UMEÅ UNIVERSITY
Institute of Information Processing – ADB

Postal address:	Tel (direct dialing):	Telefax:	Email (Internet):
S-901 87 UMEÅ (Sweden)	+46 90 166030	+46 90 166126(166688)	kivanov@cs.umu.se

Professor KRISTO IVANOV
Chairman, Administrative Data Processing

STATISTICAL DATA PROCESSING PERSPECTIVES

Draft, 2 September 1989

Introduction

The computer can be considered, among other things, as a statistical machine. This is so in the sense that a computer is often used for organizing data in order to prepare for and in order to draw inferences and make decisions about the "state" of a nation, of an organization, or of a matter of interest for a social group. In a coming era of graphic visual data processing and computer graphics it is also appropriate to consider statistics, together with analytic geometry, to be the mother-science of visualization or visual simulation of phenomena described in terms of scientific data and sense impressions (Fallati, 1843; Tarter, & Kronmal, 1976; Thomas, 1972).

Statistics is nowadays most often associated with applications of the theory of probability and so called stochastic models. It may, however, be remarked that until the end of the past century statistics often represented the early systems analysis and memory or database of organizations and work processes. This included both social and natural processes which were too complex to be be grasped by any particular theory (Meitzen, 1891; Sigwart, 1895). It is therefore natural to see statistics as one producer of later specializations such as private and national accounting (Littleton, 1981, for the private account) contributing to what in time would become business administration and political economy, sociology (Lottin, 1912, presenting the thought of Belgian astronomer and sociologist L.A.J. Quételet, born 1796; Porter, 1986), and geography melting today with so called regional economics dealing with complex computer databases and simulations of social-spatial processes and the kind of cultural criticism represented by currents in cultural geography (Olsson, 1980; Olsson, 1988a; Olsson, 1988b; Wallin, 1980; Wallin, 1986), administrative data processing (Ivanov, 1976a; Ivanov, 1976b; Ivanov, 1976c; Ivanov, 1977; Ivanov, 1979; Nilsson, 1987) and in the discipline of statistics itself (Sjöström, 1980; Sjöström, 1988).

Debates on nature and function of statistics

In contrast with the strong controversies which have marked the history of statistics (Johannisson, 1988; Porter, 1986), and their relation to anthropological and political roots (David, 1962; Sheynin, 1977), today's scene is marked by a surprisingly lame, if any, discussion. In part this may be conditioned by the fact that application of statistics in natural science apparently has been marked by far less controversies or, at least, so it can be presented (Kac, 1974, is an example of such a presentation). Because of the particular development of academic and applied statistics in the last decades, the smaller controversies which apparently have grown up have done so in terms of a technical language. This is so even in those rare cases when the whole discussion has not been clothed in the formal languages of mathematical symbols and formulas (Eisenhart, 1947a; Eisenhart, 1947b; Eisenhart, 1948; Menges, 1973; Tukey, 1960; Tukey, 1969; Tukey, 1975; Wold, 1957a; Wold, 1957b). Several discussions about the role and responsibility of the statistical consultant (Eisenhart, 1947b; Tukey, 1975, are casual

examples) are obviously relevant for the discussion of the future role of information systems analysts, operations resarchers (Stevens, 1982) and, lately, for so called knowledge engineers in the era of cooperative or co-constructive continued systems development (Ehn, 1988; Forsgren, 1988; Whitaker, & Östberg, 1988). The issues of analysis and interpretations of data obviously concern also so called users, decision makers, clients, stakeholders, systems activists or systems owners in the context of modern databases and information systems or expert systems.

The trend towards formalization and mathematization of statistics (Porter, 1986, touches upon "the mathematics of statistics" on pp. 233ff) in its extreme form seems to have followed the trends at the beginning of the century as represented mainly by the formalist school of J. Neyman and E.S. Pearson (Neyman, 1952). That work was surveyed in a review of selections of their papers (Dempster, 1968). At that early stage of the process of formalization it was still possible to note a certain philosophical depth in the disciplinary discussions concerning the concept of "state", the relation between mathematics versus reality and vague sensations, etc. (Neyman, 1960).

