied by a research team headed by A. Herzel which operated at The Istituto di Statistica e Ricerca sociale " C. Gini " della Sapienza università di Roma [9]. The study was only concerned with *ordinal scales*, since quantification cannot be applied to *nominal scales*, these data being, according to their nature, isomorphic to real numbers. A.Herzel (21 and 22) proposed a general quantification criterion based on the optimization (maximization or minimization) of some given statistical indexes, such as arithmetic mean and correlation coefficient. Different solutions can be found depending on the conditions imposed to define the set of quantifying variables. Maxima and minima signs are the same for all considered sets, and all distributions differ by a constant.
The context of research considered is one with a collective of n statistical units grouped according to $X_1, X_2, \ldots, X_{k+1}$ and arranged according to non-decreasing order of a qualitative character Y. Let f_i be the relative frequency of X_i, the problem of quantification consists in determining m+1 ($0 \leq m \leq k$ real numbers $u_1 < u_2 \ldots < u_{m+1}$ and m more natural numbers h_i ($0<h_1<h_2 \ldots < h_{m+1} <k+1$)such as to replace the given collective by another collective where modalities are u_i (1=1,2,…,m+1) and the corresponding relative frequencies are given by:

$f'_1 = \sum f_j \quad (j=1,2,\ldots, h_1)$
$f'_2 = \sum f_j \quad (j= h_{1+1}, h_{1+2},\ldots, h_2)$

…
$f'_{m+1} = \sum f_j \quad (j= h_{m+1}, h_{m+2},\ldots, k+1)$

Having replaced :

$X_{i+1} - X_{i+1} = x_i \quad (i=1,2,\ldots,k).$

The x_i can be determined considering different kinds of objective functions and constraints. For instance:

---

[8] Qualitative analysis, considered as the use of mathematical methods, does not necessarily require the quantification of modalities. The Italian School of Statistics proposed and used instruments, such as the variability and " connessione " (connections) indexes suitable for this purpose.

[9] The results of the research study, as well as the contributions given by A. Herzel himself and by B. Baldessari, G. Marbach and A. Rizzi were published in Metron, vol. XXXIII, nn 1-4, 1974

X' A X =max under the constraint **X' B X** =1 with **A** and **B** positive definite symmetric matrices.

3 Metrics

The algebraic topological structure plays a primary role within data analysis. The structure of a topological space on a set R can be induced by one or more domain and co-domain functions.
A particular type of function called distance is the preferred one among them because it reproduces the structure of real sets and other sets. This approach to data analysis, particular to the French School, is the object of formal analysis from the point of view of the mathematical structure of data.
Almost all cluster algorithms proposed in the literature which are not based on graph theory, involve the introduction of some *similarity or dissimilarity measure* among observations of units and sets of units. These measures always interact with the adopted clustering method and therefore heavily affect the results of any cluster analysis.

G. Leti, who had introduced , in 1961, (28), variability indexes as the means of dissimilarity indexes between each allocation and fixed allocation, approached distances and similarity problems in a special monograph ([34]),, by supplying both the general lines of distances in statistics – from which statistical indexes follow- and some new interpretation of the meaning of distance. In particular, he showed that when two variables are involved, Mahalanobis' distance takes into account the correlation between the variables and pointed out that the metric depends on the set whose points are considered or that the distance is relative to where the points in question are included.
The Author then approaches the relation between this metric and the Euclidean distance, and formulates the following expression for Mahalanobis' distance:

$$M=(\mathbf{X'LX})/D$$

where:

D is the determinant of correlation matrix between the variables $y_1, y_2, \ldots, y_n$.

X is the vector of deviation $[(a_h - b_h)/\sigma_h] \sqrt{D_h}$ where $\mathbf{A}\equiv(a_1, a_2, \ldots, a_n)$ and $\mathbf{B}\equiv(b_1, b_2, \ldots, b_n)$ are the two points.

D_h is the determinant of the correlation matrix having left out the variable y_h (h=1,2,…,n).

