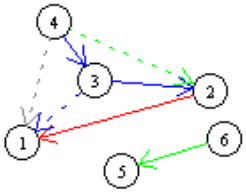
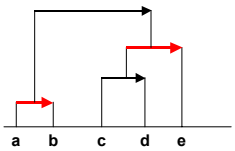
Università degli Studi di Palermo

ACTES
Cinquième Colloque International A.S.I.

PROCEEDINGS
Fifth International Conference S.I.A.



ATTI
Quinto Convegno Internazionale A.S.I.



Analyse Statistique Implicative

Objet de recherche et de formation en analyse des données, outil pour la recherche multidisciplinaire

Prolongement des débats.

Analisi Satistica Implicativa

Oggetto di ricerca e di formazione in analisi di dati, strumento per la ricerca multidisciplinare.

Prolungamento di dibatti

Implicative Statistic Analysis

Implicative Statistic Analysis as an object of research and training in data analysis, a means for multidisciplinary investigation.

Beyond of the debates

Palermo, 5-7 Novembre 2010

Supplemento n.1, N. 20, 2010 - *Quaderni di Ricerca in Didattica Matematica*
ISSN on-line 1592-4424
Editors : Jean-Claude Régnier, Régis Gras, Filippo Spagnolo, Benedetto Di Paola

Remerciements

Nous tenons à remercier les différents organismes qui nous ont apporté un soutien pour la tenue de ces Journées sur les plans scientifique, financier et logistique :

Université Lumière Lyon2 (Lyon France)
Laboratoire de recherche : Unité Mixte de Recherche UMR 5191 – ICAR
Association pour la Recherche en Didactique des Mathématiques (ARDM),
Association Extraction et Gestion des Connaissances (EGC),
Université de Palerme, en la personne de son Recteur, le Professeur Giuseppe Silvestri,
Faculté des Sciences de la Formation, en la personne du Professeur Patrizia Lendinara,
École Polytechnique de l'Université de Nantes,
Laboratoire d'Informatique de Nantes (LINA),
Gruppo di Ricerca sul l'Insegnamento della Matematiche (GRIM)
Société PerformanSe SA,Carquefou- Nantes,
ou leur parrainage scientifique *:*

Società Italiana di Statistica (SIS)
Société française de Statistique (SFdS)
Société Francophone de Classification (SFC),
International Association for Statistical Education (IASE)

Le président du comité scientifique : Jean-Claude Régnier
Le Président d'honneur : Régis Gras
Le Président du comité d'organisation : Filippo Spagnolo

Preface to the edition on line of the Proceedings of the 5th SIA Conference: communications and discussions

Palerme, November 5 to 7, 2010

Jean-Claude Régnier
Régis Gras
Filippo Spagnolo

Statistical Implicative Analysis: once more, what is it about?

In an attempt to deal with the question on the origin and development of Statistical Implicative Analysis, Régis Gras and Jean-Claude Régnier consider that, nowadays, it designs a theoretical field focused on the concept of statistical implication or, to be more precise, on the concept of quasi-implication, in order to distinguish it from the concept of logical implication as found in the domains of Logic and Mathematics. The study of this concept of quasi-implication as the mathematical object in the areas of probability and statistics, allows to build the theoretical tools that are instrumental in a data analysis method.»[1] (Gras, Régnier, 2009 p.12).

This fifth Conference on Statistical Implicative Analysis was organized within this global theoretical approach. While an important part of the analysis of the development of Implicative Statistical Analysis comes from work produced, oriented, or triggered by Régis Gras since the 70s, this development also took place through a succession of international meetings like the ones at the IUFM of Caen (France) in 2000, at the PUC in Sao Paulo (Brazil) en 2003, at the University of Palermo (Italy) in 2005, and at the University of Castellón (Spain) in 2007. These meetings constituted an opportunity for presenting work that led to discussions and debates, contributing to theory development and its many applications. This fifth edition represents an extension of the previous ones. Like in other sciences, this conference aimed at capitalizing on knowledge built by each of the S.I.A community members working within an approach based upon a fundamental tool, the software CHIC which, within this context, has been the object of a permanent evolution.

