International Federation of Classification Societies Newsletter

The International Federation of Classification Societies, founded in 1985, is composed of: British Classification Society, Classification Society of North America, Gesellschaft für Klassifikation, Japanese Classification Society, Société Francophone de Classification, Società Italiana di Statistica, Association des Sociétés Yougoslaves de Statistique, and Vereniging voor Ordinatie en Classificatie.

The IFCS is a non-profit, non-political scientific organization, the aims of which are to further classification research. Among other activities, the IFCS organises a biennial conference, and supports the Journal of Classification.

Editorial

This will be my last newsletter before I hand over the presidency of the IFCS to Prof. Day at the end of the year. Those of you who were able to attend IFCS'91 in Edinburgh last August will be well up-to-date in the affairs of the Federation by many of you, far too many in my opinion, were unable to attend. Therefore, I can take this opportunity of supplying some information.

Firstly the next two conferences are agreed. As you will already know, the Société Francophone de Classification is hosting IFCS'93 in Paris (August 31-September 3) and I am pleased to confirm that an invitation has been accepted from the Japanese Classification Society to host IFCS'95 in Tokyo. To host an international conference is no small thing and I am very grateful to these societies for their willingness to take on the burden.

The IFCS has had only three conferences and their general structure, organisation and financing remains fluid. That is another way of saying that there are no guidelines, so that organisers are sometimes unsure of the division of responsibilities between the IFCS Council, the Scientific Programme Committee and the Local Organising Committee. Then there is a difference of opinion as to the proper number of parallel sessions and how much and what kind of material is better presented in poster sessions. Following a suggestion made at the Charlottesville Council meeting, the number of poster sessions was increased at IFCS'91, but many people were unhappy with this. Again,

the best date for the meeting is a contentious issue and there have already been protests at the date of IFCS'93, even though this was chosen carefully to be outside the general Northern Hemisphere summer holiday period and immediately to precede the ISI meeting in Florence. Clearly there are many things that need ironing out and once again Hans Bock has come to our rescue. He has agreed to chair a small committee looking into these and related matters and to make recommendations and suggest guidelines. The Federation owes much to Hans Bock, who you will recall was not only our first president but also chaired the committee that laid the foundations of this newsletter. No doubt Hans will be asking the views of the Federation Member Societies, but I feel sure that he will be delighted to have the views of any of you who write to him as individuals.

Another small committee is being chaired by Professor Day to examine a proposal that the Federation might follow in the footsteps of many other scientific societies and confer honours of one kind or another on its members. These honours could recognise especially distinguished work in our field or offer encouragement to promising younger workers. Possibilities include some kind of honorific, medals, or compiling a role of honour. This committee will say whether or not it thinks the idea should

Contents

be proceeded with, and if so, in what form. Once again I am sure Bill will be glad to have the views of individuals as well as the formal responses of the Member Societies.

The Federation has not grown in the last two years but interest has been shown in several quarters. The Section on Classification and Data Analysis of the Polish Statistical Society soon expects to submit a formal application for membership. Following the successful symposium in Pushchino, last year, the USSR Classification Commission has expressed interest, and I can only hope that the recent tumultuous developments in that country will accelerate, rather than retard, its application to join the Federation. More recently, Portugal has announced its interest in joining and they sent an observer to the recent Council meeting in Edinburgh. The Federation is healthy and showing every sign of continuing its growth.

It remains for me to thank everyone who has given time to help with the many facets of Federation activities. Many of these I have already mentioned but there are too many to mention by name and, even if I tried to, there is always the risk of offending someone by forgetting them. The one exception is our Secretary-Treasurer, Pierre Legendre, whose sterling work since the first year of the Federation has been of incalculable value to us all but especially to me and my predecessor, Bob Sokal.

John Gower
Department of Data Theory
University of Leiden
President, IFCS

Journal of Classification, Number 1, 1991: Contents

Note: Book Reviews are not included.

Report on the IFCS-91 Conference —

The State of Classification Research in 1991

To have an orientation, a vector must have an origin. Let us first look at where we come from in order to assess what the future of classification research might be.

Modern methods of numerical classification, used by the members of the various Classification Societies throughout the world, have a multi-national origin. Polish anthropologist Czekanowski (1909) and ecologist Kulczynski (1928) are probably the oldest references to classifications obtained by numerical methods; they were using a technique known nowadays as seriation. Similarity measures computed from quantitative data had been pioneered by Swiss ecologist Jaccard (1901), though. Later, a group of Polish mathematicians (Florek, Lukaszewicz, Perkal, Steinhaus and Zubrzycki) led by Lukaszewicz (1951) developed nearest-neighbour analysis, which is the basis of single linkage clustering, shortly after ecologist Sørensen (1948) in Denmark had developed the method of complete linkage clustering. But it was left to Sneath (1957) in Great Britain and to Sokal & Michener (1958) in the USA to develop the first machine-based methods of numerical classification, which led to their first synthesis of the field (Sokal & Sneath, 1963) and later to the foundation of Classification Societies throughout the world. Both of these great scientists were present at the Edinburgh IFCS-91 Conference, where they presented us with important results of their respective research programmes.