σ_h is standard deviation of the variable y_h (h=1,2,…,n).

X' is the transposed matrix of **X.**

L is the matrix of partial correlation where the correlation coefficients are

calculated between two variables, coeteris paribus, with all the other variables.

The preceding expression shows that Mahalanobis' distance depends on partial correlation coefficients.

New types of distances were introduced by M. Badaloni and A. Rizzi (6).
A. Rizzi defined the distance between the two rankings **a** and b of the modalities of two variables as the number of inversions that **a** presents with respect to **b**.
M. Badaloni considered a collective C made of the following n arbitrary numbers units:

$$c_1, c_2, \ldots, c_n.$$

assuming that:

$$a_1, a_2, \ldots, a_n.$$

are the k characteristics and:

$$\mathbf{e_i} \equiv (a_{i1}, a_{i2}, \ldots, a_{in}) \qquad (i=1,2,\ldots,n)$$

where $i_1, i_2, \ldots, i_n$ denote a simple n-class combination of the natural numbers 1,2,…k, is the set of a characteristics linked to the i-th unit of c.
The number:

$$d_{ij} = n - f_{ij}$$

where f_{ij} is the absolute frequency of elements, is defined as the distance between $\mathbf{e_i}$ and $\mathbf{e_j}$.

S.Zani (44) introduced the consistency property of distance indexes. Considering two vectors **X** and **Y** in a space Ω and denoting with Z(**X, Y)** the set of Ω vectors in which each component is included among the corresponding components of **X** and **Y**, (**X, Y)** is defined as *intermediate* between **X** and **Y.** All distance indexes for which the relations :

$$d(X,Z) \leq d(Z,Y) \qquad \text{for every } z \in Z(X,Y)$$
$$d(Y,Z) \leq d(X,Y) \qquad \text{for every } X,Y \in \Omega$$

hold, is defined as coherent.

The Euclidean distance, for instance, is a coherent index which does not satisfy the ultra-metric property. On the other hand there are some distance indexes which satisfy the ultra-metric property but are not coherent. That is, neither of the two consistency and ulta-metric properties implies the other, although they do not exclude one another. Still, there are coherent distance indexes which are not triangular.

N. Lauro (23), trying to overcome the disadvantages of the Spearman's indexes – which is linked to Euclidean distance- proposed alternative indexes obtained from Minkowski's generalized distance.

The extreme sensitivity of the well known Spearman's index is corrected through weighted indexes. It is pointed out, however, that it is not always possible to obtain symmetrical indexes with respect to zero and that this depends specifically on the chosen system of weights.
The last class of indexes considered by the Author in order to overcome these disadvantages is drawn from the definition of relative distance. Among these Amato's index (1), which is based on the comparison between variability indexes, is a particular case. The allocation of the proposed indexes is approached through simulation techniques, by sampling the universe of random permutation of the first integer.

4 Cluster Analysis

Until the 1980s terminological and conceptual confusion prevailed in the research field of *cluster analysis*; this was due to the exceptional and uncoordinated proliferation of cluster algorithms encountered in almost every field of theoretical and applied scientific research during the preceding years. In fact, a great number of specialists called attention to this problem underlining the need for a coordination of theoretical work on the subject. Independently of the specific fields of application, this work should explain the logical foundations as a first step towards the formulation of a *general theory* within which the various cluster methods could be correctly framed and distinguished by their properties and limitations.
G.Lunetta (29) moved in this direction. In an interesting monograph, he approaches the relation between multi-dimensional variability and cluster analysis; he then introduces a more general class of cluster analysis criteria, which possesses relevant properties.
Let us assume we have n statistical units to be classified into p groups according to the measures of k quantitative characters; let **W** be the correlation matrix for standardised variables within groups, which corresponds to an arbitrary grouping. It is well known that **W**'s rank cannot exceed the smallest of the numbers k and n-p. Let us denote with $|\mathbf{W}_S|$ the sum of the principal minors of order s, with $s \leq \min(k, n-p)$.
A mean measurement of multi-dimensional variability within the p groups can be obtained by applying the hypervolumetric variability index expressed by:

$$q_s^2 = |\mathbf{W}_S| (n-s)!/(n!s!)$$

Given the order s of the determinants, it can be agreed that the best grouping is the one which supplies the lowest value of the mentioned index, i.e. which satisfies the condition:

$$|\mathbf{W_S}| = \min$$

It immediately follows that for s=1 the method proposed by Edward and Cavalli Sforza occurs:

$$\mathrm{tr}\,|\mathbf{W}| = \min$$

whereas, if s=k, we shall obtain the method proposed by Friedman and Rubin:

$$|\mathbf{W}| = \min$$

The usefulness of this general approach, can be found in the ability to choose the method which is the most suited to the forming of groups according to a given structure, which can be suggested by the nature of the data or by theoretical considerations. Interpreting $|\mathbf{W_S}|$
in terms of hypervolumes of (s+1)- edron within the k- dimensional space where the units to be grouped are represented, it appears that if there exists a possibility of forming groups made of points which tend to gather around the middle, it would be more convenient to set s=1, considering that $|\mathbf{W_1}|$ would be zero if all points in each group coincide with the centre.

Therefore, the criterion based on the minimization $\mathrm{tr}\,|\mathbf{W}|$ appears suitable when, within groups, characters are not correlated among themselves. If, vice-versa, within one group, the points representing units tend to be ranges around a straight line - that is if simple correlation among characters is high where partial correlations are low- we would do better to set s=2, and seek the arrangement where the sum of determinants of the second order of the correlation matrix for standardised variables takes the lowest value.

The criterion proposed by Friedman and Rubin, in Lunetta's interpretation, should be chosen when, within groups, all characters, though they are different from each group, are thought to be linked by a single multiple regression equation,. Since this condition does not actually occur very frequently, if numerous characters are considered, it is thought that good grouping can usually be obtained even with a modest value of s with respect to the number of characters. The choice of a number s lower than k is similar to deciding to replace variables with some of their principal components, which is done when there are several variables or when the rank of matrix **W** is lower than k.

Lunetta then proposed a new method of cluster analysis using minimization based on the geometric mean of the distances of each point of the set.
Among the studies on cluster analysis A. Bellacicco and, G. Storchi (10) propose a general theoretical framework specifically based on the concepts of abstract algebra

and on graph theory. The main point around which the author's reasoning revolves is expressed by the following considerations.
Data analysis consists of a set of techniques to indicate, with the support of raw data, the essential information included in the data itself. The information is in practice translated into a *form* whose interpretation depends on the particular language used and on the specification of a given criterion function. The form is abstractly defined on an object-predicate table as the set of actual realizations of a set of predicates which belong to a language L on a domain S of statistical objects or constants. A good form corresponds to a particular aspect of the table which, in the abstract language L, represents a *model* of a theory, whereas the criterion function represents predicates whose optimization leads to the definition of a given optimal form and therefore of a given model of the theory.
This point of view is probed in detail considering extensions of a table to infinity, in which the forms constitute topological invariants, and translations of the table into graphs, i.e. relational structures.

In this case data analysis sets out to emphasize special sub-graphs in which to decompose the graph representing data, by optimising a given criterion function. These sub-graphs just constitute the clusters, whose identification then corresponds to the identification of a *good* form, and therefore to a model as some given theory. It is shown that these sub-graphs or clusters, represent structurally stable forms in Thom's mining.

Among the various techniques for identifying an optimal clustering system, A. Bellacicco (11) considers a particular set-partitioning of an optimal and set covering technique as well as all optimal cut techniques which can be expressed in terms of integer optimization. Almost all methods for clustering algorithms that were known in that time fall within this field.

