AUSTRIAN JOURNAL OF STATISTICS
VOLUME 31 (2002), NUMBER 2&3, 241-244

On the Future of Data Analysis

Reinhard Viertl
Department of Statistics and Probability Theory
Vienna University of Technology

Zusammenfassung: Der Beitrag gibt eine Diskussion über die Zukunft der Datenanalyse wieder, die am letzten Tag der Tagung IDA 2000 statt fand. Außerdem sind die Initiatoren eines geplanten Netzwerks für Kooperation und Weiterbildung auf dem Gebiet der Unsicherheitsanalyse angeführt.

Abstract: This contribution is based on a discussion on the future of data analysis which took place on the last day of the symposium IDA 2000. By different discussants the main problems of current data analysis as well as future developments were discussed. Moreover a network for cooperation and continuing education in the field of uncertainty analysis was initiated and the list of scientists who initiated this is given.

1 Introduction

The last session of the IDA 2000 Conference was a discussion on the future of data analysis which took place on Friday, 22 September 2000. Eleven discussants expressed their viewpoints on future developments of data analysis (DA), ranging from initial data analysis to foundational questions concerning the quality of data, its quantitative description, and its consequences for data analysis methods.

Since the discussion took place about 3 months before the start of the 3^{rd} millennium, it is also an outlook to the development of DA in the beginning 21^{st} century.

2 Aspects of Data Analysis

Several aspects of DA were addressed during the discussion. The most important (in the opinion of the discussants) were the following:

A main question is the usefulness of DA results. An important comment was that 90 percent of all data analysis is done by non-statisticians. Concerning mathematical models only few of them are useful when looking at the problems from outside of the mathematical community.

Another aspect was that statistical methods are not the only ones for data analysis. Important is the relationship between different approaches of statistics, and fuzzy methodology.

E. Kling from SAS pointed out that many internet publications are not checked for accuracy. This is a problem for the credibility of data analysis methods distributed in the internet.

Statisticians should produce results for the private sector, where the concern about money is important. Success stories of data analysis methods should be distributed by the www.

Another comment was: Is statistical testing really needed in data analysis? Responding to this Prof. O. Hryniewicz pointed out that hypotheses in the social sciences should be formulated as fuzzy hypotheses.

2.1 Data Quality and Its Consequences

At the beginning of a DA process the kind and quality of data have to be considered. What is the meaning of bad data? An approach to data of different quality is to describe data imprecision (fuzziness) by fuzzy numbers and fuzzy vectors. When asking experts the answers are mostly fuzzy in different ways.

Another important topic is the specification of data. Interval mathematics can be helpful but a more general and useful approach is using fuzzy numbers.

Data quality problems are present in sustainability studies. For details the upcoming UNESCO Encyclopedia of Life Support Systems which will appear online, is a good reference.

Starting from data quality different uncertainties have to be combined in high quality data analysis.

2.2 Data Analysis and Decision Support

The main task of data analysis is decision support. This will continue to be also in the future. What is the type of information which is useful for decision makers? Statistical tests should not be used as automatic decision procedures.

Prof. M. Nikulin expressed the necessity for the human society to foster statistical studies for the social sciences in the future. There is a need for mathematical models for the behaviour of people on a global scale. This is related to questions of quality of life, also in connection with medical treatment. For these problems completely new models are needed. In France these topics are starting to be developed.

Also new analysis methods for high quality products concerning life times are necessary. These are for engineering applications as well as for human reliability which are both related to fuzzy data.

3 The Future of Data Analysis

In the 21^{st} century the ideas from fuzzy theory and Bayesian statistics will jointly be valuable for high quality data analysis and decision support. A main problem is statistical testing based on fuzzy evidence.

Professor H. Bandemer pointed out that there are different levels of uncertainty and expressed his optimistic view that similar to the development in mechanics a new generalized theory of data analysis will come up in the near future.

In the future statisticians should concentrate more on the analysis of real data. Otherwise statistics is in danger of loosing contact to applications. Statistics should develop into real data science. This rules out the exclusiveness of some theories, and provokes the analysis of expert knowledge and its quantitative description.

A goal for the future is merging different uncertainty models.

The following scientists participated active in the discussion:

H. Bandemer, Halle a.d. Saale, Germany
R. Dutter, Wien, Austria
K. Felsenstein, Wien, Austria
N. Fickel, Nürnberg, Germany
P. Filzmoser, Wien, Austria
O. Hryniewicz, Warsaw, Poland
S. Human, Durban, South Africa
E. Kling, SAS Systems, Israel
M. Nikulin, Bordeaux, France
J. Pilz, Klagenfurt, Austria
R. Viertl, Wien, Austria

4 A Network for Cooperation and Education

At the end of the discussion Prof. J. Pilz proposed to create a network for cooperation in modelling uncertainty and for training uncertainty analysts. The following scientists expressed their interest to participate in such a network:

O. Alexandrova, Ozersk, Russia
M. Bolgov, Moscow, Russia
G. Brunet, Niort, France
K. Felsenstein, Wien, Austria
D. Grzegorzewski, Warsaw, Poland
O. Hryniewicz, Warsaw, Poland
H. Kutterer, München, Germany
W. Näther, Freiberg, Germany
M. Oberguggenberger, Innsbruck, Austria
J. Pilz, Klagenfurt, Austria
R. Viertl, Wien, Austria

The coordinator is Professor Jürgen Pilz, whose address is

University of Klagenfurt
Division of Applied Statistics
Universitätsstr. 65-67
A-9020 Klagenfurt, Austria
E-mail: juergen.pilz@uni.klu.ac.at
Tel.: +43-463-2700-3113

References

H. Bandemer. *Ratschläge zum mathematischen Umgang mit Ungewissheit - Reasonable Computing*, Teubner, Stuttgart, 1997.

H. Bandemer and W. Näther. *Fuzzy Data Analysis*, Kluwer, Dordrecht, 1992.

M. Berthold and D. Hand (Eds.). *Intelligent Data Analysis - An Introduction*, Springer, Berlin, 1999.

B. Efron. Statistics in the 20th Century and the 21st. Festschrift 50 Jahre Österreichische Statistische Gesellschaft, Austrian Statistical Society, Wien, 2002.

R. Viertl. *Statistical Methods for Non-Precise Data*, CRC Press, Boca Raton (Florida), 1996.

R. Viertl. Statistical Inference for Non-Precise Data. In *Encyclopedia of Life Support Systems (EOLSS)*, Eolss Publishers, Oxford (to appear on-line), http://www.eolss.net

Author's address:

Prof. Reinhard Viertl
Institut für Statistik und Wahrscheinlichkeitstheorie
Technische Universität Wien
Wiedner Hauptstr. 8/107
A-1040 Wien
Austria

E-mail: R.Viertl@tuwien.ac.at