Later developments and reaction to formalization

A trend towards formalization and mathematization in methodological sciences lead in the last decades to the creation of theories of fuzzy sets with an original blend of statistics and logic (Orci, 1983, exemplify early respectively late stages of the process; Zadeh, 1965) which, at the level of basic issues of theory of science, apparently corresponds to the uprising of mathematical theories of evidence or logical theories of truth which are remarkably isolated from experimental sensuous reality (Kripke, 1975, would possibly be a counterpart in the field of logic; Shafer, 1976).

At the same time it is possible to identify a kind of reactive movement, conservative or reactionary in the original sense of the word, when compared with the modern formalizing tendencies. Such reaction seems to arise mainly from some practitioners which made contributions to the applications of statistics, e.g. in industry. One example is the work dealing with the problems of industrial quality control (Shewhart, 1939). Such work was later taken up and integrated in the efforts by practicing applied statisticians, later called operations researchers, contributing to the USA scientific war effort during world war II. The result was the development of a theory of experimental inference in the spirit of American pragmatism and experimental idealism where statistics was beginning to get re-integrated with social science and philosophy (Churchman, 1948). This amounted to a somewhat unconscious rebuilding of the classical heritage of 19th century's statistics (Shewhart, 1939).

This reactionary trend and its development of statistics into a social systems science (Ackoff, & Emery, 1972; Churchman, 1961; Churchman, 1968; Churchman, 1971) echoed the work of other statisticians who were implicitly inferring from their professional experiences the need of some kind of systems thinking: "The really critical experiment is rare and...it is frequently necessary to combine the results of numbers of experiments dealing with the same issue in order to form a satisfactory picture of the true situation" (Yates, 1951, p.33). The same kind of basic difficulties which were addressed by statistic's development towards a social systems theory for information processing were also the difficulties which appeared in the context of psychological research. This happened in the form of controversies about e.g. the interpretation and treatment of the so called null hypothesis, the treatment of the single case, clinical versus statistical prediction, etc. Relevant references to these issues are mentioned by us elsewhere in connection with the study of the interface between logic and psychology for data processing.

On the Scandinavian scene these developments lead up to formalizations of systems structures in terms of statistical data bases based on an empirical approach (Sundgren, 1975), well consonant with what has been identified as the idea of logical empiricism and statistical empiricism paving the way further to the formulation of a whole research program (Sundgren, 1973; Sundgren, 1975; Sundgren, 1982; Sundgren, Wallgren, & Wallgren, 1984). It was in turn criticized on the basis of epistemological and practical considerations (Ivanov, 1976a; Ivanov, 1976b; Ivanov, 1976c; Ivanov, 1977; Ivanov, 1979). Such criticism inspired and combined with others' coming from the insider group of professional statisticians (Rennermalm, 1981; Sjöström, 1980; Sjöström, 1983; Sjöström, 1984a; Sjöström, 1984b; Sjöström, 1988). This corresponded in part to similar critical currents in the USA (Dunn, 1974; Ivanov, 1976a, commenting Dunn; Ivanov, 1976c; Mitroff, Mason, & Barabba, 1983), and in the philosophical community (Molander, 1987).

To some extent the critique of statistical information systems echoed earlier problematizations of the technical body of statistical theory and practice (Strauch, 1970) which, by the way can be seen as concerning the process of research and publication themselves - i.e. building up of scientific databases and library collections (Branscomb, 1968; Goudsmit, 1966; MacDonald, 1972; Maddox, 1963; Walster, & Cleary, 1970).

The defective understanding and use of statistics in the behavioral and social sciences has, of course, already had serious consequences in the realm of economics (Grassman, 1985; Morgenstern, 1963; Ross, 1968). Such documented consequences in a field which is disciplinarily so close to organizational information systems may be expected to be the repository of valuable experiences that wait to be interpreted for their possible relevance to computer and information science. The same can be said about experiences statistics applied in the realm of engineering materials in the context of construction and operation of nuclear power plants.(Östberg, 1981; Östberg, 1982).