B. Chiandotto (15) proposed a subdivision of the general problem of mathematical classification (distinguishing the *allocation* problems from those of *discrimination* and *clustering*) and underlined how the problems particular to cluster analysis logically precede the problems of discrimination and allocation. He then examined the different stages covered by the process of cluster analysis, and carried out a critical survey of the most often used (hierarchical and non-hierarchical) methods of cluster analysis and illustrated their properties and limits of applicability.
A.Mineo (32, 32) called attention to the confusion that reigned within the concept of clustering, the aims pursued through it and the meaning of the word *group*. He then stated his own point of view on the concept of group and proposed a clustering method which he called *quantitative clustering*.
The proposal to subdivide the clustering methods was inserted by A. Mineo into the larger context of research of variability structures and followed the distinctions between *inductive moments* of research and inductive research.

S.Zani, (41) moving from the consideration that any clustering presupposes, implicitly at least, that it will help to single out classes (groups plainly appearing in observed data), he gives an illustration of cluster analysis which is distinguished by an *unusual* cut, with respect to the relevance of each single topic as well as to the general concept.
Having established that the concept of group is primitive indeed with respect to an algorithm, but not with respect to the empirical research which requires the use of the algorithm itself, the Author dealt with cluster analysis by moving from the principle of the simplicity of clustering, placing great stress , at each step of the method, on the the assumptions, often implicity accepted in a non critical way. The problems considered are those concerned with the choice of variables, the estimation of *distances* between pairs of elements, element grouping and the estimation of results clustering.

5 Factor Analysis

Within the field of factor analysis (which includes classical factor analysis, principal component analysis, correspondence analysis and canonical analysis) Italian scholars have mainly sought to indicate the properties and limitations of the methods proposed by the Anglo-Saxon and French Authors through the formal study of the internal relation among the various methods, and with respect to the concrete problems met with when they are employed in the analysis of real phenomena.
Attention was particularly drawn to the validity of the objectives which can actually be pursued through the application of the various techniques.

The logical and formal connections linking the various methods of statistical multivariate analysis were specifically approached by A. Naddeo (34).

S.Bolasco e R.Coppi (12), who as Naddeo approached the problems of the comparison of different techniques of multivariate statistical analysis, sought not so much to single out a standard method of analysis to which others should conform, as to probe methodological questions while safeguarding the complex dialectics between tools of analysis, data configuration and aims of the research.
The two Authors confront the most significant factorial techniques (common factors, principal component, correspondence analysis and graphical data processing), from the theoretical and experimental point of view, with reference to concrete examples.
In the comparison of the methods mentioned, the Authors approach the robustness of the results by modifying the parameters of the method, the number of variables of the system. Furthermore, the multidimensional structure is modified through the comparison with the factorial method, which starts from a typological structure deriving from the method.
In this framework, the limits of the techniques based on the matrix of variance and covariances are evident, because they do not consider interactions of an order higher than the second one. These interactions, on the other hand, are considered within the

graphical rather than the mathematical processing language. In this case the original information is not transformed (unlike with other factorial methods), and thus it can be used at the level of the final result.
Furthermore, intervention by the researcher during the operation is possible to remove or introduce new variables into the system at issue.
R. Coppi himself, together with F. Zannella (19), dealt with the factor analysis of a multivariate time series by referring to the same set of statistical units.

The methodological examination and explanation of the method of multivariate statistical analysis was undertaken by V. Amato. The Author (1,2) pointed out the usefulness of the logarithmic decomposition of a statistical matrix **A** (m,n) of observed data (m empirical observations of n given variables)according to the formula:

$$\text{Log } \mathbf{A} = \sum_{1,r} (v_i \cdot u_i') \log \lambda_i$$

where:

r	is the rank of matrix **A**
$v_1, v_2 \ldots v_r$	are the orthonormal autovectors of matrix $\mathbf{A}\mathbf{A}'$
$u_1, u_2 \ldots u_r$	are the orthonormal autovectors of matrix $\mathbf{A}'\mathbf{A}$
$\lambda_1, \lambda_2 \ldots \lambda_r$	are the auto-roots of matrix **A** which satisfy the usual decomposition into principal components: $\sum_{1,r} (v_i \cdot u_i') \lambda_i$.