The third IFCS Conference was held on August 6-9, 1991, in the Conference Centre of Heriot-Watt University in Edinburgh, Scotland. The Scientific Programme Committee was chaired by Dr. David Hand, while Dr. David Wishart competently chaired the Local Organising Committee, with the invaluable assistance and collaboration of Dr. Allan Gordon. The Travel Grant committee was chaired by Dr. Frank Critchley. There were 144 registered participants representing 22 countries. The Conference was also the opportunity for the IFCS Council to meet, under the chairmanship of IFCS President John Gower.

The programme included a balanced overview of the spectrum of present preoccupations in numerical data analysis; 80 read papers, 23 posters, and 7 computer demonstrations were presented, in addition to 9 plenary and special lectures on key issues and new fields of development of classification-related methods.

The invited session themes included: cluster analysis, permutation methods, software development, ordination methods, analysis of longitudinal data, consensus theory, diagnosis, model selection in classification, probabilistic models in cluster analysis. Contributed papers were divided into sessions entitled: correspondence analysis, factor analysis, trees, discriminant analysis, cluster analysis, applications, multidimensional scaling, nonparametric analysis, time series, neural networks, and theory and methods of classification.

From six founding Member Societies in 1987, the International Federation of Classification Societies has

already grown to eight Members, and three other national classification societies or groups have recently expressed interest in joining the Federation. The dynamism of classification scientists, as materialised by the diversity of the IFCS-91 programme, is a promise for continuing development of the field of numerical classification.

The fourth IFCS Conference will be held in Paris in August-September 1993 under the auspices of the Société Française de Classification.

References

Czekanowski, J. 1909. Zur Differentialdiagnose der Neandertalgruppe. *Korrespondenz-Blatt deutsch. Ges. Anthropol. Ethnol. Urgesch.* 40: 44-47.

Jaccard, P. 1900. Contribution au problème de l'immigration post-glaciaire de la flore alpine. *Bull. Soc. vaudoise Sci. nat.* 36: 87-130.

Jaccard, P. 1901. Étude comparative de la distribution florale dans une portion des Alpes et du Jura. *Bull. Soc. vaudoise Sci. nat.* 37: 547-579.

Kulczynski, S. 1928. Die Pflanzenassoziationen der Pieninen. *Bull. int. Acad. polonaise Sci. et Lettres. Classe Sci. math. et nat., Série B, Suppl. II* (1927): 57-203.

Lukaszewicz, J. 1951. Sur la liaison et la division des points d'un ensemble fini. *Colloquium math.* 2: 282-285.

Sneath, P. H. A. 1957. The application of computers to taxonomy. *J. gen. Microbiol.* 17: 201-226.

Sokal, R. R., and C. D. Michener. 1958. A statistical method for evaluating systematic relationships. *Univ. Kansas Sci. Bull.* 38: 1409-1438.

Sørensen, T. 1948. A method of establishing groups of equal amplitude in plant sociology based on similarity of species content and its application to analysis of the vegetation on Danish commons. *Biol. Skr.* 5: 1-34.

Pierre Legendre
Secretary-Treasurer of the IFCS
Département de sciences biologiques
Université de Montréal

A University Statistics Unit in a West-African Developing Country: the *Cellule "Analyse des Données"* of the *Université Nationale de Côte d'Ivoire*

Dominique Desbois* & Georges Vidal*

The field of activity of the *cellule "Analyse des données"* covers all phases of the computerized processing of statistical data. Its operations spread into three main directions: logistical support; training, both initial and ongoing; applied research and development.

1. Logistical support

The *cellule "Analyse des données"* offers the following logistical services.

- Hardware: The main hardware facilities are on the *Office central de la mécanographie* (OCM) site, equiped with two IBM host computers operating under two different modes: batch processing (an IBM 4381 under MVS/TSO) and conversational (an IBM 4341 under VM/CMS). A local tele-processing facility is available on the university campus through a synchroneous link (BSC) at 9600 bps for a remote cluster of IBM terminals running under an IBM 3164 controller. Acces to the EARN/Bitnet network is provided either in a synchroneous mode (9600 bps) by the west African node (CIEARN, located at the *Centre inter-régional de Côte d'Ivoire* - CIRCI) or in an asynchroneous mode (300 bps & 1200 bps) through the local X25 packet switching network (SYTRANPAC). Personal computers running under MS-DOS are available on the campus computer centre site (the *Centre universitaire de traitement de l'information* - CUTI) and some of them, equipped with a 3278 board, can be used as IBM terminals;

- Software: This service covers the purchase, utilization, maintenance and documentation of standard packages or programs libraries such as SPSS, BMDP, SAS, TSP and SPAD-N or ADDAD and MODULAD. These packages and libraries are available both on the mainframe and on the personal computers for training and research purposes. Specialized software is proposed through many programs available on the personal computers (for instance, SICLA for clustering analysis, MacSAIF for correspondence analysis, DNA Strider for molecular sequences analysis and so on);

- Technical assistance: This includes data handling, computerized processing and interpretation of results of statistical analyses. This service was primarily devoted to academic research but has been extended to some public development agencies. It may be organised in the framework of an annual contract (such as the ORSTOM[1] convention) or by an ad-hoc agreement (for some case-studies). For instance, the following subjects were dealt with among others since 1981: sickle cell anemia, strongyli-

aisis, and cattle raising[2].