Another area that can be studied for transfer of experiences is about applications of statistics to non-formalized material, typical also for administrative-organizational areas, as in text analysis and analysis of verbal reporting in its various forms (George, 1959; Janis, 1958; Lasswell, 1938; Rokkan, Verba, Viet, & Almasy, 1969; Woodward, 1934). Many of such early and conventional issues of "statistical information systems" display the advantage of containing reflections and explicit assumptions which stimulate an inquiry of corresponding methodological assumptions of today's design of computerized information systems. For the very same reason it can be fruitful to study parts of certain classics in the historical field of statistics, such as "political arithmetics" and probability (Keynes, 1952, esp. chap. 26, "The applications of probability to conduct"; Lottin, 1912; Meitzen, 1891; Petty, & (Hull, 1899), as well as controversies such as between Droysen and Buckle, and other as mentioned by various modern authors (Liedman, 1983; Porter, 1986).

Epilogue

Seen in this light, it is obvious that statistics represents at least 200 years of experience and theorizing about structuring and visualization of observational data with the purpose of drawing inferences from "databases" for supporting administrative and individual decisions, in close contact with what came to be known as political economics, political science and law, sociology and geography, completed later with natural science, especially astronomy and biology. Methods for systems development including improvement of cooperative work environment with the help of high technology should profit of an anchoring in such historical experiences and insights

beyond the pure reliance on formal science, phenomenology, modern liberal or socialistic ideas of cooperation, participation, negotiation or conversation. Bridges which already have been established over to systems and information science (Ivanov, 1976a; Ivanov, 1976b; Ivanov, 1976c; Ivanov, 1977; Ivanov, 1979) could form a platform for future research.

References

Ackoff, R. L., & Emery, F. E. (1972). *On purposeful systems: An interdisciplinary analysis of individual and social behavior as a system of purposeful events* . Chicago: Aldine-Atherton.

Branscomb, L. M. (1968). The misinformation explosion: Is the literature worth reviewing? *Scientific Research,* (27 May),

Churchman, C. W. (1948). *Theory of experimental inference* . New York: Macmillan.

Churchman, C. W. (1961). *Prediction and optimal decision: Philosophical issues of a science of values* . Englewood Cliffs: Prentice-Hall.

Churchman, C. W. (1968). *The systems approach* . New York: Delta. (Page references are to the 2nd ed., 1979.)

Churchman, C. W. (1971). *The design of inquiring systems: Basic principles of systems and organization* . New York: Basic Books.

David, F. N. (1962). *Games, gods and gambling: The origins and history of probability and statistical ideas from the earliest times to the Newtonian era* . London: Charles Griffin.

Dempster, A. P. (1968). Crosscurrents in statistics. *Science, 160*(10 May), 661-663.

Dunn, E. S., Jr. (1974). *Social information processing and statistical systems: Change and reform* . New York: Wiley.

Ehn, P. (1988). *Work-oriented design of computer artifacts. (Doctoral diss.)* . Umeå-Stockholm: University of Umeå, Arbetslivscentrum and Almqvist & Wiksell International.

Eisenhart, C. (1947a). The assumptions underlying the analysis of variance. *Biometrics,* (March), 1-21.

Eisenhart, C. (1947b). The role of a statistical consultant in a research organization. *Proc. of the Int. Statistical Conference, 1947, 3*, 308-313.

Eisenhart, C. (1948). Statistics, the physical sciences and engineering. *Am. Statistician, 2*(4, August),

Fallati, J. (1843). *Einleitung in die Wissenschaft der Statistik* . Tübingen:

Forsgren, O. (1988). *Samskapande datortillämpningar [Constructive computer applications]* (Doctoral diss., Report UMADP-RRIPCS-3.88). University of Umeå, Inst. of Information Processing. (In Swedish. Summary in English.)