The software CHIC (Classification Hiérarchique Implicative et Cohésitive), was developed in the early 90s by Saddo Ag Almouloud, then by Harrisson

1 Gras R., Régnier J.-C., Guillet F. (Eds) (2009) *Analyse Statistique Implicative. Une méthode d'analyse de données pour la recherche de causalités.* RNTI-E-16 Toulouse Cépaduès Editions. Consultable : http://www.cepadues.com/livre_details.asp?l=897

Ratsimba-Rajohn, and currently by Raphaël Couturier. Information is available from:

http://www.ardm.eu/contenu/logiciel-d-analyse-de-donnees-chic/

Now in its 5.0 version, translated into English, Spanish, French, Italian, Portuguese, and Slovaque, it allows to:

- Work with different types of variables (binary, modal, frequencies, intervals, floues),
- Quantify the significance of values attributed to quality, the consistency of the rule associated to the implicative relation between variables, between ordered classes of rules, and to the typical contribution of subjects or subject categories to certain rules.
- Graphically represent, at a chosen quality level, the rule paths and a hierarchy of rules among rules (also called generalized rules),
- Suppress, add, join variables.
- Represent through an increasing hierarchy the similarity of classes of variables.

A new function, *Réduction*, was introduced in this last version, which is fully coherent with R.Gras' theory of statistical implications (S.I.A.). This CHIC sub-program allows transforming a set of variables, which may be very large, into a reduced sub-set where one can determine the level of reduction while minimizing the hidden information. For that, one needs to define an equivalence criteria between variables based on one of two indices of implication (classic or entropic) and on more complex criteria integrating these measures as well as their support. Simultaneous consideration of the implications and their counterpart guarantees class quality and consistency, which leads to rigorous selection of equivalent elements. Thus, the program eliminates all variables within an equivalent class except for one that will represent the class. This choice takes into account minimal distances within each class.

S.I.A. – Statistical Implicative Analysis: Some topics of debate and construction.

Looking back, the first ASI1 Conference, which took place at the l'Institut de Formation des Maîtres de Caen (France) on June 23-24, 2000 and was organized by Marc Bailleul and Régis Gras, had a general theme focused upon « Mining data through the Implicative Statistical Method». This choice related to a promising orientation if one considers the development and the activities of the KEA (Knowledge **E**xtraction and Management) association. Remember that Data Mining (also called "Knowledge Discovery in Databases" in the anglo-saxon

literature) starts, in general, from the matrices of subjects (or objects) and binary, ordinal, or numerical variables (properties or attributes). The main goal is to establish conjectures about models based upon quantitative or qualitative relations and the structures detected from the data. Different methods, such as the Factorial Analysis of Correspondences (F.A.C.), the Hierarchic Ascendent Classification (C.A.H.), are usually employed in these data mining. Among them, Statistical Implicative Analysis (S.I.A) aims at extracting knowledge, invariants, and consistent symmetric inductive rules and allows to measure propositions of the type « if a is chosen, it is likely that b is chosen».

The second Conference ASI2 was organized at PUC, Pontificia Universidade Catolica in São Paulo (Brésil) on July 9-11, 2003, by Saddo Ag Almouloud,.within the general theme « O metodo estatístico implicativo utilizado em estudos qualitativos de Régis Gras de associação. Contribuição à pesquisa em Educação »

The third Conference ASI3 was organized by Filippo Spagnolo from University of Palermo (Italie) on October 6-8, 2005.

The fourth Conference ASI4 was held at Univerity Jaume I de Castellón de la Plana (Spain) on October 18-21, 2007 and was organized by Pilar Orus and Pablo Gregori.

Each of these colloquia led to publication of printed and digital proceedings accessible online, allowing for diffusion of communications that were evaluated by members of the scientific committee.