2. Training in statistical data analysis

Training is aimed at users who are neither computer scientists nor professional statisticians, but rather psychologists, teachers, sociologists, linguists, geographers, physicians, economists, plant biologists, agricultural engineers, etc.

These practical courses[3], dealing with statistics, computers and statistical packages at an introductory level, have been designed following the goal-learning approach, on the basis of needs expressed by the attendants. The ultimate objective is to help the participants to become independent in the use of statistical software installed by the *cellule "Analyse des données"* at the *Université nationale de Côte d'Ivoire* (UNCI). Starting from this global aim, the curriculum consists of a coherent hierarchy of specific operational objectives organized around intensive sessions of practical work on real life data as examples. In line with the taxonomy of operational objectives, an evaluation is made at each cognitive level to determine the degree of acquisition in terms of skill integration (task execution, retention, learning transfer - both operational and integral). The transmission of know-how involves the choice of the appropriate statistical methods and dedicated software tools, the setting up of computerized processing and the interpretation of numerical results from the statistical analysis.

- Initial training: There is no curriculum dedicated to computer science at the UNCI, so initial training in data processing is part of the standard curricula in the different disciplines. The training program is primarily developed in the Humanities School and the Economics School at the bachelor and the master degree levels: B.A. & M.A. in sociology; M.A. in communication sciences, geography, linguistics, psychology; B.A. & M.A. in applied economics - both public and corporate options); and the two years of the professional program for the educational advisor grade. The annual program accounts for 600 hours involving 700 students.

- Ongoing training: The training cycle *"Analyse des données"* is aimed at researchers and users and contains three independent and complementary modules:
• Module 1 *Introduction to statistical software;*
• Module 2 *Methods/tools of exploratory data analysis;*
• Module 3 *Advanced methods in multidimensional statistical analysis.*

The sessions are given at the rate of one or two per year, with 15 participants per session.

3. Research and development

Actions in favor of research provide both logistical support and technical counseling. The *cellule "Analyse des données"* offers services ranging from technical assistance to complete bureau studies. This logistical support for academic research takes the form of training for doctoral studies in psychology and tropical geography, where the work involves processing of survey data. These services are provided within the framework of contracts between the research institutes and the CUTI. They may end up with a complete integration of research in the context of interdisciplinary cooperation.

At its own level, the *cellule "Analyse des données"* favors two axes of research:

. Humanities: This research takes place in the field of educational sciences. It focuses on the evaluation of output from educational units in the higher education system of the Ivory-Coast through cohort studies[4] in order to find the factors which determine success;

. Biometry: The aim of a first set of studies[5] is to take stock of the tropical plant genetic resources (coffee, yam, cassava, egg-plant, etc.) through the determination of the enzyme polymorphism within a complex of plant species, by correspondence analysis of the zymmograms. The second field of research concerns the taxonomy of plant viruses through the amino-acid composition of their coat protein[6] by hierarchical clustering methods.

This research and development activity has led to interuniversity cooperation in the fields of problematics, practical pedagogic aspects, the methodology of research and evaluation, the publication of results and the use of consulting resources. In particular, we may mention the *Laboratoire de Statistiques de l'Université Paris VI* (LSUP, Prof. J.P. Benzécri), the *Groupe d'Etudes et de Méthodes appliquées à la Sociologie* (GEMAS, Prof. R. Boudon), the *Laboratoire de Biométrie* from the *Ecole Nationale Supérieure d'Agronomie de Montpellier* (ENSAM, Prof. Y. Escoufier) and the *Centre de Recherche, d'Etudes et de Documentation sur l'Observation des Conditions de vie* (CREDOC, J.P. Fénelon). In the framework of our software policy, the CUTI is a member of non-profit university associations (ADDAD[7], CISIA[8], MODULAD[9]) which distribute statistical software corresponding to our needs and financial possibilities.

1 *Institut pour la recherche et le développement en coopération,* a French applied research institute working in the developing countries.
2 Cf. *Vache de la houe, vache de la dot : élevage bovin et rapports de production en moyenne et haute Côte d'Ivoire* by Ph. Bernardet, Editions du CNRS, 227 pp, Paris, 1988. This agro-economical study analyzes the sociological factors influencing the cattle raising production process in the north (savanna) and the center (semi-wooded area) of

the Ivory-Coast.

3 Cf. *The teaching and practice of multi-dimensional data analysis: an ivory-coast case study.* by D. Desbois & G. Vidal, Proceedings of COMPSTAT 82 Volume 2 (posters), Springer-Verlag, 1982.

4 Cf. *Accès et succès à l'Université : enquête exploratoire, analyse typologique, profils et résultats des étudiants de la Faculté des Sciences Economiques,* by B. Barrère, C. Coulibaly, D. Desbois & G. Vidal, UNCI, Abidjan 1986, 370 pp (ISBN 2-7166-0289-1). This report introduces the results of an exploratory survey, the first step of research that aims at studying the internal efficiency of the Department of Economics as an educational sub-system.