George, A. L. (1959). *Propaganda analysis: A study of inferences from nazi propaganda in world war II* . Evanston: Row, Peterson & Co.

Goudsmit, S. A. (1966). Is the literature worth retrieving? *Physics Today,* (September), 52-55.

Grassman, S. (1985). *Det plundrade folkhemmet* . Stockholm: Årstidernas Förlag.

Ivanov, K. (1976a). *Från statistisk kontroll till kontroll över statistiken: Systemisk redovisning av fel i undersökningar inklusive avvägningar mellan kvalitet och medborgerlig integritet* (Research report No.1976:9, ISSN 0347-2108). University of Stockholm, Dept of Statistics.

Ivanov, K. (1976b). *Statistik för datorer: Centraliseringen av svensk statistik, konsekvenser av en organisatorisk anpassning till datoriserad statistikproduktion*

(Research report No.1976:7, ISSN 0347-2108). University of Stockholm, Dept. of Statistics.

Ivanov, K. (1976c). *Statistiska informationssystem: Framväxten av en svensk skola om data och information och dess förhållande till en kunskap om sociala informationsprocesser* (Research report No.1976:8, ISSN 0347-2108). University of Stockholm, Dept. of Statistics.

Ivanov, K. (1977). Datoriserad statistik och statistiska system [Computerized statistics and statistical systems]. *Statistisk Tidskrift,* (5), 377-388. (Summary in English.)

Ivanov, K. (1979). Datorbaserad social kommunikation: Inskränkt till ett högnivåspråk för programmering av motsägelselösa samhällsbeskrivningar [Computer-based social communication: Reduced to a high-level language for programming of unambiguous descriptions of society]. *Statistisk Tidskrift,* (3), 173-187. (Summary in English, pp.237-238.)

Janis, I. L. (1958). The psychoanalytic interview as an observational method. In G. Lindzey (Ed.), *Assessment of human motives* (pp. 149-181). New York: Rinehart.

Johannisson, K. (1988). *Det mätbara samhället: Statistik och samhällsdröm i 1700-talets Europa* . Stockholm: Norstedts.

Kac, M. (1974). Statistics and its history. In M. Kac, G. C. Rota, & J. Schwartz T. (Ed.), *Discrete thoughts: Essays on mathematics, science, and philosophy* (pp. 37-48; see also 27-36). Boston: Birkhäuser.

Keynes, J. M. (1952). *A treatise on probability* . London: MacMillan. (First published 1921.)

Kripke, S. (1975). Outline of a theory of truth. *J. of Philosophy,* , 690-716.

Lasswell, H. D. (1938). A provisional classification of symbol data. *Psychiatry, 1*, 197-204.

Liedman, S. E. (1983). *Arbetfördelning, självmord och nytta: Några blad ur samhällvetenskapernas historia från Adam Smith till Milton Friedman* (Skriftserie 8, ISBN 91-7668-039-8, 2nd ed.)). Högskolan i Örebro.

Littleton, A. C. (1981). *Accounting evolution to 1900* . Alabama: University of Alabama Press. (Reprint of original publication, 1933.)

Lottin, J. (1912). *Quételet: Statisticien et sociologue* . Paris: Félix Alcan.

MacDonald, J. R. (1972). Are the data worth owning? *Science, 176*(4042),

Maddox, J. (1963). Is the literature worth keeping? *Rockefeller Institute Review,* (February), (Reprinted in *The Graduate Journal of the University of Texas*, Winter 1964.)

Meitzen, A. (1891). *History, theory and technique of statistics* . Philadelphia: The American Academy of Political and Social Science. (R.P.Falkner, Trans. Originally published as Geschichte, Theorie und Technik der Statistik, 2nd ed. Stuttgart, 1886.)

Menges, G. (1973). Inference and decision. In D. A. S. Fraser (Ed.), *Inference and decision* . Toronto: University Press of Canada.