In the case of the current fifth Conference ASI5, it constituted an opportunity, as shown by these new proceedings, for hearing and discussing nearly 30 communications selected by the scientific committee, from those previously submitted. It is worth noting that, between the previous ASI4 colloquium and this one, a major publication in French, *Analyse Statistique Implicative. Une méthode d'analyse de données pour la recherche de causalités*, was published by Cépaduès Editions (Toulouse France)**.** It provided a general account for the 2009 stage and constituted a significant actualization of the 1996 publication *L'implication statistique*, by La Pensée Sauvage Editeur (Grenoble France). A second major publication in English appeared in **2008, *Statistical Implicative Analysis,* distributed by** Springer-Verlag, (Berlin-Heidelberg, Allemagne). Finally, a third publication, in Spanish, also appeared under the title *Teoria y Aplicaciones del Analisis Estadistico Implicativo*, distributed by Universitat Jaume-1, (Castellon Spain). The international dimension in this fifth edition was further strengthened by

the fact that five languages, English, Spanish, French, Italian and Portuguese, were accepted as languages for presentation of communications. This led to a new challenge: to avoid somewhat like a Babel tower, each participant was asked to pay particular attention and acceptance of efforts towards mutual understanding. As was the case for the other colloquia, researchers from Germany, Brazil, Chili, Spain, Cyprus, France, Gabon, Greece, Italy, Japan, U.S.A., etc. presented their work on I.S.A. as objet, as tool, or as the focus of teaching and learning.

The main themes proposed in the call for contributions were:

- SIA' fundamental concepts: statistical models, variables types, main and supplementary variables.
- New current developments, index stability, extension towards new variable types, exception rules; subject dual space – rule space, metric and topological structure of detected subjects determined by their contribution or typical behavior, vector analysis.
- Critical comparison between SIA's procedures, models, representations, and results other data analysis methods (Galois lattice, Bayesian networks, induction tree, factorial analysis, etc.)
- Use of CHIC software, current and future developments.
- ASI applications and their comparisons to other methods in the fields of didactics, education sciences, psychology, sociology, economics, art History, biology, medicine, archeology, etc.
- Graphical and digital presentation of applied results, help for interpretating results, respective and critical roles of main and supplementary variables chosen.
- Specificity of preparation for using SIA: use of CHIC software, graphical representation interpretation (implicatif graph; cohesitive hierarchic tree)
- SIA's didactical problems.

Concerning continuity, this ASI5's colloquium program attempted to maintain the previous orientation and to address a balanced view of communications about theoretical approaches or applications, and practical work with CHIC, which has been favorably viewed by experts as well as by young researchers. This choice is based on the fact that searches towards conceptualization, analysis, and interpretation clarify each other and that the debates following each intervention contributed to increased precision and specificity of certain points that clarify and guide courant and future research.

S.I.A. – Statistical Implicative Analysis: Object of research and preparation for data analysis, interdisciplinary tool. Discussion continuity.

This current publication on line is part of the constructive logic implicit in the previous proposals. It presents the communications after they were modified by the authors, who tried to take into account criticism and suggestions proposed

during debates that took place after the presentations. While the first edition of the proceedings in printed format was organized according to the language of presentation, this new edition is organized according to themes.

At the highest level, we have considered three overarching categories:

- Contributions to the foundations and the theoretical development of SIA.
- Applications of SIA.
- Questions about formation of SIA.

The second category was then sub-divided according to field of application. We identified seven domains in this colloquium:

- Mathematics and science didactics
- Study of questions related to teacher preparation and the practice of teaching
- Study of impact and use of ICT and ENT and of online teaching.
- Evaluation domain
- Art History
- Epidemiology
- Study of socio-economical phenomena

This is why we organized this publication according to the following structure:

Contributions to the foundations and the theoretical development of SIA

Applications of SIA: Mathematics and science didactics