5 These studies are funded by the International Board for Plant Genetic Resources.

6 Cf. *Classification of furoviruses based upon the amino-acid composition of their coat protein,* by C. Fauquet, D. Desbois, D. Fargette & G. Vidal, in *Viruses with fungal vectors,* J.I. Cooper & M.J.C. Asther, Eds, Development in Applied Biology 2, Association of Applied Biologists, Warwick, 1988, pp. 19-36.

7 *Association pour le Développement et la Diffusion de l'Analyse des Données,* a complete library of Fortran programs is available from ADDAD, 22 rue Charcot, 75013 Paris.

8 *Centre International de Statistique et d'Informatique Appliquée,* the CISIA (2 bis, rue Jules Breton, 75013 Paris) distributes the SPAD.N statistical software of L. Lebart, A. Morineau & Th. Lambert.

9 The MODULAD Club (INRIA - Rocquencourt, 78153 Le Chesnay, France) produces a modular Fortran library which encompasses several procedures for factorial and cluster analysis.

Conference Announcement:
The First US/Japan Conference on The Frontiers of
Statistical Modeling: An Informational Approach

May 24-29, 1992; Knoxville, Tennessee, USA.

This statistical modeling conference will bring together 35 distinguished statisticians who are involved in innovative research in applications of the informational approach to modeling. The scientific core of the conference will consist of three major areas: I. Theory and methodology of time series (organizer: T. Ozaki); II. Multivariate statistical modeling (organizer: H. Bozdogan); and III. Engineering and scientific applications (organizer: M. Ishiguro). Keynote lectures will be given by: Hirotugu Akaike, Director General, The Institute of Statistical Mathematics, Tokyo, Japan.

Further information: Judy Snow, Conference Coordinator, US/Japan Modeling Conference 1992, Department of Statistics, 331 Stokely Management Center, The University of Tennessee, Knoxville, TN 37996-0532, USA. Phone: (615) 974-2556, fax: (615) 974-3100, email: bozdogan@utkvx.utk.edu.

Software Report from Japan

The monthly journal for professional statisticians, "Statistics", in Japanese had a special issue under the title of "Information Technology Oriented Society and Statistical Education" last March. One article tackled the issue of statistical software for microcomputers in Japan, the authors of which are both JCS members, - Prof. Ohsumi and Prof. Imaizumi. Concerning the features of Japanese statistical software there are widely accepted statements which have never been discussed seriously by specialists. Some examples of these statements are the following.

There are many differences between Japan and western countries including the US with respect to statistical software. The fact that Japanese statisticians have only had weak contact with information technology seems to be a major cause.

There are only a few statisticians in Japan who hold an interest in the development of statistical software. It follows that data analysis specialists tend to design and evaluate statistical software without any expertise in software development.

Finally, and as a consequence, Japanese statistical software appears to incorporate inaccurate precision and incomplete interfaces with the various forms of data.

Hence, despite the fact that great progress has been attained in the field of computer hardware a remarkable advance is invisible in the statistical software products in Japan compared with other branches such as word processing, spreadsheet, database and communication software.

I am unable to make any statements with regard to the European and American situation regarding statistical software in general and although I would agree with the features cited above, it seems to me that they are not regional characteristics.

We have had no mature market and sufficient computing capability in Japan for the development of statistical software such as SAS, SPSS, MINITAB, P-STAT, GEN-STAT, BMDP, OSIRIS and other products. After all, in the 1980s personal computing appeared in parallel with big computers and now certain changes in the use of computing equipment are proceeding in factories and offices which are referred to as downsizing phenomena. Many products tended to go well with microcomputers and they proved themselves quite powerful. On the other hand, personal computing software products by nature have been growing, including those in the field of statistical software. Interactive operation and graphical representation are very common to microcomputer software. Readers will probably agree that microcomputers are quite effective, and there is little difference concerning the microcomputer situation between Japan and the US. The only difference relates to the varieties of types of hardware.

It is appropriate to introduce another article in the journal "Statistics", the author of which is Prof. Tarumi, a JCS member, in which his statistical microcomputer software development experience is described. The widely distributed "Statistical Analysis Handbook for Microcomputers"

consists of five volumes, more than four hundred pages in each volume, and contains BASIC program listings. Floppy disks of source programs have been also widely distributed by the publisher. The first volume was published in 1984 at the time of the appearance of the PC-8801 NEC personal computer which had 64 kilobyte main memory, black and white display, 5-inch floppy disk unit with a capacity of 320 kilobyte records, and with incomplete Chinese character handling.

It was really a big job to harmonize a multi-byte character set with the microcomputer world. In the 1980s multi-byte character systems were standardized which established the worldwide character handling system now existing in eastern countries.

Much effort of Tarumi's group was invested in porting to a number of hardware configurations other than the original machine, such as NEC PC-9801, IBM-5550, Fujitsu FM-11, and FM-16-beta.