The Author carried out the logarithmic decomposition (which could be preferable to the usual principal components one whenever the relative rather than the absolute variations are to be stressed) through the *kernel* and general identity of a rectangular matrix which can be respectively defined by the relations:

$$\mathbf{A} = (\mathbf{A}' \mathbf{A})^{1/2}$$
$$\mathbf{A}^0 = \sum_{1,r} (v_i \cdot u_i') = \mathbf{A}\,(\mathbf{A}^-)$$

where $(\mathbf{A}^-)$ is the generalised inverse, according to Penrose, of **A**.

$$\text{Log } \mathbf{A} = \mathbf{A}^0 \log \mathbf{A}$$

Through the computation of limit :

$$\lim_{k \to 0} \text{Log } \mathbf{A} = (\mathbf{A}^k - \mathbf{A}^0)/k$$

where $\mathbf{A}^k$ is a defined as symbolic k-th power of matrix **A** :

$$\mathbf{A}^k = \mathbf{A}\,\mathbf{A}^{k-1} = \sum_{1,r} (v_i \cdot u_i') \lambda^k{}_i .$$

V. Amato(2) then considered the problem of canonical correlation, proposed by Hotelling, and found his solution through the definitions of the *dominant* eigen root and of the respective eigenvectors of the matrix.

References

(1) V.Amato, *Scissione dell'autovalore di una matrice nelle sue autoradici principali e una applicazione alla correlazione interstrutturale,* Giornale degli economisti e annali di Economia, nov.dic. 1976

(2) V. Amato, *Autoradici di una matrice rettangolare e loro impiego nel calcolo dei punteggi delle variabili doppie,* Rivista di Statistica Applicara, vol.II, n. 2 1978

(3) V.Amato, *Autoradici di una matrice rettangolare e loro impiego nell'analisi canonica di Hotelling*, Studi in onore di Giuseppe De Meo, Istituto di Statistica Economica dell'università di Roma, Roma, 1978

(4) V.Amato, *Alcune possibilità di applicazioni statistiche del logaritmo di una matrice rettangolare,* Rivista internazionale di Scienze economiche e commerciali, anno XXV, n.7, 1978

(5) V. Amato, *Logarithm of a rectangular matrix with application in data analysis*,Second International Symposium on Data Analysis and Informatics, Iria, Versailles, 1978

(6) M. Badaloni, A.Rizzi, *Contributi alla cluster analysis*, Metron, XXX, 1972

(7) B. Baldessari, *Un metodo di quantificazione:aspetti generali e* campionari, Metron, XXXII, 1974

(8) B. Baldessari, *Nuovi metodi di quantificazione basati sugli indici di dissomiglianza*, Metron,XXXII, 1974

(9) Bellacicco, A. Labella, *Cluster su fuzzy sets ed invarianti algebrico-topologici su tavola oggetto-predicato,* Rendiconti del Circolo Matematico di Palermo, XXV, 1976

(10) Bellacicco, G. Storchi, *Su un metodo di clustering basato sulla condizione di bicromatismo di un grafo*, Giornate A.I.R.O, Padova, 1975

(11) A. Bellacicco, *Grafi isometrici e clustering gerarchici di grandi dimensioni*, Università di Padova, Bressanone, 1978

(12) S. Bolasco, R.Coppi, *Un'analisi comparativa di diverse tecniche fattoriali,* Statistica, n. 3, 1979

(13) E. Castagnoli, *Un'osservazione sull'analisi classificatoria,* Università di Padova, Bressanone, 1978

(14) V. Castellano, Istituzioni di Statistica, Edizioni Ilardi, Roma, 1975

(15) B. Chiandotto, *L'analisi dei gruppi:una metodologia per lo studio del comportamento elettorale*. Quaderni dell'osservatorio elettorale, n. 4, Firenze, 1978

