– sonst noch abgehandelt ist. – Das Buch ist flüssig geschrieben, außerordentlich klar und übersichtlich gedruckt und praktisch frei von Druckfehlern. Es ist dem Lernenden, aber auch dem Praktiker zu empfehlen, denn es gibt viele wertvolle Ratschläge und Einblicke.

Robert Fugmann

Dr. R. Fugmann, Wiss. Dok., Hoechst AG
Postfach 800320, D-6230 Frankfurt 80

DOBBENER, Reinhard: **Grundlagen der Numerischen Klassifikation anhand gemischter Merkmale.** (Foundations of numerical taxonomy based on mixed variables) Göttingen: Vandenhoeck und Ruprecht, 1983. (Studien zur angewandten Wirtschaftsforschung und Statistik aus dem Institut für Statistik und Ökonometrie der Universität Hamburg; Heft 15) 143 p. ISBN 3-525-11286-6, DM 49,–

The present book is neither a textbook on numerical classification nor a treatise on a specific theory of numerical classification. One gets rather the impression that the author has tried to combine several approaches and results in one paper.

In the introduction several kinds of scales of measurement are introduced and discussed which are later on assumed for the variables, by which objects are to be classified. It seems that the author knows only in part the literature on measurement theory. In particular the example on p.16 can lead to confusion. However, this lack has no effect on the approaches to classification discussed later on.

The second chapter deals with invariance properties of metrics and classification criteria for interval and ratio scales. To this aim on the one hand those metrics are characterized which are invariant with respect to the transformations which are admissible with respect to interval and ratio scales. On the other hand an invariance property is defined for classification criteria in case of a fixed number of classes, and it is proved that the common variance and determinant criteria can be characterized by means of maximum likelihood by such properties. This approach is important and interesting for a theoretical evaluation and perhaps also for the construction of classification procedures. Unfortunately several proofs (p.25, 39, 47, 50) are not given in the book but only in an unpublished discussion paper of the author.

In the third chapter classification procedures are proposed for ordinal scaled variables and for the case of the simultaneous occurrence of nominal, ordinal, interval, and ratio scales. There is an immense need for such procedures in applications. The central point is the definition of an empirical variance criterion S for ordinal variables with a finite number of possible values. This is based on a representation of an ordinal variable with p distinct values by p-1 binary variables. An entropy criterion is calculated with respect to the cumulated relative frequencies of the ordered binary variables. It is proved that this criterion has quite a few interesting properties which illustrate its great importance in the context of classification procedures. Outside this context it would be certainly helpful to consider first of all the properties of the corresponding population parameter instead of its empirical estimator. For data with variables measured on different scales it is proposed to classify interval and ratio scales in such a way that ordinal scales with a finite number of values result. In this way the problem is reduced to the evaluation of ordinal scales by means of the new entropy criterion and nominal scales by means of the usual entropy criterion. The proposal (p.98) to use normed entropy criteria in order to assign in this way equal weights to the different variables within the common variance is a bit problematic, at least in applications, since such artificial weights might lead to classifications which are difficult to interpret. But the given approach can easily be modified by introducing subjective and other weights. In particular, for ordinal scales properties of invariance and monotony are considered for a certain criterion and procedures for iterative and hierarchical classification based on this criterion are proposed.

In the fourth chapter statistical tests are discussed by which it should be tested whether a natural class structure is inherent to the data. On p.118a procedure is described which most probably will not lead to meaningful results. It is proposed to simulate data sets for which not only the vectors of measurement for each object but also the components of these vectors are independent. But it is self-evident that for concrete problems one should assume a certain dependence of the variables even under the null hypothesis of only one homogeneous class. This assumption is made in most known clustering procedures. Using the described approach in practice will have the effect that the probability of a false rejection of the null hypothesis is not controlled.

With respect to the procedure based on random permutations of the measurement vectors which is described on p.120 it seems that the author is not aware of the theory of randomization or permutation tests, respectively, originating from Fisher and Pitman, though he uses their fundamental idea. This can be concluded also from the very clumsy and time-consuming procedure on p.120 for generating random permutations. The application of such tests for evaluating classifications is by no means new, too (cf. e.g. Hubert et al., Evaluation Review 6(1982)p.505–520).

In contrast to the statement on p.122 it is possible in principle to determine the conditional distributions of the corresponding statistics by calculating for all possible permutations the value of the test statistic and assuming a discrete uniform distribution under the null hypothesis. In this way exact tests are possible without the need for a simulation study. With respect to today's computers such tests, of course, can be performed only for rather small samples and in general one will fall back on simulation studies. However, for a reliable evaluation some thousands of replications are necessary, since on the basis of only 6 permutations (cf. section 4.4) no justified conclusions are possible with respect to the structure and number of classes.

Altogether one should regard the present book rather as a kind of research report and not as a fundamental

book on numerical classification for variables with different scales of measurement. For researchers in numerical classification the book contains quite a few useful ideas and results with respect to theoretical properties of classification procedures in general and to procedures for ordinal scaled data in particular.

Joachim Krauth

Prof. Dr. Joachim Krauth
Lehrstuhl für Psychologie IV
Universität Düsseldorf. Psychol. Inst.
Universitätsstr. 1, D-4000 Düsseldorf 1

GENTLE, James E. (Ed.): **Computer Science and Statistics: Proceedings of the Fifteenth Symposium on the Interface.**
Amsterdam, New York, Oxford: North-Holland Pub. Co 1983; 379 p. ISBN 0444866884.