In these six years NEC announced many kinds of new products and in the first three years changes were related to floppy disk media and in the last three years alterations related to the central processor unit. It is reported by the author that in these modifications there has been no revision of programs for these NEC hardware configurations. Recently there appeared a new catalogue book containing 1,800 application programs related to the PC-9800 microcomputer series, which should ensure it an important place in the Japanese market.

In contrast to the microcomputer situation in the US, the Japanese software houses are faced with varieties of hardware products. The most troublesome part naturally relates to graphics operations and in that area US microcomputer software houses, it seems to us, could devote an effort to enhancement of their products' capability and usability. Our concerns would relate more to the various hardware systems.

My friend introduced to me another monthly journal "Factory Management" published last May, with a special issue on how to make use of personal computers to alter the factory. In the introduction of the issue an editor remarks that there are many prominent software products in such fields as spreadsheet handling, database systems, accounting systems and statistical analysis, and others. Then specific product names are cited such as Lotus 1-2-3, dBASE III as usual and finally the statistical product JUSE-QCAS, which is recommended to be used for the preparation of the representation sheets in quality circle meetings and also to attain quality stability by checking the inspection data, monthly, using statistics.

JUSE-QCAS is devoted specifically to statistical quality control activities and has had many users in Japan since 1985. The institution which manufactures and distributes the software product also provides an introductory seminar at a charge. The fact that the number of participants at these seminars reaches one tenth of the number of sales is quite striking.

Seminar staff say that in the early stage computer handling speed was very slow due to unfamiliarity with the keyboard but more recently faster response is not exceptional.

In that issue there is a report from a chemical products enterprise, where fifty copies of the JUSE-QCAS system were introduced for trend analysis of the amount of product and operational management in the company.

The report describes that there has been an education program for the system and already six hundred production line workers finished the course and they are utilizing the system where a number of one thousand three hundred people are on the job.

The workers who reluctantly engaged in manual statistical analysis and hand-writing of diagrams have welcomed the computing facility, getting quality circle activities more efficiently accomplished, the report tells us.

So it appears that the late-developing market for statistical software in Japan is beginning to grow.

A market for classification software products, on which I must report, also starts to spread with some delay. This area needs to avail of more mature basic statistical software products and some JCS members are working to prepare a series of classification program products. Information on this will be provided in the near future, hopefully.

Given the spread of information exchange technology the features described here should not be restricted to the national scene. It seems to me that we have been forced to acknowledge big differences between countries but also we naturally come to the conclusion that there are very common situations among the countries. With the recognition of this common situation comes the role and value of the IFCS.

Keiji Yajima, Secretary of JCS

Jugoslovenska sekcija za klasifikacije (JSK) Yugoslav Classification Society

The JSK organizes its annual meetings in Mostar (Republic of Bosnia and Herzegovina). The local organizer is Dr. Emir Veledar. Sponsorship is by Federal Statistical Office (Institute of Statistics) and the Faculty of Economics in Mostar. Presented papers are published in the proceedings (the publisher is Institute of Statistics of the Federal Statistical Office, Kneza Milosa 20, Beograd) after the meetings.

The Proceedings of the Fourth JSK Annual Meeting (Beograd 1991, 192 pages) held in Mostar on 18-19 May 1990 includes the following papers:

D. Acketa, N. Caric, Z. Horvat: "On classification of work instruments and products according to their classificational numbers"
S. Bodjanova: "Some relations between hard and fuzzy classification of multivariate statistical observations"
S. Bogosavljevic: "Multivariate stratification as a clustering problem"
V. Dragovic: "The possible use of some classification models in a dynamic analysis of statistical sets"
A. Hosek, K. Momirovic: "Taxonomic analysis of cognitive and conative characteristics in children of 11 years"
A. Kramberger: "Apriori and analytic classification of occupations"
B. Ivanovic: "Discriminative procedures in cluster analysis"
B. Mirkin: "A clustering model based on linear data reduction model"
K. Momirovic: "On the measures of distance and similarity in antiimage space"
K. Momirovic, M. Gredelj: "On the measures of distance and similarity of objects in mirror image and antimirror image space"
S. Mrkic, B. Vujnovic: "Classification of Yugoslav communes according to socio-economic and criminality structure"
V. Pompe-Kirn, A. Ferligoj: "Typology of cancer sites"
S. Stevic: "Application of I-distance method in bussiness efficiency area"
E. Veledar: "Contribution of the fuzzy set theory to clustering"
D. Vukmirovic: "The influence of social background on the structure of occupations of students' homes in Belgrade"

The Fifth JSK Annual Meeting was held in Mostar on 24-25 May 1991. Despite the political turmoil in Yugoslavia the participants from all parts of Yugoslavia came together and we had a very nice meeting. The following papers were presented:

S. Bogosavljevic, B. Vujnovic: "Typology of Yugoslav communes according to the national structure"
A. Ferligoj, V. Batagelj, P. Doreian: "Clustering methods for network analysis"
A. Hosek: "Distribution of crime in Yugoslavia estimated by the method of polar taxons"
Z. Jevtic, D. Djuric: "Classification of countries by relationship of productivity and export rates"
L. Kozlina, S. Bogosavljevic: "The analysis of student structure in Yugoslavia"
S. Kropivnik: "Elections '90 - clustering of constituencies in Slovenia according to elections' outcomes"
K. Momirovic, V. Dobric: "Some procedures for the quantification of a set of categorical variates"
T. Pogany: "Metrics, distance matrices; similarity measures"
M. Radovanovic: "On development of the classification science"
E. Veledar: "The similarity of a domestic consumption for the group of European countries"
D. Vukmirovic: "On the use of expert systems for solving clustering problems"

Anuska Ferligoj, President, JSK
(anuska.ferligoj%uni-lj.ac.mail.yu@relay.cs.net)

Vereniging voor Ordinatie en Classificatie (Dutch-Belgian Classification Society)

The 1991 spring meeting of the VOC was held at the University of Utrecht on April 24. The following papers were read:

Dr. Peter van der Heijden (University of Utrecht): Latent budget analysis of the Books of Plato.
Dr. Michel Wedel (CIVO-TNO, Zeist): A clusterwise regression method for simultaneous fuzzy market structuring and benefit segmentation.
Dr. Rian van Blokland-Vogelesang (University of Leiden): Maximum Likelihood and Minimum Inversion methods in unimodal rank order models and Unfolding.
Drs. Berrie Zielman (University of Leiden): An extension of the INDSCAL model for the analysis of asymmetric data.
Dr. Eeke Van der Burg: An analysis of vegetable soups using generalized canonical correlation.

Following the scientific part, the annual business meeting was held. Elections were due for the Board. The following persons were re-elected: Willem Heiser (President), Georgios Martakis (Secretary), and Jaap de Gruijter (Treasurer). Geert de Soete resigned, and Caspar Looman was elected as a new member. Paul de Boeck and Jacqueline Meulman continued their membership of the Board. Peter van der Heijden resigned from the auditing committee; Niels Veldhuyzen was elected as member, and Onno van Tongeren continued his membership.

The Fall meeting of the VOC will be held on 23 October 1991, at the Erasmus University in Rotterdam.

Gesellschaft für Klassifikation (German Classification Society)

Workshop on Sequence Analysis - University of Bielefeld, June 24-28, 1991.

Within the framework of the current research year on "Combinatorics and its Applications", the Center for Interdisciplinary Research (ZiF), University of Bielefeld, hosted a research group on "Combinatorial Problems in Molecular Biology", directed by Prof. A. Dress (Bielefeld) and Prof. B. Schuster (Vienna). During that time, the workshop took place. Presentations were as follows.

J.S. Nicolis: "Multifractal attractors mediating biological and cognitive categorization".

W. Ebeling: "Entropy analysis of sequences with long range order".

M. Hasegawa: "Probability of coincidence of intron positions by chance, and the origin of introns".

D.G. Arques and C.F. Michel: "A model of DNA sequence evolution based on oligonucleotides".

P.O. Degens: "Estimation of ultrametrics or additive trees by genetic distance data".

M.A. Steel: "Subtree-based approaches to tree reconstruction".

D. Penny: "The fast Hadamard transform and some general properties of tree reconstruction methods".

H.J. Bandelt: "Towards a synthesis of sequence methods and distance methods for reconstructing phylogenetic trees".

A. v.Haeseler: "Computer methods for locating kinetoplastid cryptogenes".

N. Blum: "On locally optimal alignments in genetic sequences".

J. Cavender: "Invariants of phylogenies: a review".

P. Stadler: "Molecular quasi species - concepts and applications".

N. Saitou: "Methods for constructing molecular phylogenetic trees".

A. Dress: "The statistical geometry of sequence space".

F. Melchers: "The ontogenesis of the immune system".

A.V. Valil'ev: "Using transputers in sequence analysis".

B. Lausen: "Fitting positional variation and genetic distance to sequence data".

D. Sankoff: "Genome rearrangements in mitochondrial evolution".

A. Dress: "Split decomposition - a new technique for cluster analysis admitting some overlaps, and its application to virology".

16th Annual Conference of the Gesellschaft für Klassifikation, University of Dortmund, 1-3 April 1992: Information and Classification: Concepts - Methods - Applications.

The scientific programme includes the following tentative topics: Theoretical and conceptual basis; Mathematical and statistical methods of classification; Exploratory data analysis; Information systems and knowledge processing; Graphical representation of information and data; Software for classification and data analysis; Indexing, classification and information processing in libraries and archives; Classification and information in medicine and biology; Data analytical problems in environmental sciences; Analysis of DNA- and sequence data; Data analysis and classification in economics; Typology, classification and data analysis in psychology and empirical social sciences; Information and data analysis in law and criminology; Classification and data analysis in linguistics; Data analysis in archeology and museology; Classification in engineering, administration and official statistics.

Several workshops and tutorials are scheduled, sometimes in cooperation with other societies, e.g.:
- vocational course for librarians;
- workshops: "Classification problems in medicine"; "Sequencing and the analysis of sequence data"; "Data analysis in archeology";
- tutorials: "Data analyis and cluster analysis"; "Phylogenetic analyis of molecular data"; "Classification systems for libraries and documentation".