Mitroff, I. I., Mason, R. O., & Barabba, V. P. (1983). *The 1980 census: Policymaking and turbulence* . Lexington, Mass: Lexington Books.

Molander, B. (1987). *Räkna fritt och tänka fritt* (Report from the project "Education for application of statistics"). Uppsala University, Dept. of Philosophy.

Morgenstern, O. (1963). *On the accuracy of economic observations* (2nd ed.). Princeton: Princeton University Press.

Neyman, J. (1952). *Lectures and conferences on mathematical statistics and probability.* . Washington, D.C.: U.S. Dept. of Agriculture. (Pages 1-66.)

Neyman, J. (1960). Indeterminism in science and new demands on statisticians. *J. of the American Statistical Association,* , 625ff.

Nilsson, T. (1987). *Kartor, informationssystem och geografiska informationssystem* (Report UMADP-WPIPCS 10.87). University of Umeå, Dept. of Information Processing.

Olsson, G. (1980). *Birds in egg: Eggs in bird* . London: Pion (Series on Research in Planning and Design).

Olsson, G. (1988a). Bjälken i ögat: Om tecknets kris och demokratins. *Expressen,* (August 26th),

Olsson, G. (1988b). The eye and the index finger: Bodily means to cultural meaning. In R. G. Golledge (Ed.), *A ground for common search* . Santa Barbara, Calif.: Santa Barbara Geographical Press.

Orci, I. P. (1983). *Contributions to automatic programming theory: A study in knowledge-based computing* (Doctoral diss., report TRITA-IBADB-1105, ISBN 91-85212-97-0). Royal Inst. of Technology and Univ. of Stockholm, Dept. of Information Processing and Computer Science.

Petty, W., & (Hull, C. H., Ed.). (1899). *The economic writings of Sir William Petty - Vol. 1* . Cambridge: Cambridge University Press.

Porter, T. M. (1986). *The rise of statistical thinking 1820-1900* . Princeton: Princeton University Press.

Rennermalm, B. (1981). Analys: Vad är det? *Statistisk Tidskrift,* (1), 5-10.

Rokkan, S., Verba, S., Viet, J., & Almasy, E. (1969). *Comparative survey analysis* . The Hague: Mouton.

Ross, A. M. (1968). Overblown affinity for numbers. *Washington Post (H-section).,* (30 June),

Shafer, G. (1976). *A mathematical theory of evidence* . Princeton: Princeton University Press.

Shewhart, W. A. (1939). *Statistical method from the viewpoint of quality control* . Washington, D.C.: The Graduate School, Dept of Agriculture.

Sheynin, O. B. (1977). Early history of the theory of probability. *Archives for the History of Exact Sciences, 17*(3), 201-259. (With bibliography of 118 entries.)

Sigwart, C. (1895). *Logic - (2 Vols)* (2nd ed.). New York: Macmillan. (H. Dendy, Trans. Originally published, Tübingen, 1873, 1878.)

Sjöström, O. (1980). *Svensk samhällsstatistik: Etik, policy och planering* . Stockholm: Akademilitteratur.

Sjöström, O. (1983). Vad är statistisk metod? *Statistisk Tidskrift,* (2), 109-120. (Summary in English, pp.164-165.)

Sjöström, O. (1984a). Frigörelse av statistiska synsätt i en tid av teoribrist. *Statistisk Tidskrift,* (3), 198-203.

Sjöström, O. (1984b). Vad statistikutredningen inte utredde. *Statistisk Tidskrift,* (4), 297-300.

Sjöström, O. (1988). *Vad kan vi lära av Tjernobyl? En statistisk och samhällsvetenskaplig studie. Förslag till inriktningar mot förnyelser i statistiska grundutbildningar och anknuten forskning inom högskolan* (Case study in the project 'Education for application of statistics'). University of Uppsala, Dept of Philosophy.