Es handelt sich um den Tagungsband der fünfzehnten ‚Interface'-Tagung, die im März 1983 in Houston, Texas, stattfand. Die Interface-Tagungen entsprechen im US-amerikanischen Bereich den europäischen COMPSTAT-Tagungen, d.h. man befaßt sich mit Themen, die sowohl den Bereich der Informatik, als auch den Bereich der Statistik betreffen; man befaßt sich also im wesentlichen mit rechnergestützten statistischen Auswertungen.

Die diesjährige Tagung behandelte die Themen Statistische Datenbanken, Simulation, Software Trends bei Kleinrechnern, Programme für die Analyse von Überlebenszeiten, Berechnen von Zeitreihen, numerische Algorithmen, Endbenutzerschnittstellen (‚human interface') bei Kleinrechnern, Mustererkennung und Dichteschätzung, Statistische Auswertungssysteme: Implementationstechniken, Werkzeuge für den Entwurf von Statistik-Software, nichtnumerische Algorithmen, neue Möglichkeiten der Datenanalyse, Optimierung sowie Software Metriken und Aufwandsabschätzung. Die schriftlichen Fassungen der Vorträge zu diesen Themen sind im Tagungsband enthalten. Besonders zu erwähnen sind die Gebiete des Einsatzes von Kleinrechnern und der Komplexität von Algorithmen, die in stärkerem Maße als bisher üblich behandelt wurden. Dies ist sehr zu begrüßen. Zum einen, weil Komplexitätsbetrachtungen bei den bisherigen Überlegungen zu rechnergestützten statistischen Auswertungen nur wenig bedacht wurden. Zum anderen, weil die Möglichkeiten des Einsatzes von Kleinrechnern immer größer werden – und zwar im positiven wie im negativen Sinne. Es ist wichtig, daß Wissenschaftler des Fachgebietes ‚Statistical Computing' sich frühzeitig mit dieser Entwicklung befassen. Vielleicht können sie verhindern, daß einige Fehler bei früheren Entwicklungen nochmals begangen werden.

Der Tagungsband bietet einen guten Überblick über den Stand der derzeitigen Entwicklung bei rechnergestützten statistischen Auswertungen. (For an abstract see Int. Classif. 1983-3, p. 177, No. 9752.)

R. Haux

R. Haux, Institut f. med. Dokumentation, Statistik und Datenverarbeitung, Univ. Heidelberg
Im Neuenheimer Feld 325, D-6900 Heidelberg 1

KITTREDGE, Richard;LEHRBERGER, John (Eds.): **Sublanguage. Studies of Language in Restricted Semantic Domains.** Berlin–New York: W. de Gruyter 1982. 240 p.

Der Band ist eine Sammlung von 8 Originalbeiträgen und 3 Wiederabdrucken. Sie lassen, dem Titel entsprechend, einen systematischen Zugriff auf das Phänomen der Sub- oder Teilsprachen innerhalb einer Standardsprache von einem semantischen Ansatz her erwarten. Diesen Ansatz spezifizieren die Hrsgg. in ihrer Einleitung: nicht nur der Wortschatz kennzeichnet eine Subsprache, sondern ebenso der ‚Stil'. Man könnte nun erwarten, daß – nach einer versuchsweisen Bestimmung des komplexen Begriffs ‚Stil' – alle sprachlichen Charakteristika, die den ‚Stil' konstituieren, in den ‚Subsprachen' untersucht werden. Die weiteren Erläuterungen und die einzelnen Beiträge machen aber bald deutlich, daß die Autoren von einem sehr begrenzten, grammatikalischen Ansatz ausgehen, ohne den Stilbegriff weiter zu thematisieren. Ausgangspunkt ist die Grammatik einer ‚Subsprache', wie sie von Zellig Harris in seiner transformationell-distributionalistischen Arbeit „Mathematical Structures of Language" entwikkelt wurde: „certain proper subsets of the sentences of a language may be closed under some or all of the operations defined in the language, and thus constitute a sublanguage of it"[1]. Im Zentrum des Interesses steht also die Grammatik von ‚Subsprachen', die die angenommenen Korrelationen zwischen ‚Subsprache' und entsprechendem, eingeschränktem Sachverhaltsbereich darstellen soll.

In dem wiederabgedruckten Beitrag „Syntactic formatting of science information" von 1972 hatte *Naomi Sager* im Rahmen eines Forschungsprojektes über Möglichkeiten der Informationsgewinnung einen Vorschlag gemacht, wie die Bibliotheks- und Informationsgewinnungsdienste auf der Grundlage einer maschinellen Verarbeitung natürlicher Sprachdaten mithilfe eines benutzerfreundlichen Frage-Antwort-Systems relevante Information aus wissenschaftlichen Texten ziehen könnten. Sie stellte sich die Frage, inwieweit wiss. Informationen syntaktisch formatiert werden könnte. Eine Lösung fand sie in sprachlichen Informationsmustern (Formaten), hier also in syntaktischen Mustern als Strukturen von grammatischen Wortkategorien, die durch eine transformationelle Zerlegung von Sprachäußerungen in Elementarsätze gewonnen wurden und die die Informationsstruktur der Sprachäußerungen in dem fachlichen Bereich widerspiegelten. Demonstriert wurde diese Methode am Beispiel einer speziellen Grammatik für einen pharmakologischen Sprachausschnitt. Etwa zur gleichen Zeit wurde diese Methode für die Informationsgewinnung aus deutschen (Fach)texten im Rahmen eines LDV-Forschungsprojektes für die Entwicklung und Konstruktion eines ‚Informationssystems auf linguistischer Basis' (ISLIB) theoretisch entwickelt und teilweise implementiert[2]. Solche Informationssysteme können derart anwendungsorientiert konstruiert werden, daß der Benutzer in seiner Muttersprache und ohne jegliche Programmierkenntnisse Zugang zu den aus der Textverarbeitung gewonnenen Informationen hat.