The conference languages are German and English. Deadline for the submission of contributed papers: receipt of abstract (one page, with address and references) by November 30, 1991. Information: Berthold Lausen, Fachbereich Statistik, Universität Dortmund, Vogelpothsweg, Postfach 500 500, W-4600 Dortmund 50, FRG, phone (0231) 755-3113, -3180, Fax -3918, Email ust027@ddohrz11.bitnet.

B. Lausen and P.O. Degens
Med. Inst. für Umwelthygiene, Düsseldorf
Tel: 0211/3389-286

Conference Announcements

☛ Parallel Problem Solving from Nature: Applications in Statistics and Economics. Interdisciplinary Project Center for Supercomputing, ETH Zürich, Switzerland. December 9-10, 1991.

☛ Astronomy from Large Databases II. Strasbourg Observatory, Strasbourg, France. September 14-16, 1992.

Further information: F. Murtagh (address: page 3).

Classification Society of North America

The Classification Society of North America celebrated its 23rd year of service by holding its annual meeting jointly with the Psychometric Society at the Cook campus of Rutgers University, new Brunswick, New Jersey on June 21-23, 1990. The local host was Professor Phipps Arabie, Editor of Journal of Classification and president of the Psychometric Society. The list of registrants contained about 170 names from the two societies.

The size of the technical program required parallel session for the two societies and within societies. Following is a brief summary of the CSNA program.

Six invited addresses were presented:
David Johnson, AT&T Bell Laboratories, spoke on local optimization and the traveling salesman problem. W. J. Krzanowski, University of Exeter, lectured on the location model for mixtures of discrete and continuous random variables. Bruno Leclerc, EHESS Paris, talked on the residuation model for ordinal construction of dissimilarities and other valued objects. Trevor Hastie, AT&T Bell Laboratories, reviewed generalized additive models. Marc Lipman, Office of Naval Research, explained the ONR program in discrete structured classif-ication. B.G. Mirkin, Central Economics-Mathematics Institute, Moscow outlined clustering and multi-dimensional scaling in the USSR. In addition Phipps Arabie gave the Psychometric Society Presidential Address entitled "Was Euclid an Unnecessarily Sophisticated Psychologist?"

Sessions of contributed papers were held Friday through Sunday, with 42 regular papers were presented in 11 sessions as part of CSNA's program, each being allotted 30 minutes. Only Abstracts were published. Of the 11 sessions, four were organized symposia: Ordinal and Graph—Theoretic Models for Cluster analysis (M. Janowitz); New Developments in Classification and Clustering (H. Bozdogan); Optimization in Clustering (P. Hansen and B. Jaumbard); Hierarchical Classes in Self Perception (S. Rosenberg).

The seven sessions of regular papers (with session chairs) were: Clustering I (A. Okada); Molecular and Chromosomal Analysis (D. Burdick); Statistical Analysis and Spatial Data (J. Muelman); MDS II (L. Jones); Data Analysis (C. Lewis); MDS III (W. Kuhfeld); Clustering II (W. Heiser).

The opportunity for informal discussion with some of the best people in classification is the hallmark of CSNA annual meetings.

The regular meeting of the Board of Directors of CSNA was held on Thursday evening, June 13, 1991, the evening before the meeting began. The Board expressed its gratitude to Sharon Weinberg and Pascal Rousseau whose terms end on December 31, 1991. The annual Business Meeting, open to all members of CSNA, was held late Friday afternoon, June 14. This informational meeting provided an informal forum for exchanging ideas between members and the Board of Directors.

Call for Papers: 1992 Annual Meeting, Classification Society of North America, June 11-13, 1992.

The 1992 annual meeting of CSNA will be held at the Kellogg Center on the campus of Michigan State University, East Lansing, Michigan on Friday, June 12 and Saturday, June 13, 1992. The meeting will be preceded by an optional Short Course on Classification and Clustering on Thursday, June 11.

Local Arrangements: Richard Dubes, Department of Computer Science, Michigan State University, East Lansing, MI 48824-1027, USA. Tel: (517)-355-5240, email: dubes@cps.msu.edu.

The scientific meeting will include plenary talks by invited speakers, symposia organized by CSNA members, and contributed papers. Persons wishing to present a talk should send a title and abstract of not more than 200 words on a single page to:

Program Chair: Wayne DeSarbo, Departments of Marketing and Statistics, School of Business, University of Michigan, Ann Arbor, MI 48109-1234, USA. Tel: (313)-936-2084. Email: Wayne_S._DeSarbo@um.cc.umich.edu, BITNET: gegb@umichum.

Proposals for symposia should also be sent to the Program Chair. All submissions must include the name, complete mailing address, telephone number, and institutional affiliation of each author. Electronic mail addresses will also be appreciated. In preparing abstracts, center name, address, and affiliation on the page. Identify the presenter in the case of multiple authors. List a few keywords on the bottom line to help classify the paper. Abstracts will be reproduced for conference attendees but no other proceedings will be published.