Stevens, G. C. (1982). O.R.workers, information systems analysts and the challenge of the micro. *J. of the Operational Research Soc., 33*, 921-929.

Strauch, R. E. (1970). *Some thoughts on the use and misuse of statistical inference* (Report P-3992-1). Santa Monica, Calif.: The RAND Corporation.

Sundgren, B. (1973). *An infological approach to data bases* (Doctoral dissertation). University of Stockholm, Dept. of Information Processing, & National Central Bureau of Statistics.

Sundgren, B. (1975). Infologisk utformning av statistiska databaser. *Statistisk Tidskrift,* (No. 4),

Sundgren, B. (1982). *Statistical data processing systems: Architectures and design methodologies* (Report S/SYS E11). Stockholm: National Bureau of Statistics.

Sundgren, B., Wallgren, B., & Wallgren, A. (1984). *Statistiska informationssystem: Mot ett forskningsprogram i statistiska informationssystem och statistisk analys med data från administrativa system* (ASLAB Memo 84:2). University of Linköping, Dept. of Computer and Information Science.

Tarter, M. E., & Kronmal, R. A. (1976). An introduction to the implementation and theory of nonparametric density estimation. *The American Statistician, 30*(3), 105-112.

Thomas, G. B., Jr. (1972). *Calculus and analytic geometry* . Reading, Mass.: Addison-Wesley.

Tukey, J. (1960). Conclusions vs. decisions. *Technometrics, 26*(4), (Also reprinted in Badia, et al., *Research problems in psychology.* Reading, Mass.: Addison-Wesley, 1970.)

Tukey, J. W. (1969). Analyzing data: Sanctification or detective work? *The American Psychologist, 24*, 83-91.

Tukey, J. W. (1975). *Methodology, and the statistician's responsibility for BOTH accuracy AND relevance* (Presented as talk at the annual meeting of the Am. Statistical Ass. in Atlanta, Georgia, August 1975).

Wallin, E. (1980). *Vardagslivets generativa grammatik: Vid gränsen mellan natur och kultur* . Lund: C.W.K.Gleerup. (Summary in English.)

Wallin, E. (1986). *Litteraturen om artefakter och det artificiella: Några perspektiv på den konstgjorda världen* (SALFO report, version 1, 22 June 1986). Stockholm: The Swedish Committee for Future Oriented Research FRN/SALFO.

Walster, G. W., & Cleary, T. A. (1970). A proposal for a new editorial policy in the social sciences. *The Am.. Statistician,* (April), 16-18.

Whitaker, R., & Östberg, O. (1988). Channeling knowledge: Expert systems as communications media. *AI & Society, 2*(3), 197-208.

Wold, H. (1957a). Genmäle till P. Nørregaard Rasmussen. *Ekonomisk Tidskrift,* , pp. 298-303.

Wold, H. (1957b). *Kausal inferens från icke-experimentella observationer: En översikt av mål och medel* . Uppsala and Wiesbaden: Almqvist & Wiksell and Otto Harrassowitz. (Kungl. Humanistiska Vetenskapssamfundet i Uppsala, Årsbok 1955-1956:1.)

Woodward, J. L. (1934). Quantitative newspaper analysis as a technique of opinion research. *Social Forces, 12*, 526-537.

Yates, F. (1951). The influence of statistical methods for research workers on the development of the science of statistics (With review of R.A.Fisher's book Statistical methods for research workers). *Journal Am. Statistical Ass., 46*(253-256, March), 19-34.

Zadeh, L. A. (1965). Fuzzy sets. *Inform. Control, 8*(June), 338-353.

Östberg, G. (1981). *Vad betyder 10^{-8}: En studie av innebörden av mycket låga siffror för sannolikheten för tankbrott* (Research report). Lund Institute of Technology, Dept. of Engineering Materials.

Östberg, G. (1982). *Evaluation of a design for inconceivable event* (Paper presented at the Conference on Design Policy, London, July 1982). Lund's Inst. of Technology, Dept. of Engineering Materials.