(16) B.Chiandotto, G. Ghilardi, R.Leoni, *Un modello statistico natematicodi localizzazione industriale,* Federazione Regionale delle Associazioni industriali della Toscana, Firenze, 1978
(17) R.Coppi, *Sull'analisi dell'interdipendenza nelle tabelle di contingenza multiple*, Metron, XXXXIV, 1976
(18) R.Coppi, *Riflessioni sulle tendenze della ricerca statistica nel campo dei caratteri qualitativi,* Statistica, n. 4, 1978
(19) R.Coppi, F. Zanella, *L'analisi fattoriale di una serie temporale multipla relativa allo stesso insieme di unità statistiche,* Società Italiana di Statistica, XXIX riunione, 1978
(20) A.Herzel, G.Leti, *I contributi degli italiani all'inferenza statistica*, Dipartimento statistico dell'università di Firenze, 1978
(21) A.Herzel, *Un criterio di quantificazione. Aspetti statistici,* Metron XXXII, 1974
(22) A.Herzel, *Un criterio di quantificazione. Aspetti matematici,* Metron XXXII, 1974
(23) N. Lauro**,** *Considerazioni sulla metrica degli indici di cograduazione,* Giornate A.I.R.O, 1977
(24) N.Lauro, R.Morgeluzzo, *L'analisi dei sistemi a larga scala con l'ausilio dell'elaboratore elettronico,* Informatica e documentazione, n. 1977
(25) N. Lauro, G. D'Alfonso, *L'analisi strutturale nello studio dei sistemi biomedici,* Statistica Applicata, n.3, 1978
(26) N.Lauro, P. Rostirolla, S.Vinci, *Metodologia integrata per la lettura socio-economicadel territorio e la localizzazione degli interventi,* C.S.E.I., Napoli, 1979
(27) G.Leti**,** *Nuovi tipi di distanze fra insiemi di punti e loro applicazioni alla Statistica,*Metron, XXI, 1961
(28) G.Leti, *Considerazioni sugli insiemi di distribuzioni,* Metron, serie C, vol.II, 1963
(29) A. Lunetta, *Variabilità a più dimensioni e analisi dei gruppi*, Università di Catania, 1973
(30) G. Marbach, *Sulla presunta equidistanza degli intervalli nelle scale di Valutazione, Metron XXXII, 1974*
(31) A.Mineo, *A new grouping method for the right evaluation of the chi-square test of goodness of fit,* Scand. Journal of Statistics. 6, 1979
(32) A. Mineo, *Un nuovo criterio di raggruppamento in classi,* Società Italiana di Statistica , XXIX riunione, 1978
(33) A.Mineo, *Per una classificazione della classificazione,* Università di Padova, Bressanone, 1978
(34) A. Naddeo, *Relazioni tra alcune analisi multivariate,* Università di Padova, Bressanone, 1978
(35) A. Rizzi, *Confronto tra alcuni metodi di clustering,* Università di Padova, Bressanone, 1978

(36) A.Rizzi, *Sull'uso delle variabili soggettive nei modelli econometrici.* Metron, XXXII, 1974
(37) A.Rizzi, *Un metodo di quantificazione applicato ai problemi di classificazione,* Metron, XXXII, 1974
(38) T.Salvemini, *Indipendenza statistica tra più variabili,* Società Italiana di Statistica,IX, X, XI riunione, 1951
(39) T.Salvemini, *Sulla dipendenza in media tra le componenti di una variabile tripla*, Società Italiana di Statistica, XXIII, 1963
(40) S. Zani, *Sulle proprietà degli indici di distanza nell'analisi classificatoria,* Istituto di Statistica, Università di Parma, 1975
(41) S.Zani, *L'analisi classificatoria: contributi metodologici ed impiego per l'individuazione di aree omogenee,* Istituto di Statistica, Università di Parma, 1977