If possible, each submission should be classified according to one of the following areas: Applications, Classification, Clustering, Exploratory Data Analysis, Factor Analysis, Item Response Analysis and Testing, Latent Class Models, Multidimensional Scaling, Numerical Taxonomy, Pattern Recognition, Statistical Methods, or other (please specify). Please also indicate visual aids needed in addition to overhead projectors. Papers will be allotted 20 minutes, including discussion. Longer times must be justified and will be granted at the discretion of the Program Chair.

Deadline for submission of Abstracts: March 15, 1992.

The Short Course on June 11, 1992 is designed to provide background for people desiring a broad exposure to the terms and methodology of Classification and Clustering. The instructors will be Richard Dubes, Pierre Legendre (University of Montreal) and Glenn Milligan (Ohio State University).

Michigan State University is located on a park-like campus of over 5,200 acres that is particularly beautiful in spring. East Lansing is served by several major airlines and is a reasonable automobile trip from most of the Midwest.

Richard C. Dubes, President IFCS,
Computer Science Dept., Michigan State University.

SCRIBE-Tutor: a Computer-Aided Teaching and Learning System in Exploratory and Multivariate Data Analysis

M. Jambu*, F. Guillot and J.M. Caisso**

We present a computer-aided learning system in multivariate analysis. It has been called SCRIBE-Tutor due to the role scribes played in statistics and writing in ancient Egypt.

Contents of SCRIBE-Tutor

SCRIBE-Tutor is based on courses and lectures given over ten years in Engineering Schools and Universities, and in adult or evening education classes. SCRIBE-Tutor contains 11 chapters, plus a commented bibliography.

Chapter 1: Individuals-variables data sets. Definitions and descriptions of individuals, variables and data sets that appear in multivariate data analysis. Presentation of the types of data sets associated with quantitative, qualitative and chronological data, and their recoded forms.

Chapter 2: Univariate data exploration. For quantitative variables: presentation of the statistical characteristics (arithmetic mean, median, range, quartiles, etc.) and dispersion box-plots, and other graphical tools. For qualitative variables: frequency classes and pie charts, bar-diagrams, profile-diagrams, etc.

Chapter 3: Bivariate data exploration. For quantitative variables: basic regression method, dispersion scatter plots, correlation data sets, etc. For qualitative variables: chi-squared and contingency data sets, numerical and graphical tools, etc.

Chapter 4: Elementary multivariate data analysis. Presentation of exploratory data analysis mathematical transformation of the data sets. For quantitative variables: profiles and sun-ray plots, dispersion multidiagrams, etc. For qualtitative variables: multiple contingency data sets and graphics.

Chapter 5: Factor analysis of individuals-variables data sets. Principles of factor analysis, without use of mathematical formulas, only using geometrical representations and imagination.

Chapter 6: Principal components analysis. Presentation of the basic model, with an example allowing a range of interpretation rules to be learnt.

Chapter 7: 2-D correspondence analysis. Basic model applied to correspondence data sets with a realistic and detailed example. Thorough study of the field of application of correspondence analysis.

Chapter 8: N-D correspondence analysis. The most fruitful application of 2-D correspondence analysis is presented with an example. Overview of the field of application (including survey and database analysis).

Chapter 9: Classification of individuals-variables data sets. Principles and algorithms for hierarchical clustering and partitioning, without mathematics. Basics of interpretation, using examples.

Chapter 10: Analysis and classification of proximities data sets. Presentation of proximity data sets and application of factor analysis and hierarchical clustering.

Chapter 11: Data exploration strategies. Methodological background of data analysis, plus six ways to explore data sets according to their type and the desired objectives. How to combine the different parts of the course in order to process a particular data set.

Chapter 12: Bibliography. Three levels of knowledge (beginner, junior, specialist) assumed; valuable comments on books.

Users of SCRIBE-Tutor

SCRIBE-Tutor was developed for users of multivariate data analysis: students, teachers, statisticians, scientists in any field, engineers, managers in many fields where there is data to be analyzed. No previous specific knowledge is required to understand the basics of SCRIBE-Tutor, - common sense should suffice. Material in SCRIBE-Tutor is relevant for applications in various fields such as physics, chemistry, medicine, psychometrics, sociometrics, humanities, education sciences, earth sciences, biology, astronomy, engineering, economics, marketing, management, and virtually any field where the data could be analyzed with a view towards improving basic knowledge or decision-making.

SCRIBE-Tutor Hardware Requirements and References

SCRIBE-Tutor is a 700-page multi-access electronic book. The main screen contains an iconographic menu bar with pull-down menus, a graphics left page, and a textual right page. SCRIBE-Tutor works on any IBM-compatible PC with 640 KB RAM, VGA-compatible monitor and 20 MB hard disk. A demonstration diskette is available on request. Please contact PRODIDACT.

SCRIBE-Tutor is available in English and French. For any request, please directly contact PRODIDACT.

Jambu, M. (1991): *Exploratory and Multivariate Data Analysis*, Academic Press (in press).

* CNET/PAA, 38/40 rue du Général Leclerc, 92131 Issy les Moulineaux, France.
** PRODIDACT, Espace Wagner, ZI Aix en Provence, 13763 Les Milles Cedex, France.